

**Быстров Я.В.** к. филол. н., доц.  
Киевский национальный университет имени Тараса Шевченко, м. Київ

### **ФРАКТАЛЬНАЯ МОДЕЛЬ КОНЦЕПТУАЛЬНОЙ МЕТАФОРЫ В АНГЛОЯЗЫЧНОМ БИОГРАФИЧЕСКОМ НАРРАТИВЕ**

*В статье обосновывается роль фрактальной модели концептуальной метафоры для описания динамики биографического нарратива. Фрактальная концептуализация жизненного опыта исследуется с применением концептуальной метафоры LIFE IS A STORY.*

**Ключевые слова:** биографический нарратив, фрактальная модель, концептуальная метафора, фрактальный фрагмент, самоподобие.

**Bystrov Ya.**, PhD, Associated Professor  
Kyiv National Taras Shevchenko University, Kyiv

### **FRACTAL MODEL OF CONCEPTUAL METAPHOR IN BIOGRAPHICAL NARRATIVE**

*The article deals with the role of fractal model of conceptual metaphor to prove the dynamics of biographical narrative. The conceptual metaphor LIFE IS A STORY is used to study fractal conceptualization of life experience in biographical narrative.*

**Key words:** biographical narrative, fractal model, conceptual metaphor, fractal fragment, self-similarity.

УДК 81'33

**Бобкова Т.В.**, к.філол.н., доц.,  
Київський національний лінгвістичний університет

### **ЕТАПИ РОЗРОБКИ ЛЕКСИКОГРАФІЧНОГО КОРПУСУ УКРАЇНСЬКИХ ЗАКОНОДАВЧИХ ДОКУМЕНТІВ**

*У статті проаналізовано ознаки лексикографічного корпусу текстів як джерела для укладання корпусного словника. В ролі лексикографічного корпусу розглянуто підкорпус українських законодавчих документів, в основу укладання якого покладені ключові вимоги до корпусних ресурсів. Встановлено основні етапи й принципи побудови лексикографічного корпусу.*

**Ключові слова:** корпус текстів, підкорпус, корпусна лексикографія, лексикографічний корпус, корпусний словник, колокація.

**Актуальність** дослідження пов'язана з активним впровадженням корпусного підходу в сучасній лексикографії. Це вимагає вирішення низки теоретико-методологічних питань, зокрема встановлення детермінативних ознак і принципів побудови лексикографічного корпусу текстів як джерела укладання словників нового типу. **Мета** статті – визначити основні етапи й принципи розробки лексикографічного корпусу текстів. Досягнення поставленої мети передбачає виконання таких **завдань**: 1) окреслити основні теоретичні принципи й базові поняття корпусної лексикографії; 2) дати визначення лексикографічного корпусу текстів; 3) встановити основні етапи й принципи побудови лексикографічного корпусу текстів.

У сучасній лінгвістиці корпусна лексикографія визначається як дисципліна, яка вивчає теорію і практику укладання корпусних словників на підставі кардинально нового підходу до дослідження лексичної одиниці [13, р. 21; 7, с. 281]. Якщо у фокусі уваги традиційних лексикографів знаходяться в основному системні зв'язки лексичної одиниці, то для корпусного лексикографа основний інтерес становить її поведінка, значення в контексті, регулярні зв'язки, стильові й функціональні характеристики в автентичних текстах різних жанрів. Отже, корпусна лексикографія насамперед має справу з вивченням значення або вживанням лексичної одиниці в контексті, під дискурсом при цьому розуміється мова автентичних текстів корпусу.

Вважається, що теоретичні принципи корпусної лексикографії були закладені в 1990 рр., коли вже було накопичено певний досвід розробки корпусів й укладання на їх основі корпусних словників, вирішено ряд філософських, концептуальних і методологічних питань теорії корпусу [22, р. 214]. Однак слід зауважити, що ідея створення словника безпосередньо з тексту або колекції текстів є новою, і в цьому розумінні корпусна лексикографія спирається на емпіричні

традиції текстоорієнтованих досліджень [9, р. 76]. Доелектронні колекції текстів на кшталт корпусів з середньовіччя використовувались для укладання різноманітних словників, зокрема конкордансів, укладених з метою вивчення тексту Біблії. Сучасний стан корпусної лексикографії кардинально відрізняється від традиційної внаслідок обмеженості останньої в обсязі, доступі до лексикографічного матеріалу й методах його обробки. Використання корпусів у якості інструменту опрацювання колекції текстів, прискорення систематизації великої кількості даних у реальному часі призвело до змін у методології лінгвістичного дослідження [22, р. 210].

