

УДК 81.33

Г.Г. Лук'янець
Національний університет
харчових технологій

ОСНОВНІ НАПРЯМКИ СУЧАСНИХ КОРПУСНИХ ДОСЛІДЖЕНЬ МОВИ ТА ПЕРСПЕКТИВИ ЇХ ПОДАЛЬШОГО РОЗВИТКУ

У статті визначено ключові напрямки сучасних корпусних досліджень. Зважаючи на те, що корпусна лінгвістика є перспективним напрямком сучасної прикладної лінгвістики, яка широко використовує квантитативні статистичні методи обробки мовного матеріалу, у дослідженні охарактеризовано специфіку корпусних досліджень різних рівнів мови. Оскільки корпусна лінгвістика поєднує методи дослідження різних наук, як лінгвістичних, так і комп'ютерних, прослідковано міждисциплінарні зв'язки корпусної лінгвістики. Лінгвістичні дослідження можливі на стику таких дисциплін як мовознавство та психологія, соціологія, педагогіка, історія та інші гуманітарні науки. На основі наявних напрацювань вітчизняних та зарубіжних лінгвістів описано методологію корпусного лінгвістичного дослідження, встановлено особливості якісного та кількісного аналізу мовних одиниць та явищ та визначено подальші перспективи лінгвістичних корпусних досліджень.

Ключові слова: корпусна лінгвістика, мовний корпус, рівні мови, анотування корпусів текстів.

Одним із пріоритетних напрямків сучасних прикладних лінгвістичних досліджень є корпусна лінгвістика. Корпусні студії зосереджуються на аналізі природної мови в умовах реального функціонування з використанням комп'ютерних технологій на основі великих за обсягом, ретельно відібраних та впорядкованих текстових корпусів [7, с. 48]. Базуючись головним чином на емпіричному підході до аналізу мовного матеріалу, корпусні дослідження дозволяють абстрагуватися від суб'єктивності дослідника й наблизитися до об'єктивного вивчення мови.

Пройшовши складний шлях від практично повного заперечення в 60–70-ті роки минулого століття до провідного напрямку науки про мову в сьогоденні, корпусна лінгвістика по праву визнається самостійною галуззю прикладного мовознавства [1, с. 122; 2, с. 89].

Нині спостерігається стрімкий розвиток корпусних студій у всьому світі, що підтверджується як величезною кількістю досліджень у галузі, так і різноманіттям корпусів, що постійно з'являються. Напрямок включає вивчення загальної теорії корпусної лінгвістики (Д. Байбер, Дж. Синклер, В. Тойберт), кореляції корпусної лінгвістики та інших лінгвістичних дисциплін (М. Вюн, В. Товберт), типології корпусів та методики інтерпретації корпусних даних (С. Гріс), розроблення загальних засад створення природних мов тощо (праці Б. Алтенберга, М. Баньки, У. Френсиса, Г. Кеннеді, Г. Ліча, А. Баранова, М. Михайлова, Р. Рикова, Л. Ричкової, С. Шарова та ін.).

Метою даної статті є опис етапів становлення корпусної лінгвістики та основних напрацювань вітчизняних та зарубіжних вчених у даній галузі, а також визначення ключових напрямків та подальших перспектив лінгвістичних корпусних досліджень.

Основою корпусної лінгвістики є розроблення теоретичних засад і практичних прийомів побудови, машинного опрацювання, експлуатації та аналізу мовних даних, оформлених як корпус текстів. Лінгвістичні корпуси становлять машинописне, стандартно організоване зібрання репрезентативних для певної мови, діалекту або іншої підмножини мов писемних або усних текстів в електронній формі, призначе-

них для лінгвістичного аналізу й опису. Це, свого роду, інформаційно-довідкова система, яка слугує базою для дослідження одиниць та явищ різних мовних рівнів (фонетичного, морфологічного, лексико-семантичного та синтаксичного), для вивчення особливостей використання природної людської мови у формах усного та писемного мовлення та для визначення специфіки функціонування мови в різних стилях (художньому, офіційно-діловому, публіцистичному, науковому, розмовному).

У сучасному мовознавстві лінгвістичні корпуси широко використовуються для вивчення особливостей та специфіки вживання в мовленні одиниць різних мовних рівнів. Аплікативне призначення корпусних даних — фонологічні, морфологічні, лексикологічні, лексикографічні, фразеологічні, синтаксичні, стилістичні, прагматичні тощо дослідження — детермінує тип лінгвістичної анотації корпусу [7, с. 216].

