

УДК 81'33

ББК 81

КЛАСИФІКАЦІЯ КОЛОКАЦІЙ УКРАЇНСЬКОГО ЮРИДИЧНОГО ДИСКУРСУ: ПОПЕРЕДНІ РЕЗУЛЬТАТИ

БОБКОВА Т.В.

Київський національний лінгвістичний університет

Стаття присвячена класифікації двослівних усталених сполучень на матеріалі підкорпусу українських законодавчих текстів. Запропоновано загальну бінарну типологію підходів до опису колокацій. Виявлено основні помилки автоматичного розпізнавання і морфолого-синтаксичної класифікації двослівних усталених сполучень в українському тексті. Встановлено функціональні характеристики і дистрибутивні моделі лексичних, граматичних і предикативних колокацій в українських законодавчих текстах. Застосування тексто-орієнтованого підходу забезпечило виявлення відсутніх у класифікаціях граматичних і предикативних колокацій.

Ключові слова: колокація, текст, корпус, ідентифікація, класифікація, тексто-орієнтований підхід, модель.

The article deals with the classification of two-word fixed phrases based on the Subcorpus of Ukrainian Law Acts. Binary topology of the approaches to the description of collocations is suggested. The main mistakes of the automatic recognition and morphological-syntactical classification of two-word fixed phrases in Ukrainian texts are defined. The functional characteristics and descriptive models of lexical, grammatical and predicative collocations in the Ukrainian Law Acts are determined. The use of text-oriented approach allowed revelation of the grammatical and predicative collocations missed in the classifications.

Key words: collocation, text, corpus, identification, classification, text-oriented approach, model.

На сучасному етапі розвитку прикладної лінгвістики створення систем автоматичного аналізу тексту ґрунтується на результатах опрацювання значних обсягів природно-мовного матеріалу, укладеного в електронних корпусах текстів. Вивчення корпусів свідчить про те, що 80 % текстів становлять не ізольовані лексичні одиниці, а певні граматичні структури [17, р. 62], які є важливим джерелом інформації про мову й навколишній світ [11, р. 15]. У зв'язку з цим дослідницький інтерес становить опис і класифікація високочастотних стійких сполучень, на позначення яких у сучасній лінгвістиці традиційно використовується термін “колокація” [2; 3; 5; 6; 9; 10; 11; 12; 17; 23; 24]. Необхідність зміщення акцентів у прикладних лінгвістичних дослідженнях з аналізу окремих лексем на стійкі сполучення зумовлює **актуальність** цієї статті.

Мета статті – розробити класифікацію найчастотніших колокацій сучасного українського юридичного дискурсу за даними корпусу текстів.

Досягнення поставленої мети передбачає виконання таких **завдань**:

- установити типологію підходів до класифікації стійких сполучень;
- окреслити відповідні до цих підходів основні класифікації колокацій;
- навести аргументи на користь корпусно-орієнтованого підходу до класифікації стійких сполучень;
- визначити основні критерії ідентифікації найчастотніших колокацій і розробити їх класифікацію на матеріалі підкорпусу українських законодавчих текстів.

Розробка класифікації усталених сполучень передбачає два етапи: I. Ідентифікація досліджуваних одиниць на підставі встановлених ознак. II. Класифікація різних типів колокацій. Як показує аналіз фахової літератури, основні проблеми опису полягають у визначенні критеріїв ідентифікації і класифікації колокацій за кількісними й якісними ознаками [19, р. 509–511; 22, р. 152; 24, р. 258], а також презентації цієї інформації у словниковій статті [19, р. 509]. Ідентифікація усталених сполучень в аспекті визначення інвентарю одиниць у загальній системі мови або в досліджуваній колекції текстів спирається на відповідні теоретико-методологічні підходи до визначення колокацій, які диференціюються в термінах і послідовності застосовуваних методів і процедур [7, р. 1034]. Визначення колокацій, їх статусу серед усталених сполучень та встановлення істотних ознак залежить від інтерпретації базового поняття у традиціях певної лінгвістичної школи або напряму [6].

