

УДК 81'33

ББК 81

## ЕТАПИ МЕТАРОЗМІТКИ КОРПУСУ АНГЛОМОВНИХ ТЕКСТІВ З ЛІНГВОКОГНІТОЛОГІЇ

ШКЛЯРЕВСЬКИЙ В.Г.

*Київський національний лінгвістичний університет*

У статті аналізуються етапи укладання корпусу текстів з лінгвокогнітології, в основу якого покладено ключові вимоги до корпусних ресурсів. Установлено основні принципи метарозмітки зазначеного корпусу англomовних текстів. Запропоновано методику метарозмітки корпусу за допомогою когнітивного мапування текстів.

**Ключові слова:** текст, корпус текстів, метарозмітка, когнітивне мапування, лінгвокогнітологія.

This paper examines the stages of compiling a text corpus in cognitive linguistics according to the key requirements to corpus resources. The main principles of encoding such a corpus of texts in English are determined. The methodology of corpus encoding by the technique of cognitive text mapping is suggested.

**Key words:** text, corpus of texts, encoding, cognitive mapping, cognitive linguistics.

Сучасна когнітивна корпусна лінгвістика як галузь когнітології характеризується тенденцією до розширення міждисциплінарних зв'язків і ефективного застосування корпусного інструментарію. Водночас методи когнітології також можуть бути застосовані для вдосконалення процедур корпусного аналізу. Безсумнівно, для залучення до лінгвістичних досліджень тексти корпусу мають бути розмічені за певною системою, зокрема в термінах метарозмітки. Метарозмітка – це система структурованих метаданих екстралінгвального характеру, які стосуються формальної сегментації і зовнішнього анотування тексту, а також фіксації технологічної обробки електронного файлу [10, с. 3]. Відсутність такої інформації про тексти ускладнює, а подекуди й унеможливує використання корпусу в лінгвістичних і програмних застосуваннях, тому метарозмітка, яка включає кодування, редагування й зберігання релевантних метаданих, виділяють як окремий важливий етап розробки корпусу.

Актуальність дослідження пов'язана з необхідністю впровадження корпусного підходу в сучасну лінгвокогнітологію, що потребує вирішення низки методологічних питань, зокрема поєднання корпусного й когнітивного підходів у прикладній лінгвістиці.

Мета статті – визначити основні етапи метарозмітки корпусу англomовних текстів з лінгвокогнітології. Досягнення поставленої мети передбачає виконання таких завдань:

- систематизувати підходи до здійснення метарозмітки корпусів текстів і окреслити відповідні набори метаданих;
- уточнити дефініцію поняття когнітивної мапи тексту в контексті корпусної лінгвістики;
- обґрунтувати необхідність поєднання корпусного і когнітивних методів для метарозмітки корпусу текстів з лінгвокогнітології;
- окреслити основні принципи метарозмітки корпусу англomовних текстів з лінгвокогнітології.

У сучасних корпусних дослідженнях виокремлюють структурний і ситуативний підходи до здійснення метарозмітки текстів [3; 17]. За структурним підходом класифікація текстів ґрунтується на ототожненні їх структури з такими елементами, як “назва, розділ, підрозділ, рубрика, присвята, епіграф, покликання, цитата, вживання алфавітів інших писемних систем, цифр тощо” [3, с. 61]. При цьому здійснення метарозмітки корпусу розуміється як структурне анотування, виділення елементів тексту за допомогою певної мови маркування [3, с. 62].

За ситуативним підходом загальна схема класифікації відображає структуру й основні компоненти комунікативного акту: автор, адресат, мета, предмет, умови й канал комунікації [17, с. 63].

