

ВИКОРИСТАННЯ КОНКОРДАНСІВ ТА КОРПУСНИХ ТЕХНОЛОГІЙ ДЛЯ ДОСЛІДЖЕННЯ КОНЦЕПТУ ЗДОРОВ'Я/HEALTH

Анотація. У статті розглянуто конкорданс як вид словника сполучуваності, що дозволяє аналізувати різні контексти, виводити з них відтінки значення слова, моделювати смислові поля слів-корелятивів в різних мовах. Метою статті є розкриття методики використання технологій укладання корпусів текстів та конкордансів для побудови номінативного поля концепту ЗДОРОВ'Я/HEALTH в англійських та українських медіа текстах. Матеріалом дослідження стали корпуси англійських та українських медіа текстів на тему медицини та здоров'я. Проілюстровано застосування конкордансу для визначення ключових слів-вербалізаторів досліджуваного концепту; з'ясування частотності їх вживання, аналізу їхньої семантики лексем, сполучуваності та дистрибуції. Визначено необхідність використання в сучасній лексикографії даних електронних корпусів текстів визначає актуальність дослідження концептуальних засад укладання корпусних словників.

Ключові слова: корпусна лінгвістика, корпус текстів, конкорданс, ключове слово, концепт.

Shvaiber Denys, Romanyshyn Natalia

Lviv Polytechnic National University

USING CONCORDANCE AND CORPORA TECHNOLOGIES TO RESEARCH THE CONCEPT OF HEALTH/HEALTH

Summary. The article discusses the notion of concordance as a kind of dictionary of compatibility, which allows to analyze different contexts, to deduce shades of word meaning on their basis, to model semantic fields of correlated words in different languages. The aim of the article is to reveal the methodology of using the technologies of composing corpora of texts and concordances to build a nominative field of the concept HEALTH in English and Ukrainian media texts. The study was based on media texts on medicine, health and coronavirus in English and Ukrainian media texts (newspapers and magazines: USA Today, The Wall Street Journal, The New York Times, Los Angeles Times, The Washington Post, Vysokyi zamok, Ukrainiska pravda; online edition: www.bbc.com/ukrainian/, www.bbc.com/news, time.com/, www.nytimes.com/, www.who.int; Gazeta.ua, Holos Ameryky, Hromadskyi prostir, Ukrainskyi tyzhden, Yevropeiska pravda, Hromadske.ua, Ukraina moloda and others). It is illustrated how concordance reflects all cases of use of words in this or that text, corpora of texts; the keywords of the researched concept are searched and the frequency of their use is determined. The use of concordance and corpus technologies (works by A. McCarthy and R. Carter, O'Keefe) is usually associated in linguistics with the study of the semantics of language units based on the calculation of word frequency, keywords, cluster analysis, etc., lexical and grammatical profiles. Work of users with the case is carried out by means of the specialized software – case managers who provide various opportunities on reception of the necessary information from the case. Concordance programs are used in a variety of areas: language learning and learning, language data search and analysis, translation and language engineering, corpus linguistics, lexicography, content analysis in accounting, history, marketing, music, politics, geography, journalism, and more. With concordance, you can create indexes and glossaries, calculate the frequency of use of a word, compare different uses of the word, analyze keywords, find phrases and idioms, create a concordance of one author, and so on. The main task of the concordance is to display all cases of use of the word registered in some texts. The possibility of realization of lexical meaning in the text is demonstrated, which helps to determine the role of lexical unit in the formation of textual conceptual structure.

Keywords: corpus linguistics, corpora, concordance, keyword, concept.

Постановка проблеми. Завдання розроблення й удосконалення прикладних систем автоматичного аналізу тексту й укладання словників потребують дослідження великих обсягів природно-мовних текстів електронних корпусів. Ефективність розв'язання прикладних завдань опрацювання текстової інформації значною мірою залежить від вибору підходів і теоретичних пріоритетів, визначальних для методології дослідження. Загальне визнання необхідності використання в сучасній лексикографії даних електронних корпусів текстів визначає **актуальність дослідження** концептуальних засад укладання корпусних словників.