Фонетичні корпусні дослідження є досить обмеженими, оскільки існує невелика кількість лінгвістичних корпусів, які містять фонетичну розмітку. Серед них *корпус усної англійської мови*, що містить розшифрування магнітофонних записів громадських виступів та усної розмовної мови, а також транс кріпти кінофільмів, та *акцентологічний корпус російської мови*, який несе інформацію про історію російського наголосу, включає тексти поетичного корпусу та акцентуйовані записи усного мовлення, які доступні для пошуку за місцем наголосу та просодичною структурою слова. Останнім часом особливого поширення набуло вивчення емотивності мовлення персонажів на основі аналізу інтонаційних моделей корпусів текстів.

Вивчення **граматичної** будови мови на основі лінгвістичних корпусів включає дослідження морфології, словотвору, словозміни [9, с. 42–45]. До системи автоматичного морфологічного аналізу як одного з невід'ємних компонентів програмного забезпечення корпусу текстів входить алгоритм контекстного аналізу словоформ, завдання якого полягає у правильному визначенні граматичного значення омонімічної словоформи. При реалізації морфологічної розмітки корпусу розробляються різноманітні підходи зняття морфологічної неоднозначності.

За допомогою анотованого лінгвістичного корпусу вивчають специфіку *словозміни*, аналізують *морфологічні* показники мови, наприклад, виявляють, в якій синтаксичній функції найчастіше виступають різні частини мови, наприклад, прислівник — як предикатив (*жарко*) чи прийменник (*навколо*). Також з'ясовують, з якими відмінками найчастіше використовується той чи інший прийменник, встановлюють частотність лексичних або граматичних варіантів лексем (*предмета* — *предмету*), виявляють нетипові граматичні явища чи конструкції (напр., прийменник + іменник в Н.в. *віддати в солдати*). Лінгвістичні корпуси дозволяють отримати дані про конкретні форми слова і про цілі граматичні категорії.

Широко використовують лінгвістичні корпуси і в вивченні *словотвору* завдяки можливості здійснювати пошук та статистичний підрахунок вживання заданих коренів, афіксів і флексій [3, с. 42]. Одною із важливих особливостей методу аналізу на базі лінгвістичних корпусів є можливість визначення частотності вживання граматичних конструкцій в тих чи інших жанрах чи діалектах.

Популярними є вивчення особливостей структури та використання різних типів граматичних конструкцій, наприклад, утворення минулих та перфектних форм дієслів у різні періоди розвитку мови [5, с. 100].

Корпусні дані в більшій чи меншій мірі слугують базою для написання *граматик*, наприклад *An Empirical Grammar of the English Verb: Modal Verbs* Дієтера Міндта (1995) та *Longman Grammar of Spoken and Written English* (Д. Бібер, С. Конрад). Остання побудована на базі Логнманського лінгвістичного корпусу і має ряд переваг: включає більше 3 000 прикладів реальної англійської мови (мови носіїв) та акцентує увагу на лексико-граматичних моделях мови, їх структурі, значенні, синтаксичних функціях та способах використання в різних дискурсах.

Лінгвістичні корпуси слугують матеріалом для численних *лексикологічних* досліджень [3, с. 43], серед яких вивчення контекстів, частотності вживання та

екстралінгвістичних характеристик лексичних одиниць мови, встановлення значення слів, багатозначність, омонімія, синонімія, антонімія та інші відносини між значеннями слів, а також зміни в словниковому складі мови, відображення в мові соціальних, територіальних, професійних характеристик.

Вивчення *семантики* на базі лінгвістичних корпусів є можливим завдяки існуванню різних типів семантичної розмітки: PropBank, WordNet, FrameNet. Значення лексем визначається на основі контексту їх вживання, шляхом вивчення типових словосполучень та колокацій, тобто конкордансів. Існують також спроби здійснити семантичне анотування корпусів (наприклад, семантично анотована лексична база англійської мови *WordNet*), але матеріали таких досліджень обмежені.

