

префик.+ постфик.	2	3	1,5	0,12%	0,09%	0,11%	0,08%
декомп.+ суффик.	1	3	3	0,06%	0,09%	0,05%	0,08%
префик.+ суффик.+ постфик.	1	1	1	0,06%	0,03%	0,05%	0,03%
усеч.+ редупл.	1	1	1	0,06%	0,03%	0,05%	0,03%
усеч.+ сложение	1	1	1	0,06%	0,03%	0,05%	0,03%
<b>Итого</b>	<b>1700</b>	<b>3207</b>	<b>1,89</b>	<b>100,00%</b>	<b>100,00%</b>	<b>90,14%</b>	<b>83,83%</b>
Займствования и немотивированные новые слова							
займств.	173	590	3,42			9,11%	15,36%
займств.+ суффик.	7	23	3,29			0,37%	0,60%
займств.+ слож.со втор. связ. компонен.	3	4	1,33			0,16%	0,10%
займств.+ сложение	2	2	1			0,11%	0,05%
созд. нов. корня+ префик.+ суффик.	1	1	1			0,05%	0,03%
созд.нов. корня+ суффик.	1	1	1			0,05%	0,03%
<b>Итого</b>	<b>187</b>	<b>621</b>	<b>3,33</b>			<b>9,86%</b>	<b>16,17%</b>
<b>Общ. итог</b>	<b>1887</b>	<b>3828</b>	<b>2,03</b>			<b>100,00%</b>	<b>100,00%</b>

Л. Коцюк\*  
УДК 81'373.374.33

#### ВИКОРИСТАННЯ КОРПУСІВ ТЕКСТІВ У ПРОЦЕСІ УКЛАДАННЯ СЛОВНИКІВ

*The article represents the difference that a corpus, and corresponding advances in lexicology and lexicography can make to form new generation of corpus-based dictionaries. The author raises the question of what the word is and how corpus study can influence the representation of all its features in the dictionary.*

У будь-якій діяльності можна виділити формальну чи рутинну і творчу, неформальну компоненту. Це в повній мірі стосується і роботи лексикографа. Ще В.В. Виноградов цитував слова, які стали класикою, – нема нічого важчого, ніж укласти словник і писати граматику. Не завжди можна чітко провести грань між рутинною та творчою роботою при укладанні словників. Однак можна стверджувати, що методологія корпусної лінгвістики не тільки змінила це співвідношення, але й внесла багато нового в саму лексикографію. В значній мірі потреби лексикографії стимулювали появу та розвиток корпусної лінгвістики.

**Метою** даної статті, однак, не є розібрати всі аспекти співвідношення між цими двома лінгвістичними дисциплінами, а звернути увагу на ті переваги, які отримує лексикограф при укладанні словника, використовуючи дані з корпусів текстів. Основними **завданнями** дослідження є встановлення тих характеристик лексичної одиниці, які можуть бути представлені з допомогою корпусу текстів та аналіз готових лексикографічних робіт на предмет використання даних з корпусів текстів.

---

\* © Л. Коцюк, 2006

Укладання сучасного словника, як правило, базується на деякому корпусі текстів. Це, до певної міри, одне з найбільш вдалих та впливових застосувань корпусів текстів, оскільки навіть ті люди, які ніколи не чули про корпус користуються результатами його аналізу. Дослідження щодо використання корпусів текстів у процесі укладання словників не численні, проте набувають все більшої актуальності. Цієї теми у своїх наукових доробках торкаються Сінклер (1987), Аткінс (1992), Саммерс (1996), Бо та співавтори (1996), Клер та співавтори (1996). Проте всі вони підкреслюють важливість використання результатів аналізу корпусу текстів у лексикографії.

Будь-який лінгвіст (чи студент) потребує відповіді принаймні на п'ять запитань, коли йдеться про невідоме слово: 1. Що це слово означає? 2. З чим воно асоціюється? 3. З якими значеннями воно асоціюється? 4. З якими граматичними функціями воно асоціюється? 5. Як це слово розміщується у тексті?