Аналіз останніх досліджень і публікацій. Використання конкордантних та корпусних технологій (праці А. Маккарті та Р. Картера, О'Кіффа) завжди пов'язане у лінгвістиці з вивченням семантики мовних одиниць на основі підрахунку частоти

слів, ключових слів, кластерного аналізу тощо, лексичного та граматичного профілю [4, с. 12].

У галузі прикладної лінгвістики конкорданси особливо визнані мовознавцями завдяки новим можливостям ефективного вивчення мови та обробки лексичного матеріалу на основі текстів різних видів. Конкорданси дозволяють не лише продемонструвати можливість реалізації лексичного значення в тексті, а й сприяють визначенню ролі лексичної одиниці у формуванні словесних образів [4, с. 203–211].

Мета статті є розкриття методики використання технологій укладання корпусів політичних текстів та конкордансів для побудови номінативного поля концепту ЗДОРОВ'Я/HEALTH в англійських та українських медіа текстах. Матеріалом дослідження стали корпуси англійських та українських медіа текстів на тему медицини, здоров'я та коронавірусу.

Виклад основного матеріалу дослідження. Робота користувачів з корпусом здійснюється за допомогою спеціалізованих програмних засобів – корпусних менеджерів, які надають різноманітні можливості по отриманню з корпусу необхідної інформації, а саме:

- пошук конкретних словоформ;
- пошук словоформ по лемах;
- пошук групи словоформ у вигляді розривної або нерозривної синтагми;
- пошук словоформ по набору морфологічних ознак;
- відображення інформації про походження, тип тексту тощо;
- висновок результатів пошуку із зазначеним контексту заданої довжини;
- отримання різних лексико-граматичних статистичних даних;
- збереження відібраних рядків конкордансу в окремому файлі на комп'ютері користувача тощо.

Результати пошуку зазвичай видаються у вигляді конкордансу (тому корпусні менеджери ще називають конкордансерами), де шукана одиниця представлена в її контекстному оточенні і у вигляді статистичних даних. Останні можуть фіксувати частотні характеристики окремих мовних одиниць, або граем, або можуть характеризувати частоту сполучуваності декількох лексичних одиниць.

Конкорданс – алфавітний перелік усіх слів якого-небудь тексту з наведенням контекстів їх використання [16, с. 57]. Програми-конкорданси використовуються у найрізноманітніших сферах: навчання та вивчення мови, пошук та аналіз мовних даних, переклад та мовна інженерія, корпусна лінгвістика, лексикографія, контент-аналіз у бухгалтерії, історії, маркетингу, музиці, політиці, географії, журналістиці тощо.

З допомогою конкордансу, можна створити індекси та глосарії, підрахувати частоту вживання слова, порівняти різні вживання слова, проаналізувати ключові слова, знайти фразеологізми та ідіоми, створити конкорданс одного автора, і так далі. Головним завданням конкордансу є відображення всіх зареєстрованих в деяких текстах випадків вживання слова. На відміну від словника, який ставить на чільне місце словникову статтю, конкорданс наводить приклади – праві і ліві контексти слова. У даній статті ми розглядаємо конкорданс як один з інструментів навчальної та дослідницької діяльності.

С.Ю. Толдов [9, с. 11] зазначає, що конкорданси використовуються для вирішення наступних лінгвістичних завдань:

- порівняння різних можливостей вживання одного і того ж слова;
- аналіз ключових слів;
- аналіз частоти слів та фраз;
- пошук та дослідження фраз та ідіом;
- пошук перекладу;
- створення списків слів.

Основна риса, яка робить конкорданси надзвичайно вагомими у процесі аналізу мовного матеріалу, є їх здатність показати кожне місце в тексті, де використовується певне задане слово чи словосполучення.

Найбільш поширеним видом відображення конкордансу є формат KWIC (ключове слово в контексті), за якого ключове слово розміщується по центру стрічки, а навколо нього показу-

ються приблизно 35 символів до та після. Також може показуватися посилання (як, наприклад, номер стрічки) з метою подальшого відновлення контексту в разі необхідності.