Згаданим підходам до здійснення метарозмітки відповідають два види набору метаданих: 1) стандартний і 2) спеціалізований, розроблений для розв'язання конкретних завдань. Стандартний набір метаданих регулюється рекомендаціями Text Encoding Initiative (TEI), Open Language Archive Community (OLAC), EAGLES, ISLE Metadata Initiative (IMDI) і відповідає загальноприйнятій практиці кодування текстів. Цей метаопис застосовується, як правило, для національних корпусів, орієнтованих на якісно різних користувачів [6, с. 7]. Зазвичай, стандарти включають метадані про розробника корпусу, джерела, термін створення, мову текстів, формат файлів, авторські права. До набору даних спеціалізованого метаопису включають додаткові ознаки, необхідні для здійснення певного дослідження. Так, метарозмітка навчального корпусу текстів UCLE забезпечує вивчення актуального словника студентів з урахуванням параметрів статі, віку, рідної мови, курсу навчання й факультету [13, с. 30–31]. Метарозмітка Корпусу текстів з комп'ютерної лінгвістики включає окреме кодування анотації тексту. Це дає можливість у майбутньому створити автоматичну систему реферування текстів [19, р. 36]. Поряд з функціями пошукового механізму й лінгвістичною розміткою метадані визначають функціональні можливості й сумісність корпусу з іншими програмними продуктами.

Однак проаналізовані підходи і набори метаданих не повністю відповідають завданням дослідження текстів з лінгвокогнітології, доволі специфічних за змістом, і вимагають доповнення корпусних методів когнітивними, значною мірою базованими на інтуїції. Поєднання цих підходів уважається переходом до метарозмітки нового покоління, яка дає змогу опрацювати не лише тексти з власного корпусу, але й з інших корпусів за допомогою того самого інструментарію. Подібний інструментарій, зорієнтований на переформатування корпусів на “багатокореневі” дерева, зараз знаходиться в розробці. Оскільки зазначені вище методика частково відповідають завданням обробки текстів з лінгвокогнітології, необхідним видається поєднання когнітивного мапування й корпусного аналізу. У цьому аспекті доцільно проаналізувати, як змінюється і розвивається ключове поняття лінгвокогнітології – когнітивна мапа тексту – завдяки міждисциплінарним зв'язкам з корпусною лінгвістикою і впровадженню нових технологій.

На сьогодні термін “когнітивна мапа” належить до загальнонаукових, які характеризують систему репрезентації знань про об'єкт, а методика когнітивного мапування, у свою чергу, має широке застосування в галузі психології, біології, антропології, екології, географії, менеджменту, соціології і власне лінгвокогнітології. Уперше термін “когнітивна мапа” був уведений американським психологом Е. Тольманом в експериментальній праці “Когнітивні мапи щурів і людей” у 1948 році [26, р. 45]. У загальному розумінні когнітивна мапа – це суб'єктивне уявлення про просторову організацію навколишнього середовища. Саме тому первинними різновидами когнітивних мап уважаються диграфи: мапа-шлях, мапа-огляд [там само]. З часом у загальному контексті лінгвокогнітології це поняття набуло певного розвитку, і на сучасному етапі існує кілька його визначень.

**1. Інформаційна інтерактивна модель тексту:** 1) інформаційна модель повідомлення, тексту, яка містить ментальні репрезентації змісту, представляє глобальну картину кореляції та функціонування смислових програм, зумовлену стратегіями породження, сприйняття повідомлення й інтерактивністю комунікантів; 2) жанрова модель тексту певної прототипної структури; 3) інтерактивна модель, яка фіксує задум, статуси та ролі комунікантів, їхні інтенції, інтерактивні стратегії [18, с. 14].

**2. Концептуальний простір:** 1) інтегративний простір, або бленд [8, с. 186]; 2) концептуальний простір тексту зі складною ієрархією концептів, вершиною якої є текстовий концепт [11, с. 21–22]; 3) загальний текстовий концепт і підпорядковані йому різні типи концептів інформаційного масиву тексту, топіки, пропозиції й засоби їхнього зв'язку – конектори [18, с. 14].

3. **Схема тексту:** 1) спрощений опис фрагмента картини світу індивіда, що пов'язується з конкретною проблемною ситуацією [18]; 2) концептуальна мережа, утворена пропозиціями базисних фреймів [8, с. 180]; 3) графічне відображення плану, збору, опрацювання та збереження інформації, що міститься у свідомості людини [7, с. 255].

