

КОМП'ЮТЕРНА ЛІНГВІСТИКА

УДК 81.32

Тетяна Бобкова, канд. філол. наук, доц.
Київський національний лінгвістичний університет, Київ

КОРПУСНИЙ ПІДХІД ДОСЛІДЖЕННЯ КОЛОКАЦІЙ

У статті аналізуються основні напрямки й підходи сучасних досліджень колокацій. Виявляються передумови формування корпусного підходу. Обґрунтовується необхідність застосування багатометодного підходу дослідження колокацій на базі корпусу текстів. Як джерело для укладання корпусного словника колокацій розглядається зібраний підкорпус українських законодавчих документів.

Ключові слова: колокація, словосполучення, корпус текстів, корпусний підхід, корпусний словник..

Актуальність дослідження колокацій як прояву лексико-синтаксичної сполучуваності пов'язана із загальною проблематикою автоматичного аналізу тексту. Розробка й удосконалення прикладних систем аналізу тексту потребують дослідження великих обсягів природно-мовних текстів. У процесі комунікації використовуються не окремі слова, а регулярно повторювані відтворювані конструкції – колокації, опис яких становить значний інтерес для прикладних лінгвістів. Сучасні дослідження природної мови в аспекті автоматичного аналізу тексту ґрунтуються на об'єктивних корпусних даних, вилучених і опрацьованих надійними й достовірними методами.

Мета статті – дослідити витоки й особливості корпусного підходу вивчення колокацій. Досягнення поставленої мети передбачає виконання таких завдань: 1) окреслити основні теоретико-методологічні напрямки й лінгвістичні підходи сучасних досліджень колокацій; 2) встановити теоретико-методологічні передумови корпусного підходу вивчення колокацій; 3) визначити основні відмінності корпусного підходу від текстоорієнтованого; 4) дати визначення терміна “колокація” з огляду на корпусний підхід дослідження; 5) обґрунтувати необхідність застосування багатометодного вивчення колокацій.

Основним джерелом та інструментом дослідження структур природно-мовного тексту є електронні корпуси. На позначення стійких синтагматичних послідовностей у корпусній лінгвістиці традиційно вживається термін “колокація” [Sinclair; Teubert; Tognini-Bonelli].

З самого початку формування нової корпусної парадигми феномен колокації стає ключовою концепцією, а опис зазначених “видів лексичних структур” – основною метою розробки корпусів [Teubert : 137–138]. Про це свідчить перша ґрунтовна праця в галузі корпусного аналізу – “English Collocation Studies”, 1966 (Дж. Синклер, С. Джоунз, Р. Делей). Упровадження корпусних методів призвели до поширення терміна “колокація” в багатьох галузях лінгвістики – лексикології, дериватології, синтаксисі, психолінгвістиці, лексикографії, теорії перекладу й лінгвістиці тексту [Marcinkevičienė : 77]. Ця поширеність пояснюється тими значеннями, які вкладають у поняття колокації різні лінгвісти [Meer : 315]. Безперечно, термін не є одновимірним хоч би тому, що покликаний відобразити комплекс лексичних, граматичних і статистичних ознак колокації. Ця багатоаспектність феномену протягом тривалого часу утримує увагу лінгвістів, а у зв’язку з упровадженням корпусного підходу концепція колокації набуває нового виміру в сучасній лінгвістиці.

Дискусійний характер визначення терміна “колокація” спонукав до аспектизації аналізованого поняття й формулювання широкого значення – в лінгвістичному й вузького – в статистичному аспекті [Čermak : 282–283]. На сучасному етапі в широкому сенсі під колокацією розуміють поєднання двох слів або звичне використання певних лексем разом [Лендау : 303; Sinclair : 115–116]. У вузькому значенні колокація – це слова, які зустрічаються в тексті частіше поруч, ніж за випадковою вірогідністю окремо [Sinclair : 115–116]. Залежно від широкого або вузького розуміння поняття визначається якісний і кількісний склад колокацій [Гладка : 18]. Зокрема, чи можна вважати колокацією сполучення лексичної одиниці зі службовим словом. Потрійна природа феномену визначає широкий спектр напрямків прикладних досліджень колокацій:

