

УДК 81:811.111

**Суходоля Маргарита***(м. Острозьк, Україна)**Національний університет «Острозька академія»***ЗАСТОСУВАННЯ КОРПУСНОЇ ЛІНГВІСТИКИ З МЕТОЮ  
АНАЛІЗУ АНГЛОМОВНИХ ТЕРМІНІВ НА ПОЗНАЧЕННЯ КАВИ**

*Тези присвячено опису параметризації та створення моделі корпусу текстів про каву та процеси її виготовлення. Описано етапи роботи з корпусом, а саме: розмітку та анотування. Окреслено переваги використання корпусної лінгвістики у процесі лінгвістичних досліджень.*

**Ключові слова:** *корпусна лінгвістика, корпус текстів, розмітка тексту, анотування.*

*Тезисы посвящены описанию параметризации и создания модели корпуса текстов о кофе и процессах его изготовления. Описаны этапы работы с корпусом, а именно: разметка и аннотирование. Очерчены преимущества использования корпусной лингвистики в процессе лингвистических исследований.*

**Ключевые слова:** *корпусная лингвистика, корпус текстов, разметка текста, аннотирование.*

*The research deals with the description of parameterization and creation of the corpus model of texts about coffee and its production. The stages of work with the corpus namely marking and annotation are described. Advantages of using Corpus Linguistics in the process of linguistic research are outlined in the article.*

**Key words:** *corpus linguistics, corpus of texts, text marking, annotation.*

З розвитком комп'ютерних технологій виникли нові шляхи обробки та зберігання мовного матеріалу. Це стало передумовою виникнення у мовознавстві такої галузі як корпусна лінгвістика. Розпочавши шлях свого становлення у 60-х роках ХХ ст., корпусна лінгвістика за відносно короткий проміжок часу перетво-

рилась у один з найбільш пріоритетних напрямів у мовознавстві.

У наш час корпуси текстів набувають все більшого значення у вивченні мов. Вони широко використовуються як в лінгвістиці, так і за її межами. Наприклад, лінгвісти-теоретики використовують корпуси з метою підтвердження або спростування своїх гіпотез. Прикладні лінгвісти знаходять застосування корпусів у викладанні іноземних мов та вирішенні інших професійних завдань. Врешті, корпуси відіграють важливу роль у розробці та налаштуванні різних автоматизованих систем (машинний переклад, інформаційний пошук, розпізнавання мовлення).

Вивченням корпусної лінгвістики займалися Дж. Синклер, Д. Байбер, Х. Кучера, Г. Кеннеді, Н. Іде, Т. Макенері, А. Вілсон, С. Шарова, М. Банько, А. Баранова, В. Риков. З огляду на важливість та актуальність корпусів у сучасній лінгвістиці, ми також створили корпус який репрезентує такий напрям харчової промисловості, як кава та процеси її виготовлення. З допомогою новоствореного корпусу ми матимемо змогу проаналізувати терміни безпосередньо у контексті. У цьому і полягає актуальність об'єкта дослідження – англomовних текстів про каву та процеси її виготовлення, які увійшли до корпусу.

Предметом дослідження є параметри спеціалізованого корпусу текстів про каву та процеси її виготовлення. Мета статі – опис параметризації та створення моделі корпусу текстів про каву та процеси її виготовлення та визначення переваги використання корпусної лінгвістики у процесі лінгвістичних досліджень. Постановлена мета статті зумовила вирішення таких завдань: 1) проаналізувати процес параметризації корпусу текстів про каву; 2) описати розмітку та анутовання корпусу; 3) визначити значення корпусів в лінгвістиці.

Корпусна лінгвістика як галузь прикладного мовознавства займається визначенням загальних принципів побудови, обробки та експлуатації даних лінгвістичних корпусів (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 [1].