В даний час в лінгвістичних корпусах російської мови реалізується система пошуку за лексико-семантичними ознаками, заснована на частковій семантичній розмітці текстів, яка передбачає, що більшості слів у тексті приписується одна або кілька семантичних і словотвірних ознак, наприклад, «обличчя», «речовина», «простір», «швидкість», «рух», «володіння», «властивість людини», «віддієслівний іменник» і т.п. Використовується фасетна класифікація, при якій одне слово може потрапляти в кілька класів. На першому етапі пошук здійснюється за частинами наявних у словнику ознак.

Лексикографічний аналіз на базі лінгвістичних корпусів допомагає проаналізувати контекстне вживання тих чи інших слів, особливо синонімічних (напр., *small / little*), їх частотну сполучуваність з іншими словами, регулярність їх вживання в різних стилях та чітко визначити їх семантику [6, с. 223].

n	_t	ru	UTF-8	2	1
---	----	----	-------	---	---

Більшість світових лексикографічних видавництв використовує для підготовки словників репрезентативні корпуси текстів, наприклад, видавництво Collins — Бірмінгемський корпус англійської мови (1987).

За допомогою лінгвістичних корпусів досліджують різні синтаксичні явища й конструкції: вивчають валентність дієслів, типи прийменникового керування, прямий і зворотний порядок слів, синтаксичні функції різних граматичних категорій (напр., функції інфінітиву в реченні, функції сполучників на початку і в середині речення). Корпуси дозволяють отримати дані про використання та функціях різних розділових знаків. Результатом корпусного аналізу структури складного речення є *Clause patterns in Modern British English: A corpus-based (quantitative) study* by N. Oostdijk and P. de Haan (1994).

Багато праць з корпусної граматики та синтаксису базуються на дослідженні інтуїції мовців та їх компетентності [11, с. 99]. Наявність багатомільйонних лінгвістичних корпусів полегшила вивчення мови в плані її переходу від індивідуального продукування до вироблення загальномовних норм-правил, адже кожне граматичне правило спочатку розглядалося на основі прикладів побудови граматичних структур та словоформ окремими мовцями.

Лінгвістичні корпуси не часто використовуються для аналізу дискурсу або здійснення *прагматичних досліджень*. Одним із пояснень цього є те, що досить важко знайти матеріал, який підходить для такого роду досліджень. Більшість наявних корпусів не містять інформацію про конкретні мовленнєві ситуації виголошення текстів, невербальні засоби, що їх супроводжують та прагматичні наміри мовців у мовних актах [8, с.3]. Проте у даний час складається та анотується все більше корпусів, придатних для дослідження в сфері прагматики. Прикладом таких досліджень можуть роботи норвезьких вчених в Бергені, на основі корпусу мови лондонських підлітків (*corpus of London teenage language, COLT*).

З позицій прагматики лінгвістичний корпус розглядається як зібрання зразкових текстів, орієнтованих на засвоєння і використання їх як прототипу для

подальших мовних дій в певних умовах комунікації. Таким чином, корпус текстів, орієнтований на певну прагматичну мету, або прагматично орієнтований корпус текстів може розглядатися як спосіб вирішення проблеми оптимізації мовних відносин в сучасному суспільстві [9, с. 89]. Прагматична ціль, що лежить в основі побудови корпусу, полягає не лише в навчанні правильній комунікації, але і має широкий спектр наукових досліджень та застосувань, зокрема визначення типових вербальних засобів, характерних для певних видів дискурсів.

Наявність лінгвістичних корпусів з великими підбірками текстів різних жанрів, авторів та медіа відкриває нові можливості в галузі досліджень стилістики. В останні роки, корпусна стилістика стала визнаною частиною корпусної лінгвістики [7, с. 223]. Тексти різних видів порівнюють один з одним, щоб виявити їх типові характеристики та жанрові особливості. Загалом корпус може слугувати мета системою для порівняння інших текстів. Однією з областей, де корпусна стилістика має широке використання є вивчення ідеостилію автора на основі корпусів авторських текстів. Статистичні методи застосовуються для дослідження змін стилю автора, встановлення специфіки використання виражальних засобів та стилістичних прийомів у художніх текстах.

Корпусна стилістика дозволяє здійснювати аналіз літературних текстів, використовуючи лінгвістичні корпуси. Вона поєднує в собі аналітичні методи корпусної лінгвістики та завдання стилістики, аналізуючи значення літературних текстів за допомогою обчислювальних засобів та проводячи стилістичний аналіз текстів.