Вперше електронний корпус текстів було використано для укладання частотного словника словоформ – Computational Analysis of Present-Day American English У. Френсісом і Г. Куцурою у 1967 р. Частотні словники такого типу укладаються в результаті нескладного алгоритму [9, с. 55] і є ілюстрацією найпростішого використання корпусу в лексикографії. У 1980-1990 рр. поєднання корпусної лінгвістики й лінгвостатистики призвело до появи нових типів корпусних словників у зарубіжній лексикографії: Collins Birmingham University International Language Database – проект Дж. Синклера і П. Хенкса (1987 р.) і Cambridge International Dictionary of English – П. Проктера (1993 р.). З часом у західноєвропейській лексикографії на позначення словника, укладеного на базі корпусу, закріпився термін *корпусний словник* [7, с. 282].

В українській лінгвістиці на позначення корпусних словників традиційно вживається термін *текстоорієнтовані* [9, с. 14], або текстозорієнтовані [6, с. 75]. За В. Перебийніс, на відміну від системоорієнтованих текстоорієнтовані словники відображають закономірності мовлення: частоту (частотні словники), сполучуваність (конкорданси) і позицію мовної одиниці (словопоказчики) [9, с. 14]. Зазначені властивості мовних одиниць залежать від характеру тексту, функціонального або авторського стилю. Як було вказано вище, відносна легкість укладання сприяла появі десятків частотних словників, створених групою Р. Піотровського "Статистика речи" (1968 р.) для російської, англійської, німецької, французької і молдавської мов, а в Україні –

колективом авторів на чолі з В.Перебийніс (1970-80 рр.) для української, англійської і російської мов[9, с. 53].Традиції укладання текстоорієнтованих словників було продовжено виданням Частотного словника сучасної української художньої прози (1981 р.), створеного на вибірці обсягом у 500 тис. слів, як і всі доелектронні корпуси, вручну.

Незважаючи на суто корпусне підґрунтя окреслених лінгвостатистичних досліджень, слід відзначити пріоритет укладання частотних словників надрозробкою електронних корпусів в українській лінгвістиці.Корпуси текстів розглядали в основному як джерело досліджень з лінгвостатистики абостатистичної лексикографії.Фактично, використання в лінгвістичних дослідженняхобчислювальної методології іколекцій текстівбули поштовхом для розвитку зарубіжної корпусної лексикографії, а в українському мовознавстві – статистичної лексикографії. І попри те, що підвалини корпусної лексикографії становитьтекстоорієнтований підхід, корпус і текст,як джерело даних, принципово відрізняється за якісними (цілісність, наявність ситуативного контексту, характер інформації) і кількісними параметрами (обсяг, кількість текстів) [22, р. 206]. Відмінність зазначених параметрів корпусувизначає напрямок аналізу й необхідність застосування методів дослідження мови, базованих на сучасних технологіях [22, р. 208]. Отже, корпусна лексикографія спирається на принципи текстоорієнтованого підходу вивчення мовита обов'язкового використання нових технологій.Саме поетапні технологічні зміни в практиці укладання словників покладено в основу розвитку сучасної корпусної лексикографії [7, с. 281]: 1) 1970–80 рр. – створення електронних і копій на мікрофішах; 2) середина 1980 рр. –впровадження комп'ютерних технологій у видавничій справі, збільшення фізичної та оперативної пам'яті; 3) 1980-90 рр. – нарощування потужності, мініатюризація і здешевлення комп'ютерів; 4) 1990 рр. – забезпечення вільного доступу йзбереження великих обсягів даних корпусів.

Традиції емпіричних текстоорієнтованих досліджень і виникнення корпусної лінгвістики вплинули, насамперед, на методику лексикографічного аналізу, однак в українському

мовознавстві розвиток корпусної лексикографії гальмувався через відсутність мотивації до побудови корпусів і недостатній розвиток комп'ютеризації лексикографічної практики й видавничої справи. Революційні технологічні зміни не тільки призвели до вдосконалення методології, а фактично стали основою нового теоретичного прориву в сучасній лінгвістиці [20, р. 29].