Перші два запитання здаються найбільш типовими. Всі відомі словники мають справу зі значеннями слів. З огляду на друге запитання, то ще з початку заснування науки про укладання словників і до нашого часу вважається, що правильний опис будь-якого слова буде включати інформацію про лексичне оточення цього слова, його колокацію [5]. Наприклад, слово *consequence* зустрічається з *serious*, а *consequences* – з *economic*. Напевне не існує жодного слова в англійській мові (за винятком деяких граматичних слів, і навіть цей виняток до певної міри суперечливий), яке б не мало певних правил, що стосується вживання інших слів навколо нього. Словосполучення об'єднуються у ідіоми, і іноді важко визначити межу між словами, які регулярно використовуються з якимсь іншим словом, та словами, які утворюють ідіоматичний вираз (хоча у контексті ці дві групи досить легко ідентифікуються через те значення, яке вони мають як єдине ціле чи кожного складового слова по-окремі).

Не менш важливим за сполучуваність, однак менш описаним, є термін "семантична асоціація", феномен, що стосується третього питання. Його можна визначити як тенденцію слова до вживання з певним семантичним класом чи групою слів. Так, наприклад, розглянемо, як працює семантична асоціація у випадку вживання слова *consequence*. *Consequence* має семантичні асоціації з концептами: логіки (зовсім не дивно, оскільки ядерне значення цього слова стосується логічних зв'язків): *unavoidable*, *inevitable*, *inexorable*, *inescapable*, *ineluctable*, *direct*, *ultimate*, *long-term*, *immediate*; сподівання/несподіванки: *likely*, *possible*, *probable*, *natural*, *unintended*, *odd*, *strange*, *planned-for*; чи, з негативною оцінкою: *awful*, *dire*, *appalling*, *sad* з зазначенням ступеня (не)важливості: *serious*, *important*, *dramatic*, *enduring*, *prominent*.

Як і у випадку з лексичною сполучуваністю, на поведінку слова у значній мірі впливає його вживання у одиниці/множині. Наприклад, слово *consequences*, зазвичай семантично асоціюється і відповідно сполучається з прикметниками на позначення логіки та негативної оцінки (хоча воно частіше асоціюється з останнім), та ще й додатково з таким концептом, як володіння: *biological*, *constitutional*, *economic*, *political*, *cultural*.

Як бачимо, лексична сполучуваність та семантична асоціація безпосередньо впливають на значення. Інші характеристики слова, однак також, можуть, хоча й менше, бути важливими для його вживання. Четверте запитання у вищезгаданому списку перевіряє, які граматичні асоціації, чи зв'язки, може мати слово. Це запитання видається важливим з огляду на те, що іноді зв'язки слова характеризуються певними граматичними рисами. Звернімося до вже аналізованого нами слова *consequence*. Навряд чи воно може вживатися як додаток речення на відміну від інших абстрактних іменників, таких як *preference* чи *use*. Важко зустріти приклад (можливо й дивно) такого речення: *Unfortunately it also had this tragic consequence that the baby became grossly bloated*. У той час як такі речення, як *The homeless are asked if they have a preference*. та *The minister called on schools to make more use of the colleges' vocational experience* .... досить часто вживаються. З іншого боку, *consequence* вживається як (частина) комплементу набагато частіше за інші абстрактні іменники: *It is the natural consequence of a deep recession* ....

Не важко помітити, що слово *consequence* дуже часто вживається у прийменниковій групі: *as a consequence* чи *in consequence* (чи їх варіантах). У свою чергу, ці фрази зазвичай з'являються на початку речення, на відміну від звичного розташування прийменникових груп. Ця вибірковість щодо функціонування на початку речення слугує прикладом відповіді на п'яте запитання – які типові розташування слова у тексті?

Все вищесказане свідчить про те, що "знати слово" означає набагато більше, ніж "знати значення слова". Тому лексикограф, щоб відповісти принаймні на цих п'ять запитань, змушений звернутися до аналізу корпусу текстів. З розвитком комп'ютерних

технологій цей процес набув масштабного характеру. Більшість сучасних лексикографічних праць своєю метою поставили розширити опис слова з врахуванням того, з якою частотою і в якому оточенні слово у всіх його значеннях функціонує у текстах різних стилів. Традиційні словники поряд з інформацією про значення слова показують статус іменника (злічувальний / незлічувальний), його вимову, а в деяких випадках зазначають, що це слово у певному контексті буде вживатися в офіційному стилі. Також вони можуть інформувати читача про використання даного слова у сталих виразах, таких як *in consequence* чи *as a consequence*.