Лінгвістичні корпуси характеризуються *синхронністю* та паралельною багатоплановістю, а інколи і мультимодальністю. Синхронність найкраще описується як особливий аспект *репрезентативності* одиниць всіх рівнів мови, який є ключовим аспектом корпусної лінгвістики. Екстраполявання спостережень корпусних досліджень до певної області чи рівня мови є виправданий тільки, коли корпус має достатню репрезентативну вибірку в цій галузі. Однак, оскільки для більшості рівнів мови репрезентативність лінгвістичного корпусу на практиці не може бути оцінена задовільним чином, вибірка лінгвістичного корпусу зазвичай прагне наблизитись до абсолютної репрезентативності, інтуїтивно оцінити деякі якісні та кількісні характеристики відповідного домену мови і тому вимагає від корпусу відображення цих властивостей [10, с. 34]. Таким чином, лінгвістичний корпус набуває ознаки збалансованості, або точніше, є збалансованим у відношенні до оцінок властивостей мовних одиниць.

Корпусна лінгвістика пропонує нові можливості для численних інтегрованих з лінгвістикою дисциплін, зокрема психолінгвістики, соціолінгвістики та історичної лінгвістики.

Корпусна лінгвістика здійснює статистичний аналіз частотності вживання мовних одиниць, що може слугувати також і показником процесів сприйняття, відтворення та передачі мовних даних [5, с. 99]. Так, частотність вживання лексем безпосередньо пов'язана з легкістю / складністю їх сприйняття, а вибір синонімічних слів у різних контекстах — з механізмами категоризації та відображенням зв'язків між елементами реального світу. Частотність вживання граматичних чи синтаксичних конструкцій обумовлена їх продуктивністю / непродуктивністю в мовному відтворенні.

В той час, як корпусна лінгвістика наполягає на окремому вивченні лексики та синтаксису, вчені корпусної психолінгвістики [1, с. 121] роблять припущення, що слова і синтаксичні моделі представляються як якісно однакові вузли системи, що забезпечують відповідність семантичному та прагматичному значенню висловлення.

Загальновизнаним є той факт, що мова продукується та базується на основі мисленнєвих процесів. Це твердження дозволяє використовувати лінгвістичні корпуси для вивчення порушень мисленнєвих процесів, патологій та відхилень. Для того щоб проаналізувати особливості порушення мови важливо мати дуже чітке уявлення про структурні і формальних відмінності між порушеннями і правильними формами.

Із створенням лінгвістичних корпусів, які містять соціолінгвістичну інформацію про мовців і / або авторів текстів, виникла можливість використання корпусів для соціолінгвістичних досліджень. Британський національний корпус був широко анотований щодо різних соціолінгвістичних параметрів, таких як вік, стать, соціальний стан, місце проживання письменників і т.п. Ця інформація використовувалася в ряді досліджень, наприклад, Полем Рейсоном *Social Differentiation in the Use of English Vocabulary: Some Analyses of the Conversational Component of the British National Corpus* [2, с. 89]. Інший корпус з соціолінгвістичним анотуванням — це COLT — Корпус мови підлітків Лондона (наприклад, праця *Girls' conflict talk: a sociolinguistic investigation of variation in the verbal disputes of adolescent females by A-B Stenstrom and I.K. Hasund*).

У 80-ті роки ХХ століття із покращенням технічних показників комп'ютерної техніки та збільшенням електронних варіантів історичних та літературних текстів попередніх епох, значного поширення зазнали корпусні дослідження в рамках історичної лінгвістики. Можливість отримання репрезентативної вибірки мови на різних етапах історії в машинописній формі дозволила історикам мови швидше і ефективніше проводити їх діахронічні дослідження. Отримані результати характеризуються більшою системністю та всеохопністю, дозволяють прослідкувати динаміку мовних змін, встановити закономірності розвитку національної мови та виявити специфіку функціонування діалектів. *Гельсінський корпус* являє собою відомий і часто використовуваний корпус текстів різних періодів. *Корпус Lampeter* ранньо-новоанглійської мови містить добірку брошур, виданих між 1640 і 1740 роками.