Входження цього терміна у широкий науковий обіг і використання позначуваного ним поняття фахівцями різних галузей призвело до серйозної термінологічної плутанини, пов'язаної з концепцією колокації у лінгвістиці на сучасному етапі її розвитку [11, р. 15; 12, р. 9; 13, р. 2]. Термінологічна нечіткість визначення зумовлюється, насамперед, багатоаспектністю поєднання лексичних, граматичних і статистичних ознак [10, р. 133; 17, р. 89; 25, р. 206] і у зв'язку з цим – невизначеним статусом колокацій серед усталених сполучень у мові, що не сприяє їх класифікації.

Дискусійний характер терміна “колокація” спонукав до виділення широкого лінгвістично мотивованого його значення й вузького – статистично мотивованого [12, р. 47; 17, р. 40]. У широкому смислі під колокацією розуміється поєднання, звичне використання двох слів разом [17, р. 40; 23, р. 115–116]. У вузькому значенні колокація – це сполучення слів, яке зустрічаються у тексті частіше поруч, ніж за випадковою вірогідністю окремо [23, р. 115–116]. Залежно від широкого або вузького розуміння аналізованого поняття і визначається підхід до ідентифікації і класифікації колокацій. Широкий науковий і практичний інтерес спеціалістів різних галузей сприяв формуванню в сучасній лінгвістиці кількох систем типологізації підходів до опису колокацій. Розглянемо більш детально різні типології підходів, пропоновані в лінгвістичній літературі.

Насамперед, слід виділити бінарну типологію А. Barfield та Н. Gyllstad [8, р. 3], що включає: 1) статистичний підхід, концептуальним підґрунтям якого вважається контекстуальна теорія значення (М.А.К. Halliday, J. Sinclair), і 2) фразеологічний, найбільш поширений у пострадянських країнах через особливості розвитку корпусної лінгвістики (В.М. Телія, І.О. Мельчук, А.Р. Cowie). Трикомпонентна типологія підходів до класифікації колокацій, представлена роботами В. Gitsaki [13], М. Martyńska [18], передбачає встановлення: 1) лексико-композиційного (J.R. Firth, J. Sinclair), 2) семантичного (F.J. Hausmann, M. Benson) і 3) структурного підходів. А за М.В. Хохловою [5, с. 344]: 1) контекстно-орієнтованого (J.R. Firth, J. Sinclair), 2) семантико-синтаксичного (О.Г. Борисова, А.Р. Cowie, F.J. Hausmann) і 3) підходу в межах теорії “Смисл↔Текст” (Ю.Д. Апресян, І.О. Мельчук). Чотирикомпонентна типологія, пропонована S. Handl [14, р. 48], включає: 1) тексто-орієнтований і 2) корпусно-орієнтований підходи (J. Sinclair), а також 3) фразеологічний і 4) семантичний (F.J. Hausmann, M. Benson). І нарешті, найбільш розгалужена типологія В.А. Гладкої [3] – це: 1) контекстно-орієнтований (J. R. Firth, М.А.К. Halliday, J. Sinclair), 2) корпусно-орієнтований (M. Benson, M. Netzlaff, J. Sinclair, В.П. Захаров, М.В. Хохлова), 3) лексико-граматичний (F.J. Hausmann), 4) синтаксичний (D.A. Cruse, F. Grossmann, F.J. Hausmann, A. Tutin), 5) семантичний (Th. Herbst, М.В. Владацька) і 6) конвенційний підхід (E. Aisenstadt, P. Frath, Ch. Gledhill, S. Johansson). Однак, на нашу думку, узагальнення теоретико-методологічних принципів наведених вище підходів уможливує зведення багатокомпонентних типологій до загальної бінарної:

– системо-орієнтований підхід (фразеологічний, семантичний, структурний, підхід за теорією “Смисл↔Текст”, лексико-граматичний, синтаксичний, конвенційний);

– тексто-орієнтований підхід (статистичний, лексико-композиційний підхід, контекстно-орієнтований підхід, корпусно-орієнтований).