1) вивчення природи й визначення статусу колокацій серед сталих сполучень (О. Борисова, А. Романюк, К. Ягунова, В. Atkins, S. Bartsch, F. Hausmann, A. Kilgarriff, R. Marcinkevičienė, J. Sinclair);

2) розробка й удосконалення методології вилучення колокацій та вивчення їх статистичних і лексико-граматичних ознак (М. Зацеркляний, А. Романюк, М. Хохлова, В. Шкурко, К. Ягунова, К. Church, A. Kilgarriff, M. Kren, Ch. Manning, V. Seretan, F. Smadja, M. Stubbs);

3) укладання спеціалізованих словників колокацій (І. Большаков, О. Борисова, М. Зацеркляний, А. Романюк, М. Benson, F. Čermak, A.P. Cowie, F. J. Hausmann, U. Heid, R. Marcinkevičienė, С. Müller-Spitzer).

Розглянемо докладніше основні лінгвістичні підходи в дослідженні колокацій – семантико-синтаксичний, контекстно-орієнтований і корпусно-орієнтований [Гладка : 16; Хохлова : 344].

Семантико-синтаксичний підхід передбачає аналіз феномену колокації в аспекті теорії стійких сполучень і граматики валентності. У межах семантико-синтаксичного аспекту слід виокремити власне лексико-семантичний, семантико-синтаксичний і синтаксичний підходи.

Лексико-семантичний підхід є найбільш традиційним, оскільки в лінгвістичних дослідженнях із визначення колокації частіше постулюється “лексичний пріоритет” [Marcinkevičienė : 88]. Під колокацією розуміють сполучення слів, у складі якого один компонент (базовий або зв’язаний) визначається змістом і темою повідомлення й фразеологізує сполучення, а другий – вільний компонент характеризує базове слово і є семантично ключовим [Борисова 1994 : 77]. До основних здобутків лексико-семантичного підходу вивчення сполучень можна віднести розробку класифікації фразеологізмів на підставі семантичної структури (Ш. Балли, В. Виноградов), категорії зв’язності значення (В. Виноградов), поняття лексичної функції і класифікації лексико-семантичних відношень (Ю. Апресян, І. Мельчук), побудову функціональної лінгвістичної моделі “Смисл↔Текст” (І. Мельчук).

Семантико-синтаксичний підхід (О. Борисова, О. Митрофанова, М. Хохлова, F. Hausmann) передбачає розгляд колокацій як семантико-синтаксичних одиниць, або лексично визначених елементів граматичних структур [Хохлова : 344]. Семантико-синтаксичний підхід зорієнтований на вивчення колокацій через установлені лексико-граматичні шаблони – структурні моделі мовних конструкцій із зазначенням істотних граматичних ознак лексем, які входять до мовних виразів певного класу, й синтаксичних умов їх уживання. В українському мовознавстві дослідницькі пошуки в основному спрямовано на коло питань, що стосуються системного вивчення лексико-синтаксичної сполучуваності, взаємодії внутрішніх форм слів, а також сполучуваності в аспекті лексико-морфологічної та синтаксичної деривації (І. Вихованець, К. Городенська, В. Горпинич, Н. Гуїванюк, А. Загітко, Н. Клименко, М. Кочерган, О. Снитко).

Синтаксичний підхід базується на звуженому понятті колокації як синтагматично обмеженої лексичної сполучуваності двох елементів, які підтримують синтаксичний зв’язок (В. Гладка, В. Шкурко, D. Cruse, A. Tutin, V. Seretan). При цьому враховуються дві основні ознаки колокацій – дотримання певного порядку слідування компонентів і відсутність синтаксичних трансформацій, що відрізняє колокацію

від вільних сполучень. У цьому розумінні колокація може бути визначена як комбінація з двох слів, що утворює синтаксичну структуру, або синтаксичну модель [Гладка : 17]. За синтаксичним підходом основною метою є встановлення синтаксичних та структурних ознак колокацій.

