

КОНСТРУЮВАННЯ ЛІНГВІСТИЧНОГО КОРПУСУ: ВІД ТЕОРІЇ ДО ПРАКТИКИ (НА МАТЕРІАЛІ НІМЕЦЬКОЇ ФАХОВОЇ МОВИ АРХІТЕКТУРИ ТА БУДІВНИЦТВА)

Рубана Є. С.

ВСТУП

Останні десятиліття історії вивчення характеру, специфіки існування і функціонування загальнонаціональних та фахових мов в реаліях різних мовних субстратів і географічних ареалів позначилися не лише зміною наукової парадигми, але й показали гостру потребу дослідження закономірних лінгвістичних явищ і феноменів з позицій трансдисциплінарного підходу. У рамках процесів інтенсифікації, інтеграції та глобалізації науково-технічного прогресу все частіше вчені актуалізують питання студій мовних одиниць не абстраговано, а в контексті, тобто, на перший план висувається функційний потенціал мовної одиниці у процесі міжмовної комунікації. Реалізація досліджень цього напрямку потребує використання досконалих механізмів, методів і методик лінгвістичного аналізу, а також, насамперед, удосконалення процесів вибірки і підбору матеріалу дослідження. Саме тому провідні вітчизняні та зарубіжні лінгвісти у своїх розвідках неодноразово піднімали питання доцільності використання мовних корпусів у процесі розробки та укладання нового типу словників, впровадження технологій, що базуються на використанні таких корпусів, у процес викладання іноземних мов, теоретичного обґрунтування засад лінгвістичного опису матеріалу, отриманого під час обробки різних видів мовних корпусів і т. д.

Багатовекторність проблематики використання мовних корпусів у мовознавстві і основні засади корпусної лінгвістики вивчали наступні мовознавці: Дж. Егберт, П. Бейкер, Д. Байбер, Ш. Гріс, А. Стефанович, Ю. Тао, Д. Шонефельд, В. В. Жуковська, О. М. Демська-Кульчицька, В. П. Захаров тощо. Сам термін *корпусна лінгвістика* було впроваджено в науковий обіг у 1983 році після

виходу в світ збірника праць «*Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*»¹.

Масштаби експонентного росту сучасних досліджень з корпусної лінгвістики вражають, адже прорив, який вдалося зробити в цій галузі наукового знання, сприяв тому, що сьогодні текстовий корпус називають альфою і омегою лінгвістики². Ціла когорта дослідників стверджує, що лише на основі вивчення мовних корпусів можна дати об'єктивну оцінку процесу комунікації між реальними людьми в реальних умовах і це є ключовим фактором становлення і стрімкого розвитку цієї досить молодій галузі лінгвістичної науки³.

Мовний корпус постулюється як мультифункціональне джерело емпіричних даних та ефективний інструментарій для отримання й експлікації верифікованих результатів дослідження⁴. Створення мовних корпусів сприяло розвитку не лише корпусної лінгвістики, а й суміжних наук, як-от: лексикології, лексикографії, стилістики, соціолінгвістики, етнолінгвістики, етнопсихолінгвістики, лінгвоекології, термінознавства, перекладознавства, діалектології, історії, краєзнавства і т.д.⁵.

Очевидно, що проектування і розробка мікро– та макроструктури двомовного перекладного словника у галузі архітектури та будівництва з загальним обсягом понад 35000 слів, що є одним з основних завдань нашого дослідження, не може обійтися без залучення інноваційних технологій та програмного забезпечення корпусної лінгвістики, тому питання створення корпусу писемних та усних текстів німецької фахової мови архітектури та будівництва і створення на їх основі лексикографічного джерела нового типу знаходиться на перетині двох галузей і потребує вивчення досвіду як теоретичного обґрунтування, так і практичного використання мовних корпусів, що і пояснює актуальність представленої лінгвістичної розвідки.

¹ Aarts J., Meijs W. *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Costerus New Series. Volume:45.1984. 229 p.

² Демська–Кульчицька О. *Основи національного корпусу української мови* [монографія]. Київ, 2005. 218 с.

³ Жуковська В.В. Ресурси корпусної лінгвістики у дослідженні історичної динаміки мови. *Матеріали міжнародної наукової конференції «Слово і речення: синтактика, семантика, прагматика»*. К.: Київ. унт ім. Б.Грінченка, 2013. С. 151-156.

⁴ Захаров В.П. *Корпусная лингвистика: Учебно–метод. пособие*. СПб., 2005. 48 с.

⁵ Захаров В.П., Богданова С.Ю. *Корпусная лингвистика: учебник для студентов гуманитарных вузов*. Иркутск: ИГЛУ, 2011. 161 с.

Метою нашої роботи є комплексний аналіз теоретико-методологічного базису дослідження мовних корпусів, пояснення особливостей вибірки матеріалу розробниками мовних корпусів, та встановлення переваг і обмежень щодо використання мовних корпусів у процесі викладання іноземної мови та у процесі їх лінгвістичних студій.

Виконання наступних завдань послужило реалізації поставленої мети:

- виокремити поняттєво-категоріальний апарат корпусної лінгвістики;
- висвітлити екстра- та інтралінгвістичні фактори впливу на становлення і розвиток різних видів корпусів;
- диференціювати етапи створення та використання мовних корпусів на базі німецької фахової мови архітектури та будівництва (попереднього редагування, нормалізації та стандартизації текстів, а також пост-редагування);
- експлікувати роль розробленого корпусу у плануванні та реалізації цілей занять з іноземної мови;
- описати специфіку використання зазначеного мовного корпусу у термінознавстві, зокрема, у ході розробки двомовного словника, лексикології, стилістиці, соціолінгвістиці.

Матеріалом цієї лінгвістичної розвідки послужив розроблений нами корпус текстів німецької фахової мови архітектури та будівництва.

1. Теоретичні засади конструювання і обробки мовних корпусів

1.1. Корпус як об'єкт опису корпусної лінгвістики

Історія виникнення нової предметної галузі мовознавства – корпусної лінгвістики і, зокрема, мовних корпусів як основного джерела емпіричних даних автентичного характеру датується другою половиною ХХ століття. Функційна релевантність використання усних та письмових корпусів в усіх сферах лінгвістики диктується зміною наукової парадигми, потребою пошуку й аналізу достовірного матеріалу дослідження не на основі суб'єктивних, інколи навіть, інтуїтивних тверджень, спостережень і висновків, а з опорою на нові методики, платформи, способи обробки, систематизації та класифікації лінгвістичної інформації з урахуванням різноманітних перспектив, а також інноваційного програмного забезпечення.