За системо-орієнтованим підходом усталені сполучення описуються в аспекті структури й семантики як одиниці системи мови та у співвідношенні з іншими одиницями. Вихідними положеннями класифікації колокацій є заздалегідь укладені списки неоднослівних єдностей [6], обмеження за формальними й якісними ознаками, зокрема функціональними моделями [19, р. 511], лексичними варіантами [6] і морфолого-синтаксичними моделями [5; 6]. Як показує аналіз літератури, найбільш поширеним у сучасних дослідженнях колокацій є використання морфолого-синтаксичної класифікації у термінах граматичних класів слів. Так, за F. J. Hausmann [15, р. 399] передбачено встановлення 6 типів колокацій, за M. Benson [9, р. XIX–XXX] – 7 видів лексичних і 8 – граматичних, а за P. Newmark [20] – 7 видів колокацій з урахуванням лексико-граматичних розрядів іменника таких, як збірні, зчислювальні й незчислювальні.

На відміну від зазначеного вище тексто-орієнтований підхід передбачає виявлення закономірностей поведінки одиниць безпосередньо в усному або писемному мовленні: наскільки частотною є аналізована одиниця в тексті, у яких позиціях зустрічається і які одиниці є релевантними для певного дискурсу. Відповідно до цього увагу лінгвіста зосереджено на вивченні функціональних ознак мовних одиниць, під якими розуміють [4, с. 446]: частоту, позицію у мовленнєвій послідовності, частоту реалізації системних ознак (словозмінних форм), комунікативне призначення, прагматичне й емотивне навантаження, стилістичне забарвлення. У вузькому смислі функціональні ознаки колокації інтерпретуються як статистичні характеристики усталеного сполучення. Саме тому корпусно-орієнтований підхід як різновид тексто-орієнтованого – часто ототожнюється зі статистичним [5]. Колокація розуміється як не випадкове усталене сполучення слів, характерне для мови в цілому або певного дискурсу – підмови, використовуваної для усного або писемного спілкування.

У такий спосіб, лінгвістичні показники доповнюються критерієм частоти вживання й релевантності колокацій для досліджуваного дискурсу [11, р. 17]. Використання в ролі основного критерію ідентифікації частоти вживання дозволяє виявити колокації, відсутні в наявних лексикографічних джерелах, забезпечити об'єктивну фіксацію варіативності морфологічних і функціональних ознак [22, р. 152; 24, р. 252]. Загалом застосування статистичних критеріїв сприяє більшому покриттю аналізованого тексту [7, р. 2035] і детальній класифікації вилучених колокацій. Так, опису потребують не тільки спостережувані в корпусах текстів статистично стійкі лексичні сполучення, а й граматичні реалізації, використовувані мовцями в ролі цілісних одиниць [25, р. 215]. Проте, статистичні міри асоціацій (MI, score, t-score і log-likelihood), розроблені лише для двослівних сполучень, не здатні охопити все різноманіття представлених у корпусі колокацій і диференціювати ядро й колокат [10, р. 33]. А результати автоматичного опрацювання навіть морфологічно анотованого корпусу текстів потребують ручного постредагування внаслідок чисельних повторів, помилок ідентифікації і класифікації колокацій [22, р. 152]. Отже, узагальнення результатів статистичного аналізу і диференціації структурних відношень між компонентами усталеного сполучення потребує доповнення корпусних методів лінгвістичними.

Теоретично тексто-орієнтований підхід не передбачає жодних установлених дослідником обмежень щодо якісних і формальних ознак досліджуваних одиниць і реалізується через статистичну інтерпретацію даних, отриманих з автентичних текстів у вигляді відкритого списку [6]. Однак на практиці ідентифікація і класифікація колокацій обмежується функціональними можливостями сучасних корпус-менеджерів. Найбільш поширеним є використання традиційної бінарної класифікації усталених сполучень на лексичні – колокації і граматичні – колігації, за J. R. Firth [25, р. 209], а також установлення колокацій висхідного й низхідного видів,

за J. Sinclair [23, p. 116]. Щодо статистичних критеріїв, то їх застосування уможливило класифікацію колокацій на частотні й не частотні [16], також диференціацію ідіом і колокацій [21]. Крім того, поєднання статистичного й корпусного методів аналізу дозволило F. Smadja [24, p. 256] розширити уявлення про класифікацію колокацій і, зокрема, виділити фіксовані іменні сполучення, або багатослівні найменування, предикативні і фразові шаблони.

