

теоретико-методологічна база та категоріально-термінологічний апарат, накопичується емпіричний матеріал. Однак чимало теоретичних положень потребують ще наукового осмислення, узагальнення та систематизації, а термінологія – уточнення й уніфікації. На підставі аналізу основних публікацій у кожній із галузей лінгвістики, що розглядаються, визначено актуальні проблеми та зроблено спробу прогнозування динаміки їх розвитку.

Ключові слова: лінгвоконцептологія, концепт, лінгвоперсоналогія, мовна особистість, лінгвометафорологія, метафора.

Kravtsova Yu. V. New Directions in Linguistics of the XXIst Century: Topical Problems and Prospects for Development.

The article presents a brief analytical review of new directions in linguistics of the XXIst century – linguoconceptology, linguopersonology and linguometaphorology, with theoretical and methodological base and categorical and terminological apparatus being gradually formed, empirical material being accumulated. However, many theoretical positions require repeated scientific comprehension, generalization and systematization, and the terminology needs clarification and unification. Based on the analysis of the main publications in each linguistic field considered, the topical problems are set out and attempts to forecast the dynamics of their development are made. In linguoconceptology, the debatable problems are determination of the nature, characteristics and types of concept, methods of its analysis and description, structuring of concept and conceptsphere; in linguopersonology – ascertainment of the nature and types of language personality, its modeling, methods of its analysis and description; in linguometaphorology – determination of the nature of metaphor, its structure, mechanisms of formation and perception, metaphorical motivation, modeling of metaphorization, ascertainment of the peculiarities of metaphorical conceptualization and representation. In linguoconceptology, the promising directions appear to be descriptive and comparative study of individual and national conceptspheres, conceptographic practice; in linguopersonology – study of idiolectic language personalities, as well as its different ethnocultural or social types; in linguometaphorology – study of metaphor and metaphorization in semantic, cognitive and linguocultural aspects based on the material of scientific, political and artistic discourses, metaphorical lexicographic practice.

Keywords: linguoconceptology, concept, linguopersonology, language personality, linguometaphorology, metaphor.

УДК 81'42:81'322

Леміш Н. Є.
Національний педагогічний університет
імені М. П. Драгоманова

**КОРПУСНА МЕТАРОЗМІТКА СПЕЦІАЛЬНИХ ТЕКСТІВ
З ЛІНГВОАНТРОПОГЕНЕЗУ**

Статтю присвячено визначенню критеріїв корпусної метарозмітки спеціальних текстів з лінгвоантропогенезу. Досягнення мети реалізовано через уточнення понять “лінгвоантропогенез”, “спеціальні тексти”, “метарозмітка” та пошук набору елементів для здійснення метарозмітки фахових текстів. Запропоновано включити такі групи метаданих до метарозмітки спеціальних текстів із лінгвоантропогенезу: текстові метадавні; дані про автора; лінгвістичні метадавні; тематичні питання (із функціональними сферами); технічні метадавні, що дасть змогу суттєво скоротити час технічної роботи з корпусом такого типу, розширить коло його користувачів та підвищить об'єктивність результатів досліджень із проблематики лінгвоантропогенезу.

Ключові слова: корпусна метарозмітка, спеціальні тексти, лінгвоантропогенез.

У сучасному мовознавстві відбувається зміна антропоцентричної парадигми на антропокосмічну. Це пов'язане з процесом взаємодії та взаємозв'язком лінгвокультури та ноосфери щодо нових перспектив людської еволюції – лінгвоантропогенезу у

XXI ст. З огляду на це, мовна свідомість людства виражається через мовну творчість і формування лінгвістичної культури, а текст визнається “формою прогресії системної еволюції соціального досвіду (вертикалі висхідної еволюції)”, при цьому “висхідний характер мовного розвитку людини пов’язаний із самоздійсненням духовності особистості через мовленнєвомисленнєві канали” [1, с. 99]. Отже, дослідження з лінгвоантропогенезу в науці сьогодення є особливо актуальними.

Не менш важливими є також корпусні розвідки, вивчення мови в межах яких є максимально об’єктивним через абстрагування від суб’єктивності дослідника [14, с. 127]. Таким чином, опрацювання лінгвістичного матеріалу на основі корпусів дає найбільш достовірні результати й неупереджені висновки щодо стану і проблем досліджуваних мов як на окремому етапі їхнього розвитку, так і у зіставленні.

З 60-х років XX ст. *корпусну лінгвістику* вважають емпіричною дисципліною, що спрямована на вирішення практичних завдань [2, с. 12; 23]. Із самої назви зрозуміло, що корпусна лінгвістика має справу з корпусами, а тому визначенню власне терміна *корпус* присвячено чимало праць. Наприклад, при створенні Браунівського корпусу (Brown Corpus), одного з перших електронних корпусів, що датується 1962 р. / 1963 р., за основу було використано словникову дефініцію лексеми *корпус* – “сукупність текстів, які вважаються представницькими для лінгвістичного аналізу певної мови” [2, с. 12; 7, с. 19]. Саме таке визначення є передумовою появи лінгвістичних корпусів текстів.

З-поміж розмаїття дефініцій *корпусу* особливий інтерес представляють такі: *корпус* як “різновид інформаційно-пошукової системи” [2, с. 14; 9, с. 18; 14, с. 127–128], різновид “електронної бібліотеки, побудованої за зовнішнім критерієм відповідно до певного завдання” [30, с. 36]; вибірка даних, сформована за певними правилами [1]; текстозорієнтована, повнотекстова база даних [11, с. 74]; “певний масив, точніше, закінчена колекція машиночитаних текстів, дібраних для оптимального представлення мовного різноманіття” [2, с. 14]. З огляду на те, що в українському мовознавстві терміни *корпусна лінгвістика* і *корпус текстів* з’явилися на початку XXI ст. [2, с. 15], то їх визначення перебувають ще у стадії оформлення.

Окреслюючи основні аспекти корпусу текстів, Т. В. Бобкова зазначає про вживання терміна *корпус текстів* у вузькому та широкому значенні. У широкому – це “будь-яке зібрання письмових або усних текстів, використовуване з метою дослідження мови”; у вузькому – “зібрання текстів в електронній формі, що презентує певну мову” [2, с. 11]. Також дослідниця зауважує про певну розбіжність в інтерпретації *корпусу* в різних лінгвістичних розділах / дисциплінах. У широкому розумінні *корпус* є об’єктом лінгвістичного дослідження в домені загального мовознавства, а у вузькому – об’єктом власне корпусної лінгвістики в домені корпусної лінгвістики, тобто корпус – це “колекція лінгвальних даних письмових текстів або транскрибованих записів мовлення” у загальному мовознавстві та “велика колекція зразків письмових та усних текстів” [2, с. 13] у корпусній лінгвістиці.

Трактування корпусу з позицій корпусної лінгвістики також може варіюватися: у широкому сенсі це – “зібрання, що потенційно може містити не тільки прозу, поезію, а й реєстри слів і словники”, а у вузькому (“для службового використання”) – “колекція автентичних текстів або їх фрагментів для здійснення лінгвістичних досліджень” [2, с. 13]. Під *автентичними* розуміють дані природної мови, а не штучно створені або виявлені в ході лінгвістичного дослідження. Отже, у широкому значенні корпуси репрезентують собою колекції будь-яких текстів, а у вузькому – лише автентичних.