Розуміння когнітивної мапи як схеми тексту (п. 3.) є підґрунтям для розбудови цього поняття в контексті корпусної лінгвістики. Отже, когнітивна мапа у цьому сенсі визначається як 1) мережа, побудована на підставі кількісного аналізу й встановлення відношень між словами в тексті, або 2) концептуальна структура тексту [24, р. 185], яка ґрунтується на відношеннях між частотними текстовими концептами і визначає його основний смисл. Схематично когнітивну мапу можна представити у вигляді мережі або орієнтованого графу “з упорядкованими парами вузлів” [5, с. 249], у вершинах якого розташовані об'єкти, а ребра як зв'язки між вузлами мережі вказують на характер відношень між об'єктами в досліджуваному тексті.

Наведені визначення свідчать про розширення поняття когнітивної мапи в сучасних лінгвістичних студіях унаслідок розвитку загальноприйнятого розуміння картини світу й відходу від вузького уявлення про двовимірний простір. На думку С.А. Жаботинської [8, с. 180], когнітивна мапа (в іншій термінології фрейм) розуміється в широкому сенсі як структура презентації знань, або набір ідеалізованих когнітивних моделей, які демонструють первісні, найбільш узагальнені принципи категоризації та організації інформації, вираженої за допомогою мови. Розширенню розуміння поняття когнітивної мапи сприяє впровадження у лінгвокогнітивні дослідження сучасних інформаційних технологій, зокрема представлення даних у вигляді “гіперболічних самоорганізованих мап” (Hyperbolic Self-Organizing Maps – SOMs) [9, с. 85]. Саме гіперболічні самоорганізовані мапи, створені шляхом дискретизації гіперболічного плану, можуть сприяти створенню концептуально впорядкованих мап документів, у яких об'єднуються концептуальна кластеризація, наочність і легкість доступу до інформації [9, с. 85]. Проте представлення гіперболічного простору з його концептуальним оточенням вимагає застосування відповідного методологічного інструментарію, який має забезпечити регулярність (алгоритмічність) у дискретизації, членуванні мовної інформації і її кластеризацію [там само].

Одним із засобів актуалізації концептів у наукових текстах є терміни, організовані в терміносистеми, які у своїй сукупності відображають відповідні системи концептів. У прикладному аспекті таку систему концептів можна побудувати на матеріалі словникових дефініцій, окремого тексту [5, с. 249–265] або текстового корпусу. При певному вдосконаленні відповідно до окреслених вище завдань методологічним інструментарієм для представлення системи концептів як гіперболічного простору можуть слугувати корпус-менеджери й пакети програм з автоматичного опрацювання текстів корпусів. Особливе значення в цьому мають відіграти: 1) дані мультимедійних корпусів, що містять аудіо- і відеоінформацію, і 2) стандарти анування й забезпечення доступу до даних [25, р. 552].

Як уже згадувалося вище, стандартний набір універсальних даних не відповідає уявленню про текст як набір взаємопов'язаних концептів, і тому в нашому дослідженні в основу метарозмітки корпусу покладено розуміння когнітивної мапи як концептуального простору тексту, який включає загальний концепт, підпорядковані йому частотні концепти і засоби зв'язку між ними. За подібним принципом здійснено, до прикладу, метарозмітку Корпусу грузинських діалектів [20, р. 29]. Параметри зазначеної розмітки мають відповідати двом основним вимогам, установленим розробниками: 1) демонструвати “найбільш адекватну модель конкретної лінгвістичної галузі” й 2) відтворювати за можливістю “повну етнокультурну картину цього регіону” [20, р. 29]. У корпусі інформація про текст формується через його паспорт метаданих, набір яких є вираженням суб'єктивних уявлень, або моделлю презентації

знань лінгвіста про цей текст. Отже, в нашому дослідженні основним принципом здійснення метарозмітки є моделювання інформації про тексти корпусу, об'єднаного проблематикою лінгвокогнітології.