Відомо, що вибірку мовних одиниць зазвичай проводять з використанням корпусів текстів. Їх (корпусів) зараз надзвичайно багато для багатьох мов та підмов, як наприклад, Міжнародний корпус англійської мови і Британський національний корпус (British National Corpus), Корпус українських текстів: новини, Вікіпедія, художні тексти, веб – доступні для завантаження (<http://lang.org.ua/uk/corpora/>).

З допомогою програми-конкордансу, особливого типу комп'ютерної програми, яка здійснює пошук по корпусу та представляє результати цього пошуку в особливій формі, можна виявити специфіку певних мовних та дискурсивних процесів.

Конкорданс відображає всі випадки вживання слова в тому чи іншому тексті, корпусак текстів. Таким чином, здійснюємо пошук за ключовими словами та визначаємо частоту їх вживання (див. табл. 1, рис. 3, 4).

Конкорданс – це програма, що дозволяє аналізувати великі масиви тексту на предмет виявлення закономірностей використання в мові слів або виразів. Конкорданс здійснює пошук запитуваної слова в корпусі і видає в новому вікні кілька фрагментів пропозицій з різних текстів, в яких використовується дане слово або вираз [16, с. 9].

Програми-конкордансери також можуть використовуватися як інструмент для вивчення значення слова. Проілюструємо це на прикладі роботи з програмою – конкордансером Ant.Conc. На підставі отриманих результатів конкордансу можна як зрозуміти значення даного слова з контексту, так і провести аналіз його вживання в мові. В таблиці 1 представлений приклад результатів пошуку слів *хвороба* і *virus*. Даний приклад чітко показує, що конкорданс видає кожне слово в контексті його реального використання.

Результати пошуку можна використовувати для уточнення слововживання і виведення правил використання певних слів і виразів в мові, виявлення декількох значень одного і того ж слова а також для вивчення граматичної будови мови (див. рис. 1-2).

Механізм словника-конкордансу є досить простим. Словниковим входом служить словоформа. Вона розміщується в центрі рядка і відокремляється від тексту прикладу пробілами зліва, як на рис. 1 та 2 (тобто, тут є можливе використання і інших способів виділення)» [6, с. 74].

В таблиці 1 здійснюємо пошук за ключовими словами в англійській мові – «*health*», «*disease*», «*virus*», «*covid*», «*isolation*», «*coronavirus*», «*death*», «*vaccine*», «*antibodies*», «*illness*», «*quarantine*», «*epidemic*», «*pandemic*»; в таблиці 2 – пошук за ключовими словами в українській – «*коронавірус*», «*вірус*», «*карантин*», «*хвороба*», «*здоров'я*», «*вакцина*», «*смерть*», «*ізоляція*», «*антитіла*», «*пандемія*», «*епідемія*».

За результатами аналізу всіх зареєстрованих в конкордансі слововживань ключових слів корпусу статей встановлено, що ці слова вживаються в певних значеннях, мають різну сполучуваність з іншими частинами мови, найчастіше це прикметники, які посилюють значення іменника у вигляді епітета чи порівняння, прикметники та дієслова. Наприклад:



Рис. 1. Приклад роботи програми конкордансів Ant.Conc для пошуку ключового слова «virus»