Для формування концептуальних принципів корпусної лексикографії базовими є поняття лексикографічного корпусу й колокації. Витоки корпусного підходу дослідження колокацій коріняться в контекстуальній теорії значення Дж. Фьорза й обґрунтуванні "прийомів синтагматичного аналізу природної мови" [16, р. 75]. Основи теоретичних інновацій в корпусній лексикографії було закладено роботами Дж. Синклера "з перевизначення одиниці значення в аспекті корпусних даних" [22, р. 214]. Корпусний підхід передбачає визначення колокації як сумісної вживаності слів [21, р. 109], що регулярно й послідовно спостерігається в корпусі й дає можливість ґрунтовно переглянути традиційні лінгвістичні одиниці. Завдяки статистичній інтерпретації концепція колокації Дж. Фьорза набуває нового виміру в корпусній лінгвістиці.

Сучасна корпусна лексикографія ґрунтується на визначенні корпусу як "електронної колекції текстів, зібраної відповідно до екстралінгвістичного критерію оптимального представлення певної мови або її варіантів для подальших лінгвістичних досліджень" [19, р. 47]. Однак з розвитком лексикографічної практики наведене визначення потребувало певних уточнень, оскільки до 1980 рр. жоден з корпусів не планувався з метою укладання словників. З цією метою в науковий обіг вводиться поняття *лексикографічного корпусу*:

– корпус, використовуваний у якості лексикографічного або лексичного джерела з інструментарієм для укладання словника [19, р. 47],

– колекція текстів, зібрана безпосередньо з метою укладання певного словника [11, р. 54],

– сформований за певними правилами корпус текстів для застосування лексикографами з метою створення електронної

картотеки абовідбору необхідних цитат за допомогою конкордансу [9, с.18].

В українській лексикографії на позначення лексикографічного корпусутакож використовується термін "бібліотека розписуваних текстів" [9, с. 7], який по суті є заміником терміна "анотований корпус текстів". Аналіз наведених вище дефініцій показує, що основною ознакою лексикографічного корпусу є зорієнтованість на практичне використання – слугувати джерелом укладання словників. Побудова корпусних словників ґрунтується на безпосередньому спостереженні корпусу й застосуванні програмних засобів опрацювання текстів. Загальна прагматична схема укладання корпусного словника [16, р.106] відображає основні етапи дослідницької діяльності лексикографа:

корпус → конкорданс → типова модель використання одиниці  
→ словник.

Під типовою моделлю в поданій схемі розуміється колокація, або звичний для слова контекст, встановлюваний за конкордансом. Отже, в укладанні корпусного словника слід виділити два основні етапи: 1) створення лексикографічного корпусу текстів і 2) формування реєстру й словникової статті на основі статистичних і лексико-синтаксичних характеристик одиниць. Проаналізуємо принципи та етапи побудови лексикографічного корпусу на прикладі підкорпусу законодавчих документів, створеного для укладання Словника колокацій українського юридичного дискурсу. В сучасній корпусній лінгвістиці підкорпус текстів визначається як:

– мікроструктура в макроструктурі генерального корпусу [4, с. 83];  
– статистична складова, або вибірка, організована в результатіон-лайн аналізі цілісного корпусу [19, р. 36].

Підкорпуси розробляються з метою вивчення певної форми мовлення, функціонального стилю, різновиду мови й окремих мовних явищ, дослідження яких утруднено на матеріалі національного корпусу текстів [15, р. 20–21]. Безперечно, немає підстав вважати підкорпус (як складову частину певного корпусу) не корпусом, оскільки за допомогою інструментарію (такого як SARA в BNC) або відповідних пошукових

запитів можна виокремити підкорпус з більшого корпусу [18, р. 5]. Отже, підкорпус вважається релевантним, тобто достовірним й адекватним для дослідження мови. В даному дослідженні підкорпус законодавчих документів визначається як мікроструктура у складі Корпусу текстів української мови[3]. Основні етапи розробки корпусу текстів включають: загальне планування, збирання текстів, кодування й зберігання релевантних метаданих і лінгвістичне анотування корпусу [17, р. 241; 14, р. 127–128].