Пропонований F. Smadja [24, p. 253] підхід, що передбачає аналіз статистичних даних як вихідних для морфолого-синтаксичної класифікації колокацій, використано в даному дослідженні для встановлення моделей колокацій українського юридичного дискурсу. Емпіричну базу статті становить підкорпус законодавчих текстів, що є складовою Корпусу текстів української мови. В основу планування підкорпусу покладено ключові вимоги до корпусних об'єктів: репрезентативність, збалансованість лінгвістичного матеріалу, обмеженість обсягу (1 млн. слів), стандартність і дослідницьке призначення [1, с. 22–23]. Вибір офіційних документів для класифікації колокацій зумовлюється максимальною насиченістю в досліджуваних текстах різних видів усталених сполучень [2, с. 17; 17, р. 61–62]. Це дозволить використати укладені списки потенційних колокацій для встановлення моделей, характерних не тільки для юридичного дискурсу, а й української мови в цілому. За розробленою методикою вилучення й ідентифікація колокацій ґрунтується на використанні пакету програм лематизації Корпусу української мови. Оскільки автоматичне вилучення колокацій обмежено можливостями пошукового апарату зазначеного корпусу, то для формування реєстру усталених сполучень на підставі статистичних характеристик застосовано процедуру запитів мовою SQL (автор програмного забезпечення – В.М. Сорокін). При цьому основною моделлю аналізу є двослівна колокація як найчастотніша за вживанням у текстах різних функціональних стилів [4, с. 446]. Для автоматичного виділення потенційних кандидатів колокацій встановлено статистичний поріг, відповідно до якого сполучення двох контактних розташованих слів має зустрітися в опрацьованій колекції текстів обсягом в 1 млн. слів принаймні двічі [17, р. 104; 23, р. 57]. Застосування статистичного критерію для ідентифікації і вилучення колокацій дозволяє ігнорувати обмеження щодо складових компонентів усталених сполучень. Відповідно до корпусно-орієнтованого підходу у цьому дослідженні розглядаються як лексичні, так і граматичні колокації, вилучені з підкорпусу законодавчих текстів. Для класифікації ідентифікованих в підкорпусі стійких сполучень необхідно, насамперед, відсортувати за типами зв'язків між компонентами: 1) сурядний vs. підрядний, 2) предикативний vs. непередикативний, 3) лексичний vs. граматичний.

У зв'язку з цим сполучення з сурядним зв'язком до уваги не беруться, оскільки зазначені пари слів становлять лише частину стійкої багатослівної сполуки, яка складається з двох компонентів, поєднаних сурядним сполучником ([...] *та міра*, [...] *і контейнер*). Завдяки можливості задати списком сурядні сполучники, зазначені пари слів автоматично вилучаються зі списку потенційних колокацій. Решта стійких сполучень із підрядним зв'язком розглядаються з точки зору формальної структури як вільні словосполучення. Отже, основними критеріями ідентифікації колокацій у підкорпусі законодавчих текстів виступають: 1) двокомпонентність усталеного сполучення, 2) статистичний поріг сумісного вживання слів – не менше двох разів на 1 млн. слововживань, 3) контактність компонентів і 4) наявність підрядного зв'язку. У результаті тестування розробленого програмного забезпечення для вилучення й подальшої класифікації колокацій на матеріалі підкорпусу законодавчих текстів укладено попередній список усталених сполучень, який містить 64.361 пару лем. З цього списку для класифікації колокацій українського юридичного дискурсу відібрано тисячу найчастотніших колокацій, точніше – 1011 пар з урахуванням однакового показника частоти вживання останніх у списку 20 сполучень (59 разів на 1 млн. слововживань).