Використання корпусів текстів при укладанні словника значно розширили ці можливості. Ось перелік основних акцентів, з допомогою яких корпуси текстів модифікують формування словників:

- акцент на частоту;
- акцент на сполучуваність та фразеологію;
- акцент на варіативність
- акцент на граматичні форми лексем
- акцент на автентичність [6: 96-109].

Частота вживання слів у різних значеннях та їх розподіл у тексті є, напевне єдиною сферою, на яку не поширюється людська інтуїція щодо вживання мовних форм. Саме тому важливим вкладом корпусів текстів у створення словника є включення інформації про відносні частоти певних одиниць. Таку інформацію може використати як безпосередній читач словника, так і його укладач, вирішуючи, яке зі значень певної лексичної одиниці повинно подаватися у словниковій статті першим. Наприклад, порівняємо кількість значень одного й того ж частотного слова *know*, представлену у другому виданні Longman Dictionary of Contemporary English (1987), укладеного без врахування аналізу корпусу текстів, та 3-му виданні цього ж словника 1995 року, сформованого з допомогою даних, взятих з корпусу. У виданні 1987 року це слово має 20 значень, у пізнішому виданні цих значень уже нараховують 40. Збільшення кількості значень пояснюється появою нової інформації щодо частоти вживання, включаючи поділ значення на два і більше. Також можна зафіксувати появу нових значень чи вживань, які здавалися не важливими до того, як корпус показав, наскільки вони є частотними. Одним з прикладів може слугувати фраза *I know*, включена в першому з аналізованих словників тільки як фраза на позначення раптової ідеї: *What can we get her for birthday? Oh, I know, we'll get her some flowers.* У новому виданні словника додаються ще два випадки використання цієї фрази: погодження (як в *I'm so worn out! Yeah, I know*) та попереднього непогодження (як в *It sounds silly I know, but try it anyway*).

Іншою інновацією в словниках, яка стала можливою тільки завдяки використанню корпусу текстів, є включення безпосередньої інформації про частоту. Словник Longman Dictionary of Contemporary English 1995 року видання, наприклад, відмічає слова, які особливо частотні в усному чи писемному мовленні і в деяких випадках порівнює ці дві сфери вживання.

Використовуючи корпус текстів з зазначенням частин мови, стає можливим порівняти частоту вживання одного й того ж слова, наприклад, будучи іменником чи дієсловом. Дефініції словника тоді можна розташувати у правильному порядку, зважаючи, що першими будуть показані частотніші частини мови [9: 262]. Наприклад, *gorge* як іменник (у значенні 's valley with steep sides') приблизно у чотири рази частіше вживається, ніж дієслово (у значенні 'eat greedily'). Таким чином, у словнику більш резонно буде представити значення іменника перед значеннями дієслова.

Ще більше допомогти лексикографам зможе автоматичне розпізнавання значень слова [9: 4]. Це досягнення дасть можливість ідентифікувати окремі значення слова з урахуванням частоти вживання даного слова у цьому значенні. Зважаючи на те, що слова в окремих значеннях вживаються у різних оточеннях, автоматичне розпізнавання значень теоретично можливе, але все ще не досяжне.

В "нових" словниках особливу увагу приділяють сполучуваності та фразеології, що відбиває загальну тенденцію корпусів текстів акцентувати увагу саме на цих аспектах дослідження лексичних одиниць [7: 116-136; 8: 104-115, 150-159; 9]. З цієї точки зору, словники, засновані на аналізі корпусу текстів характеризуються наступними особливими рисами. По-перше, спостерігається тенденція до подачі визначення цілій фразі, а не окремому слову (наприклад, перевага буде надаватися вислову *I know*, ніж просто *know*). По-друге, деякі словники використовують фразу вже у дефініції для ілюстрації сполучуваності. Наприклад, визначення слова *brink* у словнику Collins Cobuild English

Dictionary 1995 року видання – 'if you are *on the brink of* something, usually something important, terrible, or exciting, you are just about to experience it'. Особливістю подачі визначення таким чином є ще й доповнення інформації про лексичну сполучуваність чи семантичну асоціацію цього слова (див., наприклад, 'usually something important, terrible, or exciting'). Ще однією важливою рисою цих словників є обов'язкове представлення прикладу для підтвердження визначення та демонстрації сполучуваності.