Науковий керівник Лабораторії комп’ютерної лінгвістики інституту філології Київського національного університету імені Тараса Шевченка, доктор філологічних

наук, професор Н. П. Дарчук описує *корпус* як “зібрання текстів певною мовою, яке представлене в електронній формі і супроводжується науковим апаратом. Апарат, вбудований у корпус, називається розміткою, або анотацією. Корпус тим кращий, чим повніша і досконаліша його анотація. Наука про корпуси – це, перш за все, наука про те, як зробити хорошу розмітку корпусу” [5]. Таким чином, дослідниця не лише акцентує на залежності якості корпусу від розмітки, а й підкреслює їх неподільність.

Визначаючи основні напрями сучасних корпусних досліджень мови, Г. Г. Лук'янець виділяє лінгвістичні корпуси як базу для вивчення різних одиниць фонетичного, морфологічного, лексико-семантичного та синтаксичного рівнів. В її трактуванні, “лінгвістичні корпуси становлять машинописне, стандартно організоване зібрання репрезентативних для певної мови, діалекту або іншої підмножини мов писемних або усних текстів в електронній формі, призначених для лінгвістичного аналізу й опису” [14, с. 127–128]. Вона також припускає, що саме такі корпуси дають змогу досліджувати специфіку усного та писемного мовлення з виявленням функціональних особливостей художнього, офіційно-ділового, публіцистичного, наукового та розмовного стилів.

Інтерпретацію корпусу в суто лінгвістичних термінах також представляють як дистрибуцію слів або моделей [2, с. 14; 27, с. 5].

На лінгвістичних, або мовних, корпусах концентрують свою увагу й В. В. Жуковська та В. П. Захаров. На думку В. П. Захарова, “лінгвістичний, або мовний, корпус текстів – великий, представлений в електронному вигляді, уніфікований, структурований, розмічений, філологічно компетентний масив мовних даних, призначений для вирішення конкретних лінгвістичних завдань” [9]. За визначенням В. В. Жуковської, лінгвістичний корпус “є такою колекцією природно мовних текстів, де здійснено розмітку (маркування) хоча б за одним лінгвістичним параметром” [8]. Неважко помітити, що дослідники також не залишають осторонь важливість розмітки для існування корпусу текстів.

Отже, корпус текстів апріорно повинен мати електронну форму (що вже вважається “протокорпусом”, за В. О. Плуґяном [15]), відповідати існуючим стандартам (про що, наприклад, зазначають Лоу Бернард [23], В. Г. Шкляревський [19]), а також володіти можливістю “багатократного використання” (завдяки спеціальній розмітці / анотації / кодуванню). “За такою інтерпретацією, апріорною детермінативною ознакою корпусу текстів стає анотованість” [2, с. 16], тобто корпус текстів не може існувати без розмітки.

З появою Браунівського електронного корпусу (1962 р. / 1963 р.) корпусна лінгвістика займається “розробкою загальних принципів побудови і використання лінгвістичних корпусів (корпусів текстів) із застосуванням комп’ютерних технологій” [9]. Це свідчить на користь інтерпретації *корпусу* В. П. Захаровим: корпус = власне корпус (масив корпусних даних) + корпусний менеджер (спеціалізована пошукова система). Такої ж думки дотримується Т. В. Бобкова, зазначаючи, що до *корпусу* входять текст + корпус-менеджер (= “інструментарій лінгвістичного й статистичного аналізу текстів” [2, с. 14]). Подібний підхід зумовлює необхідність врахування як лінгвальних, так і нелінгвальних чинників при формуванні корпусу, а отже, знов-таки виправдовує виділення широкого та вузького розуміння *корпусу*: вузьке розуміння *корпусу* базується на або лінгвальних, або нелінгвальних параметрах корпусу, а широкіше – на обох.

Визначальними ознаками корпусу більшість дослідників вважає такі: 1) машинозчитуваність, 2) автентичність текстів (включаючи усні транскрибовані), 3) добірність, 4) репрезентативність. При цьому під репрезентативністю, за

О. Демською-Кульчицькою, розуміється “релевантне відображення предметної галузі в корпусних даних, де задано в пропорції, детермінованій реальною чистотою досліджуваного явища, всі властивості відтвореного в корпусі домену” [7, с. 58]. Тому планування корпусу доцільне на базі комплексу ознак, а його створення повинно включати цілий набір кроків. За В. П. Захаровим [9], формування корпусів відбувається за таким алгоритмом: проектування; забезпечення надходження текстів відповідно до зазначених джерел; підготовка “технологічного” опису; перетворення в машинозчитувану форму; конвертування й попередня обробка текстів; графематичний аналіз (токенізація); **метарозмітка**; **лінгвістична розмітка** (виділення – наше, оскільки саме наявність розмітки різних типів уможливорює оперування корпусу як інформаційно-пошукової системи для вирішення практичних завдань); коригування результатів автоматичної розмітки; завантаження розмічених текстів у структуру корпус-менеджера; забезпечення доступу до корпусу (пошук); створення документального забезпечення.

У саме такому **актуальному** векторі виконується науковий фундаментальний проект (за рахунок видатків Фонду Державного Бюджету України) факультету іноземної філології Національного педагогічного університету імені М. П. Драгоманова “Германо-романо-слов’янський корпус аутентичних наукових текстів з лінгвоантропогенезу: розробка технологій нового покоління” (з січня 2017 р.) під керівництвом доктора філологічних наук, професора Алли Валер’янівни Корольової. Першочерговим завданням у межах зазначеної теми є розробка різних типів розмітки для упорядкування збалансованого германо-романо-слов’янського корпусу оригінальних наукових текстів з лінгвоантропогенезу.

Мета статті: визначення й характеристика критеріїв для корпусної метарозмітки спеціальних аутентичних текстів із лінгвоантропогенезу.

Завдання:

– уточнити поняття “лінгвоантропогенез”, “спеціальні тексти”, “метарозмітка”;

– виділити набір елементів для метарозмітки наукових аутентичних текстів з лінгвоантропогенезу.

Найбільш результативним способом, на наше переконання, є трактування будь-якого терміна через його внутрішню форму. Таким чином, поняття *лінгвоантропогенез* можна інтерпретувати через його складові: *лінгво-* + *антропо-* + *генез*. *Linguo* (лінгво-) є гіпотетично похідним від лат. *lingua* ← двнлат. *dingua* ← PIE **dnghū* (*tongue*) [29].

Anthropo- з двнгрец. *ἄνθρωπος* – людина, антропологічна лінгвістика (антрополінгвістика) як розділ мовознавства вивчає відображення розвитку мислення людини в еволюції мови, насамперед, її лексики. Це зумовлене аксіомою, відповідно до якої “практично всі історичні зміни в людській свідомості, розвитку культури і знаннях відображаються в лексичній системі мови” (Вікіпедія).

Значення елемента *-genesis*, що використовується для творення інших слів – “народження, походження, створення (*birth, origin, creation*), є похідним від грец. *genesis* ← PIE **gene-* (*give birth, beget*) [29].