Будучи об'єктом зацікавлення значної кількості лінгвістів, проблема дефініції поняття *мовний корпус* досить швидко нівелювалася, адже сучасна корпусна лінгвістика пропонує трактувати мовний корпус як набір текстів чи уривків текстів, зібраних з метою їх використання як мовного зразка чи зразка різновиду мови⁶. Зважаючи на це, вважаємо за доцільне висувати на перший план дилему використання мовних корпусів, їх призначення, організацію вибірки, розробки, перевірки та удосконалення не на базі загальнолітературних мов, а базуючись на матеріалі спеціальних субмов конкретних загальнонаціональних мов, що репрезентують окремі сфери професійної комунікації між фахівцями та нефахівцями обраної галузі⁷.

Висвітлюючи теоретико-методологічний базис лінгвістичних досліджень мовного корпусу на базі німецької фахової мови архітектури та будівництва, слід окреслити шляхи вирішення таких важливих питань корпусної лінгвістики, як-от: визначення розміру корпусу, вибір та збір текстових документів, систематизація текстового матеріалу, спосіб організації вибірки сторінок, детермінація цільового користувача корпусу, спосіб введення даних, впорядкування та керування файлами корпусу тощо. Особливе місце у розв'язанні цих питань відводимо вивченню лінгвістичних, екстралінгвістичних та статистичних факторів, адже, наприклад, розмір мовного корпусу може суттєво впливати на валідність та достовірність висновків щодо варіантів використання мови її носіями, а також вивчення інтра– та екстралінгвістичної інформації. Загальноприйнятим розміром мовного корпусу німецьких текстів нами було визначено 15 млн. лексем.

Найбільш прагматичним методом визначення текстів, включених в мовний корпус, вважаємо рандомну вибірку зразків текстів, здійснену за допомогою методики С. Шароффа, з урахуванням трьох основних критеріїв: 1) наявність у тексті сполуки з щоразу нових чотирьох скомбінованих термінів, що належать до попередньо складеного переліку базових термінологічних одиниць НФМАБ (загалом 300 термінів, дібраних з монографій, галузевих енциклопедій, довідників, підручників тощо); 2) типовості текстів, що передбачає вибірку текстів у мережі Інтернет, обмежених певною

⁶ Duden. URL: https://www.duden.de/rechtschreibung/Korpus_Sammlung

⁷ Kontutyte E. Einführung in die Fachsprachenlinguistik. Vilnius : Vilniaus universiteto leidykla, 2017. 67 S.

тематикою; 3) кількість запитів повинна складати понад 1000⁸. Таким чином, у запропонованому корпусі НФМАБ знаходимо зразки текстів у сферах архітектури, будівництва, інженерії, технологій, мистецтва, деревообробної галузі і т.д., що вказує на поліфункціональний характер обраної галузі знань. Досліджуваний мовний корпус належить до так-званих монодисциплінарних монолінгвальних корпусів, тому однією з базових вимог до вибору текстових документів є пропорційна кількість видів тексту, представлених у корпусі.

Ще одним важливим постулатом вибірки текстів для мовних корпусів є диференціація та акцентуалізація специфіки кожного окремого виду. Наприклад, під час включення до складу корпусу зразків текстів, які передбачають наявність термінів та терміносполук (наприклад, оглядових статей, лексикографічних джерел, термінологічних словників, науково-популярної літератури) доцільно визначити обмеження в кількості слів текстів, що будуть включені до такого корпусу (50000 слів). У виборі текстів, що дозволяють фіксувати місце та роль термінів у терміносистемі і репрезентуються за допомогою термінологічних тезаурусів, підручників, посібників, довідкових матеріалів, нормативних документів, що регулюють механізми технологічних та виробничих процесів у ході реалізації будівництва та конструювання споруд, термінологічних стандартів тощо, значення надається тому, чи текст належить до інформативного типу (наприклад, описи креслень, графіків), чи представлений зразок має термінотворчий характер (проекти, патенти, винаходи, гранти тощо). При цьому, важливо пам'ятати, що кожен тип тексту характеризується певним набором лексем, термінів, терміносполук, колокацій, ідіом та інших лінгвістичних ознак.

Верифікуючи принципи розробки і укладання текстового корпусу на базі німецької фахової мови архітектури та будівництва слід зазначати гостру потребу створення полілінгвальних текстових корпусів (зокрема, бази німецько-українських текстів у галузі архітектури та будівництва), які надаватимуть матеріали для проведення порівняльних лінгвістичних досліджень між мовами, задіяними у процес фахової комунікації, націленої на відбудову та

⁸ Sharoff S. Creating General-Purpose Corpora Using Automated Search Engine Queries. Baroni, Marco; Bernardini, Silvia (eds.): WaCky! Working papers on the Web as Corpus. Bologna, 2006.

реконструкцію інфраструктури українських міст після закінчення воєнних дій. Актуальним вважаємо ініціювання спільного проєкту створення мовних корпусів на базі німецької, англійської та української фахових мов архітектури та будівництва.

1.2. Класифікаційна сітка мовних корпусів

Будучи ядром сучасних студій багатьох лінгвістичних гілок, мовні корпуси та їх класифікаційні сітки не залишаються й поза нашою увагою. Слід наголосити не тільки на перевагах створення різних типів корпусів, а й унаочнити показники їх градації. Так, крім традиційного поділу мовних корпусів на корпуси усних текстів та корпуси письмових текстів, з метою проведення комплексного дослідження німецької фахової мови архітектури та будівництва виокремлюємо такі можливі для створення типи корпусів: спеціальні, анотовані, моніторингові (динамічні), порівняльні корпуси тощо.

Залежно від вибору часового зрізу розрізняємо синхронічний та діахронічний корпуси текстів. Синхронічний корпус НФМАБ містить зразки текстів за певний часовий період і використовується для аналізу характеристик мови під час встановленого часового періоду⁹. Так, наприклад, ми змогли створити базу фахових текстів, що датуються періодом входження Буковини до складу Австро-Угорщини (1849-1918 рр.). На основі цієї бази текстів ми можемо дослідити закономірності фахової комунікації у процесі створення архітектурного ансамблю Резиденції митрополитів Буковини і Далмації (м. Чернівці), Чернівецького міського театру та інших доробок архітектури тієї доби. Діахронічний корпус дозволяє вивчати мову в ретроспективі і включає зразки текстів різних історичних періодів, будучи безцінним базисом під час етапів проєктування і розробки словника.

Слід провести паралелі також між мономовними та полімовними корпусами з урахуванням відмінностей у цілях їх створення. Наприклад, двомовні корпуси стають у нагоді лінгвістам при розробці двомовних перекладних словників, створенні термінологічних тезаурусів, машинному перекладі та у викладанні

⁹ Пинтюк Є.С. Етимологічні характеристики інтернаціоналізмів і термінів іншомовного походження німецької фахової мови архітектури та будівництва. *Закарпатські філологічні студії*. 2021. Вип. 17. Том 1. С. 162-166. http://zfs-journal.uzhnu.uz.ua/archive/17/part_1/34.pdf.