Результати перевірки автоматично вилучених потенційних колокацій показали, що для програмного опрацювання українських текстів суттєві проблеми становлять: а) багатослівні усталені сполучення з трьох і більше компонентів, два з яких помилково ідентифікуються програмою як двослівна лексична або граматична колокація (*банк України, орган державний, в іноземний*); б) нерозв'язана в автоматичному режимі граматична (*вона територія, вони використання*) і лексико-граматична омонімія (*рад Україна, до скласти*); в) відповідне статистичному порогу сумісне вживання в тексті двох слів, не пов'язаних між собою підрядним зв'язком (*Україна від, громадянин яка*). Максимальний відсоток (10,78 %) помилкової ідентифікації колокацій у підкорпусі законодавчих текстів становлять багатослівні стійкі сполуки, зокрема найменування й терміни, основні дистрибутивні моделі яких представлені в Таблиці 1. Знаком (*) позначено передбачуваний відсутній компонент три- і чотирикомпонентні усталені сполучення.

Таблиця 1

Види помилок ідентифікації багатослівних сполучень

Тричленні сполучення			Чотиричленні сполучення		
модель	к-ть помилок	приклад	модель	к-ть помилок	приклад
PrpAN*	35	<i>в іноземний*</i>	N*N*NN	1	<i>**здоров'я України</i>
NAN*	33	<i>орган місцевий*</i>	N*A*NN	1	<i>**справа Україна</i>
AAN*	20	<i>вищий навчальний*</i>	PrpAN*N*	1	<i>в автономний**</i>
A*NN	6	<i>*сила Україна</i>			
PrpNN*	4	<i>*межа Україна</i>			
VAN*	3	<i>скласти верховний*</i>			
N*NN	2	<i>*міністр України</i>			
N*AN	1	<i>*об'єднаний нація</i>			
NPpN*	1	<i>раз на*</i>			
PrpNumN*	1	<i>протягом три*</i>			
ВСЬОГО:	106		ВСЬОГО:	3	

Як показує аналіз Табл. 1, проблеми автоматичної ідентифікації колокацій виникають при аналізі як лексичних, так і граматичних усталених сполучень. Максимальний відсоток помилок становить ідентифікація більш коротких трьохслівних усталених сполучень. Найчастіше (35 разів на 1011 аналізованих пар) програма помилково ідентифікувала як двослівне сполучення трьохслівну граматичну колокацію з прийменниковим керуванням PrpAN* (*в установлений**, *на фондовий**). Проте в загальному списку вилучених пар слів лексичні колокації частіше ніж граматичні помилково ідентифікуються як двослівні. При цьому з двослівними усталеними сполученнями змішуються 6 моделей трьохслівних лексичних (NAN*, AAN*, A*NN, VAN*, N*NN, N*AN) і 2 моделі чотирьохслівних лексичних колокацій (N*N*NN, N*A*NN), а також 3 моделі трьохслівних граматичних (PrpAN*, PrpNN*, NPpN*) і 2 чотирьохслівних граматичних колокацій (PrpNumN*, PrpAN*N*). За попередніми даними аналізу тисячі найчастотніших у законодавчих текстах усталених сполучень можна зробити висновки, що у відношення контамінації частіше вступають лексичні двослівні й трьохслівні колокації. Гіпотетично можна припустити, що зв'язки у двослівних граматичних колокаціях характеризуються більшим ступенем жорсткості і фразеологізованості. Крім того, зі зменшенням частоти потенціальних колокацій спостерігається закономірне збільшення кількості помилково ідентифікованих багатослівних сполучень (див. Табл. 2), що також підтверджується даними про максимальну частоту вживання сполучень із найменшою кількістю складових [4, с. 446].

