

УДК 81-119

*Яницька О. М.,**викладач кафедри практики німецької та французької мов Рівненського державного гуманітарного університету*КОРПУСНИЙ АНАЛІЗ ДЛЯ ВИЗНАЧЕННЯ КЛЮЧОВИХ ЛЕКСЕМ-НОМІНАНТІВ КОНЦЕПТУ *VIPA* НА МАТЕРІАЛІ АНГЛІЙСЬКОЇ, УКРАЇНСЬКОЇ ТА ФРАНЦУЗЬКОЇ МОВ

Анотація. У статті описано етапи корпусного аналі-технологій дав можливість інформатизувати велику кількість зу для визначення ключових лексем-номінантів концепту текстів. Для можливості роботи з такими обсягами текстів, по-
VIPA в англійській, українській і французькій мовах. Про-шуку в них потрібної інформації почали створюватися лінгві-аналізовано концепт *VIPA* в Британському національному стичні корпуси, тобто колекції текстів, спеціально відібраних,
 корпусі, Національному центрі текстуальних і лексичних ресурсів французької мови та корпусі текстів української розмічених за різними лінгвістичними параметрами й забезпе-
 мови, лабораторії комп'ютерної лінгвістики Інституту фі-чених системою пошуку. У дослідженні використовуватимемо логії Київського національного університету імені Тара-корпусний аналіз для визначення ключових лексем-номінантів са Шевченка.
 концепту *VIPA* в англійській, українській і французькій мовах.
Ключові слова: корпус, корпусна лінгвістика, текст, Одна з найважливіших властивостей корпусу текстів – його синонім.
 репрезентативність, тобто здатність відбивати всі властивості проблемної ділянки. Репрезентативність визначається фоне-
Постановка проблеми. Корпусна лінгвістика є надзвичай- тичними, морфологічними, синтаксичними, стильовими па-
 но важливою галуззю сучасної лінгвістичної науки. Корпусний раметрами. Головною метою корпусної лінгвістики є лінгвіс-
 аналіз дав змогу уточнити результати й висновки раніше прове- тичний опис мовної системи (конкретне вивчення комунікації
 денних досліджень мови, проводити нові, ширші та системніші людей), особливий спосіб відображення мовного матеріалу в
 за кількістю охоплення емпіричного мовного матеріалу лінгві- корпусі текстів, який, у свою чергу, може використовуватися
 стичні дослідження. У центрі уваги корпусного лінгвістичного іншими лінгвістичними дисциплінами. Корпус текстів – це
 аналізу опинилася мовна особа, тобто її мовна діяльність, ма- певний філологічний об'єкт; організована словесна множи-
 сова комунікація, проблема її опису.
 на, елементами якої є відібрані тексти; організована певним
 Корпусна лінгвістика швидко розвивається в англосаксон- чином словесна єдність, елементами якої є тексти або спеці-
 ській науці, провідні лабораторії Європи та США працюються альним чином відібрані уривки з текстів [7]. Тексти в корпус

над розвитком, збагаченням і поповненням мовних корпусів. відбираються згідно з проблемною ділянкою, тобто ділянкою

Дослідженням загальної теорії корпусної лінгвістики присвячені реалізації мовних явищ, які цікавлять дослідника. Однією з

важчих праць таких науковців, як А. Люделінг, М. Кіто; кореляцій- найважливіших вимог, що висувається до складу корпусу, є

цією корпусної лінгвістики й інших лінгвістичних дисциплін, вимога повноти.

Мовний корпус є ні що інше як збір конкрет-

типологією корпусів, методикою інтерпретації корпусних них уживань певних мовних аспектів. Отже, корпусний аналіз

даних, створенням текстових корпусів займаються лінгвісти: дає змогу зробити висновки про феномени мови, ґрунтуючись

А. Реноуф, В. Жуковська, О. Демська. Новизна статті полягає у на конкретному мовному матеріалі. Корпусна лінгвістика не є

спробі дослідити концепт ВІРА в різних корпусах. Оскільки ра- альтернативою традиційній лінгвістиці, а, навпаки, доповнює і

ніше такого типу порівняльний аналіз не проводився, він дасть збагачує її [6].