іноземних мов. Монолінгвальні корпуси є унікальним джерелом мовних даних, містячи важливу інтра– та екстралінгвістичну інформацію, необхідну для створення теоретичних граматики, досліджень у галузі лексикології, лексичної семантики, соціолінгвістики, психолінгвістики, стилістики, історії мови, лінгвоекології, культурології, етнографії і т.д.¹⁰

Особливістю монолінгвальних анотованих корпусів є їх здатність містити показники екстра– та інтралінгвістичної інформації. Екстралінгвістична анотація передбачає наявність даних про вид тексту, рік публікації тексту, ім'я та вік автора, його гендерну приналежність, інформацію про сфери використання тексту та про розміщення тексту у реєстрі. Інтралінгвістична анотація включає аналітичні маркери, а саме: маркери на позначення частин мови слів чи словосполучень, категорію лексичної одиниці, дані про тип та часові ознаки речення, семантику слів та ідіом, ідентифікацію власних і загальних назв, а також дискурсивні характеристики тексту¹¹.

Слід зауважити, що розвиток корпусної лінгвістики сприяв появі і такого типу як діалектний корпус. На перший погляд, потреба створення такого корпусу для дослідження фахових мов може здаватися не актуальною, але згаданий тип корпусу не лише презентує інформацію про звуки і фонемі, які використовуються в діалекті, а й містить величезну кількість інформації, пов'язаної з морфемами, словами, ідіоматичними виразами, реченнями та іншими лінгвістичними елементами певного регіонального варіанту мови. Важливо розуміти, що для оптимізації процесу фахової комунікації між фахівцями та нефхівцями галузі такий корпус стане ключем до оволодіння будь-якого архітектурного стилю та напряму, що є генетичним кодом окремого народу і, зокрема, певної регіональної спільноти як його частини. Крім того, з плином часу лінгвістичні дані, а разом з ними і знання, звичаї, картина світу окремої групи людей втрачаються і безслідно зникають. Діалектний корпус ж уможливиле зберігання, обробку і аналіз даних та інформації про особливості використання мови у певному

¹⁰ Schmidt T., Wörner K. *Multilingual corpora and multilingual corpus analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2012. 407 p.

¹¹ P. Shrestha, Ch. Jacquin, B. Daille. Reduction of Search Space to Annotate Monolingual Corpora. Proceedings of 5th International Joint Conference on Natural Language Processing. Chiang Mai, Thailand. Asian Federation of Natural Language Processing, 2011. P. 1457–1461.

географічному регіоні і служить матеріалом соціокультурних, соціоекономічних та соціополітичних досліджень. Цей тип корпусів дозволяє оцінити мову, забарвлену унікальними лінгвістичними особливостями і порівняти її із стандартним варіантом.

Якщо мова йде про встановлення генеалогічної близькості та типологічної спорідненості двох мов, тоді створюються порівняльні корпуси, які укомплектовуються однаковими текстами, написаними двома чи більше мовами. Обов'язковою умовою порівняльних корпусів є тотожне число даних та вибірки, взятих з кожної мови¹².

Огляд різних видів корпусів показав, що потреба здійснення комплексних трансдисциплінарних студій нерідко зумовлює появу нових гібридних видів мовних корпусів, що служать для реалізації поставлених завдань.

1.3. Стандартизація та уніфікація усних і письмових текстів в корпусі німецької фахової мови архітектури та будівництва

Перш ніж корпус текстів стане придатним для використання у різних галузях людського знання, розробники кожного мовного корпусу здійснюють його попереднє редагування, нормалізацію, стандартизацію та стратифікацію. Нормалізований варіант усного чи письмового тексту зменшує робоче навантаження під час подальшого використання корпусу і полегшує доступ до такого корпусу як з боку людини, так і машини.

Попереднє редагування корпусу текстів НФМАБ – це процес коригування фахових текстів у текстовому корпусі з метою покращення якості вихідних текстових даних для подальшого практичного застосування у машинному навчанні, граматичних анотаціях, машинному перекладі, у ході здійснення вибірки мовних даних, обробки тексту, виокремлення технічних термінів та терміносполук, пошуку екстра- та інтралінгвістичної інформації тощо. Попереднє редагування разом зі стандартизацією текстів є основними складовими процесу нормалізації фахових текстів у мовному корпусі НФМАБ. Попереднє редагування передбачає усунення типографічних помилок, невідповідності пунктуації, лексичної та синтаксичної невідповідності, а також керування метаданими, спрощення текстового формату, маркування

¹² B. Maia. Training Translators in Terminology and Information Retrieval using Comparable and Parallel Corpora. Corpora in Translator Education. ST JEROME PUBLISHING, 2003. P. 43-54.

ідіоматичних виразів і видалення нетекстових елементів¹³. На противагу етапу попереднього редагування, процес стандартизації текстового матеріалу спроектованого нами корпусу полягає в транслітерації назв архітектурних об'єктів та архітекторів, перевірці дотримання граматичних норм, вимірюванні індексу використання термінів тощо.

Однією з основних цілей стандартизації текстів у нашому корпусі є зробити тексти максимально читабельними, тобто оптимізувати зміст текстів так, щоб кінцевий користувач зміг без зайвих зусиль зрозуміти зміст та ідею тексту. Ця мета досягається за умови врахування таких аспектів: 1) керування науково-технічними термінами, що використовуються у тексті; 2) фіксація і забезпечення виконання стандартних граматичних правил мови в текстах; 3) дотримання структурної узгодженості текстів; 4) маркування ідіоматичних фраз і терміносполук у текстах.

Після стандартизації та уніфікації текстів для мовного корпусу НФМАБ кожному текстовому файлу присвоюється певний шифр, в якому перший знак служить для позначення категорії тексту, а наступні цифри є показником серійного номеру текстового документа. Далі слідує обробка текстів із застосуванням численних методик (наприклад, частотного підрахунку термінів, термінологічних колокацій, детермінації ключових слів в контексті, локального групування слів тощо) та інноваційного програмного забезпечення.

Отож, окресливши теоретичний базис для проєктування та конструювання мовних корпусів усних і письмових текстів на базі німецької фахової мови архітектури та будівництва, у цьому розділі здійснено проєкцію на основні питання сучасних лінгвістичних студій не тільки у галузі корпусної лінгвістики, а й у всіх сферах людських знань, які визначають своїм об'єктом опису мовний корпус. Менш ніж за сто років питання розміру корпусу, вибору, збору, класифікації, уніфікації та систематизації текстового матеріалу, способу організації вибірки джерел, авторів, уривків і сторінок, способу введення даних, порядку їх номінації, впорядкування й керування тестовими документами корпусу стають ключовими положеннями у дослідженнях з лексикології, лексичної семантики, соціолінгвістики, психолінгвістики, стилістики,

¹³ P. Pym. Pre-editing and the use of simplified writing for MT. Translating and the Computer 10: The Translation Environment 10 Years. Aslib, London, 1990. P. 80–95.