Витоки **контекстно-орієнтованого** підходу коріняться в емпіричній текстозорієнтованій традиції британського контекстуалізму Дж. Ферса й обґрунтуванні контекстуалізації як прийому синтагматичного аналізу природної мови [Marcinkevičienė : 75]. Результатом контекстуалізації на лексичному рівні є встановлення колокацій – типового й постійного оточення слова [Хохлова : 344]. Ферсіанський підхід, або контекстуальна теорія значення, ґрунтується на визнанні колокації цілісною одиницею значення, дослідження якої передбачає виокремлення складників. Під колокаціями розуміють характерні стійкі сполучення слів, поява яких поруч зумовлюється регулярним характером взаємного очікування внаслідок дії семантичних факторів. У цьому розумінні основним здобутком контекстуалізму Дж. Р. Ферса є встановлення пов'язаних між собою формальних ознак контексту: 1) колокація – супровід слова, матеріал, у якому воно найчастіше зустрічається, і 2) колігація – взаємозв'язок граматичних категорій в синтаксичній структурі [Tognini-Bonelli : 209]. Окреслений вище ферсіанський підхід застосовується в рамках контекстуальної теорії значення, за якою ідентифікація значення слова або виразу базується на формалізації контекстної моделі.

Хоч контекстуальна теорія значення формувалась у межах текстозорієнтованого підходу та з метою з'ясування текстових ознак, встановлені Дж. Ферсом характеристики можуть бути також використані в корпусному підході вивчення колокацій. Так, функціональні можливості вбудованого в корпус пошукового апарату, зокрема KWIC-конкордансу, є оптимальним інструментом ідентифікації формальних моделей колокацій і колігацій на матеріалі великої кількості текстів. Незважаючи на декларування слідом за Дж. Ферсом текстозорієнтованого підходу, слід зауважити, що текст і корпус відрізняється за кількома істотними параметрами [Tognini-Bonelli : 207–208]:

1) єдність тексту як мовної події між учасниками комунікації vs. зібрання текстів у корпусі, яке неможливо ідентифікувати як єдину мовну подію;

2) цілісність тексту vs. фрагментарність KWIC-формату конкордансу;

3) горизонтальний напрямок аналізу тексту vs. вертикальний напрямок KWIC-формату конкордансу корпусу;

4) специфічний ситуативний і культурний контекст тексту vs. відсутність унікального контексту корпусу, формальний характер аналізованих параметрів;

5) текст як мовленнєвий зразок vs. корпус як аплікативний зразок мови;

6) змістовний характер текстової інформації vs. узагальнений на мовному матеріалі характер інформації корпусу, відсутність безпосереднього зв'язку з конкретним контекстом.

Отже, попри те, що в основу корпусного підходу покладено засади текстозорієнтованого, корпус і текст принципово відрізняється за якісними (цілісність, наявність ситуативного контексту, характер інформації) і кількісними ознаками (обсяг, кількість текстів), що є визначальними для аналізу й методології дослідження мовного матеріалу. На думку Е. Тоніні-Бонеллі, це дозволяє стверджувати, що в сучасній лінгвістиці використовуються два типи емпіричних даних: текст як зразок мовлення й корпус як потенційна мовна модель [Tognini-Bonelli : 208–212]. Відмінність ознак корпусу вимагає застосування нових методів й уточнення визначення колокації, оскільки якість та обсяг отриманих даних значно виходять за межі тексту.

Упровадження корпусних методів і статистичних обчислень дало змогу зробити теоретичні узагальнення. Зокрема, на підставі емпіричних даних спостереження корпусів встановлено, що 80 % тексту складають не ізольовані лексичні одиниці, а регулярно відтворювані структури, які є важливим будівельним мовним матеріалом [Marcinkevičienė : 62]. Узагальнення спостережень великої кількості текстів дозволило дійти висновків, що породження текстової інформації на рівні структури й значення зумовлено дією двох взаємодоповнювальних принципів ідіоматичності: використання напівпобудованих фраз для одноразового вибору мовцем і необмеженого вибору, покладеного в основу граматики [Sinclair : 109–110]. Найбільш конкретним проявом принципу ідіоматичності, або фразеологічності, у природно-мовному тексті вважаються колокації.