Об'єктом дослідження корпусної лінгвістики є корпус текстів. Науковці по-різному підходять до визначення цього терміну. Мейер, наприклад, подає наступне визначення поняття «корпус»: організована певним чином словесна єдність, елементами якої є цілі тексти чи спеціальним чином відібрані уривки з текстів, що доступні для лінгвістичного аналізу [4]. Френсіз трактує поняття «корпус» як зібрання текстів, яке вважається репрезентативним стосовно даної мови, діалекту або іншої ділянки мови й призначене для використання в лінгвістичних дослідженнях [3].

Для створення корпусу текстів про каву та процеси її виготовлення було здійснено вибірку текстів різного типу, що задовольнили вимоги до репрезентативності тексту. Наступним етапом стало визначення кількості слів, що увійшли до окремих частин підрозділу. Дотримуючись заданих параметрів ми обрали 10 текстів з теми «Кава та процеси її виготовлення» із загальною кількістю 12 000 слів. Оскільки усі тексти різні, вони повинні були містити приблизно однакову кількість слів. Зробивши обчислення, ми визначили, що середня кількість слів для кожної групи текстів – 4000.

Було обрано наступні жанри текстів:

1. Довідкова література / REFERENCE (словники, глосарії) – у корпусі вони умовно позначатимуться REFERENCE\_Dictionary; REFERENCE\_Glossary

2. Навчальна література / EDUCATION (підручники, інструкції) – у корпусі вони умовно позначатимуться EDUCATION\_Handbook; EDUCATION\_Instruction

3. Наукова література / SCIENCE (статті, монографії, тези, дисертації) – у корпусі вони умовно позначатимуться SCIENCE\_Article; SCIENCE\_Monograph; SCIENCE\_Thesis; SCIENCE\_Dissertation.

Наступним етапом нашої роботи стало анотування обраних текстів, що увійшли до корпусу.

Анотування – процес опрацювання інформації, мета якого – отримання узагальненої характеристики документа, що розкриває логічну структуру і зміст. Анотація дозволяє користувачеві скласти достатнє й об'єктивне попереднє уявлення про незнайо-

му для нього наукову публікацію і тим самим допомагає в пошуку, відборі та систематизації потрібної інформації [2, с. 553–554]. Анотування тексту, використане нами, носить довідковий характер, тобто у нього входять відомості про автора, назву, заголовки та власне сам текст.

Для позначення розмітки ми використали такі теги:

<title>...</title>; <header>...</header>; <author>...</author>;  
<text>...</text>;

<p>...</p> (на позначення параграфу); <headword>...</headword>;  
<definition>...</definition>; <illustration>...</illustration>

Створений нами корпус може широко використовуватись в лінгвістиці, включаючи найновітніші її напрямки. Його призначення – показати функціонування лінгвістичних одиниць в їх природньому контекстному оточенні. На основі корпусу можна отримати дані про частоту словоформ, лексем та граматичних категорій.

Отже, створення корпусів текстів дає змогу розширити і автоматизувати аналіз мовного матеріалу, який відіграє ключову роль у лінгвістичному дослідженні будь-якого виду. Комп'ютерні програми дозволяють швидко відшукати потрібні приклади з корпусів текстів, які зберігаються в електронному вигляді. Як наслідок, з'являється можливість зекономити значну кількість часу порівняно зі збиранням матеріалу вручну.

### **Література:**

1. Жуковська В. В. Вступ до корпусної лінгвістики: Навчальний посібник / В. В. Жуковська. – Житомир : Вид-во ім. І. Франка, 2013. – 142 с.
2. Шевчук С. В. Українська мова за професійним спрямуванням / С. В. Шевчук, І. В. Клименко. – Київ : Алерта, 2011. – 696 с.
3. Francis W. Language Corpora B. C. / W. Francis // In: Jan Svartvik (ed.) Directions in Corpus Linguistics: Proceedings of Nobel 82, Stockholm, 4–8. August, 1991. – Berlin – New York : Mouton de Gruyter. – P. 17–35.
4. Meyer Ch. P. English Corpus Linguistics. An introduction / Ch. P. Meyer. – Cambridge University Press, 2004. – 168 p.