Фразеологія здається особливо важливою у випадках частотних слів, більшість з вживань яких можуть бути представлені у сталих виразах [7; 9]. Саммерс, наприклад, відмічає, що слово *day* найчастіше зустрічається у таких фразах, як *one day, the other day, some day* [9], а Сінклер зауважує, що у словниках недостатньо приділяється увага саме такій фразеології, особливо це стосується граматичних слів, на противагу лексичним (наприклад, слово *a* зустрічається у таких фразах, як *come to a head* як альтернатива артиклю *the*, проте словники часто ігнорують таке його вживання) [7].

Інший важливий акцент при укладанні словника, на який корпус текстів зробив можливість наголосити – це варіативність мовлення в залежності від регіону, статі, соціальної групи, часового періоду чи стилю. Особливістю словників, які беруть інформацію з корпусу текстів, є подача випадків варіативності щодо стилю мовлення. Згідно спостережень Байбера, наприклад, частотні дієслова англійської мови *say, get, go, know* та *think* набагато частіше вживаються у розмовному стилі, ніж у інших [3]. Разом вони складають 40% всіх лексичних одиниць в усному мовленні, у той час, як на письмі відсоток їх вживання становить тільки 20 % від усіх лексем.

Традиційно у лексикографії словосполучення розглядали тільки з лексичної точки зору. Різниця між лексемами та грамами базується на принаймні трьох твердженнях: по-перше, граматичні слова частіше вживаються за прості лексеми; по-друге, граматичні слова утворюють парадигми, а лексичні значення слів – синтагми; по-третє, граматичні слова не мають власного значення на відміну від лексичних слів. Останнім часом кожне з цих тверджень можна піддати сумніву за допомогою доказів з корпусу текстів. Наприклад, у згаданому вже нами вислові *come to a head* статус артикля *a* як граматичного слова дуже сумнівний, оскільки його парадигматичні характеристики згладжуються тим фактом, що це слово не може бути замінене артиклем *the*, і відповідно *a* відіграє таку ж роль як і слово *head*, тобто його можна вважати таким, що має лексичне значення.

У процесі укладання словника за допомогою корпусу текстів, можна запропонувати приклади для ілюстрації найбільш типового вживання слова чи фрази, і, якщо приклади взяті безпосередньо з корпусу, автентичність гарантується. Важливо розрізняти поняття автентичності та типовості. Будь-який корпус текстів включає безліч автентичних прикладів вживання певного слова, вони є частиною реальних текстів, проте в той же час – продуктом інновацій, гри слів, або просто певних обставин, і таким чином вони не можуть бути типовими. На противагу, деякі укладачі словників вигадують речення, які показують типові випадки вживання, але вони не є автентичними.

Зважаючи на той факт, що всі лексикографи погоджуються, що типовість є важливим фактором, не всі з них наполягають на повній автентичності. Бо стверджує, що "більшість цитат не підходять для учнівських словників тому що вони надто складні граматично, включають важкі слова та ідіоми, чи звертаються до специфічних контекстів" [2]. У вступі до Longman Dictionary of Contemporary English 1995 року видання говориться, що всі приклади у цьому словнику взяті з усних та письмових матеріалів корпусу Longman Corpus Network. Деякі приклади беруться безпосередньо з корпусу, а інші дещо видозмінені з метою уникнути складних слів, а деякі були створені спеціально для певного лексичного входу. Укладачі Collins Cobuild English Dictionary 1995 року видання наголошують на виключно автентичності прикладів. Як бачимо думки лексикографів розходяться у цьому питанні і передбачаємо, що це питання буде підніматися у процесі розвитку комп'ютерної лексикографії.

Отже, словники, які використовують результати аналізу корпусу текстів відрізняються від своїх попередників. Дослідження автентичних текстів дає змогу зробити сучасний словник "живим", тобто таким, який відображає реальне мовлення, усне чи то письмове, офіційне чи неформальне. Такі словники будуть належне оцінені як іншими лексикографами, так і користувачами: студентами чи науковцями.

#### Література

1. Atkins V.T.C. Tools for Computer-Aided Corpus Lexicography: the Hector Project. // Acta Linguistica Hungarica 41 (1-4), 1992-93. – P. 5-71.