Завдяки опрацюванню мовних даних статистичними методами сформульована в межах контекстуалізму концепція колокації [McIntoshas 1966; Halliday 1966; Sinclair 1991] набула нового виміру в корпусному підході. Заслугою Дж. Синклера було застосування окреслених вище засад теорії Дж. Ферса в лексикографічній практиці (COBUILD, 1980), що власне й завершило виокремлення напрямку нео-ферсіанства в корпусній лінгвістиці [Teubert : 137]. На підставі спостереження корпусних даних щодо невідповідності меж значення і лексичної одиниці Дж. Синклером розроблено “модель значення”, суть якої полягає

в “перевизначенні одиниць значення у світлі корпусних доказів” [Tognini-Bonelli : 214]. Значення не зосереджується в межах однієї лексичної одиниці, воно поширюється на послідовність слів у словосполученні, а колокація розуміється як лінгвістична одиниця, що регулярно й послідовно спостерігається в корпусі.

Теоретичні інновації Дж. Синклера в галузі корпусної лексикографії відкрили шлях до нового розуміння словосполучення в цілому [Marcinkevičienė : 44] й уможливили теоретико-методологічні засади нового підходу вивчення колокацій. Отже, у західноєвропейській лінгвістиці передумовою дослідження колокацій стали емпіричні традиції британського контекстуалізму й корпусний метод аналізу мовного матеріалу. В українському мовознавстві дослідження колокацій ґрунтуються в основному на засадах семантико-синтаксичного підходу. За ступенем фразеологізації елементів колокації дістали проміжний статус між вільними й фразеологізованими сполуками, їх визначають як “невільні сполучення”, або “фразеологічні звороти, значення яких з погляду сучасної мови не мотивується, не виводиться із суми значень слів-компонентів” [Дарчук : 120].

Запропонована Дж. Синклером концепція колокації стає для європейських корпусних лінгвістів моделлю, яка не тільки обґрунтовує використання, а й теорію природної мови [Marcinkevičienė : 62]. Фактично, Дж. Синклером встановлено корпусну модель колокації та процедури її опису. Однак багатоаспектність феномена призвела до формування в наступні роки “достатньо конфліктного набору дефініцій”, що навряд чи сприяло з’ясуванню природи колокації і власне визначенню терміна [Marcinkevičienė : 89]. Для вирішення цієї проблеми було запропоновано два шляхи: або відмовитися від використання невизначеного терміна, або спиратися на звужену концепцію колокації [Šerņak : 282–283].

Слід зауважити, що корпусний підхід потребує визначення колокації не в теоретичному аспекті, а з огляду на потреби використання статистичного апарату й лексикографічної практики [Marcinkevičienė : 89]. Застосування статистичних методів вимагає “точного, по можливості формального, структурного визначення” лінгвістичної одиниці підрахунку, що спирається на “ознаки і критерії, які б максимально звільняли від суб’єктивізму і неперевірюваності” [Перебийніс : 9]. А формування статті корпусного словника базується на принципах включення типових випадків, які ілюструють слововживання певної одиниці [Kilgarriff : 115; Marcinkevičienė : 89].

Отже, ферсіанська модель, встановлення якої починається з формального визначення текстової функції, а завершується висновками про значення, набуває спрощення в корпусному підході, в основу якого покладено статистичні ознаки колокацій. Слідом за Дж. Синклером [Sinclair : 57], під колокацією розуміється комбінація слів, що зустрілася в тексті принаймні двічі. При цьому обсяг тексту обмежується мільйоном слововживань, оскільки більший обсяг вимагає більшої кількості виявів і обов'язкового зазначення, що це колокації, які зустрічаються разом частіше [Marcinkevičienė : 104].

На сучасному етапі розвитку корпусної лінгвістики колокація відіграє роль основної одиниці значення, ужитку [Tognini-Bonelli : 206] й сприйняття [Большакова : 22], яку неможливо описати традиційними лінгвістичними методами через необхідність урахування потрійної природи феномену. Функціональні можливості корпус-менеджера корпусу текстів дозволяють виявити зв'язки лексичної одиниці одночасно в кількох аспектах й отримати реальні дані щодо статистичних, морфологічних і синтаксичних характеристик на великому репрезентативному матеріалі. На відміну від словників і цитатних картотек, корпуси текстів дають можливість не тільки виявити колокації, а й оцінити ступінь зв'язку компонентів [Лендау : 303; Романюк : 162]. Слід відзначити, що основною перевагою корпусного підходу є одночасне представлення індивідуальних зразків мовлення на горизонтальній вісі й соціальної практики слововживання у повторюваних моделях на вертикальній вісі KWIC-конкордансу.