Отже, *лінгвоантропогенез* визначаємо як відображення в мові змін, які відбулися або відбуваються в людській свідомості, розвитку культури і знань.

В. О. Карлова описує *лінгвоантропогенез* як науковий напрям, що “продовжує досліджувати питання взаємозв’язків мови і давньої свідомості, які домінують у сучасних працях лінгвістів, зокрема і когнітивного спрямування” [10, с. 225], тобто, це вивчення індивідуального (з позицій антропоцентризму) і колективного (з позицій антропокосмізму) розвитку людської особистості, її свідомості через їхнє відображення в мові, зокрема у когнітивному вимірі.

Проблема антропогенезу досліджується у філософському та релігійному аспектах, з позицій психології та лінгвістики. Сумуючи результати міждисциплінарних розвідок, можна стверджувати, що на ранніх стадіях розвитку свідомості людини характер її мислення безпосередньо залежав від природного середовища та його відчутті й сприйняття, тобто детермінується синкретичним зв'язком людини з природою, а саме відсутністю розмежування науки, релігії, мистецтва з провідною роллю природи [17, с. 38–39; 18, с. 273; 10, с. 226]. Також необхідно відмітити “злитість індивіда з колективом, одноплемінниками, з “ми”” [18, с. 570; 10, с. 226], до чого наразі повертається антропокосмізм. Отже, перехід до наукової антропокосмічної парадигми зумовлює своєчасність дослідження відображення процесу взаємодії та взаємозв'язку лінгвокультури та ноосфери відносно нових перспектив лінгвоантропогенезу на матеріалі корпусів різних текстів.

З огляду на той факт, що однією з визначальних ознак корпусу текстів є його репрезентативність, то “він повинен містити **всі типи текстів**, представлені у даній мові в даний історичний період, і при цьому містити їх у правильній пропорції” (виділення. – наше) [5, с. 9].

Репрезентативність і збалансованість корпусу визначається за його розмірами (відсотковим співвідношенням) й обсягом. В. П. Захаров [9], наприклад, зауважує, що корпуси можуть відноситися до всієї мови, до підмови (окремого жанру, стилю, мови певної вікової або соціальної групи, мови окремо взятого письменника або вченого, т.д.); параметри класифікації корпусів можуть бути різними залежно від їхнього призначення. Важливо врахувати приналежність текстів до різних жанрів, стилів, періодів, хронологічних меж, джерел й різним авторам. Тексти можуть братися з книг, періодики, можуть бути раніше виданими або лише підготовленими до видання.

До *спеціальних* текстів, що є невід'ємною складовою будь-якого представницького корпусу, слідом за Н. П. Дарчук [5, с. 9], відносимо наукові, науково-популярні і навчальні тексти. Отже, спеціальні тексти з лінгвоантропогенезу належать до наукових або науково-популярних видань.

У наукових виданнях, як правило, представлено результати теоретичних / експериментальних досліджень одного / групи авторів у тій чи іншій галузі знань [3]. Це:

- монографія (всебічне дослідження певної теми / проблеми, що належить одному чи кільком авторам);
- автореферат дисертації (авторський виклад основних положень самостійно проведеного дослідження, що подається на здобуття наукового ступеня кандидата чи доктора наук);
- тези доповідей (повідомлень) наукової конференції, з'їзду або симпозіуму (друкуються у збірнику напередодні відкриття роботи наукового заходу);
- матеріали конференції, з'їзду або симпозіуму (друкуються доповіді, резолюції, рекомендації, які відображають підсумки роботи цього заходу);
- збірник наукових праць (індивідуальні або колективні статті, авторами яких виступають науковці, викладачі, аспіранти, здобувачі, студенти наукових установ, навчальних закладів чи наукових товариств);
- рецензія на наукову або навчальну працю;
- відгук про автореферат.

Близькими до попередніх є науково-популярні видання, призначені для більш широкого кола читачів, які не є спеціалістами з конкретної галузі знань. Стиль викладу інформації в таких виданнях є доступним, а отже, й популярним. Це можуть бути серійні видання, типу “Бесіди про науку”, “Популярно про складне” [3] тощо.

Проміжну ланку, на нашу думку, складає група історичних документів, пам'яток культури, літератури, науки (напр., “Пам'ятки історичної думки України”, “Літературні пам'ятки України” [3]), які можна використовувати як джерельну базу в науково-дослідницькій діяльності.

Для збалансованості корпусу можна також включити групу навчальних видань, покликаних допомогти учням, студентам, аспірантам, викладачам отримати уявлення і в подальшому оволодіти конкретним навчальним предметом. Це можуть бути: 1) навчальні програми дисциплін, тематика яких будь-яким чином торкається лінгвоантропогенезу і включає списки рекомендованої літератури тощо; 2) підручники і навчальні посібники із систематизованим викладом окремих тем, дотичних до проблеми вивчення; 3) хрестоматії, де в скороченому варіанті представлено історичні / літературні твори, на матеріалі яких можна прослідкувати еволюцію свідомості та мислення людини; 4) конспекти / тексти лекцій з повним викладом тем, що безпосередньо розкривають різні аспекти лінгвоантропогенезу.

Доцільним також може бути включення до корпусу матеріалів довідкових видань, а саме: статей з *енциклопедій*, що подають основні відомості щодо того чи іншого поняття, предмета, терміна, події, персоналій, пов'язаних з лінгвоантропогенезом; *довідників* (наукових, навчальних, популярних), де розкрито дотичні теми; *словників* (енциклопедичних, термінологічних, тлумачних, лінгвістичних) із переліком слів, словосполучень, термінів, імен з відповідними тлумаченнями чи коментарями.

Усі видання, включені до корпусу, повинні бути текстовими (за знаковою природою інформації), опублікованими у вигляді книг, журналів, газет або електронних видань (за матеріальною конструкцією), оригінальними / автентичними, тобто написаними мовою оригіналу (за мовною ознакою), із зазначенням першості, перевидання, доповненості, виправленості, переробленості, розширеності, переглянутості (за повторністю випуску).

Уточнивши поняття *лінгвоантропогенезу* та *спеціальних текстів*, необхідно з'ясувати специфіку та критерії *метарозмітки*, оскільки створення корпусу без розмітки не є можливим [8, с. 132].

Складовими терміна *метарозмітка* є *мета-* та *розмітка*. Першим значенням префіксу *meta-* є “після, поза” (“after, behind”) від грец. прийменника *meta* на позначення “in the midst of; in common with; by means of; between; in pursuit or quest of; after, next after, behind”, що у свою чергу походить від PIE **me-* “in the middle” [29], *розмітку*, зазвичай, розуміють як маркування чогось. Таким чином, *метарозмітка* (в нашій інтерпретації) – проміжне виділення певних параметрів для отримання чогось більшого.

