

**Яницька О. М.,**  
викладач кафедри практики німецької та французької мов  
Рівненського державного гуманітарного університету

## КОРПУСНИЙ АНАЛІЗ ДЛЯ ВИЗНАЧЕННЯ КЛЮЧОВИХ ЛЕКСЕМ-НОМІНАНТІВ КОНЦЕПТУ *VIRA* НА МАТЕРІАЛІ АНГЛІЙСЬКОЇ, УКРАЇНСЬКОЇ ТА ФРАНЦУЗЬКОЇ МОВ

**Анотація.** У статті описано етапи корпусного аналізу для визначення ключових лексем-номінантів концепту *VIRA* в англійській, українській і французькій мовах. Проаналізовано концепт *VIRA* в Британському національному корпусі, Національному центрі текстуальних і лексичних ресурсів французької мови та корпусі текстів української мови, лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка.

**Ключові слова:** корпус, корпусна лінгвістика, текст, синонім.

**Постановка проблеми.** Корпусна лінгвістика є надзвичайно важливою галуззю сучасної лінгвістичної науки. Корпусний аналіз дав змогу уточнити результати й висновки раніше проведених досліджень мови, проводити нові, ширші та системніші за кількістю охоплення емпіричного мовного матеріалу лінгвістичні дослідження. У центрі уваги корпусного лінгвістичного аналізу опинилася мовна особа, тобто її мовна діяльність, масова комунікація, проблема її опису.

Корпусна лінгвістика швидко розвивається в англосаксонській науці, провідні лабораторії Європи та США працюють над розвитком, збагаченням і поповненням мовних корпусів. Дослідженням загальної теорії корпусної лінгвістики присвячені праці таких науковців, як А. Люделінг, М. Кіто; кореляцією корпусної лінгвістики й інших лінгвістичних дисциплін, типологією корпусів, методикою інтерпретації корпусних даних, створенням текстових корпусів займаються лінгвісти: А. Реноуф, В. Жуковська, О. Демська. Новизна статті полягає у спробі дослідити концепт *VIRA* в різних корпусах. Оскільки раніше такого типу порівняльний аналіз не проводився, він дасть змогу поглянути під іншим кутом на мовні реалії і зробити нові висновки.

**Мета статті** – описати особливості корпусного аналізу, проаналізувати концепт *VIRA* у британському, французькому та українському корпусах.

Предмет – концепт *VIRA* у британському, українському та французькому корпусах.

Матеріалом для дослідження послуговували дані Британського національного корпусу, Національного центру текстуальних і лексичних ресурсів французької мови й корпусу текстів української мови, лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка.

**Виклад основного матеріалу дослідження.** Корпусна лінгвістика – це розділ прикладної лінгвістики, що займається розробкою загальних принципів побудови й використання лінгвістичних корпусів (корпусів текстів). Дослідник досліджує текст, письмовий або усний. Розвиток комп'ютерних

технологій дав можливість інформатизувати велику кількість текстів. Для можливості роботи з такими обсягами текстів, пошуку в них потрібної інформації почали створюватися лінгвістичні корпуси, тобто колекції текстів, спеціально відібраних, розмічених за різними лінгвістичними параметрами й забезпечених системою пошуку. У дослідженні використовуватимемо корпусний аналіз для визначення ключових лексем-номінантів концепту *VIRA* в англійській, українській і французькій мовах. Одна з найважливіших властивостей корпусу текстів – його репрезентативність, тобто здатність відбивати всі властивості проблемної ділянки. Репрезентативність визначається фонетичними, морфологічними, синтаксичними, стильовими параметрами. Головною метою корпусної лінгвістики є лінгвістичний опис мовної системи (конкретне вивчення комунікації людей), особливий спосіб відображення мовного матеріалу в корпусі текстів, який, у свою чергу, може використовуватися іншими лінгвістичними дисциплінами. Корпус текстів – це певний філологічний об'єкт; організована словесна множина, елементами якої є відібрані тексти; організована певним чином словесна єдність, елементами якої є тексти або спеціальним чином відібрані уривки з текстів [7]. Тексти в корпусі відбираються згідно з проблемною ділянкою, тобто ділянкою реалізації мовних явищ, які цікавлять дослідника. Однією з найважливіших вимог, що висувається до складу корпусу, є вимога повноти. Мовний корпус є ні що інше як збір конкретних уживань певних мовних аспектів. Отже, корпусний аналіз дає змогу зробити висновки про феномени мови, ґрунтуючись на конкретному мовному матеріалі. Корпусна лінгвістика не є альтернативою традиційній лінгвістиці, а, навпаки, доповнює і збагачує її [6].

