

УДК 81'33

РІЗНОВИДИ КОРПУСУ ТЕКСТІВ У ПРОЦЕСІ ПЕРЕКЛАДУ ДОКУМЕНТІВ ОФІЦІЙНО-ДІЛОВОГО СТИЛЮ

Дем'янчук Ю. І.

Львівський державний університет безпеки життєдіяльності

У статті висвітлено класифікацію корпусу текстів відповідно до їхніх лінгвостилістичних і семантико-синтаксичних ознак. У процесі зіставлення багатокomпонентної наукової класифікації вказується на доцільність застосування під час перекладу багатомовних законодавчих документів паралельного, суміжного та порівняльного корпусів тексту. Пропонується алгоритм пошуку опорних слів (термінів, лексико-синонімічного ряду, окремих морфологічних одиниць) у суміжному зіставленні багатомовних колокацій.

Ключові слова: корпусна лінгвістика, колокація, паралельний корпус тексту, суміжний корпус тексту, порівняльний корпус тексту, офіційно-ділові документи.

Дем'янчук Ю. И. Разновидности корпуса текстов в процессе перевода документов официально-делового стиля. В статье рассматривается классификация корпуса текстов в соответствии с лингвостиллистическими и семантико-синтаксическими признаками. В процессе сопоставления многокомпонентной научной классификации указывается на целесообразность применения во время перевода многоязычных законодательных документов параллельного, смешанного и сравнительного корпусов текста. Предлагается алгоритм поиска опорных слов (терминов, лексико-синонимического ряда, отдельных морфологических единиц) в смешанном сопоставлении многоязычных коллокаций.

Ключевые слова: корпусная лингвистика, коллокация, параллельный корпус текста, смешанный корпус текста, сравнительный корпус текста, официально-деловые документы.

Demianchuk Y. I. Text corpora types in official documents translation. The article considers the classification of text corpus in accordance with their stylistic and semantic-syntactical characteristics. In comparing of multicomponent scientific classification indicated the expediency of application of parallel, mixed and comparative corpora in the translation of multilingual legislative documents. Here also an algorithm of pivot words searching (terms, lexical-synonymic row, separate morphologic units) in mixed comparison of multilingual collocations has been proposed.

Key words: corpus linguistics, collocation, parallel text corpus, mixed text corpus, comparative text corpus, official documents.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями. Переклад документів офіційно-ділового стилю в системі корпусної лінгвістика набуває особливої **актуальності**, відповідно до розвитку технічних засобів в створення нових корпусів тексту. Найбільш важливими типами корпусів для перекладу офіційно-ділових документів є багатомовні суміжні корпуси текстів (multilingual corpora), паралельні корпуси текстів (parallel corpora) та порівняльні корпуси текстів (comparable corpora). Багатомовний корпус текстів поєднує кілька аналогічних за структурою одномовних корпусів текстів. Тому порівняльний корпус і його різновиди вміщують у себе одномовні субкорпуси, що мають подібну структуру й складені за однаковими принципами.

Отже, складання багатомовних корпусів тексту законодавчої тематики, за допомогою яких здійснюються дослідження окремих мовних одиниць і лексичних похибок, мотивує до вдосконалення лінгвістичного забезпечення різноманітних інформаційних систем.

Аналіз останніх досліджень і публікацій. Дослідженням лінгвістичних корпусів і їхніх лексико-граматичних особливостей займалися Т. В. Бобкова, А. А. Віландербек, Д. В. Січінава, Д. О. Добровольський та інші. Проте проблема стилістичних властивостей багатомовних корпусів, які б надавали можливість якісно опрацьовувати офіційно-ділові документи, залишається недослідженою.

Метою статті є вивчення видів багатомовних корпусів тексту, які сприяють якісному перекладу текстів офіційно-ділового стилю.

Згідно з метою дослідження, поставлено такі **завдання:** 1) визначити лінгвістичні та стилістичні особливості міжнародних офіційно-ділових текстів та особливості їх корпусного багатомовного перекладу; 2) розглянути класифікацію одномовних і багатомовних корпусів та її колокаційних змін; 3) у контексті параметрів і ретроспекції багатомовних корпусів тексту з'ясувати перспективу їхнього застосування в офіційно-діловому стилі.