Методом паспортизації текстів корпусу обираємо метод когнітивного мапування, який має на меті моделювання тексту в термінах структуризації його концептуального простору. Під моделюванням розуміється дослідження текстів “шляхом побудови й вивчення їх моделей; використання моделей для визначення чи уточнення характеристик і раціоналізації способів побудови нових за конструкцією об'єктів” [15, с. 92]. У роботі застосовано жорсткий різновид когнітивного мапування, що ґрунтується лише на даних текстів [7, с. 254]. Моделювання даних про текст використовується для презентації інформації, яка відображає систему тексту через індуктивні або дедуктивні моделі [15, с. 10].

У сучасній лінгвістиці існує кілька визначень поняття “модель”:

– конструкція, структура, зразок, за яким побудована певна одиниця мови з одиниць нижчого рівня [15, с. 93];

– притаманна конкретній мові парадигматична схема елементів (складників) її структури; компактне символічне зображення цієї системи (або якихось її частин чи елементів), яка становить основу моделювання мови як одного з методів її вивчення [2, с. 238];

– абстрактне представлення ключових характеристик певної мовної одиниці, яке полегшує розуміння структури і використання цієї одиниці [23, р. 2];

– конструкція, логічно виведена за допомогою певного математичного апарату [1, с. 99].

У цьому дослідженні модель розуміється як структура, паспорт метаданих або схема метарозмітки корпусу текстів. За методами й прийомами побудови використовується модель паспортизації текстів корпусу є індуктивною, за змістом – когнітивною. Аналізуючи текст за напрямом – від часткового до загального, дослідник установлює класи й підкласи одиниць за певними диференційними ознаками, при цьому результати і точність аналізу залежать від повноти охоплення різних уживань або різновидів модельованого явища в базі даних і від ступеня деталізації аналізу [15, с. 339]. За типологією Є.А. Карпіловської [12, с. 33], за структурними властивостями модель паспортизації текстів корпусу є класифікаторною, тобто такою, що відображає устрій оригінала мовного об'єкта чи об'єктів у статистиці. Отже, метарозмітка корпусу текстів з лінгвокогнітології у дослідженні здійснюється шляхом побудови індикативної класифікаторної моделі – когнітивної мапи, застосовуваної для паспортизації текстів.

Відомо, що емпіричною базою даних у корпусних дослідженнях виступає репрезентативний корпус автентичних текстів. Укладання корпусу англomовних текстів з лінгвокогнітології принципово відрізняється від окреслених раніше етапів розробки корпусів текстів [4; 10; 14], які включають загальне планування, збір текстів, кодування і збереження релевантних метаданих і лінгвістичне анотування корпусу. Здійснення зовнішнього анотування створюваного корпусу текстів виконується у такі етапи: I. Загальне планування й збирання корпусу текстів. II. Когнітивне мапування й метарозмітка корпусу.

I. **Загальне планування** корпусу визначається потребами лінгвістів [14, с. 318] і завданнями конкретної розвідки. Оптимізація характеристик корпусу залежать від його обсягу, необхідних людських і грошових ресурсів, а також інформаційних потреб. Мінімізація різних типів помилок при плануванні дозволяє досягти репрезентативності створюваного корпусу [4, с. 95–96]. Матеріал дослідження становить корпус сучасних англomовних текстів з лінгвокогнітології, укладання якого базується на чинних вимогах до планування корпусів [6, с. 56–59]. За О.М. Демською-Кульчицькою [6, с. 7], основні завдання укладання корпусу “як моделі організації емпіричних даних для лінгвістичних студій і програмних застосувань” передбачають (1) формулювання лінгвістичної концепції корпусу; (2) визначення предметної

галузі та парадигми метаданих корпусу; (3) проектування архітектури корпусу; (4) визначення параметрів анування даних і (5) лінгвістичне забезпечення програмного оброблення мовних даних.

Створений корпус текстів з лінгвокогнітології побудовано на принципах репрезентативності, автентичності, відібраності й збалансованості. Репрезентативність корпусу визначається як здатність відображати всі властивості предметної галузі. При цьому під предметною галуззю розуміється “рівень реалізації мовної системи, яка містить мовні феномени, що підлягають лінгвістичному описові” [6, с. 56–57]. У цьому дослідженні предметною галуззю є англомовні письмові тексти з лінгвокогнітології. Адекватна фіксація закономірностей обраної для дослідження підмови досягається дотриманням експліцитних та імпліцитних принципів побудови корпусу [4, с. 95].