1. **Загальне планування** визначає дослідницьку мету лінгвістів-розробників й залежно від цього детермінативні параметри корпусу [17, р. 241; 7, с. 318]. Детермінативні параметри лексикографічного корпусу залежать від обсягу, тематики, призначення словників, необхідних людських і грошових ресурсів, а також потенційного адресанта. Адекватне планування підкорпусу передбачає встановлення розміру (загальної довжини, кількості складових елементів) і діапазону текстових категорій (жанрів). Це запобігає появі помилок двох типів – "випадкової помилки" і "помилки відхилення", загрозливих "зовнішній валідності" – обсягу, який дозволяє переносити характеристики вибірки на всю генеральну сукупність [12, р. 219]. Випадкова помилка виникає через недостатній обсяг вибірки, а помилка відхилення – при суттєвій відмінності організованої вибірки від генеральної сукупності. Мінімізація цих типів помилок дозволить досягти репрезентативності створюваного корпусу.

Підкорпус законодавчих документів є складовою частиною Корпусу текстів сучасної української мови і має **фіксований обсяг** у 1,25 млн. слів. Фіксований за обсягом підкорпус дозволяє заощадити зусилля при вивченні певної предметної галузі й детермінувати поріг її відображення, оскільки економніший корпус забезпечує вищий поріг відображення [1, с. 116–117; 4, с. 56; 7, с. 332]. Незручні в застосуванні "надвеликі текстові корпуси ускладнюють роботу користувача", тому ефективніше "працювати з підкорпусом або удосконалювати пошукову програму"[4, с. 56]. Метою розробки підкорпусу

законодавчих документів є укладання словника колокацій українського юридичного дискурсу.

На сьогодні виявлення колокацій на базі статистичних методів займає провідне місце в лексикографічній практиці, про що свідчить поява спеціальних словників [2; 16] і лексикографічних систем [10, с. 32]. За корпусним підходом, в основу якого покладено статистичну інтерпретацію, під колокацією розуміється комбінація слів, що зустрілася в тексті принаймні двічі [19, р. 47; 16, р. 104]. При цьому обсяг текстового ресурсу обмежується 1 млн. слововживань, оскільки більший обсяг вимагає більшої кількості появ і обов'язкового зазначення, що це колокації, які зустрічаються разом частіше [16, р. 104]. У даному дослідженні під *колокацією* розуміються типові, регулярно відтворювані синтагматичні лексичні послідовності, що зустрілися в корпусі принаймні двічі внаслідок дії статистичних, семантичних і синтаксичних чинників. Відповідно до дослідного призначення в основу планування підкорпусу законодавчих документів покладено детермінативні параметри скінченності обсягу, репрезентативності, збалансованості, анотованості й повноти текстів, які становлять ключові вимоги до корпусних об'єктів.

2. **Збирання текстів** підкорпусу здійснено за встановленими експліцитними й імпліцитними принципами [14, р. 127–128; 17, р. 241; 7, с. 318]. Документальним джерелом підкорпусу є "Зібрання законодавства України". Вибір інформативно насичених законодавчих документів зумовлено метою розробки підкорпусу – укладанням словника колокацій: тексти офіційних документів характеризуються максимальною кількістю колокацій на тисячу слів [2, с. 17]. Стиль законодавчих документів вимагає чіткості, точності, стандартності формулювань, що виражається в частотності різного роду стійких сполучень: найменувань – *Паризька конвенція, асамблея союзу, Гаазька угода*; терміносполучень – *міжнародна реєстрація, промисловий зразок, юридична особа*; мовленнєвих формул і кліше – *перешкоджати застосуванню, в установленому порядку, додержуватись положень*; похідних



прийменників і сполучників – у відповідності з, у зв'язку з, у силу того, що.

Укладений підкорпус законодавчих документів є одномовним й однорідним в аспекті функціонального стилівідібраних текстів. Результати аналізу текстів при відборі до підкорпусу свідчать, що значний відсоток "Зібрання законодавства України" становлять документи російською мовою. Зокрема, документи з грифом "*Секретно*", "*Цілком таємно*" й "*Для службового користування*" представлені в базі вихідними даними російською мовою за відсутності самого тексту. Не відповідають принципу лінгвістичної однорідності також тексти декларацій і договорів, окремі фрагменти яких представлені українською й російською мовою. У результаті відбору до тіла підкорпусу були включені законодавчі документи українською мовою, що відповідають **принципу лінгвістичної однорідності**.