Виклад основного матеріалу дослідження. Багатомовний корпус текстів – це великий, представлений в електронному вигляді на декількох мовах, уніфікований, структурований, розмічений, філологічно компетентний масив мовних даних, призначений для вирішення конкретних лінгвістичних завдань.

Таблиця 1

Класифікація корпусів відповідно до стилістичних та морфологічних ознак

Поділ за ознакою	Типи корпусів
Тип даних	Писемні Мовні Змішані
Мова текстів	Український Російський Англійський Німецький тощо.
Паралельність	Одномовні Двомовні Багатомовні
Художні властивості	Художні Діалектні Розмовні Термінологічні Змішані
Жанр тексту	Художні Фольклорні Драматичні Публіцистичні
Доступність	Доступні Комерційні Закриті
Призначення	Дослідницькі Ілюстративні
Динамічність	Динамічні (моніторні) Статичні
Розмітка	Розмічені Нерозмічені
Характер розмітки	Морфологічні Синтаксичні Семантичні Стилістичні
Обсяг текстів	Повнотекстові Фрагментарні
Хронологічний аспект	Синхронічні Діахронічні
Структура	Центральні й архівні Ядерні та периферійні

У поняття багатомовного корпусу текстів уходить також система управління кількома стилістичними й лінгвограматичними даними, які керуються корпусним менеджером (Corpus manager). Це спеціалізована пошукова система, яка вміщує програмні засоби для пошуку даних в окремому корпусі, отримання статистичної інформації та надання результатів (із похибками) користувачеві у зручній формі.

Дослідник Е. Е. Віладеберк визначає передумови й особливості використання багатомовних корпусів [2, 11], а саме: 1) досить великий (репрезентативний) обсяг корпусу гарантує типовість даних і забезпечує повноту уявлення всього спектра багатомовних явищ; 2) дані різного типу знаходяться в багатомовному корпусі в певній контекстній і стилістичній формі, що створює можливість їхнього усебічного й об'єктивного вивчення; 3) одноразово створений і підготовлений масив даних може використовуватися багаторазово, багатьма дослідниками, з метою порівняння та зіставлення колокацій.

Популярні одномовні корпуси тексту (Brown Corpus, Lancaster-Oslo-Bergen Corpus, LOB, British National Corpus, International Corpus of English, Bank of English) вирішують лінгвістичні завдання лише за наявності одномовного масиву текстів [8, 248]. Потрібно також, щоб тексти вміщували різну додаткову лінгвістичну й екстралінгвістичну інформацію. Важливу роль у таких текстах відіграє розмітка (tagging, annotation), яка полягає у позначенні спеціальних міток (tag, tags): зовнішніх, екстралінгвістичних (відомості про автора й відомості про текст: автор, назва, рік і місце видання, жанр, тематика; відомості про автора можуть включати не тільки його ім'я, а також вік, стать, роки життя тощо). До того ж необхідна вказівка на структурні (глава, абзац, пропозиція, словоформа) і власне лінгвістичні дані, що описують лексичні, граматичні та інші характеристики елементів тексту. Такі корпуси орієнтовані на художній, публіцистичний і науковий стилі.

Незважаючи на різноманітність корпусів, можна виділити два основних способи їхнього поділу на класи [11]: 1) зіставлення корпусів, що стосуються всієї мови (або мови певного періоду); 2) поділ корпусів, що належать до певної мовознавчої галузі (жанру, стилю, мови певної вікової або соціальної групи); 3) поділ корпусів за типом лінгвістичної розмітки (див. таблиця 1). Більшість з них усе ж таки належить до корпусів морфологічного або синтаксичного типу (так званого treebanks, «банку синтаксичних структур»). Натомість корпус із синтаксичною розміткою вміщує в себе й морфологічні характеристики лексичних одиниць.

Можна погодитися із С. Я. Вадяєвим, що корпус тексту, який становить добірку текстів предметної галузі знань, можна назвати спеціальним, і використовуватися він має з метою, для якої спроектований [1, 4]. Такі багатомовні корпуси підходять для перекладу наукових та офіційно-ділових текстів.