термінології і т.д. Корпус текстів постає безцінним ресурсом для отримання валідних та достовірних висновків з урахуванням специфіки використання мови її носіями, а також інтра- та екстралінгвістичної інформації. Поліфункціональність мовних корпусів є поштовхом для появи нових гібридних видів корпусів, які поєднують в собі ознаки кількох традиційних типів. Незважаючи на багатовекторність у вивченні мовних корпусів, процес конструювання будь-якого мовного корпусу є сталим і поетапним (етапи попереднього редагування, стандартизації, нормалізації, уніфікації текстового матеріалу, обробки текстів та постредагування текстового корпусу).

2. Методологія дослідження лінгвістичних корпусів

2.1. Трансдисциплінарний підхід до вивчення текстових корпусів

Ідентифікація та аналіз певних мовних явищ та феноменів з опорою на корпус текстів не могли б повною мірою бути реалізованими без розробки нової методології, наднових методів та методик дослідження у руслі корпусної та комп'ютерної лінгвістики. Субстратом для впровадження інноваційних технологій у процес вивчення текстових корпусів є трансдисциплінарний підхід. Такий комплексний та багатогранний аналіз текстового матеріалу корпусів репрезентується уже на стадії використання статистичних методів та прийомів. З урахуванням характеру тексту, який використовується в корпусі, схема проведення квалітативного (якісного) аналізу може частково відрізнитися від процедури здійснення квантитативного (кількісного) аналізу, а враховуючи різносторонність мовних корпусів, мовознавці схильні вважати, що в довгостроковій перспективі найефективнішим шляхом об'єктивної оцінки отриманих даних є поєднання обох типів аналізу.

Кількісні результати, отримані на основі обробки корпусів усних та письмових текстів, знайшли також застосування у трансдисциплінарних галузях, на кшталт, математичної лінгвістики¹⁴, судової лінгвістики, стильометрії. Використання даних корпусної лінгвістики у цих галузях знань послужило поштовхом для проведення нових досліджень, формулювання нових гіпотез,

¹⁴ Основи статистичної лінгвістики: Навчально-методичний посібник / Відп. ред. проф. Ф.С. Бацевич. Видавничий центр ЛНУ імені Івана Франка, 2008. 124 с.

розвитку та удосконалення матеріалів для викладання мови і створення додатків для обробки текстових даних.

Все більшої актуальності серед лінгвістів в останні роки набуває так-звана методика подвійного фокусу, яка якраз об'єднує квантитативні та квалітативні підходи як у дослідженні корпусу, так і загалом мови¹⁵.

Для спостереження особливостей розподілу властивостей мови на базі невеликих за обсягом корпусів доцільним буде використання методик кількісного аналізу. Таких методик буде достатньо, щоб встановити закономірності диференціації різних лінгвістичних ознак у фаховій мові та їх розподіл за видами тексту. Кількісний аналіз репрезентує кількісну оцінку даних та обробку їх кількісних показників. Цей методологічний напрям дозволяє класифікувати різні типи елементів відповідної мови чи мовного різновиду на основі певних наперед визначених параметрів. Завдяки методиці кількісного аналізу, наприклад, окреслюємо індекс частоти вживання термінів НФМАБ в мовному корпусі, що належать до тієї чи іншої частини мови (78% – іменники, 19% – дієслова, 2% – прикметники, 1% – всі інші морфологічні категорії).

Для усесторонньої об'єктивної оцінки фахової мови з боку широкого загалу вчені здійснюють подальший якісний аналіз отриманих кількісних результатів. Квалітативний підхід уможливує аналіз поведінки мовних феноменів і, як результат, дозволяє привідкрити завісу щодо їх функційної ролі у розумінні і використанні мови. Основною метою якісного аналізу є обґрунтування та опис лінгвістичного явища, зафіксованого у ході проведення кількісного аналізу. Тоді як кількісний аналіз пояснює природу певного мовного феномену, якісний аналіз вказує на причини та фактори, що сприяли його виникненню в мовному корпусі¹⁶.

У дослідженні ми дотримуємося погляду, що систематичне поєднання кількісних результатів з інтерпретацією їх якісних показників є джерелом взаємозалежного інтерфейсу, який використовується для кращого розуміння лінгвістичних явищ конкретної фахової мови.

¹⁵ Biber, D., S. Conrad, and R. Reppen. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press. 1998. 300 p.

¹⁶ Кашеева А. В. Квантитативные и качественные методы исследования в прикладной лингвистике. Социально-экономические явления и процессы. 2013. № 3(049). С. 155-162.

2.2. Експлікація методів статичного аналізу мовних корпусів

Одним з ключових пунктів нашої розвідки є аспект методології студій корпусів фахових мов, тому на цьому етапі доречним буде зробити короткий огляд методик статистичного аналізу мовних корпусів усних та письмових текстів¹⁷. Експлікація методів та методик статистичного аналізу є важливим завданням корпусної лінгвістики, враховуючи континуальне збільшення кількості друкованих та електронних корпусів.

Далі нам слід більш детально розглянути найпоширеніші методи статистичної обробки результатів, отриманих на основі аналізу мовних корпусів (тобто, описувальний метод, методи статистичного висновку та методи оцінки статистичних даних).

Одразу декілька статистичних методик об'єднують під егідою описувального статистичного методу. Однією з типових методик цього методу є статистика підрахунку частоти, яка базується на стратифікації всіх лінгвістичних елементів з урахуванням попередньо визначених схем класифікації та подальшому арифметичному підрахунку кожного з елементів відповідного класу зі схеми класифікації. Оскільки підрахунок частоти надає інформацію лише про кількість вживань слів, запропонована методика не акцентує увагу дослідника на поширеності слова з точки зору частки загальної кількості лексем, зафіксованих у корпусі. Якщо корпуси, які порівнюються, мають однаковий обсяг лексем, то ніяких проблем не виникатиме, але якщо все-таки існуватиме різниця у кількості слів, то, проводячи будь-який простий підрахунок частоти, можна зіштовхнутися із численними труднощами¹⁸.

Методи статистичного висновку використовуються здебільшого, щоб перевірити значущість спостережуваних моделей використання мови в рамках загального вивчення лінгвістичних явищ. Існує чимало методик методу тестування значущості, які використовуються для аналізу даних, отриманих з корпусу текстів. Серед них: χ^2 -квадрат, дисперсійний аналіз, Т-тест (критерій Ст'юдента), критерій ступеня кореляції Пірсона¹⁹. Так, наприклад, методика χ^2 -квадрату використовується зазвичай, щоб перевірити, чи дійсно моделі спостережуваного розподілу слів мають суттєві

¹⁷ Статистичні методи для лінгвістів : посібник / Перебийніс В. І. Вид. 2, випр. і допов. Вінниця : Нова Книга, 2013. 176 с.

¹⁸ Глисон Г. Введение в дескриптивную лингвистику. М., 1959. 486 с.