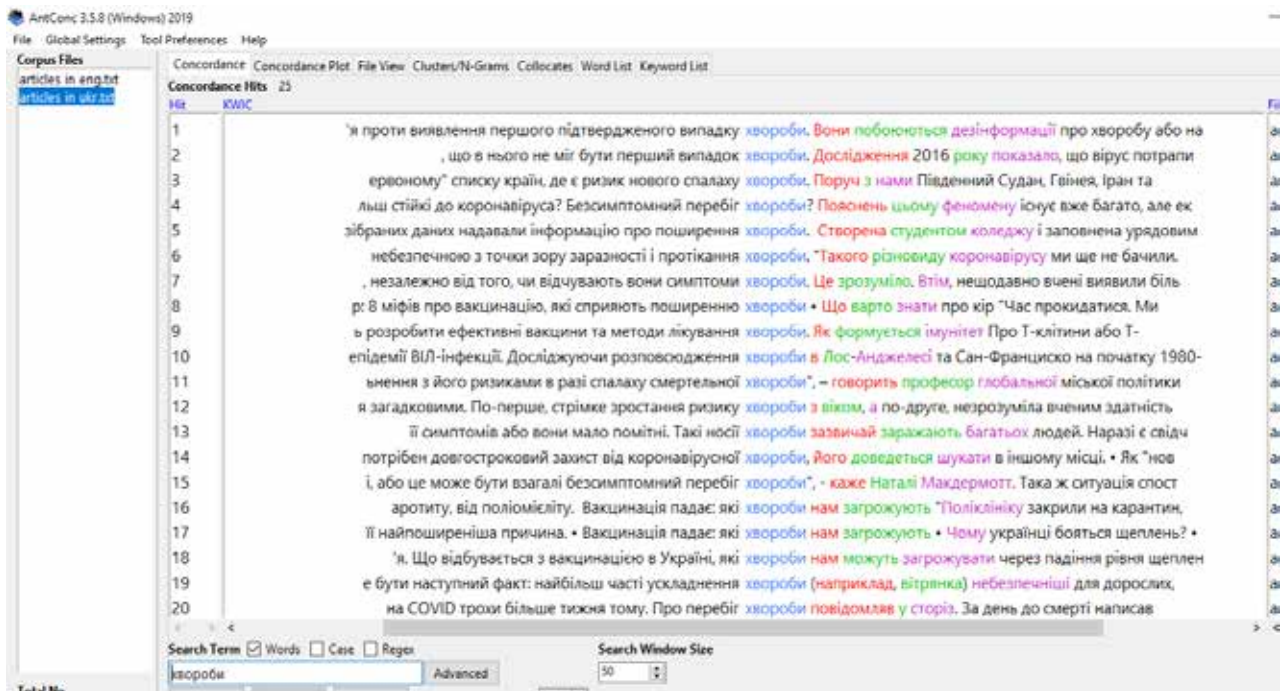


Рис. 2. Приклад роботи програми конкордансів Ant.Conc для пошуку ключового слова «хвороба»

— прикметник: *predicting, unexpected, unrecorded, swedish, statistical, relating, rampant, confirmed, experimental, global, soaring, positive, original, far, local, effective, undergoing, positive, weeks-long, negative, self-isolating, severe* та ін.; *шалений, черговий, специфічний, страшніший, дивовижний, китайський, стримінговий, позитивний, хронічний, ключовий* тощо.

— дієслова: *to lose, to damage, to hospitalize, to contact, to issue, to count, to rise, to fight, to disrupt, to tackle, to harm, to claim, to be infected with, to boost, to keep, to collapse* та багато інших; *promi-*

cati, знищувати, рекомендувати, виділити, заражати, інфікуватися, кашляти, втомлюватись, припускати, зупинити, відстежити, госпіталізувати, попереджати тощо.

Під час пошуку за ключовими словами у конкордансі, спостерігаємо частотність метафоричних способів актуалізації концепту, наприклад: *hate from the left, be from mild to severe, be assigned to an outbreak, proceed with no symptoms, the news has prompted record share* та ін.; *прилипати до поверхні віруса, прогнозувати стан хворого, по-долали спалах кору, протікати безсимптомно,*

Таблиця 1

Кількісний аналіз ключових слів концептосфери ЗДОРОВ'Я/ HEALTH

| Ключове слово в англійських статтях | Частота вживання | Ключове слово в українських статтях | Частота вживання |
|-------------------------------------|------------------|-------------------------------------|------------------|
| health | 68 | здоров'я | 45 |
| disease | 4 | хвороба | 35 |
| virus | 60 | вірус | 29 |
| covid | 148 | covid | 106 |
| isolation | 4 | ізоляція | 5 |
| coronavirus | 200 | коронавірус | 74 |
| death | 19 | смерть | 12 |
| vaccine | 155 | вакцина | 42 |
| antibodies | 8 | антитіла | 26 |
| quarantine | 13 | карантин | 16 |
| epidemic | 3 | епідемія | 21 |
| pandemic | 55 | пандемія | 30 |