Таблиця 2

Залежність кількості помилок від частоти

№	Частота сполучення	К-ть допущених помилок
1	2361-2000	1
2	1999-1000	2
3	999-500	4
4	499-300	6
5	299-200	17

Ефективність ідентифікації колокацій залежить також і від якості лематизації компонентів двослівних усталених сполучень, здійсненої програмними засобами Корпусу текстів української мови. За попередніми даними аналізу тисячі найчастотніших пар, загальний відсоток не розв'язаної в автоматичному режимі омонімії сягає 2,67 %. При цьому основну проблему для системи становить автоматичне зняття омонімії п'яти видів, представлених нижче в Табл. 3.

Таблиця 3

Основні помилки лематизації компонентів колокацій

Лексико-граматична омонімія			Грамматична омонімія		
вид	к-ть помилок	приклад	вид	к-ть помилок	приклад
Prn-N	6	<i>той-том</i>	Prn-Prn	4	<i>їй-вона</i>
N-V	6	<i>склад-скласти</i>			
N-A	4	<i>рада-рад</i>			
V-V	4	<i>бути-бутити</i>			
N-N	2	<i>стаття-стать</i>			
Conj-N	1	<i>того-тога</i>			
ВСЬОГО:	23		ВСЬОГО:	4	

За наведеними вище даними, максимальний відсоток помилок виявляється при ідентифікації частотних колокацій з прийменниковим керуванням (*у том, до скласти*). Крім поданих у Табл. 3 даних, спостерігається також суміщення або накладання кількох типів помилок розпізнавання. Зокрема, помилкова ідентифікація частини багатослівного усталеного сполучення внаслідок нерозв'язаної граматичної омонімії PrpPrnN* – *в вона – в її (складі*)* або лексико-граматичної омонімії (Prn-Conj) – Prp*PrnN – (*y**) *тому число*. Найменший відсоток (1,09 %) помилок ідентифікації двослівних усталених сполучень становлять випадки встановлення синтаксичного зв'язку між словами, контактено розташованими в тексті, але не зв'язаними між собою синтаксично. У цьому розумінні основною помилкою є ідентифікація двослівних граматичних колокацій NPrp – 9 (*Україна від*), NConj – 1 (*громадянин яка*), ConjPrn – 1 (*яке вони*).

Застосований у цій статті підхід до класифікації колокацій дозволяє виявити безпосередньо спостережувані в текстах моделі двослівних колокацій, релевантні для українського юридичного дискурсу (див. Табл. 4). При цьому не всі встановлені на матеріалі підкорпусу законодавчих текстів дистрибутивні моделі наявні у відомих класифікаціях колокацій [9; 15; 20], зокрема граматичні й предикативні моделі. Це, в першу чергу, пояснюється характеристиками досліджуваного текстового матеріалу, зокрема функціональним стилем законодавчих документів. Так, серед граматичних колокацій максимальними показниками частоти вживання характеризуються складені прийменники – PRP: *у разі* – 1.406, *згідно з* – 1.131.

Таблиця 4

Класифікація колокацій

Лексичні колокації		Граматичні колокації		Предикативні колокації	
модель	к-ть	модель	к-ть	модель	к-ть
AN	252	PrpN	215	ConjV	24
NN	134	NPrp	130	NV	5
VN	14	VPrp	40	NV(-но)	6
VV	5	PartV	11	V(-но)	5
VAdv	3	APrp	10	PrnN	3
AV	1	PRP	10	NA	2
AdvV	2	PrpA	5		
		AdvPrp	5		
		PartAdv	5		
		PartA	2		
		NPart	2		
ВСЬОГО:	411	ВСЬОГО:	435	ВСЬОГО:	45