Змогу поглянути під іншим кутом на мовні реалії і зробити нові

У дослідженні для аналізу англійського корпусу використову-

висновки.

ватимемо Британський національний корпус (BNC), типологіч-

Мета статті – описати особливості корпусного аналізу, ними характеристиками якого є фрагментарність, синхронність,

проаналізувати концепт ВІРА у британському, французькому та загальномовність, мономовність, змішаність. Основними сфе-

українському корпусах.

рами використання є видавнича справа; академічні лінгвістичні

Предмет – концепт ВІРА у британському, українському та студії, вивчення мови; штучний інтелект, оброблення природної

французькому корпусах.

мови (NLP); пошук інформації. Обсяг корпусу – 100 млн. сло-

Матеріалом для дослідження послуговували дані Британ- вовживань. Багато людей використовують корпуси і працюють

ського національного корпусу, Національного центру тексту- із ними через програми *Concordance*. Лінії відповідності зводять

альних і лексичних ресурсів французької мови й корпусу тек- разом багато зразків використання слова або фрази, даючи до-

ствів української мови, лабораторії комп'ютерної лінгвістики слідникові можливість вивчати регулярність. Корпусний аналіз

Інституту філології Київського національного університету також допомагають аналізувати концептуальні картини, піз-

імені Тараса Шевченка.

навальні структури, мову думки. Концептуальна картина світу

Виклад основного матеріалу дослідження. Корпусна всебічно глобальніша, ніж будь-яка картина мови. Контекстуаль-

лінгвістика – це розділ прикладної лінгвістики, що займаєть- но-корпусне дослідження Британського національного корпусу

ся розробкою загальних принципів побудови й використання дає нам змогу аналізувати мову об'єктивно і глибоко. Програма

лінгвістичних корпусів (корпусів текстів). Дослідник дослі- перераховує слова, готує списки слів, визначає частоти вживання

джує текст, письмовий або усний. Розвиток комп'ютерних слова й індекси.

Результати мають широке застосування в лінгві-

160

ISSN 2409-1154 Науковий вісник Міжнародного гуманітарного університету. Сер.:
Філологія. 2016 № 25 том 1

стичному аналізу, лексикографії, перекладі, численних комерцій-
у словнику фольклору – дані відсутні.

них галузях, академічних дисциплінах. Програма *Concordance*

Довір'я, релігія, віровизнання, вірування

дає можливість також сортувати слова в багатьох відношеннях,

у словнику публіцистики – дані відсутні;

шукати фрази, робити пошук спорідненості, типових слів і регу-

у словнику художньої прози – дані відсутні;

лярного вживання.

у словнику фольклору – дані відсутні.

Використовуючи дані словника с *иномімів Oxford thesaurus*

Конфесія

dictionary of synonyms знаходимо *faith* [8, с. 536], проводимо

у словнику публіцистики:

корпусний аналіз кожного іменника в програмі й визначаємо

Слово – **конфесія**. Частина мови – **ім. ж. р.** Абсолютна час-

частотність уживання.

тота – **71**;

Faith (5045)

у словнику художньої прози – дані відсутні;

1) **belief (5014)**, *credence* (176), **confidence (6766)**, *conviction*

у словнику фольклору – дані відсутні.

(1977), **trust** (9909), *certainty* (1338), *certitude* (24), *assurance* (1797),

Отже, визначаємо ключові лексеми концепту ВІРА в україн-

assuredness (6), *sureness* (29), *reliance* (855), *dependence* (1257);

ській мові: ВІРА й ДОВІРА – за частотою вживання в корпусах.