Одним із важливих складників “емпіризації корпусу” [6, с. 57] є дотримання вимоги автентичності відібраних текстів. Автентичність корпусу англомовних текстів з лінгвокогнітології забезпечується включенням до нього створених носіями мови автентичних текстів, зокрема праць Джорджа Лакоффа, Марка Джонсона, Марка Тернера, Джерома А. Фельдмана. Для побудови корпусу необхідно організувати певну вибірку на основі “чітких правил екстрагування даних, що відповідають обраній стратегії побудови корпусу, мотивовані типом корпусу і метою його створення та специфікою експлуатації” [6, с. 59].

**2. Збирання текстів** корпусу здійснюється за встановленими лінгвістами-розробниками екстралінгвістичними й суто лінгвістичними принципами [14, с. 318]. Відібраність даного корпусу ставить вимоги організації коректної вибірки, що базуватиметься на індивідуальних критеріях відбору емпіричного матеріалу, обмеженого лінгвістичними факторами – формою мовлення (письмові тексти), мовою (англійська), предметною галуззю (лінгвокогнітологія) і екстралінгвальними чинниками – хронологією (друга половина ХХ ст.). Створюваний корпус текстів з лінгвокогнітології має бути одномовним і однорідним в аспекті функціонального стилю відібраних текстів. З метою дотримання принципу хронологічної однорідності вибірки, де представлено проблематику лінгвокогнітології, до корпусу включені тексти сучасних американських когнітологів – лінгвістів (Дж. Лакофф, М. Тернер), філософів (М. Джонсон) та нейровчених (Дж. А. Фельдман). Отже, в результаті відбору до корпусу були включені тексти, які відповідають принципу лінгвістичної і хронологічної однорідності.

Збалансованість корпусу зазвичай полягає у введенні пропорційної кількості відібраних текстів. Традиційно на практиці збалансованість корпусу досягається в результаті використання різних методик відбору текстового матеріалу і є одним із доволі складних завдань. Незначний згідно з сучасними корпусними вимогами обсяг корпусу відібраних текстів з лінгвокогнітології потребує дотримання принципу збалансованості: для цього в корпус включені тексти різних авторів. Корпус містить повні тексти, що забезпечує їхню структурну, лексичну й синтаксичну завершеність. Відбір текстів згідно з принципами побудови корпусів дозволяє класифікувати укладений корпус текстів як репрезентативний для дослідження стану розвитку лінгвокогнітології у США на кінець ХХ ст. При цьому під репрезентативністю розуміється достатнє й пропорційне представлення в діапазоні корпусу текстів з лінгвокогнітології оригінальних творів різних авторів.

**II. Когнітивне мапування й метарозмітка корпусу.** Як правило, здійснення етапу метарозмітки корпусу текстів передбачає кодування, редагування й зберігання релевантних метаданих, необхідних для розв’язання різних лінгвістичних завдань. Парадигму корпусних метаданих становить інформація про конститутивні текстові дані – вміст електронного заголовка, основне призначення якого – інформувати якісно різних користувачів про тексти на предмет їх авторства, стилістично-жанрової специфіки, тематики, дати й місця написання тощо [6, с. 89]. Надання інформації про тексти, які входять до складу корпусу, оптимізує

використання корпусного ресурсу в лінгвістичних дослідженнях і програмних застосуваннях. Процес метарозмітки полягає у приписуванні текстам і їхнім компонентам кодів бібліографічного й екстралінгвістичного опису [10, с. 7]. Набір метаданих, що зберігається як окрема від текстового масиву база даних, визначає дослідні можливості корпусу, його сумісність з іншими програмними продуктами.