За попередніми результатами класифікації усталених сполучень, найбільший відсоток автоматично ідентифікованих і вилучених з текстів пар слів становлять граматичні колокації (48,82 %), хоча моделі лексичних колокацій реалізуються в законодавчих текстах частіше AN (*договірний сторона* – 1902, *верховний рада* – 911, *національний банк* – 886); NN (*конституція Україна* – 1301, *закон Україна* – 1032, *президент Україна* – 821). У досліджуваному підкорпусі граматичні колокації представлені 11 дистрибутивними моделями, при цьому найбільш частотними є моделі прийменникового керування (PrpN, NPrp, VPrp). Лексичні колокації, які складають 46,13 % ідентифікованих усталених сполучень, представлені 7 дистрибутивними моделями сполучуваності. Максимально вживаними в аналізованих текстах є субстантивні моделі сполучуваності, які відповідають багатослівним найменуванням (*кабінет міністр*, *новий редакція*) і термінологічним сполученням (*цінні папери*, *охорона здоров'я*), характерним для підмови даної предметної галузі законодавчих текстів. Незначний відсоток вилучених із підкорпусу колокацій становлять предикативні структури (5,05 %), реалізовані в законодавчих текстах 6 дистрибутивними моделями, найбільш вживаною з яких є ConjV зі сполучним словом у ролі суб'єкта: *яке бути* – 423, *яке мати* – 268.

Отримані результати можуть бути використані для розв'язання прикладних завдань з автоматичного аналізу українського тексту – морфолого-синтаксичної розмітки, зняття омонімії і для машинного навчання системи розпізнавання колокацій, а за умови верифікації результатів на матеріалі текстів різних функціональних стилів – для укладання загальномовного словника колокацій. Здійснене дослідження дозволяє дійти таких **висновків**: 1. За теоретико-методологічними засадами слід виділити системно-орієнтований і тексто-орієнтований підходи до класифікації колокацій. 2. Основні помилки автоматичної ідентифікації колокацій в українських текстах становить розпізнавання багатослівних усталених сполучень, нерозв'язана омонімія і сумісне вживання синтаксично не пов'язаних слів. 3. Існує закономірність між частотою вживання потенціальних колокацій у корпусі й кількістю помилково ідентифікованих сполучень. 4. Тексто-орієнтований підхід забезпечує виявлення відсутніх у встановлених класифікаціях граматичних і предикативних колокацій. 5. Найуживанішими в аналізованому підкорпусі є лексичні субстантивні і граматичні прийменникові колокації, релевантні для юридичного дискурсу.