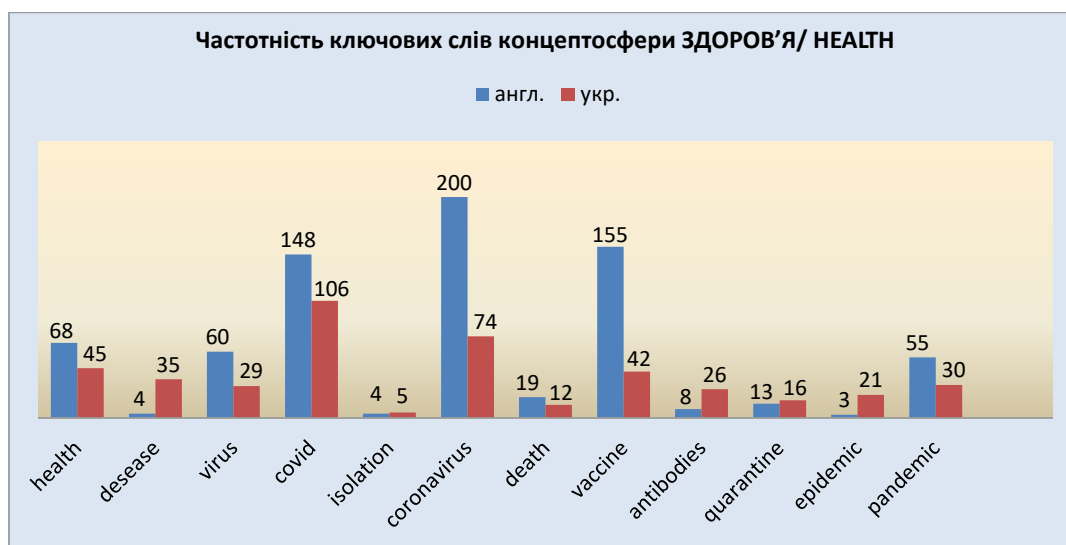


Рис. 3. Порівняльна діаграма вживання найбільш вживаних ключових слів концептосфери ЗДОРОВ'Я