¹⁹ Левицкий В. В. Квантитативные методы в лингвистике: учебное издание. Вінниця: Нова Книга, 2007. 264 с.

розбіжності з рівномірним розподілом властивостей у наборі мовних даних. Методику Т-тесту можна використовувати: 1) для встановлення відмінностей між лексико-семантичними групами термінів щодо однієї характеристики на основі середнього арифметичного показника цієї властивості для кожної окремої групи; 2) для порівняння середнього арифметичного значення частоти вживання лексико-семантичних груп термінів та терміносполук у порівнянні з лексемами загальнонаціональної мови; 3) для зіставлення середніх арифметичних показників однієї і тієї ж групи термінів та терміносполук за їх частотним потенціалом. Дисперсійний аналіз – це методика статистичного обрахунку даних корпусу, яка дозволяє оцінити різницю потенціалів залежної змінної та змінної номінального рівня, що має дві або більше категорій. Коефіцієнт Пірсона – це критерій ступеня кореляції, який зазвичай використовується в лінійній регресії і є показником лінійного зв'язку між двома наборами даних.

Метод статистичної оцінки теж об'єднує декілька статистичних методик, як-от: кластерний, факторний види аналізу, багатовимірне масштабування (шкалювання), лінійно-логіфічні моделі тощо, які мають сталий характер використання в лінгвістичних дослідженнях з метою вилучення прихованих шаблонів реляційних інтерфейсів лінгвістичних ознак і властивостей з текстового матеріалу, отриманого на основі обробки мовного корпусу. У процесі вивчення мовних даних методика кластерного аналізу може стати у нагоді під час поділу даних на лексико-семантичні групи термінів (етап кластеризації) та присвоєння окремих ознак гіперонімам цих груп (етап класифікації). Методикою факторного аналізу послуговуються для скорочення великої кількості лінгвістичних змінних до меншої кількості факторів спостереження. Методика багатовимірного масштабування (шкалювання) є альтернативною до методики факторного аналізу. Відмінності між ними полягають у тому, що за допомогою факторного аналізу можна виявити подібність між об'єктами (наприклад, змінними) у кореляційній матриці, а за допомогою багатовимірного масштабування стає можливим аналіз подібних чи відмінних ознак будь-якого типу матриці на додаток до матриць кореляції²⁰.

²⁰ Грицюк П.М., Остапчук О.П. Аналіз даних: Навчальний посібник. Рівне: НУВГП, 2008. 218 с.

Таким чином, можна зробити висновок, що трансдисциплінарний характер актуальних лінгвістичних досліджень вимагає не лише залучення спеціалістів різних предметних сфер, а й розробки і реалізації нового програмного забезпечення, новітніх прийомів, методик, методів і загалом нової методології лінгвістичного дослідження. На часі актуалізується використання методики подвійного фокусу, тому розвідки зарубіжних лінгвістів акцентують увагу на потребі комбінації методик квантитативного та квалітативного аналізів. Історичний огляд засвідчує стрімкий приріст у використанні передового технічного обладнання саме у руслі корпусної лінгвістики. У поєднанні з методами статистичної обробки даних: дескриптивним, методами статистичного висновку та методами статистичної оцінки даних – лінгвістам вдається отримати об'єктивні та надійні результати, а методики хі-квадрату, дисперсійного аналізу, Т-тесту, критерію ступеня кореляції Пірсона вже увійшли в історію лінгвістичних вчень як такі, без яких не можливо уявити сучасна лінгвістичну парадигму.

3. Проблеми і перспективи використання мовних корпусів

3.1. Інтеграція мовних корпусів в процес вивчення німецької мови за професійним спрямуванням

Процеси глобалізації та інтенсифікації науково-технічного прогресу зумовлюють потребу інтеграції корпусів фахових мов як основного та допоміжного джерела мовних даних і інформації в процес викладання та вивчення німецької фахової мови як першої та другої іноземної мови, отримання студентами технічних спеціальностей доступу до мовних та текстових даних корпусу фахових текстів в класі або лінгафонному кабінеті, розширення можливостей учнів завдяки використанню емпіричних лінгвістичних ресурсів, а також перспективи покращення мовленнєвих навичок та ефективності оволодіння НФМАБ з опорою на мовні корпуси.

Підґрунтям для популяризації використання мовних корпусів у процесі викладання німецької фахової мови у галузі архітектури та будівництва є ряд об'єктивних причин: 1) традиційні текстові матеріали для викладання німецької мови досить часто містять інтуїтивно вигадані приклади, які, зазвичай, ігнорують важливі аспекти використання мови і висувають на перший план рідше вживані стилістичні варіанти; 2) матеріали для викладання німецької фахової мови у галузі архітектури та будівництва на основі мовних

корпусів є більш надійними та автентичними, оскільки вони розроблені з тих корпусів, які містять дані, інформацію та приклади реального використання досліджуваної фахової мови; 3) зазначені матеріали націлені на підготовку майбутніх архітекторів та здочих до взаємодії та комунікації в реальних життєвих ситуаціях; 4) посилання на фактичне використання слів, фраз, лексичних колокацій і речень завдяки мовним корпусам надаються майбутнім спеціалістам зазначеної галузі знань, щоб навчити їх моделям використання цих елементів сучасної НФМАБ; 5) у мовних корпусах представлена інформація про частоту використання різних мовних одиниць, що є особливо актуальним під час лексичного вибору на етапі вивчення ступеневого словникового запасу НФМАБ; 6) німецько-українські корпуси текстів оптимізують процес засвоєння майбутніми фахівцями різних питань граматики, стилістики, синтаксису НФМАБ, наприклад, принципів використання полісемічних термінологічних одиниць чи правил розподілу лексичних одиниць та їх семантичного співвідношення тощо; 7) у процесі здійснення інтерсеміотичного перекладу креслень та будівельних планів в усні чи письмові тексти студенти матимуть конкретні приклади вибору тих чи інших мовленнєвих засобів.

Якщо розглядається методика використання мовних корпусів як первинного ресурсу для викладання НФМАБ, то наш фокус концентрується на впровадженні в навчальну діяльність: 1) інтерактивного методу, згідно з яким майбутні архітектори грають керівну роль у процесі вивчення іноземної мови; 2) методики, відповідно до якої концепт студента розглядається з погляду майбутнього фахівця запропонованої галузі; 3) методики коригування помилок з опорою на мовний корпус та інших прогресивних методик викладання німецької мови за професійним спрямуванням.

