

Анастасія КОЗАКЕВИЧ
Житомирський державний університет
імені Івана Франка
Науковий керівник:
канд. філол. наук, доцент, Чумак Л.М.

**ЗАСТОСУВАННЯ КОРПУСНОГО МЕНЕДЖЕРУ AntConc ДЛЯ
УКЛАДАННЯ ТЕРМІНОЛОГІЧНОГО ГЛОСАРІЮ В
КОМП'ЮТЕРНІЙ ТА ПРИКЛАДНІЙ ЛІНГВІСТИЦІ**

Постановка проблеми. У сучасній лінгвістиці швидке та точне укладання термінологічних глосаріїв має важливе значення для збереження консистентності термінології в наукових текстах. Ця проблема є особливо актуальною в умовах великих текстових корпусів, де кількість спеціалізованих термінів постійно зростає. Наприклад, у сфері комп'ютерної лінгвістики постає питання ідентифікації нових термінів, що виникають у зв'язку з розвитком нових

технологій і методологій (наприклад, “штучний інтелект, “обробка природної мови, машинне навчання). Традиційні аналітичні методи аналізу текстової інформації стають занадто трудомісткими та неефективними. Автоматизовані інструменти, такі як, наприклад, AntConc, можуть значно прискорити процес роботи з термінологічною лексикою, дозволяючи дослідникам обробляти великі обсяги даних і точно акцентувати ключові терміни.

Аналіз останніх досліджень і публікацій. Поряд із роботами В. В. Жуковської (2013), Д. Бибера та С. Конрада (1998), вивчення застосування корпусної лінгвістики розвивається також завдяки дослідженням таких авторів, як І. М. Сінкевич та Д. М. Лейхтман (2020), які досліджували можливості використання корпусних інструментів для аналізу спеціалізованих текстів у технічній термінології. Науковці вказують, що аналіз співвживань термінів є надзвичайно важливим для точного розуміння та перекладу спеціалізованої лексики. Крім того, дослідження П. Майєра (2014) про важливість контекстуальних даних у термінологічному аналізі підтверджують, що використання корпусів може забезпечити глибше розуміння значення термінів у специфічних галузях.

Обґрунтування актуальності дослідження. Застосування інструменту AntConc при аналізі спеціалізованих текстових корпусів надає можливість автоматизованого аналізу термінів, що є особливо актуальним для сучасної комп'ютерної лінгвістики. Наприклад, розвиток технологій машинного навчання та штучного інтелекту призвів до появи нових термінів і понять, які вимагають систематизації. Таке дослідження стає актуальним, оскільки пропонує інноваційні рішення для укладання термінологічних глосаріїв, що відповідають сучасним вимогам наукового прогресу.

Формулювання мети дослідження. Основною метою проведення дослідження є встановлення та обґрунтування можливостей і меж використання корпусного менеджера AntConc для автоматизованого створення термінологічних глосаріїв. Особлива увага приділяється аналізу частотності термінів, їх співвживань, контекстуального використання та лемматизації, що дозволить забезпечити більш точне розуміння термінологічної структури в комп'ютерній та прикладній лінгвістиці.

Виклад основного матеріалу з обґрунтуванням отриманих наукових результатів. Огляд можливостей AntConc. AntConc дозволяє дослідникам:

- Проаналізувати частотність слів. За допомогою частотних списків можна швидко ідентифікувати ключові терміни у корпусі, що є основою для подальшого аналізу.

- Здійснити Пошук співвживань (collocations). Визначення частотних співвживань допомагає зрозуміти, як терміни використовуються разом з іншими словами. Це корисно для аналізу фразеологізмів, сталих виразів або специфічних термінологічних поєднань. Наприклад, у комп'ютерній лінгвістиці термін "модель" часто співживається з такими словами, як "дані", "семантичний" та "структура".

- Застосувати морфологічний аналіз Лемматизації в алгоритмах пошукових систем. AntConc дозволяє аналізувати різні форми одного й того самого слова (лемми), що важливо для точності термінологічного аналізу, особливо в мовах зі складною морфологією.

- Упорядкувати в алфавітному переліку всіх слів якого-небудь тексту з наведенням контекстів їх використання, тобто у відповідності до типу словника Конкорданс (Keyword-in-Context, KWIC). Цей особливий тип словника Дозволяє досліднику бачити конкретні приклади використання терміна у тексті, що допомагає аналізувати його контекст і визначати точне значення у різних ситуаціях.

Практичне застосування. Для аналізу термінологічної лексики було обрано корпус текстів із журналів та конференцій з комп'ютерної лінгвістики (приблизно 500 тисяч слів). Було застосовано інструмент AntConc для виділення ключових термінів, їх співживань та аналізу лемматизації. Основні результати дослідження включають ідентифікацію таких термінів, як "алгоритм", "корпус", "обробка природної мови", "машинне навчання". Співживання цих термінів вказувало на важливі зв'язки між ними, наприклад, "корпус даних", "алгоритм обробки". Це дало змогу побудувати глосарій, що включає терміни з їхніми визначеннями, прикладами використання та контекстуальними поєднаннями.

Висновки. AntConc є незамінним інструментом для укладання термінологічних глосаріїв завдяки здатності автоматизувати важливі аспекти корпусного аналізу. Інструменти частотного аналізу, пошуку співживань й вивчення контексту забезпечують точне і ефективно використання термінологічних одиниць відповідних сфер діяльності людини. Обмеженням може бути необхідність ручного спеціального втручання дослідника в процес виділення термінів, але це компенсується можливістю гнучкого налаштування аналізу для різних типів текстів.

Список використаних джерел:

1. Жуковська В. В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир: Вид-во ЖДУ ім. І. Франка, 2013. 140 с.
2. Майер П. Контекстуальний аналіз термінів у корпусах технічних текстів. Технічна термінологія. №4, 2014. С. 23-35.
3. Сінкевич І. М., Лейхтман Д. М. Корпусний підхід до аналізу технічної термінології: Проблеми і перспективи. Мовознавство, №3, 2020. С. 45-60.
4. Anthony L. AntConc (Version 4.0.11) [Computer Software]. Waseda University. 2022. URL: <https://www.laurenceanthony.net/software/antconc/>
5. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press. 1998. URL: <http://dx.doi.org/10.1017/CBO9780511804489>
6. McEnery T., & Hardie A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press. 2011. 312 p.