2. Baugh D., Harley A. and Jellis S. The Role of Corpora in Compiling the Cambridge International Dictionary of English. // International Journal of Corpus Linguistics 1, 1996. – P. 39 – 59.
3. Biber D., Johansson S., Leech G., Conrad S., Finegan E. Longman Grammar of Spoken and Written English. – London: Longman, 1999.
4. Clear J., Fox G., Francis G., Krishnamurthy R. and Moon R. Cobuild: The State of the Art. // International Journal of Corpus Linguistics 1, 1996. – P. 303 – 314.
5. Hoey M. What's in a word. // MED Magazine. – Issue 10, August 2003. – <http://www.macmillandictionary.com/med-magazine.htm>
6. Hunston S. Corpora in Applied Linguistics. – Cambridge University Press, 2002. – 241 p.
7. Sinclair J.M. Looking Up: An Account of the Cobuild project. – London: HarperCollins, 1987.
8. Sinclair J.M. A Way with Common Words. – In: Hasselgard H., Oksefjell S. Out of Corpora: Studies in Honour of Stig Johanson. – Amsterdam: Rodopi, 1999. – P. 157 – 179.
9. Summers D. Computer Lexicography: The Importance of Representativeness in Relation to Frequency. – In: Thomas J. and Short M. Using Corpora for Language Research. – London: Longman, 1996. – P. 260-265.

**Т. Сидорко\***

Інститут філології КНУ ім. Т. Шевченка (Київ)  
УДК 161.2.81'322

### РЕКЛАМНИЙ ТЕКСТ ЯК ОСОБЛИВА ОДИНИЦЯ КОРПУСУ ДАНИХ

*У статті йдеться про актуальність та принципи створення корпусу рекламних текстів, збалансованого за жанрами, моделями тощо. Укладання корпусів рекламних текстів з їх наступним лінгвістичним аналізом дозволяє створити надійну базу для проведення лінгвістичних та психолінгвістичних досліджень, використовуючи різні методи для отримання статистично надійних результатів.*

Створення корпусу рекламних текстів як особливого виду сугестивних одиниць є дуже важливим. Рекламні тексти мають високий ступінь впливовості на свідомість споживача (у порівнянні зі звичайним текстом), у них використовують специфічні мовні й психолінгвістичні прийоми для досягнення кінцевої мети реклами (купівлі рекламованого продукту). Для опису рекламного тексту можуть бути використані різні схеми, що містять як загальнолінгвістичні, так і специфічні для рекламного тексту характеристики: фонетичні, морфологічні, синтаксичні.

*Матеріалом* для створення корпусу є рекламні тексти з популярних газет і журналів. *Предметом* вивчення в роботі був корпус рекламних текстів. *Мета нашого дослідження* – прослідкувати специфіку функціонування мовних засобів рекламного тексту та різних методів лінгвістичного та психолінгвістичного маніпулювання свідомістю споживача. *Завдання роботи*: на основі лінгвістичного аналізу бази даних рекламних текстів визначити її вплив на масову свідомість і колективну поведінку.

Інформація про рекламний текст, яка міститься в корпусі, надає можливість наочно побачити весь масив проблемного матеріалу. Проте формуванню корпусу текстів повинен передувати філологічний аналіз фактичного матеріалу.

Насамперед потрібно паспортизувати усі рекламні тексти: слід навести джерело, означити рекламований предмет, визначити довжину тексту (як одного із факторів дієвості). Отже, базу рекламних текстів для зручності користування можна розділити на такі масиви: **джерело рекламного тексту, цільова аудиторія, предмет рекламування, вплив** (лінгвістичний, психолінгвістичний).

Одна з найважливіших умов репрезентативності текстів передбачає, що склад корпусу має бути деяким чином збалансований за жанрами:

- ✓ Реклама-діалог: рекламне звернення представлено у формі діалогу;
- ✓ Порівняльна реклама: у ній зіставляються споріднені предмети з метою пояснення одного за допомогою іншого;
- ✓ Реклама-інструкція: у цьому випадку рекламна аргументація подана у формі інструкції;
- ✓ Реклама-тест: пропонується відповісти на запитання, а далі іде пояснення правильної відповіді;
- ✓ Реклама-загадка: у цій моделі рекламний текст починається з питання;