Незначний згідно з сучасними корпусними вимогами обсяг підкорпусу вимагає дотримання **принципу збалансованості** текстів. При відборі законодавчих документів до підкорпусу виявлено, що довжина текстів різних жанрів є нерівномірною і коливається від 2 слів (*Цілком таємно*) до 44,020 тис. слововживань. Саме тому для дотримання принципу збалансованості при плануванні було вирішено використати діапазоном текстових категорій. На сьогодні лінгвістичний матеріал підкорпусу представлено всіма сорока п'ятьма текстовими категоріями "Зібрання законодавства України". Відповідно до кінцевої мети розробки підкорпусу – укладання словника колокацій у ході відбору текстів перевага віддавалася повнотекстовим документам обсягом не менше 2 тис. слововживань. Кожен з жанрових типів представлений найбільшими за обсягом тексту документами. У підкорпус включені повні тексти законодавчих документів, що забезпечує їхню структурну, лексичну й синтаксичну завершеність.

З метою дотримання **принципу хронологічної однорідності** вибірки й максимального представлення сучасного українського юридичного дискурсу було вирішено обмежити хронологічні рамки підкорпусу – з 1992 р. й по теперішній час. Це

пояснюється екстралінгвістичними факторами – проголошенням в 1992 р. державної незалежності України і, як наслідок цього – активного розвитку законодавства. Відібрані згідно до експліцитних та імпліцитних принципів тексти забезпечують адекватну "фіксацію закономірностей підмови" [23, р. 55] й дозволяють класифікувати розроблений підкорпус законодавчих документів як **репрезентативний** для дослідження сучасного стану українського юридичного дискурсу. При цьому під репрезентативністю розуміється необхідно достатнє й пропорційне представлення в діапазоні текстових категорій підкорпусу різних жанрових типів законодавчих документів.

3. **Метарозмітка** корпусу текстів передбачає кодування, редагування й зберігання релевантних метаданих [14, р. 127; 17, р. 241], необхідних для розв'язання різних лінгвістичних завдань. Парадигму корпусних метаданих становить інформація про конститутивні текстові дані – вміст електронного заголовка, основне призначення якого – інформувати якісно різного користувача про тексти на предмет їх авторства, стилістично-жанрової специфіки, тематики, дати й місця написання тощо [4, с. 89]. Надання інформації про тексти, що входять до складу корпусу, оптимізує використання корпусного ресурсу в лінгвістичних дослідженнях і програмних застосуваннях.

Процес метарозмітки становить приписування текстам і їхнім компонентам кодів бібліографічного й екстралінгвістичного опису [5, с. 7]. На сьогодні набір метаданих корпусу встановлюється стандартами метарозмітки за рекомендаціями Text Encoding Initiative, Open Language Archive Community і ISLE Metadata Initiative відповідно до призначення корпусу й потреб розробників, які вносять до стандарту додаткові ознаки. Набір метаданих, що зберігається як окрема від текстового масиву база даних, визначає дослідні можливості корпусу, його сумісність з іншими програмними продуктами.

Як було зазначено вище, включенню текстів до тіла корпусу також передують розробка системи категорій текстів різних типів. За загальною систематикою *категорії текстів* визначаються як типи джерел матеріалів корпусу (в даному випадку – нехудожні тексти); термін *жанр* використовують до предметних категорій

(законодавчі документи) в загальному сенсі, а термін *текстовий тип* – стосовно конкретних типів документів [7, с. 320]. Однак у вживанні окреслених вище термінів бракує узгодженості, зокрема щодо визначення жанру й типу текстів. Як складова Корпусу української мови підкорпус законодавчих документів забезпечено міжнародною метатекстовою розміткою (стандарт EAGLES Дж. Синклера з додатками С. Шарова) [3, с. 47]. Корпус української мови побудовано так, щоб користувач міг одержати різноаспектну інформацію про мовні одиниці, в тому числі:

– зовнішню – про анування тексту (автор, джерело, рік видання тощо);

– внутрішню – стосовно структуризації тексту (номер розділу, абзацу, речення, слова тощо), його жанрові особливості й тип тексту.

Підкорпус законодавчих документів, як категорія нехудожніх текстів супроводжується вказівкою на сферу функціонування, пов'язану з мовленнєвими особливостями, – *офіційно-ділова*, або жанр – *законодавчі тексти*, і визначенням типу тексту (*договір, інструкція, кодекс, коментар* тощо) [3, с. 47]. У даному дослідженні застосовано статистичний підхід до визначення колокації: генеральною сукупністю для зональної вибірки слугує однорідний масив текстів підкорпусу, а зоною – "сукупність одиниць, встановлена за певною ознакою" [8, с. 14], у даному випадку – за статистичним параметром частоти появи сполучення слів у текстах підкорпусу.