На застосуванні статистичних і корпусних методів ґрунтується укладання словників колокацій, створених українськими розробниками для автоматичного аналізу природно-мовного тексту: лексикону багатослівних сполук [Романюк : 163], словника колокацій для бази знань криміналістичних інформаційних систем [Зацеркляний : 184]. При цьому увагу дослідників зосереджено не на розв'язанні лінгвістичних проблем фразеологізації еквівалентних слову стійких сполучень, а на ефективності автоматичного виявлення зазначених сполук у текстах [Зацеркляний : 184; Романюк : 158–159].

Однак застосування лише корпусного підходу для опису природно-мовних структур, безперечно, є недостатнім: виявлені зазвичай невидимі, очікувані й характерні моделі лише окреслюють контури теорії мови [Мцппінк : 227; Teubert : 144]. Поєднання корпусних методів із лінгвістичними створює перспективи виходу не тільки за межі простої статистики, а й деяких традиційних уявлень про мову [Marcinkevičienė : 53–57]. Саме тому вивчення колокацій має ґрунтуватися на багатоме-

тодному підході – використанні корпусного методу як доповнення до традиційних лінгвістичних, а не як єдино вірного методу дослідження [Міцнінк]. Так, статистичне опрацювання виявлених у корпусі колокацій необхідно доповнити лінгвістичним аналізом граматичних характеристик кожного з членів словосполучення, встановленням моделей і типів синтаксичного зв'язку [Дарчук : 177–180], а також лексико-граматичних і семантичних ознак через визначення семантичних преференцій і просодії [Tognini-Bonelli : 215]. Отже, методика лексикографічного опису колокацій передбачає вилучення колокацій через застосування корпусного методу, а для аналізу їх лексико-граматичних і семантичних ознак – дистрибутивного, контекстологічного аналізу й моделювання.

Як було зазначено вище, укладання корпусних словників ґрунтується на безпосередньому спостереженні корпусу й застосуванні програмних засобів опрацювання текстів. Загальна методична схема укладання корпусного словника [Marcinkevičienė : 106] відображає основні етапи дослідницької діяльності лексикографа: корпус → конкорданс → типова модель використання одиниці → словник. При цьому під типовою моделлю розуміється колокація – звичний контекст слова, встановлюваний за конкордансом. На етапі конкордансу автоматичний аналіз корпусу закінчується, й починається робота лінгвіста з виявлення типових сполучень, синтаксичних структур, різних значень багатозначного слова [Sinclair : 105–106]. Отже, в укладанні корпусного словника колокацій є два основні етапи: 1) створення лексикографічного корпусу текстів і 2) формування й редагування реєстру на основі статистичних і лексико-синтаксичних характеристик колокацій.

Проаналізуємо основні принципи планування підкорпусу законодавчих документів, зібраного для укладання корпусного словника колокацій українського юридичного дискурсу. У сучасній корпусній лінгвістиці підкорпус вважається релевантним, тобто достовірним і адекватним для дослідження певної форми мовлення, функціонального стилю, різновиду мови й окремих мовних явищ, дослідження яких утруднено на матеріалі національного корпусу текстів. Підкорпус законодавчих документів є складовою Корпусу текстів сучасної української мови. Фіксований обсяг підкорпусу (1,25 млн. слів) дозволяє заощадити зусилля при вивченні предметної галузі й детермінувати поріг її відображення [Большакова : 116–117; Лендау : 332].

Укладання корпусного словника колокацій українського юридичного дискурсу передбачає формування реєстру на підставі статистичних ознак: колокація розуміється як комбінація двох слів, зафіксована в тексті обсягом в 1 млн. слововживань принаймні двічі [Marcinkevičienė :

104]. У цьому разі генеральною сукупністю для зональної вибірки слугує масив підкорпусу законодавчих документів, а зоною – “сукупність одиниць, встановлена за певною ознакою” [Перебийніс : 14], тобто за частотою появи сполучення в текстах підкорпусу. В основу планування підкорпусу законодавчих документів покладено детермінативні параметри скінченності обсягу, репрезентативності, збалансованості, лінгвістичної та хронологічної однорідності, анованості й повноти текстів, що становлять ключові вимоги до корпусних об’єктів.