В англійській мові існує кілька неоднозначних відповідників для укр. терміна *розмітка* / *метарозмітка*: *metadata, annotation, tagging, mark-up*. А. Давидов, як і С. Савчук, використовують термін *метатекстова розмітка* [25; 16]. В. П. Захаров ототожнює *розмітку* та *анотацію* [9]. Цими термінами він позначає приписування текстам та їхнім компонентам спеціальних міток, так само, як і В. В. Жуковська, яка зазначає, що “процес розмітки (*tagging, annotation*) полягає в приписуванні текстам і їх компонентам спеціальних міток (*tag, tags*) [8, с. 76]. Але обсяг значення *tagging* та *annotation* в англійській мові не збігається, оскільки існує декілька видів розмітки. *Розмітка* може бути екстралінгвістичною (*метарозмітка (metadata)*): автор, назва тексту, рік і місце видання, жанр, тематика тощо), структурною (позначення глав, розділів, абзаців, речень, словоформ) і власне лінгвістичною (морфологічна, синтаксична, семантична, анафорична, просодична тощо). У *Corpus Mark-up and*

Annotation (University of Huddersfield) [24] анотація (*annotation*) представлена дещо ширше, включаючи дискурсивну (анафоричну, репрезентовану займенниками), прагматичну (тип мовленнєвого акту, рівень ввічливості тощо), стилістичну, частиномовну і семантичну розмітки. Отже, у *Corpus Mark-up and Annotation* розмежовано розмітку – *mark-up* і розмітку – *annotation*. Під першим терміном розуміються “дані про, напр., опис того, що репрезентує собою текст; або дані про характеристики формату та структури тексту”, під другим – “дані, отримані в результаті лінгвістичного аналізу корпусу (лінгвістичні метадані)” [24]. Наявне попередження про можливість взаємозаміни цих двох термінів, а також вживання лексеми *tagging* (розмітка) на позначення “будь-якого додавання додаткової інформації до корпусу”. Термін *mark-up* включає коди, які вказують на структурні елементи тексту, такі, як параграф, речення, початок глави, кінець, розрив сторінки, заголовки тощо, що уможливило пошук окремих слів в окремих структурних елементах текстів. Власне “дані про щось” наводять інформацію про місце видання або запису тексту, автора (прізвище, стаття, т.д.), формуючи так званий *хедер* (*header*), або *метадані* (*metadata*). Основним призначенням метаданих є спрямувати пошук у спеціальну сферу, тобто висвітлити, наскільки текст є спеціальним (= описати, на якій сфері спеціалізується конкретний спеціальний текст).

Лоу Бернард (Lou Burnard – консультант University of Oxford Computing Services, який стояв у витоків корпусної лінгвістики), пропрацювавши 36 років над комп’ютерною обробкою даних, підкреслює важливість і необхідність збору і формування метаданих для роботи з корпусами. Цікаво, що власне метадані, як правило, називають “даними про дані” [23] (“метадані (*metadata*) – це дані про дані корпусу (бібліографічні, типологічні)” [8, с. 135]), але саме метадані дають можливість досліднику визначитись із відповідністю цифрового джерела вимогам конкретного завдання, релевантністю для проведення певної розвідки на його матеріалі.

Безсумнівним є також той факт, що існування корпусної лінгвістики неможливе без метаданих. Це зумовлюється тим, що будь-яке лінгвістичне дослідження пов’язане з ідентифікацією мовленнєвих зразків у природній мові, репрезентованими в корпусах емпіричної науки – корпусної лінгвістики. Лоу Бернард вдало порівнює ілюстрації мовного використання з лабораторними зразками: для останніх важливою є партія, з якої вони відібрані, а для мовних ілюстрацій – контекст, який співвідносить мовленнєвий зразок з “природним середовищем”. Отже, без метаданих матеріал дослідження втрачає свою автентичність, точність і, врешті-решт, позбавляє отримані на його базі результати об’єктивності та достовірності.

Лоу Бернард виділяє 4 типи метаданих: редакторські (*editorial metadata*), аналітичні (*analytic metadata*), дескриптивні / описові (*descriptive metadata*), адміністративні (*administrative metadata*) [23]. *Редакторські метадані* – інформація про відношення складових корпусу та їх оригінальних джерел; *аналітичні* – шляхи інтерпретації та аналізу складових корпусу; *дескриптивні* – класифікаційна інформація із внутрішніх або зовнішніх характеристик складових корпусу; *адміністративні* – документальна інформація про власне корпус (назва / заголовок, наявність, статус перегляду тощо).

Виділяючи корпусний менеджер (спеціалізовану пошукову систему) у корпусі текстів, В.П. Захаров інформує, що він (корпусний менеджер) відображає інформацію про походження, тип тексту тощо у тому числі [9]. До бібліографічних і типологічних ознак документа, з якого вибирають контекст, відносять такі: автор, назва, джерело, рік видання, тип тексту і т.ін., що є також різновидом розмітки.

Неоднозначність понять *метарозмітка* / *розмітка*, а також практична спрямованість їхнього використання (“Хороша розмітка, зокрема, дозволяє швидко й ефективно знайти у корпусі ті слова і конструкції, які потрібні користувачеві” [5, с. 8]) свідчать про нагальність одного з завдань корпусної лінгвістики – розробки та удосконалення розмітки, тобто процесу приписування текстам та їх компонентам спеціальних міток (*tag, tags*). Такі мітки В. В. Жуковська поділяє на три різновиди:

– *зовнішні, екстралінгвістичні* (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, але також вік, стать, роки життя й багато чого іншого (це кодування інформації має назву *метарозмітка*);

– *структурні* (розділ, абзац, речення, словоформа);

– власне *лінгвістичні*, що описують лексичні, граматичні та інші характеристики елементів тексту [8, с. 76].

Ототожнюючи розмітку й анотацію, В. В. Жуковська розрізняє анотацію (*tagging / annotation*) та структурне маркування (*mark-up*) корпусу / корпусних даних [8, с. 76]. Слідом за Дж. Лічем, дослідниця визначає *tagging / annotation* як: “процес анотування корпусних даних – це додавання інтерпретованої, лінгвістичної інформації до електронного корпусу усного чи / або писемного мовлення” [26, с. 2; 8, с. 77]. Терміносполуку ж *структурна анотація / маркування (mark-up)* у спеціальній літературі використовують “на позначення і структури тексту, і зовнішньої стосовно нього інформації (його бібліографічний опис, дані про мовців тощо)” [28, с. 81; 8, с. 77]. Така дефініція є найширшою. У вужчому значенні *структурна анотація*, за Г. Астоном і Л. Бернардом, включає “межі глав, розділів, абзаців, речень, і т.д., а також особливу роль заголовків, переліків, приміток, посилань, супровідних підписів, покликів та ін.” [20, с. 24; 8, с. 77].

Серед лінгвістичних типів розмітки, як правило, виділяють розмітку на різних рівнях мови: морфологічну, синтаксичну, семантичну, анафоричну, просодичну. Попри те, що “жодна розмітка не може апіорно вважатися стандартом”, вона є корисним інструментом, зокрема при виборі метаданих “необхідно керуватися цілями дослідження й потребами лінгвістів, а також можливостями по внесенню в текст тих або інших додаткових ознак” [8, с. 78–82]. До того ж саме визначення “типології метаданих сприяє запобіганню дослідницьких непорозумінь у процесі зіставлення і опрацювання корпусів різних текстів” [19, с. 300].