4. **Лінгвістична розмітка** корпусу полягає в приписуванні граматичних, лексичних та інших характеристик елементам тексту [14, р.128; 17, р. 241; 5, с. 3] й покликана забезпечити будь-які потреби користувача стосовно інформації про мову текстів.

Анування корпусу є необхідним підґрунтям сучасної лексикографічної практики, оскільки виконання запитів лексикографа краще забезпечується пошуковою системою лінгвістично анованого корпусу [14, р. 149]. Тип лінгвістичної розмітки детермінується аплікативним призначенням корпусу [4, с. 111]. Основною серед лінгвістичних типів вважається *морфологічна* – частиномовна розмітка, яка включає приписування кодів частин мови, їхніх граматичних категорій і

значень [5, с. 7; 4, с. 94]. Морфологічна розмітка слугує базою для синтаксичної і семантичної, а, отже, є обов'язковою для сучасних корпусів текстів. Результатом *синтаксичної розмітки* є опис синтаксичних зв'язків між лексичними одиницями й синтаксичних конструкцій [5, с. 7]. На сучасному етапі розвитку корпусної лінгвістики відсутнє єдине теоретичне підґрунтя для *семантичної розмітки*, протев загальному розумінні це–приписування кодів семантичних категорій і підкатегорій певній лексичній одиниці або словосполученню [5, с. 7].

Як складова Корпусу української мови підкорпус законодавчих документів забезпечується набором програм морфологічної, синтаксичної і лексико-семантичної розмітки. Наявність зазначених комп'ютерних програм опрацювання передбачає лінгвістичну розмітку тексту в автоматичному режимі з подальшим зняттям омонімії. Завдяки цьому користувач має змогу одержувати словники з морфологічною, синтаксичною, лексичною та енциклопедичною інформацією, укладати за текстом алфавітні й частотні словники, морфологічно індексувати текст або окреме слово й здійснювати синтаксичний аналіз речення [3, с. 46–47]. Анотація корпусних даних здійснюється шляхом введення формалізованої лінгвістичної інформації в аналізовані тексти, зокрема:

– *морфологічна розмітка* передбачає визначення частиномовної приналежності й категоріальних ознак кожної словоформи тексту;

– *синтаксична розмітка* – виділення й визначення типів словосполучень, видів синтаксичного зв'язку та семантичного відношення; представлення в автоматизованому режимі синтаксичної структури на рівні членів речення;

– *лексико-семантична розмітка* полягає в приписуванні інформації щодо розряду, типу за таксономічною класифікацією і дериваційних характеристик.

Окреслені вище функціональні можливості програм лінгвістичної анотації Корпусу української мови є інструментарієм для встановлення лексико-граматичних ознак вилучених з підкорпусу колокацій.

Завершальні етапи розробки лексикографічного корпусу текстів включають:

5. **Конвертування** зібраних розмічених текстів у структуру спеціалізованої лінгвістичної інформаційно-пошукової системи – корпус-менеджер, який забезпечує швидкий багатоаспектний пошук і статистичне опрацювання корпусних даних [5, с. 8]. Підкорпус законодавчих документів інтегрується в інформаційно-пошукову систему Корпусу української мови, яка забезпечує здійснення пошуку за параметрами словоформ, лем і граматичних кодів, а також збереження проміжних результатів опрацювання корпусних даних. Отже, дослідні можливості розробленого підкорпусу законодавчих документів визначаються метаописом, лінгвістичною розміткою даних і функціями вбудованого корпус-менеджера.

6. **Забезпечення доступу** до корпусу реалізується в межах дисплейного класу, поширення на CD і в режимі глобальної мережі, при цьому різним категоріям користувачів можуть надаватися різні права й можливості [5, с. 8]. Як складова частина Корпусу української мови розроблений підкорпус знаходиться в вільному для користувачів доступі, режим доступу: <http://www.mova.info/corpus2.aspx>. Практичне застосування даної роботи, в першу чергу, може бути знайдено в лексикографічній практиці й написанні лексичних граматик, а також у прикладній лінгвістиці для укладання онтологій, розв'язання завдань машинного перекладу, налагодження лінгвістичних процесорів та інформаційного пошуку.