У дослідженні для аналізу англійського корпусу використовуватимемо Британський національний корпус (BNC), типологічними характеристиками якого є фрагментарність, синхронність, загальнономовність, мономовність, змішаність. Основними сферами використання є видавничі справи; академічні лінгвістичні студії, вивчення мови; штучний інтелект, оброблення природної мови (NLP); пошук інформації. Обсяг корпусу – 100 млн. слововживань. Багато людей використовують корпуси і працюють із ними через програми *Concordance*. Лінії відповідності зводять разом багато зразків використання слова або фрази, даючи дослідникові можливість вивчати регулярність. Корпусний аналіз також допомагають аналізувати концептуальні картини, пізнавальні структури, мову думки. Концептуальна картина світу всебічно глобальніша, ніж будь-яка картина мови. Контекстуально-корпусне дослідження Британського національного корпусу дає нам змогу аналізувати мову об'єктивно і глибоко. Програма перераховує слова, готує списки слів, визначає частоти вживання слова й індекси. Результати мають широке застосування в лінгві-

стичному аналізу, лексикографії, перекладі, численних комерційних галузях, академічних дисциплінах. Програма *Concordance* дає можливість також сортувати слова в багатьох відношеннях, шукати фрази, робити пошук спорідненості, типових слів і регулярного вживання.

Використовуючи дані словника синонімів *Oxford thesaurus dictionary of synonyms* знаходимо *faith* [8, с. 536], проводимо корпусний аналіз кожного іменника в програмі й визначаємо частотність уживання.

#### **Faith (5045)**

1) **belief (5014)**, credence (176), **confidence (6766)**, conviction (1977), **trust (9909)**, certainty (1338), certitude (24), assurance (1797), assuredness (6), sureness (29), reliance (855), dependence (1257);

2) belief (5014), religion (4305), creed (485), persuasion (514), dogma (252), teaching (8915), doctrine (1689), denomination (110), sect (244);

3) duty (7806), allegiance (650), obligation (2186), promise (3760), faithfulness (60), loyalty (1585), fidelity (323), devotion (585), consecration (107), dedication (609), fealty (51), obedience (635).

Провівши аналіз і порівнявши дані з даними тлумачних словників, словників-синонімів, двомовних перекладацьких словників (англійська-французька, англійська-українська), визначаємо ключові лексеми-номінанти концепту ВІРА в англійській мові – FAITH, BELIEF, TRUST, CONFIDENCE.

Переходимо до аналізу концепту ВІРА в українському корпусі. Досліджуватимемо корпуси текстів української мови, розроблені співробітниками лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка Н. Дарчук і В. Сорокіним (керівниками проекту). Цей тип аналізу дає змогу зробити дослідження статистично обґрунтованим і поглибленим.

Опираючись на дані Практичного словника синонімів української мови, укладеного С. Караванським [3, с. 56], виводимо синонімічний ряд.

ВІРА, упевненість (до кого), довіра, довір'я (Христова релігія), віровизнання; (у містичні істоти) вірування, конфесія. Для отримання вибірки слів за вказаним діапазоном частот використовуватимемо частотні словники публіцистики, художньої прози та фольклору.

**Віра** у словнику публіцистики:

Слово – **віра**. Частина мови – **ім. ж. р.** Абсолютна частота – **443**;

у словнику художньої прози:

Слово – **віра**. Частина мови – **ім. ж. р.** Абсолютна частота – **606**;

у словнику фольклору – дані відсутні.

**Упевненість** у словнику публіцистики:

Слово – **упевненість**. Частина мови – **ім. ж. р.** Абсолютна частота – **66**;

у словнику художньої прози:

Слово – **упевненість**. Частина мови – **ім. ж. р.** Абсолютна частота – **17**;

у словнику фольклору – дані відсутні.