Для цілей пропонованої розвідки особливий інтерес представляє саме *метарозмітка* як необхідний інструмент для подальшої роботи з лінгвістичним корпусом. Під *метарозміткою* ми розуміємо зовнішні, екстралінгвістичні 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, але й вік, стать, роки життя тощо.

На увагу також заслуговує дослідження *метарозмітки* В. Г. Шкляревським. Визначаючи передумови формування стандартів метарозмітки, окреслюючи етапи еволюції стандартів метаданих і простежуючи їхню реалізацію в англійських корпусах, розроблених в Україні, дослідник називає *метарозмітку* інтелектуальною розміткою. Така розмітка, описуючи позамовні властивості тексту, “дає змогу не лише виявити фактори впливу на словесну тканину текстів, виявити залежність мовних (лінгвальних) характеристик текстів від середовища їх побутування, а й окреслити тенденції розбудови тієї чи іншої дисципліни за допомогою аналізу відповідних фахових текстів” [19, с. 300]. Такий підхід також свідчить про розмежування власне розмітки корпусу (надання додаткової екстралінгвістичної інформації про тексти

корпусів – позамовної / екстралінгвальної) й анотування текстових елементів (усе, що стосується мовних / лінгвальних параметрів – “лінгвістична розмітка”) [19, с. 300]. С. Св. Волков [4] також розуміє під *метарозміткою* екстралінгвістичну розмітку, що “описує обставини створення тексту”. Обидва дослідники виділяють з-поміж елементів метарозмітки “елементи “формальної” структурної розмітки (текст, розділ тощо)” і “елементи “зовнішньої”, екстралінгвістичної метарозмітки, які характеризують текст в цілому – інформація про автора, умови створення текстів, інформація про тематику, стиль, жанр текстів” [Волков, Шкляревський]. Саме елементи екстралінгвістичної метарозмітки є складовими інтерпретації метарозмітки у пропонованій розвідці.

Цікавим є спостереження В. Г. Шкляревського стосовно появи терміна *описова розмітка* (*descriptive markup* – [31, с. 219]) на позначення метарозмітки Браунівського корпусу: “метарозмітка розумілась як загальний опис формальної композиційної структури письмових текстів: “виділення частин тексту, назви, рядка тощо” [19, с. 300]. Це був проміжний етап між лінгвістичною розміткою та власне метарозміткою.

За міжнародний варіант метарозмітки визнають класифікацію Дж. Сінклера-Шарова [33], що “передбачає виділення зовнішніх (позамовних) чинників, серед яких розрізняють три групи – походження тексту, зовнішні ознаки й мету написання тексту, та внутрішніх (змістовних і мовних) чинників – предметної області й стилістичних особливостей тексту. В результаті укладається паспорт з чотирма групами елементів метаопису (бібліографічні дані, відомості про жанр, автора, інформація про структуру, особливості розмітки й модифікації тексту)” [19, с. 302]. Це дійсно так, але для **наукових** текстів (наприклад, спеціальних з лінгвоантропогенезу) зазначений вище набір елементів не є достатнім для визначення їхньої поняттєвої структури. Слушним у цьому зв'язку є зауваження В. Г. Шкляревського щодо використання “індивідуалізованого стандарту розмітки”, або “тематичної метарозмітки”.

Спираючись на результати, отримані В. Г. Шкляревським на основі вивчення проектів, розроблених в українській корпусній лінгвістиці із застосуванням метарозмітки англійських текстів, акцентуємо на інформації, яку було включено до метарозмітки таких проектів залежно від призначення корпусів. Це: 1) дані про автора (ім'я, місце роботи, чи є англійська рідною мовою) і текст (тема, назва, обсяг, джерело, дата публікації) [6; 12, с. 32–33]; 2) назва тексту, прізвище та ініціали автора, назва видавництва й рік видання, анотація тексту [22, с. 36]; 3) назва серіалу, номер сезону та епізоду, жанр серіалу, дата виходу серії на екрани [13, с. 37]. Незважаючи на той факт, що таку метарозмітку можна вважати базовою для опису фахових текстів, вона все ж таки не є достатньо інформативною для роботи зі спеціальними текстами з лінгвоантропогенезу. Отже, для опрацювання нашої теми необхідний власний варіант метарозмітки.

З цього приводу Г. Лук'янець, слідом за Д. Бібером [21, с. 216], зауважує, що тип (лінгвістичної) анотації корпусу зумовлено прикладним призначенням корпусних даних [14, с. 128]. Лінгвістичні корпуси “містять соціолінгвістичну інформацію про мовців і / або авторів текстів”, наприклад, “Британський національний корпус був широко анований щодо різних соціолінгвістичних параметрів, таких як вік, стать, соціальний стан, місце проживання письменників і т.п.” [14, с. 131], що свідчить на користь індивідуального підходу до формування метарозмітки фахових текстів.

Працюючи над створенням корпусу Манде, А. Давидов виділяє критерії для відбору текстів. На відміну від підходу Дж. Сінклера, який виділяє критерії двох типів: зовнішні (пов'язані з комунікативною функцією тексту) та внутрішні (що

відображають мовні подробиці) [32], А. Давидов наголошує на залежності побудови корпусу від зовнішніх чинників, “тобто зміст корпусу необхідно відбирати незалежно від його мови” [25, с. 59]. За А. Давидовим, *метатекстова розмітка* повинна включати кілька груп за п'ятьма параметрами, а саме: текстові метадані; дані про автора; лінгвістичні метадані; тематичні питання; технічні метадані [25, с. 60].

Щодо *текстових метаданих*, то це назва / заголовок тексту (за наявності); точна або приблизна дата створення тексту (не видання); розмір тексту (кількість слів); вказівка на автентичність тексту або перекладний варіант (у цьому випадку зазначення мови тексту оригіналу); канал (письмовий або усний); бібліографічні дані для виданих текстів (місце і рік опублікування, назва видавництва, кількість сторінок; для періодичних видань – том, номер, серія); місце і дата створення для рукописів (якщо відомо), а також форма – написаний від руки, машинописний, електронний; місце і дата запису, якщо це усне мовлення; *дані про автора*: повне ім'я автора / авторів (якщо відомо), псевдонім (за наявності); вік автора на момент створення тексту; стать автора (чол., жін., невідома); авторське володіння мовою (носій рідної мови, іноземної); *лінгвістичні метадані*: зазначення мовного варіанту / діалекту / ідіолекту у випадку мультилінгвального корпусу (з мовним різноманіттям); вказівка на систему письма (різні графічні системи за умови їхніх відмінностей у текстах корпусу); *тематичні питання* можуть бути репрезентовані разом з функціональними сферами. Зазвичай поділяються на групи. Наприклад, для створення корпусу Манде А. Давидовим запропоновано 6 груп: “фольклор: казки, анекдоти, епічні твори, приказки / прислів'я, тексти традиційних пісень; література: прозові твори, сценарії фільмів, театральні п'єси, поетичні твори, тексти популярних пісень; релігія: християнські, ісламські, т.п.; навчальні: шкільні підручники, мови та грамотність, здоров'я, сільське господарство, наука; особисте: особисті записи, кореспонденція; інше: журналізм, спорт, історія, реклама, бізнес, побут” [25, с. 62]; *технічні метадані*: ім'я члена проекту, який додав до корпусу певний текст; дата додавання тексту до корпусу.