Документальним джерелом підкорпусу є “Зібрання законодавства України”. Вибір інформативно насичених законодавчих документів зумовлений метою укладання корпусного словника: тексти офіційних документів характеризуються максимальною кількістю колокацій на тисячу слів [Борисова 1995 : 17]. Стиль законодавчих документів вимагає чіткості, точності, стандартності формулювань, що виражається в частотності різного роду стійких сполучень: найменувань – *асамблея союзу, Гаазька угода, Паризька конвенція*; терміносполучень – *міжнародна реєстрація, промисловий зразок, юридична особа*; мовленнєвих формул і кліше – *перешкоджати застосуванню, в установленому порядку*; похідних прийменників і сполучників – *у відповідності з, у силу того, що*.

Укладений підкорпус законодавчих документів є одномовним й однорідним в аспекті функціонального стилю включених текстів. При відборі законодавчих документів виявлено, що довжина текстів різних жанрів є нерівномірною і коливається від 2 слів (*Цілком таємно*) до 44,020 тис. слів. Незначний згідно з сучасними корпусними вимогами обсяг підкорпусу збалансовано за діапазоном текстових категорій: лінгвістичний матеріал представлено сорока трьома жанровими різновидами “Зібрання законодавства України”. Відповідно до кінцевої мети розробки підкорпусу – укладання корпусного словника колокацій у ході відбору перевага надавалася повнотекстовим документам обсягом не менше 3 тис. слововживань. Кожен із жанрів презентований найбільшими за обсягом повними текстами законодавчих документів, що забезпечує їх структурну, лексичну й синтаксичну завершеність.

Із метою дотримання принципу хронологічної однорідності вибірки й максимального представлення сучасного українського юридичного дискурсу хронологічні рамки підкорпусу обмежено з 1991 р. і по теперішній час. Це пояснюється екстралінгвістичними факторами – проголошенням державної незалежності України і, як наслідок цього, – активним розвитком законодавства. Відповідний обсяг, хронологічна однорідність і збалансованість за жанрами забезпечують “сутнісну

норму репрезентативності” [Лендау : 325–326] підкорпусу законодавчих документів. Підкорпус інтегровано в інформаційно-пошукову систему Корпусу текстів сучасної української мови й забезпечено набором програм морфологічної, синтаксичної і лексико-семантичної розмітки, які слугують інструментарієм для встановлення ознак колокацій. Розроблений підкорпус знаходиться у вільному для користувачів доступі: <http://www.mova.info/corpus2.aspx>.

Отже, відібрані відповідно до експліцитних та імпліцитних принципів корпусної побудови тексти законодавчих документів дозволяють класифікувати розроблений підкорпус як репрезентативний для дослідження сучасного стану українського юридичного дискурсу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *Большакова Е. И.* Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : [учеб. пособие] / Е. И. Большакова и др. – М. : МИЭМ, 2011. – 272 с.

2. *Борисова Е. Г.* Фразеологическое значение в устойчивых словосочетаниях / Е. Г. Борисова, О. В. Захарова // Филологические науки. – М. : 1994. – № 4. – С. 77–84.

3. *Борисова Е. Г.* Коллокации. Что это такое и как их изучать / Е. Г. Борисова. – М. : Филология, 1995. – 49 с.

4. *Гладка В. А.* Структурно-синтаксичний підхід у вивченні колокацій (на матеріалі французької мови) / В. А. Гладка // Наукові записки. Серія “Філологічна”. – Острогор : Вид-во нац. ун-ту “Острозька академія”. – Вип. 39. – 2013. – С. 16–20.

5. *Дарчук Н.* Комп’ютерне анотування українського тексту : результати і перспективи : монографія / Н. Дарчук. – К. : Освіта України, 2013. – 544 с.

6. *Зацеркляний М. М.* Об’єктно-орієнтований тезаурус і словник колокацій для бази знань криміналістичних інформаційних систем / М. М. Зацеркляний, Д. Ю. Узлов // Системи обробки інформації. – Харків : Вид. Харківського ун-ту Повітряних Сил ім. І. Кожедуба, 2013. – Вип. 2. – С. 183–186.