За загальною систематикою [14, с. 320] категорії текстів визначаються як типи джерел матеріалу корпусу (в цьому разі – наукові тексти); термін жанр використовують до предметних категорій (тексти з лінгвістики та суміжних дисциплін) в загальному сенсі, а термін текстовий тип – стосовно конкретних типів текстів (тут – монографії). Однак у вживанні окреслених вище термінів бракує узгодженості, зокрема щодо визначення жанру й типу текстів. За пропонуваною С. Лендау [14, с. 322–323] схемою корпусних текстових типів, тексти з лінгвокогнітології належать до домену “Письмо”, текстової категорії “Нехудожня література”, до жанру “Освітня література (університетські підручники чи високорівневі інформаційні праці)”.

Проте, як було зазначено вище, включенню текстів до тіла корпусу має передувати розробка системи категорій текстів різних типів за власними критеріями. Відповідно до побудови індуктивної моделі шляхом зіставного аналізу вихідних даних відібраних текстів необхідно виділити ті ознаки, які можуть слугувати основою для класифікаторної моделі метарозмітки. Виділені ознаки мають відповідати загально визначеним вимогам: 1) релевантності інформації [22, р. 492–493] та 2) входженню до структурованого й усталеного лексикону “словника метаданих” [22, р. 493]. На основі узагальнення виділених ознак можна сформулювати параметри метарозмітки й укласти алгоритм класифікації для відповідної комп’ютерної програми.

Як уже було зазначено, зовнішнє анотування корпусу англomовних текстів з лінгвокогнітології здійснюється через побудову індуктивної моделі [15, с. 101] шляхом: а) формування бази даних, тобто представницького масиву, який охоплює якомога більше випадків уживання модельованих одиниць; б) зіставного аналізу ознак, які можуть слугувати основою для класифікації вихідних даних; в) формулювання правил установлення параметрів метарозмітки; г) визначення класів, створення алгоритму класифікації; д) укладання відповідної комп’ютерної програми метарозмітки. У нашому дослідженні для розробки індуктивної моделі використовується комплексна методика, яка об’єднує когнітивне мапування і статистичний аналіз.

Методика когнітивного мапування [21] передбачає проходження трьох етапів, узагальнених і доповнених нижче відповідно до завдань цього дослідження: 1) визначення набору концептів, які складатимуть когнітивну мапу кожного з текстів корпусу. Формування набору концептів здійснюється з використанням статистичного аналізу для вибору лексики [16, с. 68], важливої для вирішення поставлених завдань; 2) установлення типів попередньо визначених відношень між частотними концептами в кожній когнітивній мапі [28, р. 2559] та зіставлення наборів концептів “семантично близьких текстів” [16, с. 68]; 3) побудова когнітивної мапи й кодування текстів [27, р. 102] корпусу програмним шляхом; 4) аналіз побудованої когнітивної мапи.

4. У дослідженні не передбачено проведення власне **лінгвістичної розмітки**, тобто приписування граматичних, лексичних та інших характеристик елементам текстів корпусу [10, с. 3], яке має забезпечити будь-які потреби користувача стосовно інформації про мову текстів. Це пояснюється метою укладання корпусу текстів з лінгвокогнітології – отримання інформації не про мову текстів, а про їхню зовнішню, концептуальну структуру, на базі якої створюється індуктивна класифікаторна модель паспортизації. Однак окремі аспекти лінгвістичної розмітки (зокрема, семантичної) при побудові зазначеної моделі передбачені. На сучасному етапі розвитку корпусної лінгвістики немає єдиного теоретичного підґрунтя для здійснення семантичної розмітки, проте в загальному розумінні це – приписування кодів семантичних категорій і підкатегорій певній лексичній одиниці або словосполученню [10, с. 7]. Функціональні можливості розробленої програми метарозмітки мають слугувати інструментарієм для подальших лінгвокогнітивних досліджень прикладного характеру.

У сучасній прикладній лінгвістиці серед завершальних етапів розробки корпусу текстів виділяють [10; 4] конвертування і забезпечення доступу до корпусу.

5. Етап **конвертування** зібраних текстів у структуру спеціалізованої лінгвістичної інформаційно-пошукової системи, інакше, корпус-менеджеру, має забезпечити швидкий багатоаспектний пошук і статистичне опрацювання корпусних даних [10, с. 8]. Дослідні можливості розроблюваного корпусу текстів з лінгвокогнітології визначаються моделлю метарозмітки і функціями вбудованого корпус-менеджеру.