А. Давидов упевнений, що саме така метатекстова розмітка “допоможе контролювати процес наповнення корпусу новими текстовими даними та оцінювати баланс корпусу” [25, с. 62].

Оскільки для цілей нашого дослідження будуть обрані лише готові автентичні письмові тексти, спонтанно генеровані або перекладені тексти розглядатися не будуть, то за основу формування власної метарозмітки беремо критерії А. Давидова.

Підсумовуючи, зазначимо, що метарозмітка спеціальних текстів з лінгвоантропогенезу включатиме такі елементи:

– *текстові метадані*: назва / заголовок тексту (за наявності); ключові слова (за наявності); дата створення / видання тексту; розмір тексту (кількість слів); бібліографічні дані для виданих текстів (місце і рік опублікування, назва видавництва, кількість сторінок; для періодичних видань – том, номер, серія); місце і дата створення для рукописів (якщо відомо), а також форма – друкований, електронний;

– *дані про автора*: повне ім'я автора / авторів (якщо відомо); вік автора (на момент створення тексту, дати життя); стать автора (чол., жін., невідома); авторське володіння мовою (носій рідної мови, іноземної);

– *лінгвістичні метадані*: зазначення мови (у межах германських, романських, слов'янських), мовного варіанту / діалекту / ідіолекту (за потреби); вказівка на систему письма (різні графічні системи за умови їхніх відмінностей у текстах корпусу);

– тематичні питання (із функціональними сферами): 1) історія / культура / семіотика / психологія; 2) наукові / науково-популярні / навчальні тексти; 3) внутрішньожанрові форми (монографія, автореферат тощо);

– технічні метадані: ім'я керівника проекту, імена членів проекту, які додали до корпусу певний текст; дата додавання тексту до корпусу.

Саме такий набір елементів метарозмітки, на нашу думку, дасть змогу суттєво скоротити час технічного опрацювання спеціальних текстів з лінгвоантропогенезу, відкриваючи шляхи для “мікроісторичної” лінгвістики та вияву нових, раніше невідомих проблем, що потребують подальшого вирішення.

Література:

1. *Беляцкая А. А.* Лингвокультура и ноосфера: новые перспективы человеческой эволюции (Лингвоантропогенез в XXI веке) / А. А. Беляцкая // Ученые записки Таврического национального университета имени В. И. Вернадского. Серия “Философия. Культурология. Политология. Социология”. – 2013. – Т. 26 (65). – № 5. – С. 98–105.
2. *Бобкова Т. В.* Корпус текстов: основные аспекты назначения / Т. В. Бобкова // Науковий вісник кафедри Юнеско КНЛУ. Серія Філологія. Педагогіка. Психологія. – 2014. – Вип. 29. – С. 11–20.
3. Видання ... [Електронний ресурс]. – Режим доступу : <http://journ.univ.kiev.ua/vr/wp-content/uploads/види-виднь>
4. *Волков С. Св.* Метаразметка в историческом корпусе XIX века / С. Св. Волков // Корпусная лингвистика-2004: Труды международной конференции. – СПб. : Изд-во СПб. ун-та, 2004. – С. 86–98.
5. *Дарчук Н. П.* Корпусна лінгвістика: проблеми, методи, перспективи (робоча навчальна програма для аспірантів) / Н. П. Дарчук. – К. : КНУ імені Тараса Шевченка, 2013. – 11 с.
6. *Дарчук Н. П.* Дослідницький корпус української мови: основні засади і перспективи / Н. П. Дарчук // Вісник Київського національного університету імені Тараса Шевченка. Серія: Літературознавство. Мовознавство. Фольклористика. – К. : Видавничо-поліграфічний центр “Київський університет”, 2010. – № 21. – С. 45–49.
7. *Демська-Кульчицька О.* Основи національного корпусу української мови : монографія / Оріся Демська-Кульчицька. – К. : Інститут української мови НАНУ, 2005. – 219 с.
8. *Жуковська В. В.* Вступ до корпусної лінгвістики : навчальний посібник / В. В. Жуковська. – Житомир : Вид-во ЖДУ ім. І. Франка, 2013. – 142 с.
9. *Захаров В. П.* Корпусная лингвистика : учебно-методическое пособие / Виктор Петрович Захаров. – СПб. : РОПИ СПб. ун-та, 2005. – 48 с.
10. *Карлова В. О.* Лінгвоантропогенез як еволюція давньої художньої свідомості / В. О. Карлова // Філософія мови та нові тенденції в перекладознавстві й лінгвістиці : матеріали Міжнар. наук.-практ. конф., (Київ, 6–7 квіт. 2017 р.) / М-во освіти і науки України, Київський національний університет імені М. П. Драгоманова. – 2017. – С. 225–229.
11. *Карпіловська Є. А.* Вступ до комп'ютерної лінгвістики: комп'ютерна лінгвістика : підручник / Євгенія Анатоліївна Карпіловська. – Донецьк : ТОВ “Юго-Восток, ЛТД”, 2006. – 188 с.
12. *Коломієць В.* Корпус анотацій наукових статей із комп'ютерної лінгвістики / В. Коломієць, В. Орел // Комп'ютерна лінгвістика: сучасне і майбутнє : матеріали Міжнародної науково-практичної конференції. – К. : Вид. центр КНЛУ, 2012. – С. 32–34.
13. *Лебедев К.* Створення Багатомовного корпусу паралельних текстів / К. Лебедев // Комп'ютерна лінгвістика: сучасне і майбутнє : матеріали Міжнародної науково-практичної конференції. – К. : Вид. центр КНЛУ, 2012. – С. 36–37.
14. *Лук'янець Г. Г.* Основні напрямки сучасних корпусних досліджень мови та перспективи їх подальшого розвитку / Г. Г. Лук'янець // Наукові праці НУХТ. – 2012. – № 44. – С. 127–133.
15. *Плунгян В. А.* Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики / В. А. Плунгян // Русский язык в научном освещении. – М. : Языки славянской культуры, 2008. – № 2 (16). – С. 7–20.
16. *Савчук С. О.* Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. – М., 2005. – С. 62–88.
17. *Тайлор Э. Б.* Первобытная культура / Эдвард Бернетт Тайлор. – М. : Знання, 1990. – 548 с.