Проте під час дослідження спеціального корпусу варто враховувати, що спеціальний корпус не завжди може відображати об'єктивну дійсність; він призначений для використання лише в галузевих структурах. Тому створення корпусу офіційно-ділових

текстів базується на мінімальних ознаках: розташуванні на магнітному носії; наявності єдиної процедури відбору; репрезентативності; єдності розмітки або подання корпусу; кінцівки розміру корпусу.

Варто наголосити, що документальні тексти характеризуються вираженою книжково-письмовою мовою. У них використовуються стійкі (для тих чи інших їхніх різновидів) словосполучення й терміни, що відображають ту сферу діяльності, з якою пов'язані тексти. Така спільність функції зумовлюється великим ступенем спільності у стилі документальних текстів різними мовами. Ця спільність найяскравіше проявляється в дипломатичних документах міжнародного масштабу, в договорах та уго-

дах між країнами Європейського Союзу. Переклад офіційно-ділових матеріалів, які вміщують дипломатичні документи (статуту, тексти міжнародних угод, договори, заяви), урядові постанови, матеріали різних конференцій, вимагає від перекладача знання характерних особливостей їхньої побудови.

Для офіційних документів характерна чіткість побудови, що забезпечує логічну послідовність викладу, відокремлення однієї думки від іншої. Зазвичай великі за розміром офіційно-ділові документи діляться на розділи (sections); підрозділи (subsections); глави (chapters); статті (articles); пункти, параграфи (clauses, items, points, counts, paragraphs); підпункти (subparagraphs).

Такі терміни, як «*clause, item, point, count, paragraph*», часто не мають чіткої диференціації значень, і під час перекладу міжнародних документів необхідно лише дотримуватися того, щоб до одного й того самого англійського (німецького) терміна був український (російський) відповідник. Тому, перш ніж розпочати переклад офіційного документа, перекладач повинен виписати з нього всі англійські терміни для позначення поділу тексту й вибрати для них українські (російські) відповідники. Крім того, в англійському тексті офіційних документів часто використовуються алфавітні зображення розділів або пунктів: a), b), c), d) тощо.

Під час роботи з офіційно-діловими документами потрібно враховувати точність перекладу, оскільки на основі достовірного перекладу часто приймаються важливі політичні рішення, забезпечується автентичність текстів договорів, угод. Указане нами правило стосується не лише адекватної передачі змісту документа, а й конотацій. Наприклад, особливість вступної частини офіційно-ділового документа (преамбули) – це максимально стисла форма викладення змісту документа, внаслідок чого в одному реченні чи декількох колокаціях викладається більше п'яти законодавчих термінів. Цей аспект створює появу всередині однієї пропозиції численних інфінітивних оборотів.

Паралельний корпус можна використовувати для вирішення багатьох проблем, пов'язаних із перекладом офіційних документів, а саме: 1) виявлення типових перекладацьких прийомів і трансформацій конотацій документів; 2) навчання статистичних систем автоматичного перекладу міжнародних текстів (наприклад, заходів, Google Translate); 3) створення одномовних і багатомовних словників юридичної термінології; 4) підбір можливих еквівалентів.

Розглянувши дослідження Д. Е. Січінаєва, можемо стверджувати, що параметризація паралельного корпусу офіційно-ділових текстів уміщує такі виміри [6, 320]: 1) широту (охоплення максимальної кількості ділових текстів першого рівня); 2) глибину (виокремлення підкатегорій на нижчих рівнях); 3) пропорційність (заповнення окремих «клітинок» офіційними текстами, дібраних у певній пропорції); 4) часовий вимір.

Головними ж підходами, методами та критеріями до здійснення параметризації корпусу є такі: орієнтування на авторитетний корпус; параметризація на

основі фахової оцінки; параметризація із застосуванням формальних методів; параметризація з посиланням на списки джерел; параметризація з позицій отримувачів тексту; параметризація на основі внутрішньомовних критеріїв; параметризація відповідно до потреб користувачів; динамічний підхід; циклічний підхід; використання ресурсів Інтернету; метод стратифікації; метод пропорційної вибірки; метод випадкової вибірки; метод цілеспрямованого доповнення; доповнення спеціалізованими корпусами; критерій культурної значущості тексту тощо.