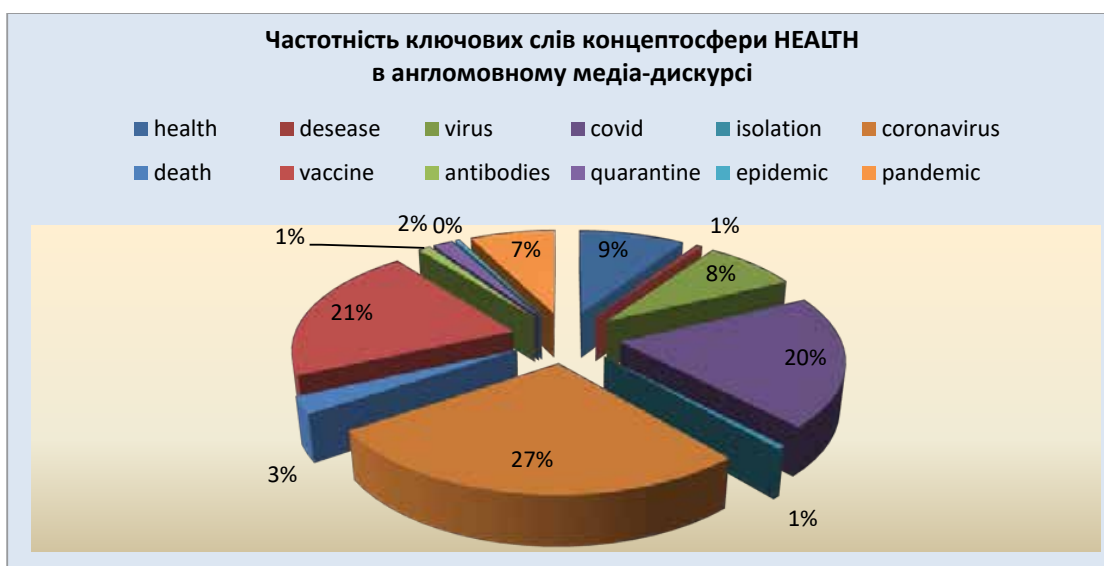


Рис. 4. Частотний аналіз найбільш вживаних ключових слів концептосфери HEALTH в англійському медіа-дискурсі

повернутись з минулого, спровокувати масовий спалах, не віщувати біди, «оголосити війну» фейкам, вбивати лише багатих, триматися в повітрі, залишатися назавжди, корона вірус вбиває лише багатих, злягати з корона вірусом тощо.

Конкорданси можуть бути представлені і в інших формах, зручних для користувачів і відповідати конкретним завданням. При необхідності на екрані може бути відображена додаткова інформація (про джерело прикладу – найменування, тип, варіант англійської / української мови).

Висновки. Конкорданс як особливий тип словника представляє інтерес для лінгвістів, перш за все, як джерело інформації. Це словник, який візуалізує впорядковану базу контекстів вживання слова в межах аналізованого тексту. Структура

конкордансу, як і будь-якого іншого типу словника, повинна варіюватися в залежності від типу тексту, його родової, жанрової приналежності.

Конкорданс є надійною інформаційною базою, що дозволяє досліднику в залежності від поставлених завдань обрати свій метод аналізу та інтерпретації. За допомогою методик стиснення конкордансу і концептуального аналізу словникової статті конкордансу можна зробити висновок про місце лексеми в структурі актуалізованого лексикону, її семантичних і граматичних особливостей, синтагматических зв'язках, а також виявити імпліцитні смисли, вербалізовані лексемою, у моделюванні концепту. Лінгвістичний корпус і програма конкорданс можуть також значно оптимізувати процес вивчення іноземних мов.

Список літератури:

1. Аскольдов С. А. Концепт и слово / Русская словесность: антология / под. ред. В.Н. Нерознака. Москва, 1997. С. 267–280.
2. Загорійчук А., Орловська О. Методи лінгвістичних досліджень. URL: <http://intkonf.org/zagoriychuk-a-kpn-orlovska-ov-metodi-lingvistichnihdosl-idzhen>
3. Лабенко О. В. Концептуалізація образу ХВОРОБИ в текстах зачинів українських, англійських та французьких замовлянь: структура, семантика, символізм. Теоретична і дидактична філологія : збірник наукових праць. Вип. 25. Переяслав-Хмельницький, 2017. С. 218–226.
4. Мартінек С. Концептосфера здоров'я в українській мовній картині світу. Вісник Львівського університету. Серія філологічна. Вип. 70. 2019. С. 234–247.
5. Петкау А. Ю. Концепт здоровье в советской социальной рекламе. Лингвистический, социальный, историко-культурный, дидактический контексты функционирования русского языка как государственного языка Российской Федерации : сб. материалов межрегион. конф. Волгоград, 2014. С. 521–526.
6. Петкау А. Ю. Концепт здоровье: перцептивный образ. Русский язык: исторические судьбы и современность : тр. и материалы междунар. конгресса исследователей русского языка. Москва, 2014. С. 126.
7. Печенікова Л. М. Концепт здоров'я в системі мови (на матеріалі праць ХНПУ ім. Г. С. Сковороди). Вип. 31. 2011. С. 10–13.
8. Соснина Е. П. Корпусная лингвистика и корпусный подход в обучении иностранному языку. URL: http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_linguistics_language_teaching/
9. Толдова С. Ю. Конкорданс. URL: <http://www.lomonosovfund.ru/enc/ru/encyclopedia:0127200>
10. Ball C. N. Tutorial: Concordances and Corpora. URL: <http://www.george.town.edu/cball/corpora/tutorial.html>
11. Corpus Concordance English. URL: <http://lertextutor.ca/conc/eng/>
12. Kennedy G. (1998). An Introduction to Corpus Linguistics. London.
13. Leech G. (1991). The State of Art in Corpus Linguistics. English Corpus Linguistics / Aijmer K., Altenberg B. (eds.). London. P. 8–29.
14. McEnery A., Wilson A. (1996). Corpus Linguistics. Edinburgh.
15. Oakes M.P. (1998). Statistics for Corpus Linguistics. Edinburgh University Press, Edinburgh.
16. Sinclair J. (1991). Corpus, Concordance, Collocation, Oxford University Press.