У ході використання інтерактивного методу із залученням мовних корпусів на заняттях німецької фахової іноземної мови студенти безпосередньо знайомляться з усними текстами корпусу, що зберігаються на комп'ютері, і в інтерактивному режимі вони отримують доступ, можуть використовувати і звертатися до різноманітних мовленнєвих даних, щоб вивчити моделі вербального спілкування, використання наголосу в мовленні, зразків інтонації, використання різних надсегментних властивостей, таких як паузи, правил посередництва у процесі фахової комунікації, моделей

варіацій мовлення на основі завдань, діалогічних взаємодій, використання контекстно-орієнтованих варіантів професійного мовлення, комунікативної компетенції, динамічного використання мовлення, спонтанного продукування мовлення, особливостей вживання термінологічної лексики в мовленні, специфіку граматики конкретної фахової мови, моделі комунікації між фахівцями та нефхівцями досліджуваної галузі, закономірності архітектурного дискурсу, урахування регіональних особливостей та культурних традицій певного народу чи його частини, управління процесом мовленнєвої діяльності і т.д. Якщо викладач використовує текстові дані корпусу під час викладання НФМАБ, то студенти мають змогу ознайомитися з правилами побудови письмової комунікації, використанням стандартної граматики на письмі, специфікою термінологічного різноманіття та універсуму в текстах, використанням терміносполук, складанням письмового фахового тексту, різницею між формальним і неформальним письмом, монологічною організацією текстів, синоптичною структурою змісту фахових текстів НФМАБ, моделями структурування аргументів, процесом побудови логічної композиції писемного фахового тексту та способами презентації та градації інформації в тексті.

Методика використання даних з корпусу текстів на заняттях з німецької мови за професійним спрямуванням сприяє тому, що студенти можуть самостійно досліджувати, як, наприклад, конкретний гіпонім використовується у фахових текстах, складених носіями німецької мови, виявляти помилки, якщо такі є, і виправляти їх з прямим посиланням на використання в мовному корпусі без участі викладача. Ця методика покликана заохочувати студентів розробляти власні стратегії оволодіння фаховою мовою з опорою на корпус, а не керуватися стратегіями, розробленими викладачами.

Методика викладання німецької мови, що базується на стратегії *студент як майбутній фахівець галузі*, передбачає самостійне покращення майбутніх архітекторів своїх мовленнєвих та лінгвістичних навичок. Практичні завдання, зазвичай, будуть пов'язані як з актуальними питаннями синтаксису та граматики, так і з основними мовними моделями НФМАБ.

Ми пропонуємо розробити комплекс німецько-українських корпусів фахових текстів, що розглядатиметься також як одна з найбільш автентичних репрезентативних колекцій сучасної німецької та української фахових мов у запропонованій галузі, звідки

ми можемо отримати необхідні лінгвістичні дані, відповідну інформацію та приклади для розробки основних, а також додаткових матеріалів та ресурсів для викладання німецької мови за професійним спрямуванням, наприклад: підручників, довідкових матеріалів, навчальних програм, граматичних підручників, словників, а також баз даних з відповідною термінологією. У цьому випадку текстовий корпус виступає допоміжним ресурсом для викладання НФМАБ.

Використання мовного корпусу під час укладання підручників з німецької фахової мови вносить корективи у роботу їх розробників. Розробник підручників, замість того, щоб вибирати конкретні приклади німецьких текстів для довідок, опису та ілюстрації, може вибирати цікаві текстові зразки з мовного корпусу, враховуючи баланс загальної колекції зразків, щоб переконатися, що найчастотніші терміни та їх типове використання відповідають загальним вимогам навчальної програми. Якщо підручник містить глосарій термінів НФМАБ, то це – список найбільш вживаних термінологічних одиниць, знайдених в текстовому корпусі, з посиланням на їх типові приклади використання.

Укладачі двомовних перекладних словників, використовуючи мовні корпуси на етапах проєктування та розробки, дотримуються наступних вимог: 1) перекладний словник повинен включати найчастіше вживані слова мовного корпусу; 2) матеріал, отриманий на основі вибірки з мовного корпусу, слід упорядкувати в алфавітному порядку; 3) слова слід класифікувати за смисловим критерієм із зазначенням їх використання в контексті; 4) кожне слово вихідної мови повинно мати щонайменше одне концептуально еквівалентне слово у цільовій мові. Основна педагогічна мета двомовного перекладного словника – надати студентам необхідну інформацію про форми та функції слів НФМАБ, необхідних для засвоєння цієї мови²¹. Студентам не потрібно володіти попередніми лексичними знаннями з німецької мови, щоб почати використовувати двомовний словник як допоміжний інструмент. Роль викладача іноземної мови у цьому контексті полягає в тому, щоб представити ресурс майбутнім архітекторам. Викладач повинен

²¹ Кійко С. В., Пинтюк С. С. Німецько-український термінологічний словник архітектури і будівництва: принципи розробки й укладання. *Науковий вісник Міжнародного гуманітарного університету. Серія: Філологія*. Одеса: Видавничий дім „Гельветика”, 2019. Вип. 43. С. 121-125.

слідкувати за тим, щоб словник не дав надмірного лексичного навантаження на студентів, і забезпечити їх перехід на етап навчання, на якому вони самі зможуть досліджувати складніші моделі вживання та значення німецьких термінів.

Надання автентичної інтра– та екстралінгвістичної інформації багатослівних виразів, зокрема, ідіом та термінологічних сполучень на заняттях німецької фахової іноземної мови може вимагати інтерактивного інтерфейсу студентів з двомовним словником. Щоб задовольнити цю потребу, нам, як лексикографам, слід розробити двомовний термінологічний словник в електронній формі та завантажити його в Інтернет так, щоб просте звернення до певного виразу могло відобразити рядок еквівалентних форм, що зберігаються в словнику. Викладач на таких заняттях виконуватиме роль «координатора групи» і це змінить вектор традиційного підходу до викладання фахових мов.

Мовні корпуси можуть бути надзвичайно корисним ресурсом для надання морфологічної та синтаксичної інформації різних типів, необхідної для написання граматик фахових мов. Такі граматики можна ефективно використовувати для українських студентів технічних спеціальностей, оскільки вони можуть надати інформацію про різні граматичні проблеми, що спостерігаються в сучасній німецькій професійній мові, а також можуть надати емпіричні дані для перевірки попередніх гіпотез щодо різних граматичних правил НФМАБ. Підручники з граматик фахових мов на основі мовних корпусів можуть стати чудовим ресурсом для початківців і студентів просунутого мовного рівня, які прагнуть досягти рівня володіння мовою їх носіїв.

3.2. Лінгвістичний корпус і його використання у студіях НФМАБ

Мовний корпус є новим загально визнаним джерелом для проведення досліджень у галузях прикладної та описової лінгвістики. Поява та імплементація мовних корпусів у лінгвістичних дослідженнях дозволили по-новому інтерпретувати традиційні дослідження в лексикології, лексичній семантиці, соціолінгвістиці, психолінгвістиці, стилістиці та термінознавстві. Здійснення неоціненного внеску ресурсів корпусної лінгвістики у нову течію термінознавчих досліджень визначається наступними аспектами: 1) завдяки використанню мовних корпусів лінгвістам

вдалося з нових перспектив підійти до опису фахових мов та їх особливостей (за допомогою нових наборів даних, інформації та прикладів); 2) мовні корпуси послужили для теоретиків засобом перевірки, підтвердження чи спростування вікових теорій та моделей використання фахових мов. 3) дослідники термінознавства отримали можливість безпосереднього використання мовних даних у різних лінгвістичних розвідках; 4) спеціалісти у зазначеній вище сфері отримали доступ до різноманітних та великих за обсягом баз мовних даних, що зробило можливим формування нових спостережень щодо фахових мов, учасників фахової комунікації та процесу її реалізації.