Здійснена розвідка дозволяє дійти таких висновків: 1. Корпусна лексикографія ґрунтується на принципах текстоорєнтованого підходу, контекстуальної теорії значення і використання нових технологій опрацювання природно-мовного тексту. 2. Базовими поняттями корпусної лексикографії є лексикографічний корпус і колокація. 3. Лексикографічний корпус визначається як електронний текстовий ресурс, використовуваний у якості лексикографічного джерела укладання словника. 4. Під колокацією розуміються характерні, часто відтворювані синтагматичні лексичні послідовності, поява яких зумовлюється статистичними, семантичними й

синтаксичними чинниками. 5. Етапи розробки корпусу включають: планування, збирання текстів, кодування метаданих, анотування, конвертування в структуру корпус-менеджеру й забезпечення доступу. 6. Підкорпус законодавчих документів, побудований на принципах скінченності обсягу, репрезентативності, збалансованості й анотованості текстів, доцільно вважати представницьким для укладання Словника колокацій юридичної підмови.

### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *Баранов А.Н.* Введение в прикладную лингвистику: учеб. пособ. / А.Н. Баранов. – М. : Эдиториал УРСС, 2001. – 360 с.
2. *Борисова Е.Г.* Коллокации. Что это такое и как их изучать / Е.Г. Борисова. – М. : Филология, 1995. – 49 с.
3. *Дарчук Н.П.* Дослідницький корпус української мови: основні засади і перспективи / Н.П. Дарчук // Вісник Київського нац. ун-ту ім. Т. Шевченка. Серія: Літературознавство. Мовознавство. Фольклористика. – К. : ВПЦ "Київський університет", 2010. – № 21. – С. 45–49.
4. *Демська-Кульчицька О.* Основи національного корпусу української мови: монографія / О. Демська-Кульчицька. – К. : Ін-т укр. мови НАНУ України, 2005. – 219 с.
5. *Захаров В.П.* Корпусная лингвистика: учеб.-метод. пособ. / В.П. Захаров. – СПб. : РОПИ С.-Петербургского ун-та, 2005. – 48 с.
6. *Карпіловська Є.А.* Вступ до комп'ютерної лінгвістики: комп'ютерна лінгвістика : підручник / Є.А. Карпіловська. – Донецьк : ТОВ "Юго-Восток, ЛТД", 2006. – 188 с.
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: посіб. / В.І. Перебийніс. – Вінниця : Нова Книга, 2002. – 171 с.
9. *Перебийніс В.І., Сорокін В.М.* Традиційна та комп'ютерна лексикографія: навч. посіб. / В.І. Перебийніс, В.М. Сорокін. – К.: В-во Київського нац. лінгвістичного ун-ту, 2009. – 218 с.
10. *Шкурко В.В.* Лексикографічний агент екстракції колокацій у природно-мовному тексті / В.В. Шкурко // Вісник Київського нац. ун-ту ім. Т. Шевченка. Серія: Літературознавство. Мовознавство. Фольклористика. – К. : ВПЦ "Київський університет", 2012. – № 28. – С. 31–35.
11. *Atkins B.T.S., Rundell M.* The Oxford Guide to Practical Lexicography / B.T.S. Atkins, M. Rundell. – Oxford : Oxford University Press, 2008. – 552 p.
12. *Biber D.* Using Register-diversified Corpora for General Language Studies / D. Biber // Computational Linguistics, 1993. – Vol. 19 (2). – P. 219–241.
13. *Biber D.* Corpus Linguistics: Investigating Language Structure and Use / D. Biber. – Cambridge : Cambridge University Press, 1998. – 310 p.