2) *belief* (5014), *religion* (4305), *creed* (485), *persuasion* (514),

Переходимо до аналізу концепту ВІРА у французькому

dogma (252), *teaching* (8915), *doctrine* (1689), *denomination* корпусі. Будемо

використовувати базу даних *Національного*

(110), *sect* (244);

центру текстуальних і лексичних ресурсів/Centre National de

3) *duty* (7806), *allegiance* (650), *obligation* (2186), *promise* *Ressources Textuelles et*

Lexicales. Центр створений у 2005 році

(3760), *faithfulness* (60), *loyalty* (1585), *fidelity* (323), *devotion* як єдиний портал,

сукупність інформатизованих лінгвістич-

(585), *consecration* (107), *dedication* (609), *fealty* (51), *obedience* них ресурсів і

обробки мови. Центр інтегрує різноманітні бази

(635).

лінгвістичних даних, а також дослідження корпусу французь-

Провівши аналіз і порівнявши дані з даними тлумачних кої мови *Corpus français* –
700 мільйонів слів он-лайн.

словників, словників-синомімів, двомовних перекладацьких

Користуючись даними словника синонімів *Le Robert*,

словників (англійська-французька, англійська-українська), ви- *Dictionnaire des*

synonymes, des nuances et des contraires sous la

значаємо ключові лексеми-номінанти концепту ВІРА в англій- *direction de*

Dominique Le Fur, визначаємо синонімічний ряд **foi**

ській мові – FAITH, BELIEF, TRUST, CONFIDENCE.

n.f. [4, с. 513], аналізуємо частотність уживання кожної одиниці.

Переходимо до аналізу концепту ВІРА в українському кор-
Foi n.f. (31014)

пусі. Досліджуватимемо корпуси текстів української мови, роз-
1) croyance (4446) – conviction (9256);

роблені співробітниками лабораторії комп'ютерної лінгвістики

2) religion (26442) – confession (3094) – culte (16518) – dogme

Інституту філології Київського національного університету (1752) – église (49373);

імені Тараса Шевченка Н. Дарчук і В. Сорокіним (керівниками

3) confiance (76538) – crédit (38717);

проекту). Цей тип аналізу дає змогу зробити дослідження ста-

4) conscience (36612) – honneur (43663) – loyauté (2668) –

тистично обгрунтованим і поглибленим.

probité (627) – sincérité (2835).

Опираючись на дані Практичного словника синонімів укра-

Розгорнуті дані корпусного дослідження лексем-номінантів

їнської мови, укладеного С. Караванським [3, с. 56], виводимо концепту:

синонімічний ряд.

FOI n.f.

ВІРА, упевненість (до кого), довіра, довір'я (Христова релі-

mot-clé: foi nombre d'occurrences: 31014 classe de fréquence:

гія), віровизнання; (у містичні істоти) вірування, конфесія. Для 10.

отримання вибірки слів за вказаним діапазоном частот вико-

CROYANCE n.f.

ривуватимемо частотні словники публіцистики, художньої

mot-clé: croyance nombre d'occurrences: 4446 classe

прози та фольклору.

de fréquence: 13.

Віра у словнику публіцистики:

CONFIANCE n.f.

Слово – **віра**. Частина мови – **ім. ж. р.** Абсолютна

mot-clé: confiance nombre d'occurrences: 76538 classe

частота – **443**;

de fréquence: 9.

у словнику художньої прози:

CONSCIENCE n.f.

Слово – **віра**. Частина мови – **ім. ж. р.** Абсолютна

mot-clé: conscience nombre d'occurrences: 36612 classe

частота – **606**;

de fréquence: 10.

у словнику фольклору – дані відсутні.

Проаналізувавши та порівнявши дані з даними тлумачних

Упевненість у словнику публіцистики:

словників, словників-синонімів, двомовних перекладацьких

Слово – **упевненість**. Частина мови – **ім. ж. р.** Абсолютна словників (англійська-
французька, французька-українська), ви-

частота – **66**;

значаємо ключові лексеми-номінанти концепту ВІРА у французь-
у словнику художньої прози:

кій мові – **FOI, CROYANCE, CONSCIENCE, CONFIANCE.**

Слово – **упевненість.** Частина мови – **ім. ж. р.** Абсолютна

Висновки. Підсумовуючи викладене, можна резюмувати,
частота – **17;**

що висока частотність уживання концепту в корпусах показує

у словнику фольклору – дані відсутні.