18. Токарев С. А. Религия в истории народов мира / Сергей Александрович Токарев. – М. : Политиздат, 1964. – 623 с.
19. Шклярєвський В. Г. Метарозмітка корпусів текстів: стандарти і реалізація / В. Г. Шклярєвський // Наукові записки Національного університету Острозька академія. Серія: Філологічна. – 2014. – Вип. 49. – С. 300–303.
20. Aston G. The BNC handbook: exploring the British National Corpus with SARA / G. Aston, L. Burnard. – Edinburgh: Edinburgh University Press, 1998. – 256 p.
21. Biber D. Corpus linguistics: Investigating language structure and use / D. Biber, S. Conrad, R. Reppen. – Cambridge : Cambridge University Press, 2001. – 312 p.
22. Bobkova T. Corpus of computational linguistic texts / T. Bobkova // Computer Treatment of Slavic and East European Languages. – Bratislava : Tribun, 2009. – P. 35–40.
23. Burnard Lou. Metadata for corpus work [Electronic Resource]. – 2004. – Mode of Access : <http://users.ox.ac.uk/~lou/wip/metadata.html>
24. Corpus Mark-up and Annotation [Electronic Resource] // University of Huddersfield. – Mode of Access : http://www.pala.ac.uk/uploads/2/5/1/0/25105678/04_markup_and_annotation_handout.pdf
25. Davydov A. Towards The Manding Corpus: Texts Selection Principles and Metatext Markup / A. Davydov // Proceedings of the Second Workshop on African Language Technology – AFLAT 2010. – P. 59–62.
26. Leech G. Introducing corpus annotation / G. Leech // Corpus Annotation: Linguistic Information from Computer Text Corpora. – Addison Wesley Longman, 1997. – P. 1–19.
27. McEnery T. Corpus-based language studies: an advanced resource book / T. McEnery, R. Xiao, Y. Tono. – Taylor & Francis, 2006. – 386 p.
28. Meyer Ch. F. English corpus linguistics. An introduction / Charles F. Meyer. – Cambridge : Cambridge University Press, 2002. – 168 p.
29. Online Etymology Dictionary [Electronic Resource]. – Mode of Access : <http://www.etymonline.com>
30. Ooi V. B. Y. Computer corpus lexicography / Vincent B. Y. Ooi. – Edinburgh : Edinburgh University Press, 1998. – 243 p.
31. Renear H. A. Text Encoding / A. Renear // A Companion to Digital Humanities / eds S. Schrelbman, R. Siemens, J. Unsworth. – Oxford : Blackwell, 2007. – P. 218–239.
32. Sinclair J. Corpus and Text – Basic Principles / J. Sinclair // Developing Linguistic Corpora: a Guide to Good Practice / ed. M. Wynne. – Oxford : Oxbow Books, 2005. – P. 1–16.
33. Sinclair J. Preliminary recommendations on text typology [Electronic Resource] / J. Sinclair // EAGLES Document EAG-TCWG-TTYP/P, 1. – Mode of Access : <http://www.ilc.cnr.it/EAGLES96/texttyp/texttyp.html>

References:

1. Belyatskaya A. A. Lingvokul'tura i noosfera: novyye perspektivy chelovecheskoy evolyutsii (Lingvoantropogenez v KHKHÍ veke) [Linguoculture and noosphere: new perspectives of human evolution (Linguoanthropogenesis in the 21st century)] / A. A. Belyatskaya // Uchenyye zapiski Tavricheskogo natsional'nogo universiteta imeni V. I. Vernadskogo. Seriya "Filosofiya. Kul'turologiya. Politologiya. Sotsiologiya". – 2013. – T. 26 (65). – № 5. – S. 98–105.
2. Bobkova T. V. Korpus tekstiv: osnovni aspekty vyznachennya [Body of texts: main aspects of definition] / T. V. Bobkova // Naukovyy visnyk kafedry Yunesko KNLU. Seriya Filolohiya. Pedagogika. Psykholohiya. – 2014. – Vyp. 29. – S. 11–20.
3. Vydannya ... [Edition] [Elektronnyy resurs]. – Rezhym dostupu : <http://journ.univ.kiev.ua/vr/wp-content/uploads/vydy-vydn>
4. Volkov S. Sv. Metarazmetka v istoricheskom korpuse KHÍKH veka [Meta-markings in the historical building of the nineteenth century] / S. Sv. Volkov // Korpusnaya lingvistika-2004: Trudy mezhdunarodnoy konferentsii. – SPb. : Izd-vo SPb. un-ta, 2004. – S. 86–98.
5. Darchuk N. P. Korpusna linhvistyka: problemy, metody, perspektyvy (robocha navchal'na prohrama dlya aspirantiv) [Corpus linguistics: problems, methods, perspectives (work syllabus for graduate students)] / N. P. Darchuk. – K. : KNU imeni Tarasa Shevchenka, 2013. – 11 s.
6. Darchuk N. P. Doslidnyts'kyy korpus ukrayins'koyi movy: osnovni zasady i perspektyvy [The Ukrainian language research corpus: the main principles and perspectives] / N. P. Darchuk // Visnyk Kyivs'koho natsional'noho universytetu imeni Tarasa Shevchenka. Seriya: Literaturoznavstvo. Movoznavstvo. Fol'klorystyka. – K. : Vydavnycho-polihrafichnyy tsentr "Kyivs'kyy universytet", 2010. – № 21. – S. 45–49.