7. *Лендау С. І.* Словники : мистецтво та ремесло лексикографії / С. І. Лендау; [пер. з англ.]. – К. : К.І.С., 2012. – 480 с.

8. *Перебийніс В. І.* Статистичні методи для лінгвістів : [посібник] / В. І. Перебийніс. – Вінниця : Нова Книга, 2002. – 171 с.

9. *Різник С. М.* Об’єктна сполучуваність іменних девербативів у сучасній українській мові : автореф. дис. ... канд. філол. наук : 10.02.01 / С. М. Різник; Київський нац. ун-т ім. Т. Г. Шевченка. – К. : [б.в.], 2009. – 17 с.

10. *Романюк А.* Розпізнавання багатослівних конструкцій / А. Романюк, Г. Кваснюк, М. Романишин // Вісник Нац. ун-ту “Львівська політехніка”. Комп’ютерні системи проектування. Теорія і практика. – Л. : Вид. Львівської політехніки, 2011. – № 711. – С. 158–165.

11. *Хохлова М. В.* Экспериментальная проверка методов выделения коллокаций / М. В. Хохлова // Slavica Helsingiensia 34. Инструментарий русистики : Корпусные подходы. – Хельсинки, 2008. – С. 343–357.

12. Čermak F. Linguistic units and text entities : theory and practice / F. Čermak // EURALEX'98 Proceedings. – Liege : University of Liege, 1998. – P. 109–118.
13. Kilgarriff A. Tickbox lexicography / A. Kilgarriff // Proc. e-Lexicography Conference. – Louvain-la-Neuve : Universit  catholique de Louvain-la-Neuve, 2009. – P. 115–117.
14. Marcinkevičienė R. Lietuvių kalbos kolokacijos : [monograija] / R. Marcinkevičienė. – Kaunas : Vytauto Didžiojo universiteto leidykla, 2010. – 212 s.
15. Meer V. G. Collocations as one particular type of conventional word combinations. their definition and character / V. G. Meer // EURALEX'98 Proceedings. – Liege : University of Liege, 1998. – P. 313–322.
16. Munnink I. Using corpus evidence and experimental data : a multimethod approach / I. Munnink // Corpus-based Studies in English : Papers from ICAME 17. – Amsterdam : Rodopi, 1996. – P. 227–244.
17. Sinclair J. Corpus, Concordance, Collocation (describing English Language) / J. Sinclair. – Oxford : Oxford University Press, 1991. – 200 p.
18. Teubert W. Linguistique de corpus: un alternative / W. Teubert // Semen. Critical Discourse Analysis I. – Besanzon : PUFC, 2009. – T. 27. – P. 130–152.
19. Tognini-Bonelli E. Corpus classroom currency / E. Tognini-Bonelli // Darbai ir Dienos. – Kaunas : Vytauto Didžiojo Universitetas, 2000. – No 24. – P. 205–243.

Стаття надійшла до редакції 14.04.14

Татьяна Бобкова, канд. филол. наук, доц.
Киевский национальный лингвистический университет, Киев

Корпусный подход в исследовании коллокаций

В статье анализируются основные направления и подходы современных исследований коллокаций. Устанавливаются предпосылки формирования корпусного подхода. Обосновывается необходимость применения многометодного подхода в исследовании коллокаций на основе корпуса текстов. В качестве источника составления корпусного словаря коллокаций анализируется под-корпус украинских законодательных документов.

Ключевые слова: коллокация, словосочетание, корпус текстов, корпусный подход, корпусный словарь.

Tetiana Bobkova, Ph D, Doc
Kyiv National Linguistic University, Kyiv

Corpus-based Approach toward Collocation Study

The article deals with the main trends and approaches toward modern collocation studies. The background of corpus-based approach toward collocation study is determined. A necessity of use of multimethod approach in corpus-based collocation study is proved. As a source for corpus collocation dictionary compiling the Sub-corpus of Ukrainian Law Acts is analyzed.

Keywords: collocation, phrase, corpus of texts, corpus-based approach, corpus dictionary.