семантичну щільність концепту. Цей аналіз дав змогу визна-

Довіра у словнику публіцистики:

чити ключові лексеми-номінанти концепту ВІРА в англійській

Слово – **довіра.** Частина мови – **ім. ж. р.** Абсолютна (FAITH, BELIEF, TRUST,
CONFIDENCE), українській (ВІРА,

частота – **626;**

ДОВІРА) та французькій (FOI, CROYANCE, CONSCIENCE,

у словнику художньої прози:

CONFIANCE) мовах. На перспективу залишається аналіз фра-

Слово – **довіра.** Частина мови – **ім. ж. р.** Абсолютна зоологічної об'єктивації
концепту й лінгвокультурний аналіз

частота – **105;**

значення ключових лексем-номінантів.

161

ISSN 2409-1154 Науковий вісник Міжнародного гуманітарного університету. Сер.:
Філологія. 2016 № 25 том 1

Література:

Яницкая Е. Н. Корпусный анализ для определения

1. Жуковська В.В. Вступ до корпусної лінгвістики : [навчальний по-
ключевых лексем-номинантов концепта ВЕРА на мате-
сібник] / В.В. Жуковська. – Житомир : Вид-во ЖДУ ім. І. Франка,
риале английского, украинского и французского языков
2013. – 142 с.

Аннотация. В статье описаны этапы корпусного ана-

лиза для определения ключевых лексем-номинантов кон-

2. Демська О.М. Текстовий корпус: ідея іншої форми / О.М. Демська. –

цепта ВЕРА в английском, украинском и французском

К. : Національний університет «Києво-Могилянська академія»,

языках. Проанализирован концепт ВЕРА в Британском

2011. – 282 с.

национальном корпусе, Национальном центре текстовых

3. Караванський С. Практичний словник синонімів української мови /

и лексических ресурсов французского языка и корпусе тек-

С. Караванський. – К. : Українська книга, 2000. – 480 с.

стов украинского языка, лаборатории компьютерной линг-

4. Dominique le Fur, Le Dictionnaire des synonymes, nuances et

вистики Інститута філології Київського національного

contraires, Paris, 2005. – 498 p.

університета імені Тараса Шевченка.

5. Dubois Jean, Giacomo Mathée, Guespin Louis, Marcellesi Christiane.

Ключевые слова: корпус, корпусная лінгвістика,

текст, синонім.

Dictionnaire de linguistique Larousse, 2002. – 514 p.

6. Lüdeling A., Kytö M. Corpus Linguistics. An International Handbook

**Ianytska O. Corpus analysis to identify key lexemes that
Walter de Gruyter, 2009. – 1402 pp.
nominate the concept of FAITH in the English, Ukrainian
7. Meyer Charles F. English Corpus Linguistics: An Introduction.
and French languages**

Cambridge University Press, 2004. – 168 p.

Summary. The article describes the stages of corpus anal-

8. Oxford Thesaurus – An A-Z Dictionary of Synonyms. – 2062 p.

ysis to identify key lexemes of the nominees of concept FAITH

9. Renouf Antoinette, Kehoe Andrew (Editors). Corpus Linguistics:

in the English, Ukrainian and French languages. The concept

Refinements and Reassessments : Rodopi, 2009. – vii, 462 p.

FAITH is analyzed in the British national corpus, the National

Center for Textual and Lexical Resources of the French lan-

10. [Електронний ресурс]. – Режим доступу : <http://www.mova.info/>

guage and the corpus of the Ukrainian language of the labo-

corpus.aspx.

ratory of computer linguistics of the Institute of Philology of

11. [Електронний ресурс]. – Режим доступу : <http://www.cnrtl.fr/>

Kyiv national University named after Taras Shevchenko.

12. [Електронний ресурс]. – Режим доступу : <http://corpus.byu.edu/bnc/>

Key words: corpus, corpus linguistics, text, synonym.

162