7. *Dems'ka-Kul'chysts'ka O.* Osnovy natsional'noho korpusu ukrayins'koyi movy: monohrafiya [Fundamentals of the National Corps of the Ukrainian Language: monograph] / Orysya Dems'ka-Kul'chysts'ka. – K. : Instytut ukrayins'koyi movy NANU, 2005. – 219 s.
8. *Zhukovs'ka V. V.* Vstup do korpusnoyi linhvistyky: navchal'nyy posibnyk [Introduction to Core Linguistics: Textbook] / V. V. Zhukovs'ka. – Zhytomyr : Vyd-vo ZHDU im. I. Franka, 2013. – 142 s.
9. *Zakharov V. P.* Korpusnaya lingvistika: uchebno-metodicheskoye posobiye [Corpus linguistics: a teaching aid] / Viktor Petrovich Zakharov. – SPb. : ROPI SPb. un-ta, 2005. – 48 s.
10. *Karlova V. O.* Linhvoantropohenez yak evolyutsiya davn'oyi khudozhn'oyi svidomosti [Lingvoantropogenesis as the evolution of the ancient artistic consciousness] / V. O. Karlova // Filosofiya movy ta novi tendentsiyi v perekladoznavstvi y linhvistytsi : materialy Mizhnar. nauk.-prakt. konf., (Kyiv, 6–7 kvit. 2017 r.) / M-vo osvity i nauky Ukrainy, Kyivskyy natsional'nyy universytet imeni M. P. Drahomanova. – 2017. – S. 225–229.
11. *Karpilovs'ka YE. A.* Vstup do komp'yuternoyi linhvistyky: komp'yuterna linhvistyka: pidruchnyk [Introduction to Computer Linguistics: Computer Linguistics: A Textbook] / Yevheniya Anatoliyivna Karpilovs'ka. – Donetsk : TOV “Yuho-Vostok, LTD”, 2006. – 188 s.
12. *Kolomiyets' V.* Korpus anotatsiy naukovykh statey iz komp'yuternoyi linhvistyky [Corpus of annotations of scientific articles on computer linguistics] / V. Kolomiyets', V. Orel // Komp'yuterna linhvistyka: suchasne i maybutnye : materialy Mizhnarodnoyi naukovo-praktychnoyi konferentsiyi. – K. : Vyd. tsentr KNLU, 2012. – S. 32–34.
13. *Lebedyev K.* Stvorenniya Bahatomovnoho korpusu paralel'nykh tekstiv [Creation of Multilingual Corps of Parallel Texts] / K. Lebedyev // Komp'yuterna linhvistyka: suchasne i maybutnye : materialy Mizhnarodnoyi naukovo-praktychnoyi konferentsiyi. – K. : Vyd. tsentr KNLU, 2012. – S. 36–37.
14. *Luk'yanets' H. H.* Osnovni napryamky suchasnykh korpusnykh doslidzhen' movy ta perspektyvy yikh podal'shoho rozvytku [The main directions of modern case studies of language and prospects for their further development] / H. H. Luk'yanets' // Naukovi pratsi NUKHT. – 2012. – № 44. – S. 127–133.
15. *Plungyan V. A.* Korpus kak instrument i kak ideologiya: o nekotorykh urokakh sovremennoy korpusnoy lingvistiki [The Corps as an Instrument and as an Ideology: On Some Lessons of Contemporary Corpus Linguistics] / V. A. Plungyan // Russkiy yazyk v nauchnom osveshchenii. – M. : YAzyki slavyanskoy kul' tury, 2008. – № 2 (16). – S. 7–20.
16. *Savchuk S. O.* Metatekstovaya razmetka v Natsional'nom korpusе russkogo yazyka: bazovyye printsipy i osnovnyye funktsii [Metatext marking in the National building of the Russian language: basic principles and basic functions] // Natsional'nyy korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektyvy. – M., 2005. – S. 62–88.
17. *Taylor E. B.* Pervobytnaya kul'tura [Primitive culture] / Edvard Bernett Taylor. – M. : Znannya, 1990. – 548 s.
18. *Tokarev S. A.* Religiya v istorii narodov mira [Religion in the history of the peoples of the world] / Sergey Aleksandrovich Tokarev. – M. : Politizdat, 1964. – 623 s.
19. *Shklyarevs'kyy V. H.* Metarozmitka korpusiv tekstiv: standarty i realizatsiya [Metadatura of text corpus: standards and realization] / V. H. Shklyarevs'kyy // Naukovi zapysky Natsional'noho universytetu Ostroz'ka akademiya. Seriya: Filolohichna. – 2014. – Vyp. 49. – S. 300–303.
20. *Aston G.* The BNC handbook: exploring the British National Corpus with SARA / G. Aston, L. Burnard. – Edinburgh : Edinburgh University Press, 1998. – 256 p.
21. *Biber D.* Corpus linguistics: Investigating language structure and use / D. Biber, S. Conrad, R. Reppen. – Cambridge : Cambridge University Press, 2001. – 312 p.
22. *Bobkova T.* Corpus of computational linguistic texts / T. Bobkova // Computer Treatment of Slavic and East European Languages. – Bratislava : Tribun, 2009. – P. 35–40.
23. *Burnard Lou.* Metadata for corpus work [Electronic Resource]. – 2004. – Mode of Access : <http://users.ox.ac.uk/~lou/wip/metadata.html>
24. *Corpus Mark-up and Annotation* [Electronic Resource] // University of Huddersfield. – Mode of Access : http://www.pala.ac.uk/uploads/2/5/1/0/25105678/04_markup_and_annotation_handout.pdf
25. *Davydov A.* Towards The Manding Corpus: Texts Selection Principles and Metatext Markup / A. Davydov // Proceedings of the Second Workshop on African Language Technology – AFLAT 2010. – P. 59–62.
26. *Leech G.* Introducing corpus annotation / G. Leech // Corpus Annotation: Linguistic Information from Computer Text Corpora. – Addison Wesley Longman, 1997. – P. 1–19.
27. *McEnery T.* Corpus-based language studies: an advanced resource book / T. McEnery, R. Xiao, Y. Tono. – Taylor & Francis, 2006. – 386 p.

28. Meyer Ch.F. English corpus linguistics. An introduction / Charles F. Meyer. – Cambridge : Cambridge University Press, 2002. – 168 p.
29. Online Etymology Dictionary [Electronic Resource]. – Mode of Access : <http://www.etymonline.com>
30. Ooi V. B. Y. Computer corpus lexicography / Vincent B. Y. Ooi. – Edinburgh : Edinburgh University Press, 1998. – 243 p.
31. Renear H. A. Text Encoding / A. Renear // A Companion to Digital Humanities / eds S. Schrelbman, R. Siemens, J. Unsworth. – Oxford : Blackwell, 2007. – P. 218–239.
32. Sinclair J. Corpus and Text – Basic Principles / J. Sinclair // Developing Linguistic Corpora: a Guide to Good Practice / ed. M. Wynne. – Oxford : Oxbow Books, 2005. – P. 1–16.
33. Sinclair J. Preliminary recommendations on text typology [Electronic Resource] / J. Sinclair // EAGLES Document EAG-TCWG-TTYP/P, 1. – Mode of Access : <http://www.ilc.cnr.it/EAGLES96/texttyp/texttyp.html>

Леміш Н. Е. Корпусная метаразметка специальных текстов по лингвоантропогенезу.

Статья посвящена определению критериев корпусной метаразметки специальных текстов по лингвоантропогенезу. Достижение цели реализуется путём уточнения понятий “лингвоантропогенез”, “специальные тексты”, “метаразметка” и поиском набора элементов для метаразметки специальных текстов. Предлагается включить в метаразметку специальных текстов по лингвоантропогенезу такие группы метаданных: текстовые метаданные; данные про автора; лингвистические метаданные; тематические вопросы (с функциональными сферами); технические метаданные, что позволит существенно сократить время технической работы с корпусом, расширит круг его пользователей и повысит объективность результатов исследований по проблематике лингвоантропогенеза.

Ключевые слова: корпусная метаразметка, специальные тексты, лингвоантропогенез.

Lemish N. Ye. Corpus Mark-up of Specialized Texts on Linguoanthropogenesis.

The paper deals with peculiarities and criteria defining for building corpus mark-up of specialized texts on linguoanthropogenesis. The aim was achieved through specifying of the concepts “linguoanthropogenesis”, “specialized texts”, “mark-up”, as well as describing the set of components to form the mark-up of such specialized texts. In the end the author suggested the following groups of metadata to be included into the mark-up of specialized texts on linguoanthropogenesis: textual metadata; data about authors; linguistic metadata; thematic issues (including their functional areas); technical metadata, which would help to save time on technical processing of the corpus, expand the range of its users and increase the objectivity of the obtained results on the problems of linguoanthropogenesis.

Keywords: corpus mark-up, specialized texts, linguoanthropogenesis.

УДК 81'367.625

Леута О. І.
Національний педагогічний університет
імені М. П. Драгоманова

СУЧАСНІ ІНТЕРПРЕТАЦІЇ ДІЄСЛІВНОГО ПРЕДИКАТА В СТРУКТУРІ РЕЧЕННЯ

У статті проведено аналіз сучасного етапу розвитку вітчизняної та зарубіжної граматичної науки в царині тлумачення природи та принципів організації речення. Особливу увагу звернено на становлення та здобутки семантичного синтаксису як нового етапу інтерпретації закономірностей породження речення на засадах вербоцентризму. Проаналізовано роль та конструктивні здатності предиката як основної ланки структури речення, розглянуто найпоширеніші класифікації дієслівних предикатів у вітчизняному та зарубіжному мовознавстві. Акцентовано увагу на перспективах розвитку синтаксису через дослідження предиката.