Сьогодні лінгвістичні корпуси письмових чи усних текстів успішно використовуються для навчання іноземним мовам та в лінгвістичній педагогіці. На базі корпусів формуються списки автентичної активної лексики студентів, частотні списки термінів для використання в професійних курсах [6, с. 231]. Практика викладання іноземної мови базується на правилах і зразках реального вживання тих чи інших граматичних конструкцій. На відміну від традиційних граматик, корпусні враховують стильову специфіку вживання лексичних одиниць та граматичних конструкцій, що є надзвичайно важливим для людей, що вивчають мову.

Міжнародний корпус учнів англійської мови (International Corpus of Learner English) містить мовні тексти осіб, що вивчають мову, як іноземну, та має на меті допомагати педагогам відслідковувати та аналізувати типові помилки, які роблять учні під час навчання.

Отже, нині корпусна лінгвістика має широке використання не лише в рамках лінгвістики, а й в інших галузях гуманітарної науки, зокрема соціології, психології, педагогіці та ін.

Створення, розвиток і використання електронних лінгвістичних корпусів — це один із найбільш передових та перспективних напрямків сучасної лінгвістики. Саме в рамках цього напрямку найбільш вірогідні інноваційні результати як в галузі теоретичної лінгвістики (одержання нових знань про структуру мови), так і в області прикладної лінгвістики (розвиток технологій нового покоління для автоматичної обробки текстів, машинного перекладу [9, с. 223] і модернізації методів лінгвістичних досліджень).

Дослідники вже розробили ряд принципів конструювання та використання лінгвістичних корпусів для забезпечення проведення успішних лінгвістичних досліджень мовних даних [11, с. 46]. На часі створення все більш досконалої та розвиненої системотехніки корпусів, створення віртуальних систем професійної взаємодії в лінгвістиці, розширення та вдосконалення інтелектуальних функцій такого знаряддя мовознавчого дослідження як лінгвістичний корпус.

Перспективними є корпусні дослідження мови на стику дисциплін, а саме вивчення когнітивних механізмів, емотивності, комунікативної спрямованості,

соціальної визначеності та історичної зумовленості вживання та функціонування мовних одиниць та явищ на базі лінгвістичних корпусів.

Певною мірою корпусні студії змінюють пріоритети сучасних лінгвістичних досліджень і демонструють виразну переорієнтацію об'єкта дослідження з «системи» на «узус», з «мови» на «мовлення».

Ранні корпусні дослідження семантики багатозначних мовних елементів — морфем, слів, синтаксичних конструкцій — базувалися на побудові семантичної бази значень, що критикувалася за свій суб'єктивний характер. Нещодавно, із появою семантично анотованих лінгвістичних корпусів, когнітивна семантика почала використовувати метод корпусного аналізу. Досліджень у цій галузі ще не достатньо, але це є перспективним напрямком, адже значення одиниць мови виводиться на основі їх сполучуваності та частотності вживання в певних мовних контекстах. Такі дослідження також сприяють втіленню новітніх підходів до лексикографії та лінгвістичної педагогіки.

Об'єктом постійних дискусій серед корпусних лінгвістів є розмір зразка (прикладу), а також саме питання про доцільність складання корпусів текстів із зразків. На сьогодні домінує тенденція до складання корпусів з повних текстів [7, с. 281–284]. Багато дослідників констатують, що робота з фрагментами заданої величини накладає певні обмеження при дослідженні тексту як єдиного цілого, зв'язності тексту тощо. Крім того, на основі повнотекстового корпусу достатньо легко генерується традиційний корпус зразків, у той час, як обернена операція неможлива [10, с. 23].

Тенденція до складання корпусів повних текстів створює нові перспективи для вивчення стилістики тексту, визначення ознак різних типів дискурсу та виявлення прагматичних особливостей мовної комунікації.

Останніми інноваціями в корпусній лінгвістиці є створення мультилінгвальних, паралельних та мультимодальних корпусів, наприклад, MULTEXT (Multilingual Text Tools and Corpora). Використовуючи напрацювання мовних корпусів та спеціального програмного забезпечення, розробники цього корпусу паралельно аналізують вербальні тексти та аудіо дані 6 європейських мов. Увесь корпус маркований за логічними та структурними рисами, підгрупи даних містять маркування граматичних та семантичних ознак у межах текстів. Корпус сприяє проведенню порівняльних лінгвістичних досліджень різних мов та вивченню різних засобів передачі та відображення даних об'єктивної реальності в мові.

