

4. Meibauer, J. (2001). *Pragmatik*. Tübingen: Starffenburg-Verl. 208 s.

С. А. Матвєєва
доктор філологічних наук, доцент
професор кафедри прикладної лінгвістики,
порівняльного мовознавства та перекладу
НПУ імені М. П. Драгоманова

КОРПУС ТЕКСТІВ ЯК ІНСТРУМЕНТ ОБРОБКИ МОВНОГО МАТЕРІАЛУ

Останнім часом особливу увагу дослідників різних галузей лінгвістичної науки привертають питання формалізації мовних явищ та одиниць всіх рівнів із метою створення універсальних ресурсів зберігання, обробки, аналізу та використання мовного матеріалу. Одним із результатів розвитку автоматизованої обробки інформації стала поява нового напрямку мовознавства – *корпусної лінгвістики*. Сьогодні дані корпусів активно використовуються в багатьох галузях гуманітарного знання, як то лексикографія, перекладознавство, соціолінгвістика, педагогіка тощо, а вивченню теоретичних і прикладних аспектів корпусної лінгвістики приділяється велика увага в світовому мовознавстві, що говорить на користь особливої **актуальності** зазначеного напрямку. **Метою** цієї роботи є окреслення основних шляхів використання ресурсів корпусної лінгвістики при роботі з мовним матеріалом.

За твердженням В. В. Жуковської, корпусна лінгвістика є “галуззю прикладного мовознавства, що займається визначенням загальних принципів побудови, обробки та експлуатації даних лінгвістичних корпусів (або корпусів текстів) із використанням сучасних комп’ютерних технологій, розробленням методики збору реальних мовних явищ – писемних та усних текстів, а також способів їх збереження та аналізу” (Жуковська, 2013).

Мовний корпус є однією з найсучасніших технологій, яка відкриває широкий спектр теоретичних та практичних можливостей дослідження. Це зумовлюється тим, що мовознавство як наука шукає нові шляхи підвищення

об'єктивності й достовірності результатів, які вона отримує при обробці мовних даних. У цьому відношенні саме корпусно-базований підхід дозволяє максимально швидко обробити великий обсяг інформації. Більше того, використання цифрових корпусів мінімізує суб'єктивний вплив дослідника на отримані результати та надає підґрунтя для виявлення й аналізу раніше неврахованих мовних явищ. Такий антропокосмічний вектор сучасної лінгвістики (Леміш, 2017) спрямовує наукові дослідження на перспективний шлях вивчення корпусу як такого, включаючи створення різних типів текстових корпусів (Н. П. Дарчук, Ю. І. Дем'янчук, Н. Є. Леміш та ін.).

Під корпусом текстів розуміється певним чином організоване електронне зібрання письмових і усних текстів природної мови, якому притаманні обов'язкові ознаки і яке призначене для вирішення конкретних лінгвістичних завдань (Демська, 2011). Визначаючи поняття “корпус” в мовознавстві, ми погоджуємось із І. В. Мейзерською, яка вважає корпус складною цілісною системою зі своєю логічною структурною організацією та диференціальними ознаками” (Мейзерська, 2014). Отже, у загальному розумінні лінгвістичний корпус – це представлений у цифровому форматі, великий за обсягом, уніфікований, структурований, розмічений і філологічно компетентний масив текстів природною мовою, доповнений системою керування – універсальними програмними засобами для пошуку та опрацювання різноманітної лінгвістичної інформації.

Новий підхід до систематизації й обробки текстів вимагає розроблення нової методології аналізу. І. В. Мейзерська зазначає, що текстові принципи аналізу не працюють для такої системи. На її думку, “корпус укладається на основі великої репрезентативної групи текстів, оброблених таким чином, що мовний матеріал розміщується за принципом конкордансу” (Мейзерська, 2014). При укладанні корпусу кожен текст маркується та поділяється на фрагменти, що ілюструють одиниці і групуються у вибірку контекстів. На відміну від звичайних текстів, що характеризуються лінійною, горизонтальною організацією, жанровою та стильовою специфікою, авторською

індивідуальністю, структурною та смисловою цілісністю, наявністю певної комунікативної мети, корпус не має спільної комунікативної мети (комунікативна мета простежується лише на рівні одного речення). Контексти в корпусі, “представлені в конкордансі, засвідчують функціонування мовної одиниці в різних стилях, сферах уживання, даючи широку картину реальної мовної практики” (Мейзерська, 2014). Таким чином, застосування корпусного підходу надає можливість вивчення будь-яких мовних одиниць у різних мовленнєвих жанрах, а також дозволяє визначити специфіку використання цих одиниць у різних видах дискурсу.

Частіше лінгвістичні корпуси використовують у морфологічних, лексикологічних та синтаксичних дослідженнях. Так, наприклад, корпусні бази дозволяють здійснювати пошук та статистичний підрахунок вживання заданих коренів, афіксів і флексій, завдяки чому вивчаються способи творення мовної одиниці. Лінгвістичні корпуси дозволяють отримати дані про конкретні форми слова і про цілі граматичні категорії (наприклад, системно-структурне дослідження морфології дієслова дозволяє виявити обсяг реалізації дієслівної парадигми та особливості використання дієслівних форм у сучасних дискурсах; категорії виду та видових протиставлень в сучасних текстах; категорії часу та явищ транспозиції тощо).