References:

1. Askoldov S.A. (1997). Kontsept y slovo. Russkaia slovesnost: antolohiya / pod. red. V.N. Neroznaka. Moscow, pp. 267–280. (in Russian)
2. Zahoriichuk A., Orlovska O. Metody linhvistychnykh doslidzhen. URL: <http://intkonf.org/zagoriychuk-a-kpn-orlovska-ov-metodi-lingvistichnihdoslidzhen> (in Ukrainian)
3. Labenko O. V. (2017). Kontseptualizatsiia obrazu khvoroby v tekstakh zachyniv ukrainskykh, anhliiskykh ta frantsuzkykh zamovliany: struktura, semantyka, symvolizm. Teoretychna i dydaktychna filolohiia: zbirnyk naukovykh prats, vol. 25. Pereiaslav-Khmelnitskyi, pp. 218–226. (in Ukrainian)
4. Martinek S. (2019). Kontseptosfera zdorovia v ukrainskii movnii kartyni svitu. Visnyk Lvivskoho universytetu. Seriia filolohichna, vol. 70, pp. 234–247. (in Ukrainian)
5. Petkau A. Yu. (2014). Kontsept zdorove v sovetskoii sotsyalnoi reklame. Linhvystycheskyi, sotsyalnyi, ystoryko-kulturnyi, dydaktycheskyi konteksty funktsyonyrovannya russkoho yazyka kak gosudarstvennoho yazyka Rossyiskoi Federatsyy: sb. materyalov mezhrehyon. konf. Volhohrad, pp. 521–526. (in Russian)
6. Petkau A. Yu. (2014). Kontsept zdorove: pertseptyvnyi obraz. Russkyi yazyk: ystorycheskye sudby y sovremennost: tr. y materyaly mezhdunar. konhressa yssledovatelei russkoho yazyka. Moskva, p. 126. (in Russian)
7. Pechenikova L. M. (2011). Kontsept zdorovia v systemi movy (na materialy prats KhNPU im. H. S. Skovorody), vol. 31, pp. 10–13. (in Ukrainian)
8. Sosnyina E. P. Korpusnaia linhvystyka y korpusnyi podkhod v obuchenii ynostrannomu yazyku. URL: http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_linguistics_language_teaching/ (in Russian)
9. Toldova S. Yu. Konkordans. URL: <http://www.lomonosovfund.ru/enc/ru/encyclopedia:0127200> (in Russian)
10. Ball Catherine N. Tutorial: Concordances and Corpora. URL: www.george.town.edu/cball/corpora/tutorial.html
11. Corpus Concordance English. URL: <http://lertextutor.ca/conc/eng/>
12. Kennedy G. (1998). An Introduction to Corpus Linguistics. London.
13. Leech G. (1991). The State of Art in Corpus Linguistics. English Corpus Linguistics / Aijmer K., Altenberg B. (eds.). London, p. 8–29.
14. McEnery A., Wilson A. (1996). Corpus Linguistics. Edinburgh.
15. Oakes M.P. (1998). Statistics for Corpus Linguistics. Edinburgh University Press, Edinburgh.
16. Sinclair J. (1991). Corpus, Concordance, Collocation, Oxford University Press.