Реалізують усі критерії й параметри якісного перекладу офіційно-ділових документів, на нашу думку, Corpus of Contemporary American English, COCA та ПКТ «DER» [1; 9]. Зокрема, параметризацію Корпусу сучасної американської англійської мови (Corpus of Contemporary American English, COCA) уклав М. Дейвіс [9, 457], і це єдиний моніторинговий корпус англійської мови. Пропорційність дотримано щодо жанрів, підкатегорій текстів і часу. Щороку до корпусу додається приблизно 4 мільйони слововживань до кожного жанру, й наразі на кожен із них припадає по 80 мільйонів слововживань. Корпус удає поєднує параметризацію за зовнішніми критеріями, пропорційність, динамічний підхід, уможливає порівняння між кількома корпусами, втілює загальну тенденцію до репрезентативності й збалансованості, що дає йому незаперечну перевагу над менш систематично організованими великими корпусами.

Паралельний корпус «DER» може слугувати як основа та інструментарій для подальших стилістичних, лексикологічних, граматичних і порівняльних досліджень німецькомовних документів. Корпус дає змогу проводити порівняльний аналіз особливостей німецької та російської (української) мов [1, 7–8], що стосуються лексичного складу, морфології й синтаксису (статистичний аналіз і порівняльний аналіз лексичних структур і синтаксичних конструкцій у паралельних текстах); здійснювати пошук еквівалентів для слів і колокацій, а також прикладів уживання лексичних одиниць та їх перекладу іншою мовою, залежно від контексту вживання; проводити дослідження у галузі стилістики (розгляд способів передачі стилістичних особливостей у текстах різними мовами, а також визначення стилістичних засобів і прийомів за допомогою яких перекладач передає зміст оригінального тексту); розробляти й удосконалювати програмні засоби для сортування текстів, самостійного пошуку та вилучення еквівалентів для окремих лексем і колокацій.

Найважливішим лінгвістичним принципом побудови цього корпусу текстів є забезпечення його репрезентативності. Значне збільшення обсягу досліджуваного корпусу не означає пропорційного зростання в ньому кількості унікальних лексем (тобто автоматичного підвищення репрезентативності). Динаміка зростання кількості словоформ зменшується разом зі зростанням кількості слововживань у корпусі, тому репрезентативність корпусу повинна забезпечуватися не стільки за рахунок збільшення його обсягу, скільки за рахунок більш ретельної вибірки текстів.

Статистичні дані про структурні елементи тексту можуть стати надійною основою для виявлення паралельних сегментів у текстах німецькою, англійською, російською та українською мовами й використовуватися під час розробки паралельних корпусів текстів. Оскільки міжнародні документи Європейського Союзу та НАТО посідають особливе місце серед політичних типів текстів у складі офіційно-ділового стилю, то їх лексичний склад зумовлений зовнішньополітичним характером, інтернаціоналізацією німецької й англійської мов.

Між загальною комунікативно-функціональною спрямованістю тексту й основними засобами вираження модальності існує безпосередня залежність, що виражається для модальних дієслів німецької мови в зміні їх частотно-розподільчих характеристик залежно від функціонально-стильової належності тексту. Мовний фон уживання окремих модальних дієслів німецької мови, частотність і вибір синтаксичних конструкцій для їх уживання тісно пов'язані із загальною комунікативно-функціональною спрямованістю текстів.

Загалом у розглянутих паралельних корпусах тексту (Corpus of Contemporary American English і паралельний корпус тексту «DER») виділяється низка підтипів: тексти мовою А і їх переклади мовою В; тексти мовами А і В і їх переклади, відповідно, мовами В й А; лише перекладні тексти мовами А, В, С, ..., Х, якщо оригінальні тексти були написані мовою D. Також існує поділ на «гучні» ПКТ (noisy

parallel corpora; з пропусками в перекладі, без точної відповідності між оригіналом і перекладом), «дзеркальні» ПКТ (mirror corpora), що складаються з текстів мовами А і В і перекладів цих текстів, відповідно, мовами В й А.