Нам слід наголосити на особливостях використання мовних корпусів у дослідженні лексико-семантичних характеристик термінів та терміносполук. Для цього потрібно розуміти, що термінознавчі дослідження корпусів фахових текстів включатимуть і сфери дослідження інших прикладних галузей мовознавства, зокрема, семантики, лексикографії, семасіології, семіотики, ономасіології, етнолінгвістики та інших наукових дисциплін. Новий напрям розвитку запропонованої науки пов'язаний з прямим посиленням її використанням даних та інформації, зібраних із загальних, а також історичних корпусів. Наприклад, спосіб семантичної зміни деяких абстрактних іменників НФМАБ досліджується з посиленням на розроблений нами корпус текстів НФМАБ. У дослідженнях уточнюються механізми, задіяні в природі семантичної зміни іменників, щоб висвітлити когнітивні процеси, які в ній беруть участь. На базі історичного корпусу фахових текстів здійснюється ретроспективний аналіз семантики термінів НФМАБ у діахронійному зрізі, який дозволяє, наприклад, верифікувати семантичну еволюцію тропоніму «*der Platz*» вже в українській фаховій мові архітектури та будівництва з урахуванням специфіки запозичення цієї лексеми з НФМАБ. Такі дослідження є свідченням того, що посилення на історію корпусних досліджень у дослідженні лексико-семантичних та тематичних рядів НФМАБ має надзвичайно важливе значення. Вони дозволяють простежити еволюцію значення слів мови в історичному зрізі. Інформація, отримана в результаті аналізу корпусу, показує, що семантичні відмінності слів пов'язані з характерними контекстами, позначеними образними, морфологічними, синтаксичними, просодичними та ідіоматичними рамками. Хоча метод дослідження значень слів на основі корпусу є актуальним в лексикології та історичній лінгвістиці німецької мови,

його використання в дослідженні фахових мови ще не було зафіксовано.

Зразки письмового та усного тексту, що зберігаються в корпусі, фактично вилучаються з реального соціального та текстового контекстів, саме тому є надійною базою для проведення соціолінгвістичних досліджень. Так, на матеріалі мовного корпусу можна констатувати спосіб кодування соціолінгвістичної (наприклад, стать, регіон, професію, освіту, посаду, культурну та етнічну приналежність, місце проживання) та дискурсивної інформації (наприклад, стадії виробничо-технічного циклу, їх час та учасників, контексти і ситуації міжкультурної фахової комунікації).

Доцільним є використання мовних корпусів і для аналізу психолінгвістичних особливостей фахової комунікації між фахівцями та нефахівцями галузі (наприклад, у дослідженні виникнення труднощі у процесі комунікації і, зокрема, діалогічної взаємодії).

Підбиваючи підсумки, варто зазначити необхідність висвітлення проблем інтеграції мовних корпусів у процес вивчення та викладання німецької мови за професійним спрямуванням як першої та другої іноземної мови в Україні, а також перспектив використання текстових корпусів як основних та допоміжних ресурсів на заняттях німецької фахової іноземної мови. На практиці мовні корпуси використовуються як основне джерело у ході реалізації методик інтерактивного методу; методики «студент як майбутній фахівець галузі» та методики коригування помилок. Як допоміжне джерело мовний корпус розглядається під час укладання матеріалів для викладання іноземної мови за професійним спрямуванням (наприклад, підручників, довідкових матеріалів, навчальних програм, навчально-методичних комплексів, граматичних підручників, словників тощо).

Крім того, виникає гостра потреба в оновленні теоретико-методологічної бази досліджень традиційного термінознавства на матеріалі мовних корпусів.

ВИСНОВКИ

Мовний корпус як набір текстів чи уривків текстів, зібраних з метою їх використання як мовного зразка чи зразка різновиду мови виступає об'єктом опису корпусної лінгвістики і є центральним поняттям розробленого нами корпусу текстів НФМАБ, а також

запропонованого нами проєкту комплексу німецько-українських корпусів текстів у галузі архітектури та будівництва. У основу цієї лінгвістичної розвідки закладено питання теоретичного обґрунтування етапів розробки та укладання корпусу текстів НФМАБ та перспектив реалізації згаданого вище проєкту, що передбачає створення мультидисциплінарного білінгвального загального текстового корпусу німецької та української фахових мов.

Ключовими факторами, що впливають на процес розробки і створення мовних корпусів, визначаємо лінгвістичні, екстралінгвістичні та статистичні показники. Основними критеріями для здійснення рандомної вибірки текстів у розробленому нами корпусі верифіковано: 1) наявність у тексті сполуки чотирьох скомбінованих термінів з попередньо укладеного реєстру базових термінологічних одиниць НФМАБ; 2) типовість текстів, що передбачає вибірку текстів у мережі Інтернет, обмежених певною тематикою; 3) кількість запитів повинна складати понад 1000.

У лінгвістичній розвідці акцентуємо увагу на стратегії поетапного конструювання мовних корпусів. Пропонуємо виокремлювати етапи попереднього редагування, стандартизації текстового матеріалу, обробки текстів та постредагування текстового корпусу у розробці комплексу німецьких фахових текстів у галузі архітектури та будівництва.

Експлікація характеру інтра– та екстралінгвістичної інформації, отриманої на основі аналізу мовних корпусів, передбачала реалізацію трансдисциплінарного комплексного підходу у розробці нового програмного забезпечення, новітніх прийомів, методик, методів та методології лінгвістичного дослідження із залученням спеціалістів різних суміжних предметних галузей знань. Об'єктивними та статистично надійними результатами обробки мовних корпусів фахових текстів вважаємо показники, отримані у ході використання дескриптивних методів, методів статистичного висновку та статистичної оцінки. На наш погляд, найбільш доцільними у термінознавчих студіях мовних корпусів фахових мов є використання методик χ^2 -квадрату, дисперсійного аналізу, Т-тесту (критерію Ст'юдента) та критерію ступеня кореляції Пірсона.