14. *Kilgarriff A.* Efficient Corpus Development for Lexicography: Building the New Corpus for Ireland / A. Kilgarriff // Language Resources and Evaluation. – Dordrecht : Springer, 2006. – Vol. 40 (2). – P. 127–152. 15. *Leech G.* The State of the Art in Corpus Linguistics / G. Leech // English Corpus Linguistics, eds K. Aijmer, B. Altenberg. – L. : Longman, 1991. – P. 8–29. 16. *Marcinkevičienė R.* Lietuvių kalbos kolokacijos: monograija / R. Marcinkevičienė. – Kaunas : Vytauto Didžiojo universiteto leidykla, 2010. – 212 p. 17. *McEnery T., Hardie A.* Corpus Linguistics: Method, Theory and Practice / T. McEnery, A. Hardie. – Cambridge: Cambridge University Press, 2012. – 294 p. 18. *Meyer Ch.F.* English Corpus Linguistics. An Introduction / Ch.F. Meyer. – Cambridge : Cambridge University Press, 2002. – 168 p. 19. *Ooi V.B.Y.* Computer Corpus Lexicography / V.B.Y. Ooi. – Edinburgh : Edinburgh University Press, 1998. – 243 p. 20. *Renouf A.* Corpus Development 25 years on: from Super-corpus to Cyber-corpus/ A. Renouf // Corpus Linguistics 25 Years On. Language and Computers, ed R. Facchinetti. – Amsterdam : Rodopi, 2007. – Vol. 62. – P. 27–49. 21. *Sinclair J.* Corpus, Concordance, Collocation (describing English Language) / J. Sinclair. – Oxford : Oxford University Press, 1991. – 200 p. 22. *Tognini-Bonelli E.* Corpus Classroom Currency / E. Tognini-Bonelli // Darbai ir Dienos. – Kaunas: Vytauto Didžiojo Universitetas, 2000. – No 24. – P. 205–243. 23. *Tognini-Bonelli E.* Corpus Linguistics at Work / E. Tognini-Bonelli. – Amsterdam : Benjamins, 2001. – 244 p.

Стаття подана до редакції 14.10.14

**Бобкова Т.В.**, к. филол. н., доц.,

Киевский национальный лингвистический университет

## ЭТАПЫ РАЗРАБОТКИ ЛЕКСИКОГРАФИЧЕСКОГО КОРПУСА УКРАИНСКИХ ЗАКОНОДАТЕЛЬНЫХ ДОКУМЕНТОВ

*В статье анализируются особенности лексикографического корпуса текстов как источника для составления корпусного словаря. В качестве лексикографического корпуса рассматривается подкорпус украинских законодательных документов, в основу составления которого положены ключевые требования к корпусным ресурсам. Определяются основные этапы и принципы составления лексикографического корпуса.*

**Ключевые слова:** корпус текстов, подкорпус, корпусная лексикография, лексикографический корпус, корпусный словарь, коллокация.

**Bobkova T. V.**, PhD, Associate Professor  
Kyiv National Linguistics University

**UKRAINIAN LAW ACTS LEXICOGRAPHICAL CORPUS:  
STAGES OF COMPILING**

*The article deals with lexicographical corpus features as the source for corpus dictionary compiling. As an example of lexicographical corpus the Subcorpus of Ukrainian Law Acts is analyzed, which is compiled according to the requirements to the corpus resources. The main stages and principles of lexicographical corpus compiling are determined.*

**Keywords:** *corpus of texts, subcorpus, corpus lexicography, lexicographical corpus, corpus dictionary, collocation.*

УДК 811.11: 81'42'2-83

**Булкіна А.В.**, асп.,  
Інститут філології КНУ імені Тараса Шевченка

**ДІАЛОГІЧНИЙ ЛЮБОВНИЙ ДИСКУРС: РЕАЛІЗАЦІЯ  
СТРАТЕГІЇ ВВІЧЛИВОСТІ В МОВЛЕННЄВОМУ ЖАНРІ  
"ЗНАЙОМСТВО" (на матеріалі творчості Т.Гарді)**

*Статтю присвячено розгляду особливостей реалізації стратегії ввічливості в межах мовленнєвого жанру "знайомство" любовного фікціонального дискурсу. Розкрито місце та роль жанру в континуумі любовного дискурсу. Виділено низку базових тактик, що сприяють реалізації стратегії та здійснено детальний аналіз кожної із них.*

**Ключові слова:** *діалог, знайомство, комунікативна стратегія ввічливості, комунікативна тактика, любовний дискурс, мовленнєвий жанр, позитивна ввічливість.*

Сфера міжособистісної комунікації індивідів, її комунікативно-функціональні та прагма-когнітивні характеристики викликають значний інтерес з боку лінгвістичних досліджень в контексті антропоцентричного підходу до дослідження мовних явищ.

Особливості міжособистісної комунікації яскраво опроявляються в сфері любовного (романтичного) спілкування.