ЛІТЕРАТУРА

1. Бобкова Т.В. Лексикографический корпус как источник для словарей нового типа / Т.В. Бобкова // *Ύmogus kalbos erdvėje* / [eds A. Diomidova, L. Kamičaitytė, J. Radavičiūtė]. – 2015. – NR. 8. – S. 15–26.
2. Борисова Е.Г. Коллокации. Что это такое и как их изучать / Елена Григорьевна Борисова. – М. : Филология, 1995. – 49 с.
3. Гладка В.А. Структурно-синтаксичний підхід у вивченні колокацій (на матеріалі французької мови) / В.А. Гладка // *Наукові записки нац. ун-ту “Острозька академія”*. Серія “Філологічна”. – 2013. – Вип. 39. – С. 16–20.
4. Перебийніс В.І. Частота мовних одиниць як відображення їхніх системних характеристик / В.І. Перебийніс, Т.В. Бобкова // *Проблеми загального, германського та слов'янського мовознавства*. – Чернівці : Книги-XXI, 2008. – С. 446–453.
5. Хохлова М.В. Экспериментальная проверка методов выделения коллокаций / М.В. Хохлова // *Slavica Helsingiensia* 34. Инструментарий русистики: корпусные подходы. – 2008. – С. 343–357.
6. Ягунова Е.В. От коллокаций к конструкциям [Электронный ресурс] / Е.В. Ягунова, Л.М. Пивоварова // *Acta linguistica petropolitana. Труды Института лингвистических исследований РАН*. – 2011. – Режим доступа : http://www.webground.su/data/lit/pivovarova_yagunova/Ot_kollokatsiy_k_konstruktsiya.pdf
7. Ananiadou S. A Methodology for Automatic Term Recognition / S. Ananiadou // *Proceedings of the 15th Conference on Computational Linguistics*. – 1994. – Vol. 2 – P. 1034–1038.
8. Barfield A. Introduction : Researching L2 Collocation Knowledge and Development / A. Barfield, H. Gyllstad // *Researching Collocations in Another Language. Multiple Interpretations* / [eds A. Barfield, H. Gyllstad]. – Houndsmill : Palgrave macmillan, 2009. – p. 1–18.
9. Benson M. The BBI Combinatory Dictionary of English : Your guide to collocations and grammar / M. Benson, E. Benson, R. Ilson. – Amsterdam : John Benjamins Publishing Company, 2009. – P. XIX–XXX.
10. Durrant Ph. Are high-frequency collocations psychologically real? / Ph. Durrant, A. Doherty // *Corpus Linguistics and Linguistic Theory*. – Berlin–New York : De Gruyter, 2010. – Vol. 6, No 2. – P. 125–155.
11. Evert S. The Statistics of Word Cooccurrences Word Pairs and Collocations / S. Evert. – Ph.D. thesis, IMS. – Stuttgart : University of Stuttgart, 2005. – Mode of Access : [http://elib.uni-stuttgart.de/opus/\[...\]](http://elib.uni-stuttgart.de/opus/[...])
12. Fontenelle Th. What on earth are collocations? / Th. Fontenelle // *English today : the International Review [of the English Language]*. – 1994. – Vol. 10(40), No. 4. – P. 42–48.
13. Gitsaki Ch. Second Language Lexical Acquisition. A study of the development of collocational knowledge / Ch. Gitsaki. – San Francisco-London : International Scholars Publications, 1999.
14. Handl S. Essential collocations for learners of English / S. Handl // *Phraseology in Foreign Language Learning and Teaching* / [eds F. Meunier, S. Granger]. – Amsterdam: John Benjamins, 2008. – P. 43–66.
15. Hausmann F.J. Wortschatzlernen ist Kollokationslernen / F.J. Hausmann // *Praxis des neusprachlichen Unterrichts*. – 1984.– 31. – P. 395–406.
16. Lewis M. Implementing the Lexical Approach: Putting Theory into Practice / M. Lewis. – Hove : Language Teaching Publications, 1997.
17. Marcinkevičienė R. Kolokacija : tyrimo kryptys, metodai, mokyklos / R. Marcinkevičienė // *Lituanistica*. – 1995. – No 2. – P. 40–54.
18. Martyńska M. Do English language learners know collocations? / M. Martyńska // *Investigationes Linguisticae*. – 2004. – Vol. XI. – P. 1–12.

19. McKeown K.R. Collocations / K.R. McKeown, D.R. Radev // *A Handbook of Natural Language Processing* / [eds R. Dale, H. Moisl, H. Somers]. – 2000. – P. 507–523.
20. Newmark P. *Approaches to Translation* / P. Newmark. – London-New York-Toronto : Phoenix ELT, 1981. – P. 114–115.
21. Patsala P. On the Distinction and Presentation of Idioms and Collocations in Bilingual English-Greek Dictionaries / P. Patsala // *Phraseology and Collocation*. – Vol. 2 (2). – 2004. – P. 1029–1035.
22. Seljan S. First Steps in Term and Collocation Extraction from English-Croatian Corpus / S. Seljan, A. Gašpar // *Computational Language Analysis, Computer-Assisted Translation and e-Language Learning* / [ed S. Seljan]. – Zagreb : Zavod za informacijske studije, 2012. – P. 149–156.
23. Sinclair J. *Corpus, Concordance, Collocation* / John McHardy Sinclair. – Oxford : Oxford University Press, 1991. – 200 p.
24. Smadja F.A. Automatically Extracting and Representing Collocations for Language Generation / F.A. Smadja, K.R. McKeown // *Proceedings on the 28-th Annual Meeting of the ACL*. – Pittsburg : PA, 1990. – P. 252–259.
25. Tognini-Bonelli E. *Corpus Classroom Currency* / E. Tognini-Bonelli // *Darbai ir Dienos*. – 2000. – No 24. – P. 205–243.

Дата надходження до редакції 05.06.2015