Все ж, необхідно пам'ятати, що незважаючи на революційні можливості корпусної лінгвістики, вона є лише частиною численного методологічного інструментарію сучасної лінгвістики. Навіть великі за обсягом корпуси не в змозі відобразити все можливе в мові, а натомість і незначні за обсягом корпуси можуть надати інформацію, яку неможливо було б отримати не звертаючись до корпусного дослідження.

ЛІТЕРАТУРА

1. Демська-Кульчицька О. Основи національного корпусу української мови : монографія / Оріся Демська-Кульчицька. — К.: Інститут української мови НАН України, 2005. — 218 с.
2. Рыков В.В. Прагматически ориентированный корпус текстов / В.В. Рыков // Тверской лингвистический меридиан. — 1999. — Вып. 3. — С. 89–96.
3. Соснина Е.П. Перспективы корпусного подхода к исследованию грамматических аспектов иностранного языка // Вопросы теории языка и методики преподавания иностранных языков: сборник трудов международной научной конференции. — Изд-во Таганрог. гос. пед. ин-та, 2007. — С. 42–45.
4. Широков В.А. Корпусна лінгвістика / В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна та ін. — К.: Довіра, 2005. — 471 с.
5. Aarts J. Theory and Practice in Corpus Linguistics / Jan Aarts, Willem Meijs // Journal Computational Linguistics. — 1991. — Volume 17. — Issue. — P. 99–103.

6. *Baker M.* Corpora in translation studies. An overview and suggestions for future research / Mona Baker // *Target*. — 1995. — № 7(2) . — P. 223–243.
7. *Biber D.* Corpus linguistics: Investigating language structure and use / D. Biber, S. Conrad, R. Reppen. — Cambridge: Cambridge University Press, 2001. — 312 p.
8. *McCarthy M.* What are corpora and how have they evolved? / M. McCarthy, A. O’Keefe // *The Routledge Handbook of Corpus Linguistics*. — Routledge, 2010. — P. 3-14.
9. *McEnery T.* Corpus linguistics / T. McEnery, A. Wilson — Edinburgh: Edinburgh University Press, 2001. — 235 p.
10. *Meyer Ch.* English corpus linguistics. An introduction / Charles F. Meyer. — Cambridge : Cambridge University Press. — 2002. — 188 p.
11. *Sinclair J.* Corpus, Concordance, Collocation / J. Sinclair — Oxford: Oxford University Press, 1991. — 170 p.

Г.Г. Лукьянец

Основные направления современных корпусных исследований языка и перспективы их дальнейшего развития

В статье определены ключевые направления современных корпусных исследований. Учитывая то, что корпусная лингвистика является перспективным направлением современной прикладной лингвистики, которая широко использует количественные статистические методы обработки языкового материала, в исследовании охарактеризовано специфику корпусных исследований единиц различных уровней языка. Поскольку корпусная лингвистика объединяет методы исследования различных наук как лингвистических, так и компьютерных, прослежены междисциплинарные связи корпусной лингвистики. Лингвистические исследования возможны на стыке таких дисциплин как: языкознание и психология, социология, педагогика, история и другие гуманитарные науки. На основе имеющихся наработок отечественных и зарубежных лингвистов описано методологию корпусного лингвистического исследования, установлены особенности качественного и количественного анализа языковых единиц и явлений, также определены дальнейшие перспективы лингвистических корпусных исследований.

Ключевые слова: корпусная лингвистика, языковой корпус, уровни языка, аннотирования корпусов текстов.

G. Lukyanets

Main directions of contemporary corpus linguistic researches and perspectives of further development

The article outlines the key areas of contemporary corpus linguistics. Presupposing that corpus linguistics is a promising area of modern applied linguistics that makes extensive use of quantitative statistical methods of processing linguistic material, the study characterized the specificity of corpora research of different language levels. Since corpus linguistics combines research methods of various sciences, both linguistic and computer ones, the article retraces interdisciplinary linguistic peculiarities. Linguistic research is possible at the intersection of such disciplines as linguistics and psychology, sociology, pedagogy, history and other humanities. Based on the current developments of domestic and foreign linguists the methodology of linguistic corpus research is described. Moreover, the peculiarities of qualitative and quantitative analysis of linguistic units are specialized and future perspectives of linguistic corpus research are defined.

Key words: corpus linguistics, text corpus, language level, annotation of corpora.

e-mail: jimp@ukr.net

Одержана редколегією 21.03.2012 р.