6. **Забезпечення доступу** до корпусу текстів зазвичай реалізується в межах дисплейного класу, поширення на CD і в режимі глобальної мережі, при цьому різним категоріям користувачів можуть надаватися різні права й можливості [10]. Розробником такого корпусу текстів планується забезпечення доступу для зареєстрованих користувачів всесвітньої мережі Інтернет.

Здійснене дослідження дозволяє дійти таких висновків: 1. Метарозмітка корпусу текстів визначається метою й завданнями розробки корпусу і здійснюється в два етапи: 1) загальне планування та збирання текстів корпусу й 2) паспортизація цих текстів шляхом встановлення набору й відношень між ключовими концептами тексту методом когнітивного мапування як способу тестового моделювання. 2. Поняття моделі інтерпретується в цьому дослідженні як паспорт метаданих корпусу текстів. За методами й прийомами побудови використовується модель паспортизації текстів корпусу є індуктивною, за змістом – когнітивною, а за структурними властивостями – класифікаторною. 3. Когнітивна мапа тлумачиться як мережа, побудована на підставі даних статистичного аналізу, яка ґрунтується на відношеннях, встановлених між частотними концептами в тексті. 4. Укладання корпусу англомовних текстів з лінгвокогнітології здійснено на принципах репрезентативності, автентичності, відібраності й збалансованості. Дотримання зазначених принципів дозволяє класифікувати розроблюваний корпус як репрезентативний для дослідження стану розбудови лінгвокогнітології США у визначальний для неї період кінця ХХ століття.

#### ЛІТЕРАТУРА

1. Апресян Ю.Д. Идеи и методы современной структурной лингвистики (краткий очерк) / Юрий Дереникович Апресян. – М. : Просвещение, 1966. – 305 с.
2. Ахманова О.С. Словарь лингвистических терминов / Ольга Сергеевна Ахманова. – [2-е изд.]. – М. : УРСС : Едиториал УРСС, 2004. – 571 с.
3. Бук С. Структурне анотування у корпусі текстів (на прикладі прози Івана Франка) / С. Бук // Українська мова. – 2009. – № 3. – С. 59–71.
4. Бобкова Т.В. Етапи розробки лексикографічного корпусу українських законодавчих документів / Т.В. Бобкова // Мовні і концептуальні картини світу : [наук. вид.]. – К. : ВПЦ “Київський університет”, 2014. – Вип. 48. – С. 89–104.
5. Дарчук Н.П. Комп’ютерна лінгвістика (автоматичне опрацювання тексту) : [підручник] / Наталія Петрівна Дарчук. – К. : ВПЦ “Київський університет”, 2008. – 351 с.
6. Демська-Кульчицька О. Основи національного корпусу української мови : [монографія] / Орися Демська-Кульчицька. – К. : НАНУ, 2005. – 219 с.
7. Єсіпович К.П. Феномен когнітивного картування в сучасній лінгвістичній парадигмі / К.П. Єсіпович // Studia Linguistica. – 2013. – Вип. 7. – С. 254–257.
8. Жаботинська С.А. Посесивна конструкція і концептуальні трансформи / С.А. Жаботинська // Мова. Людина. Світ : до 70-річчя проф. М.П. Кочергана : [зб. наук. статей] / [ред. О.О. Тараненко]. – К. : Видавничий центр КНЛУ, 2006. – С. 178–192.
9. Жаботинская С.А. Онтологии для словарей тезаурусов: лингвокогнитивный подход / С.А. Жаботинская // Філологічні трактати. – 2009. – № 2. – Том 1. – С. 71–87.