Дослідні мовні корпуси є не лише надійною джерельною базою для вивчення стану мови на сучасному етапі її розвитку, а й основним і допоміжним ресурсом у ході навчання та викладання фахових іноземних мов. Використання мовних корпусів на заняттях німецької

фахової іноземної мови потребує від викладача значної підготовки і трансформації його ролі в роль інструктора. Імплементація методик інтерактивного методу, методики «студент як майбутній фахівець галузі» та методики коригування помилок дозволяють інтегрувати дані мовних корпусів у процес вивчення іноземної мови та полегшувати засвоєння майбутніми архітекторами та будівельниками обраної мови. Укладачі матеріалів для викладання фахової іноземної мови (наприклад, підручників, довідкових матеріалів, навчальних програм, навчально-методичних комплексів, граматичних підручників, словників тощо) теж досить часто у своїй діяльності користуються зразками, взятими з текстових корпусів, для врахування актуальних аспектів використання мови, а не інтуїтивно вигаданих прикладів. Матеріали для викладання іноземних мов, складені на основі мовних корпусів, є більш надійними та автентичними, оскільки вони розроблені з тих корпусів, які містять дані, інформацію та приклади реального використання фахової мови.

Корпусна лінгвістика зосереджена на «мові, що використовується» з посиланням на користувачів і контекст. Корпуси є унаочненням того, як фахівці використовують мову для вираження унікальної ідентичності, солідарності, генерування нових ідей, поширення інформації, культивуації знань, кодування культури своєї країни, збереження її історії, розвитку комунікативних спільнот та диференціації спільних і відмінних рис – національного й інтернаціонального.

Перспективним вважається вивчення текстових корпусів інших фахових мов, тому пропонуємо у майбутньому створити перший паралельний перекладний корпус українських та німецьких фахових мов, а також перший цифровий архів німецьких та українських фахових мов, який уможливить збір, зберігання, керування і поширення мовних ресурсів та інструментів.

АНОТАЦІЯ

Ця лінгвістична розвідка спрямована на комплексне дослідження процесу формування текстових корпусів німецької фахової мови архітектури та будівництва. Наукова новизна цього лінгвістичного дослідження полягає в тому, що вперше процес генерації корпусу фахових текстів НФМАБ є предметом лінгвістичного аналізу. У дослідженні представлено специфіку моделювання сучасних фахових мов у формі корпусу з їх багаторівневою класифікаційною

структурою, описано найбільш суттєві лінгвостатистичні питання та фактори в запропонованій галузі, а також висвітлено особливості використання корпусів у таких важливих сферах термінознавчих студій, як-от: навчання та викладання фахових мов, створення словників, вивчення регіональних особливостей фахових мов.

Теоретичне значення цього дослідження визначається вивченням внеску лінгвістичної інформації та прикладів, отриманих на базі корпусу, у становлення та розвиток нової парадигми термінознавства. Аналіз сучасних засобів, методів і прийомів корпусної лінгвістики доповнює методологічну теорію застосування мовних корпусів для вивчення специфіки функціонування фахових мов.

Практичне значення роботи зумовлене тим, що її матеріали можуть бути використані в лекціях і практичних курсах з корпусної лінгвістики, термінознавства, загального мовознавства, порівняльної лексикології, когнітивної лінгвістики, перекладознавства. Висновки цієї розвідки можуть бути корисними для викладачів та студентів в педагогічній практиці викладання та вивчення фахових іноземних мов.

Література

1. Aarts J., Meijs W. *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Costerus New Series. Volume:45.1984. 229 p.

2. Демська–Кульчицька О. Основи національного корпусу української мови [монографія]. Київ, 2005. 218 с.

3. Жуковська В.В. Ресурси корпусної лінгвістики у дослідженні історичної динаміки мови. *Матеріали міжнародної наукової конференції «Слово і речення: синтактика, семантика, прагматика»*. К.: Київ. унт ім. Б.Грінченка, 2013. С.151-156.

4. Захаров В.П. Корпусная лингвистика: Учебно–метод. пособие. СПб., 2005. 48 с.

5. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. Иркутск: ИГЛУ, 2011. 161 с.

6. Duden. URL: https://www.duden.de/rechtschreibung/Korpus_Sammlung

7. Kontutyte E. *Einführung in die Fachsprachenlinguistik*. Vilnius : Vilniaus universiteto leidykla, 2017. 67 S.

8. Sharoff S. Creating General-Purpose Corpora Using Automated Search Engine Queries. Baroni, Marco; Bernardini, Silvia (eds.): WaCky! Working papers on the Web as Corpus. Bologna, 2006.

9. Пинтюк Є.С. Етимологічні характеристики інтернаціоналізмів і термінів іншомовного походження німецької фахової мови архітектури та будівництва. *Закарпатські філологічні студії*. 2021. Вип. 17. Том 1. С. 162-166. http://zfs-journal.uzhnu.uz.ua/archive/17/part_1/34.pdf.

10. Schmidt T., Wörner K. *Multilingual corpora and multilingual corpus analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 2012. 407 p.

11. P. Shrestha, Ch. Jacquin, B. Daille. Reduction of Search Space to Annotate Monolingual Corpora. Proceedings of 5th International Joint Conference on Natural Language Processing. Chiang Mai, Thailand. Asian Federation of Natural Language Processing, 2011. P.1457–1461.

12. B. Maia. Training Translators in Terminology and Information Retrieval using Comparable and Parallel Corpora. Corpora in Translator Education. ST JEROME PUBLISHING, 2003. P. 43-54.

13. P. Pym. Pre-editing and the use of simplified writing for MT. Translating and the Computer 10: The Translation Environment 10 Years. Aslib, London, 1990. P. 80–95.

14. Основи статистичної лінгвістики: Навчально-методичний посібник / Відп. ред. проф. Ф.С. Бацевич. Видавничий центр ЛНУ імені Івана Франка, 2008. 124 с.

15. Biber, D., S. Conrad, and R. Reppen. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press. 1998. 300 p.

16. Кащеева А. В. Квантитативные и качественные методы исследования в прикладной лингвистике. Социально-экономические явления и процессы. 2013. № 3(049). С. 155-162.

17. Статистичні методи для лінгвістів : посібник / Перебийніс В. І. Вид. 2, випр. і допов. Вінниця : Нова Книга, 2013. 176 с.

18. Глисон Г. Введение в дескриптивную лингвистику. М., 1959. 486 с.

19. Левицкий В. В. Квантитативные методы в лингвистике: учебное издание. Вінниця: Нова Книга, 2007. 264 с.

20. Грицюк П.М., Остапчук О.П. Аналіз даних: Навчальний посібник. Рівне: НУВГП, 2008. 218 с.

21. Кійко С. В., Пинтюк Є. С. Німецько-український термінологічний словник архітектури і будівництва: принципи розробки й укладання. *Науковий вісник Міжнародного гуманітарного університету. Серія: Філологія*. Одеса: Видавничий дім „Гельветика”, 2019. Вип. 43. С. 121-125.

Information about the author:

Rubana Yevheniia Stepanivna,

Postgraduate Student at the Department of Germanic, General and
Comparative Linguistics

Chernivtsi National University

Research Assistant at the Department of Variational Linguistics and
German as a Second / Foreign Language

University of Augsburg

2, Universitätsstraße, Augsburg, 86159, Germany