Головною передумовою для того, щоб корпус став корисним інструментом, є його відповідність дослідницькому завданню. Більше того, він повинен мати якомога глибшу розмітку, розгалужену анотацію для надання досліднику чи перекладачу всіх необхідних для досягнення цілей даних. Класифікація корпусів може відбуватися за різними показниками й параметрами, наприклад:

- корпуси повних текстів або фрагментів;
- дослідницькі корпуси: ілюстративні чи інтерпретаційні;
- письмові, розмовні або змішані корпуси;
- одномовні або багатомовні корпуси;
- синхронічні або діяхронічні корпуси;

- статичні або динамічні корпуси;
- збалансовані або моніторингові корпуси;
- паралельні або порівняльні корпуси;
- малі, середні або великі корпуси,
- та інші.

Ю. І. Дем'янчук пропонує класифікацію корпусів за стилістичними та морфологічними ознаками:

Таблиця 1

Класифікація корпусів текстів

відповідно до стилістичних та морфологічних ознак (Дем'янчук, 2016)

<i>Ознака</i>	<i>Тип корпусу</i>
Тип даних	Писемні / мовні / змішані
Мова текстів	Український / англійський / 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 / 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

В узагальненому вигляді процедура корпусного аналізу включає в себе три основні кроки: ідентифікацію мовних даних за допомогою категоріального аналізу, співвідношення мовних даних за допомогою статистичних методів та інтелектуальну інтерпретацію результатів (Жуковська, 2013). Оскільки корпусні дослідження базуються переважно на емпіричному підході до аналізу мовного матеріалу, це дозволяє досягнути максимальної об'єктивності вивчення мови, унеможливаючи або мінімізуючи суб'єктивні погляди

дослідника. Корпусно-базований аналіз є унікальним інструментом вивчення мови, завдяки якому можна здійснювати пошук у великих текстових масивах, одержувати дані про мовні одиниці та явища різних мовних рівнів (фонетичного, морфологічного, лексико-семантичного та синтаксичного): досліджувати частоту словоформ, лексем, граматичних категорій, синтаксичних конструкцій, визначати нетипові граматичні явища та конструкції, встановлювати найближче лексичне та граматичне оточення слова, за допомогою якого можна проаналізувати вживання слова в усіх характерних для нього колокаціях, колігаціях та синтаксемах.

Висновки. Таким чином, корпусно-базований підхід до аналізу мовного матеріалу демонструє широкий спектр можливостей дослідження окремих мовних одиниць і цілих текстів (від формальних компонентів до комунікативно-прагматичних особливостей). **Перспективами** подальших розвідок є напрацювання методологічного апарату корпусної лінгвістики в перекладознавчих дослідженнях та розроблення практичного інструментарію для використання корпусів перекладачами різних мов.

Список використаних джерел

1. Демська, О. М. (2011). *Текстовий корпус: ідея іншої форми* (с. 96). К.: НаУКМА.
2. Дем'янчук Ю. І. (2016). Різновиди корпусу текстів у процесі перекладу документів офіційно-ділового стилю. *Науковий вісник ДДПУ імені І. Франка. Серія "Філологічні науки". Мовознавство. № 5* (с. 105).
3. Жуковська, В. В. (2013). *Вступ до корпусної лінгвістики* (с. 9, 20). Житомир: Вид-во ЖДУ ім. І. Франка.
4. Леміш, Н. Є. (2017). Корпусна метарозмітка спеціальних текстів з лінгвоантропогенезу. *Науковий часопис національного педагогічного університету імені М.П. Драгоманова. Серія 9. Сучасні тенденції розвитку мов. Вип. 15* (с. 139, 143).
5. Мейзерська, І. В. (2014). Корпусний підхід у сучасній лінгвістиці:

перспективи і можливості застосування. *Науковий вісник кафедри Юнеско КНЛУ Серія Філологія. Педагогіка. Психологія*. Вип. 28 (с. 56, 57).

Н. В. Маштакова,
кандидат філологічних наук, доцент,
доцент кафедри прикладної лінгвістики,
порівняльного мовознавства та перекладу
НПУ імені М. П. Драгоманова

ПРИНЦИПИ ЛІНГВОКУЛЬТУРНОГО ТА КОГНІТИВНОГО ПІДХОДІВ ДО ВИВЧЕННЯ ПРИСЛІВ'ІВ ІЗ СЕМАНТИКОЮ ПРИКМЕТ

Лінгвокультурологічний підхід до вивчення мовних явищ базується на відомій теорії лінгвальної відносності Сепіра-Уорфа й передбачає вивчення мовного відображення культури певного народу.

В. фон Гумбольдт, вважаючи мову світом, що заповнює простір між внутрішнім світом людини й зовнішнім об'єктивним, наголошував: “Мова є ніби зовнішній прояв духу народів: мова народу є його дух, і дух народу є його мова, і важко уявити собі що-небудь більш тотожне” (Гумбольдт, 1984).

Особливо **актуалізувалися** лінгвокультурологічні дослідження наприкінці ХХ століття й пов'язані в слов'янському мовознавстві з працями В. Н. Телія, В. В. Воробйова, Н. Д. Арутюнової, Ю. С. Степанова, В. А. Маслової.

Метою нашого дослідження є опис лінгвокультурного та когнітивного підходів до вивчення прислів'їв із семантикою прикмет.

Так, В. А. Маслова пропонує три взаємодоповнюючі визначення лінгвокультурології, як-от: а) “галузь лінгвістики, що виникла на стику лінгвістики і культурології”; б) “гуманітарна дисципліна, що вивчає втілену в живу національну мову і виявлену в мовних процесах матеріальну й духовну культуру”; в) “інтеграційна галузь знань, що увібрала результати досліджень з культурології й мовознавства, етнолінгвістики й культурної антропології” (Культурологія, 1998).