10. Захаров В.П. Корпусная лингвистика : [учеб.-метод. пособ.] / Виктор Петрович Захаров. – СПб. : РОПИ СПб. университета, 2005. – 48 с.
11. Кагановська О.М. Текстові концепти художньої прози (на матеріалі французької романістики середини ХХ сторіччя) : [монографія] / Олена Марківна Кагановська. – К. : ВЦ КНЛУ, 2002. – 292 с.
12. Карпіловська Є.А. Вступ до комп'ютерної лінгвістики : комп'ютерна лінгвістика : [підручник] / Євгенія Анатоліївна Карпіловська. – Донецьк : ТОВ “Юго-Восток, ЛТД”, 2006. – 188 с.
13. Коломієць В. Спеціальний навчальний корпус текстів UCLE : сучасний стан і перспективи використання / В. Коломієць, С. Котик // Комп'ютерна лінгвістика : сучасне і майбутнє : [матер. Міжнар. наук.-практ. конф.]. – К. : Вид. центр КНЛУ, 2012. – С. 29–32.
14. Лендау С.І. Словники: мистецтво та ремесло лексикографії / Сидні І. Лендау; [пер. з англ.]. – [2-е вид.]. – К. : К.І.С., 2012. – 480 с.
15. Перебийніс В.І. Математична лінгвістика : навч. пос. [для студентів відділень структурно-математичної, прикладної та комп'ютерної лінгвістики] / Валентина Ісидорівна Перебийніс. – К. : Вид. центр КНЛУ, 2014. – 125 с.
16. Попова С.В. Извлечение ключевых словосочетаний / С.В. Попова, И.А. Ходырев // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. – 2012. – № 1 (77). – С. 68–72.
17. Савчук С.О. Метатекстовая разметка в Национальном корпусе русского языка : базовые принципы и основные функции / С.О. Савчук // Национальный корпус русского языка : Результаты и перспективы. – М., 2005. – С. 62–88.
18. Селіванова О.О. Лінгвістична енциклопедія / Олена Олександрівна Селіванова. – Полтава : Довкілля-К, 2011. – 844 с.
19. Bobkova T. Corpus of computational linguistic texts / T. Bobkova [etc.] // Computer Treatment of Slavic and East European Languages / [eds J. Levická, R. Garabik]. – Bratislava : Tribun, 2009. – P. 35–40.
20. Beridze M. The Corpus of Georgians Dialects / M. Beridze, D. Nadaria // Computer Treatment of Slavic and East European Languages / [eds J. Levická, R. Garabik]. – Bratislava : Tribun, 2009. – P. 25–34.
21. Carley K. Extracting, representing, and analyzing mental models / K. Carley, M. Palmquist // Social Forces. – 1992. – Vol. 70 (3). – P. 601–636.
22. Corpus Linguistics : [an international handbook] / [eds A. Lüdeling, M. Kytö]. – Berlin : Walter de Gruyter, 2008. – Vol. 1. – 776 p.
23. Crystal D. The Cambridge Encyclopedia of the English Language / [ed. D. Crystal]. – Cambridge : Cambridge University Press, 1995. – 489 p.
24. Glynn D. Polysemy, syntax, and variation: A usage-based method for Cognitive Semantics / D. Glynn // Directions in Cognitive Linguistics / [eds V. Evans, S. Pourcel]. – Amsterdam-Philadelphia : John Benjamins Publishing, 2009. – P. 77–106.
25. Newman J. Corpora and Cognitive Linguistics / J. Newman // RBLA, Belo Horizonte. – 2011. – Vol. 11. – No 2. – P. 521–559.
26. Özesmi U. Ecological models based on people's knowledge: A multi-step fuzzy cognitive mapping approach / U. Özesmi, S.L. Özesmi // Ecological Modelling. – 2004. – No. 176. – P. 43–64.
27. Stenetorp P. BRAT: A web-based tool for NLP-assisted text annotation / P. Stenetorp [etc.] // European Chapter of the Association for Computational Linguistics : [proceedings of the 13th Conference], (April 23–27, 2012). – Avignon : EAACL, 2012. – P. 102–107.
26. Tsuruoka Y. FACTA : A text search engine for finding associated biomedical concepts / Y. Tsuruoka, J. Tsujii, S. Ananiadou // Bioinformatics. Applications Note. – 2008. – Vol. 24. – No. 21. – P. 2559–2560.