Інший дослідник Д.О. Добровольський указує на доцільність використання з метою перекладу міжнародних офіційно-ділових документів російсько-німецького корпусу паралельних текстів, який створений Австрійською академією наук у рамках проекту «Корпус Австрійської академії» (ААС), для дослідження динаміки лексичних зв'язків і розгляду залежності «між властивостями слова та його семантичною структурою» [3, 272]. На думку дослідника, зіставлення слова з його перекладними еквівалентами, що наявні в корпусі, «допомагає розкрити нетривіальні семантичні особливості відповідних лексем обох мов» (якщо це стосується німецькомовних документів) [3, 284].

Висновки та перспективи подальших досліджень. Паралельний корпус тексту є універсальним засобом не тільки для порівняльних досліджень кількох мов (наприклад, німецько-англійсько-російсько-української) і пошуку еквівалентів, але й для досліджень кожної мови корпусу окремо. У цьому разі паралельний корпус слугує так званою «точкою перетину» двох або більше мовних культур, складаючись із декількох субкорпусів. Паралельний і похідні від нього багатомовні корпуси дають змогу якісно визначити належність тексту до певної предметної сфери і стилю.

ЛІТЕРАТУРА

1. Вадяев С. Е. Лингвистические принципы построения и использования корпуса текстов для исследования официально-делового стиля современного немецкого языка (на материале электронного корпуса «DER») : автореф. дисс. ... канд. филол. наук : спец. 10.02.04 / С. Е. Вадяев ; Нижегород. гос. лингвист. ун-т им. Н.А. Добролюбова. – Нижний Новгород, 2005. – 22 с.
2. Виландеберк А. А. Принципы и методы гармонизации терминологии на основе корпуса специальных параллельных текстов (на материале документов ООН) : автореф. дисс. ... канд. филол. наук / А.А. Виландеберк. – СПб., 2005. – 22 с.
3. Добровольский Д. О. Корпус параллельных текстов: архитектура и возможности использования / Д. О. Добровольский, А. А. Кретов, С. Шаров // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. – М. : Индрик, 2005. – С. 263–296.
4. Гурина Е. Е. Асимметрия членения текста в параллельном корпусе (на материале романа Т. Драйзера «Сестра Кэрри») / Е. Е. Гурина, Д. О. Добровольский, А. А. Кретов // Вестник Воронежского гос. университета. Серия «Лингвистика и межкультурная коммуникация». – 2005. – № 1. – С. 67–74.
5. Лингвистические ресурсы автоматизированного рабочего места филолога : [монография] / [Л. Н. Беляева, А. А. Виландеберк, Л.А. Девель, И.Н. Ларченков, С.В. Молчанова, В.Р. Нымм, Т.Н. Петрова-Маслакова]. – СПб. : Инфо-да, 2004. – 181 с.
6. Сичинава Д. В. Комплексное исследование одноязычного и параллельного корпусов в грамматических исследованиях / Д. В. Сичинава // Труды Международной конференции «Корпусная лингвистика-2011». – СПб., 2011. – С. 316–332.
7. Сичинава Д. В. Паралельні українсько-російський та російсько-український корпуси / Д. В. Сичинава, О. О. Тищенко-Монастирська, М. О. Шведова // Лексикографічний бюлетень, 20. – К., 2011. – С. 35–38.
8. Biber D. Representativeness in corpus design / D. Biber // Literary and Linguistic Computing. – 1993. – № 8 (4). – P. 243–257.
9. Davies M. The Corpus of Contemporary American English as the first reliable monitor corpus of English / M. Davies // Literary and Linguistic Computing. – 2010. – Vol. 25. – № 4. – P. 447–464.
10. Sinclair J.M. Corpus typology – A framework for classification / J.M. Sinclair // Studies in anglistics. – Stockholm : Almqvist & Wiksell, 1995. – P. 17–33.
11. Баранов А. Н. Проблема репрезентативности корпуса текстов / А. Н. Баранов // Диалог : тр. Междунар. конф. [Электронный ресурс]. – Режим доступа : <http://www.dialog-21.ru/archivearticle>.