**Довіра** у словнику публіцистики:

Слово – **довіра**. Частина мови – **ім. ж. р.** Абсолютна частота – **626**;

у словнику художньої прози:

Слово – **довіра**. Частина мови – **ім. ж. р.** Абсолютна частота – **105**;

у словнику фольклору – дані відсутні.

**Довір'я, релігія, віровизнання, вірування**

у словнику публіцистики – дані відсутні;

у словнику художньої прози – дані відсутні;

у словнику фольклору – дані відсутні.

**Конфесія**

у словнику публіцистики:

Слово – **конфесія**. Частина мови – **ім. ж. р.** Абсолютна частота – **71**;

у словнику художньої прози – дані відсутні;

у словнику фольклору – дані відсутні.

Отже, визначаємо ключові лексеми концепту ВІРА в українській мові: ВІРА й ДОВІРА – за частотою вживання в корпусах.

Переходимо до аналізу концепту ВІРА у французькому корпусі. Будемо використовувати базу даних *Національного центру текстуральних і лексичних ресурсів/Centre National de Ressources Textuelles et Lexicales*. Центр створений у 2005 році як єдиний портал, сукупність інформатизованих лінгвістичних ресурсів і обробки мови. Центр інтегрує різноманітні бази лінгвістичних даних, а також дослідження корпусу французької мови *Corpus français* – 700 мільйонів слів он-лайн.

Користуючись даними словника синонімів *Le Robert, Dictionnaire des synonymes, des nuances et des contraires sous la direction de Dominique Le Fur*, визначаємо синонімічний ряд **foi n.f.** [4, с. 513], аналізуємо частотність уживання кожної одиниці.

Foi n.f. (31014)

1) croyance (4446) – conviction (9256);

2) religion (26442) – confession (3094) – culte (16518) – dogme (1752) – église (49373);

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

4) conscience (36612) – honneur (43663) – loyauté (2668) – probité (627) – sincérité (2835).

Розгорнуті дані корпусного дослідження лексем-номінантів концепту:

**FOI n.f.**

mot-clé: foi nombre d'occurrences: 31014 classe de fréquence: 10.

**CROYANCE n.f.**

mot-clé: croyance nombre d'occurrences: 4446 classe de fréquence: 13.

**CONFIANCE n.f.**

mot-clé: confiance nombre d'occurrences: 76538 classe de fréquence: 9.

**CONSCIENCE n.f.**

mot-clé: conscience nombre d'occurrences: 36612 classe de fréquence: 10.

Проаналізувавши та порівнявши дані з даними тлумачних словників, словників-синонімів, двомовних перекладацьких словників (англійська-французька, французька-українська), визначаємо ключові лексеми-номінанти концепту ВІРА у французькій мові – **FOI, CROYANCE, CONSCIENCE, CONFIANCE**.

**Висновки.** Підсумовуючи викладене, можна резюмувати, що висока частотність уживання концепту в корпусах показує семантичну щільність концепту. Цей аналіз дав змогу визначити ключові лексеми-номінанти концепту ВІРА в англійській (FAITH, BELIEF, TRUST, CONFIDENCE), українській (ВІРА, ДОВІРА) та французькій (FOI, CROYANCE, CONSCIENCE, CONFIANCE) мовах. На перспективу залишається аналіз фразеологічної об'єктивації концепту й лінгвокультурний аналіз значення ключових лексем-номінантів.

*Література:*

1. Жуковська В.В. Вступ до корпусної лінгвістики : [навчальний посібник] / В.В. Жуковська. – Житомир : Вид-во ЖДУ ім. І. Франка, 2013. – 142 с.
2. Демська О.М. Текстовий корпус: ідея іншої форми / О.М. Демська. – К. : Національний університет «Києво-Могилянська академія», 2011. – 282 с.
3. Караванський С. Практичний словник синонімів української мови / С. Караванський. – К. : Українська книга, 2000. – 480 с.
4. Dominique le Fur, Le Dictionnaire des synonymes, nuances et contraires, Paris, 2005. – 498 p.
5. Dubois Jean, Giacomo Mathée, Guespin Louis, Marcellesi Christiane. Dictionnaire de linguistique Larousse, 2002. – 514 p.
6. Lüdeling A., Kytö M. Corpus Linguistics. An International Handbook Walter de Gruyter, 2009. – 1402 pp.
7. Meyer Charles F. English Corpus Linguistics: An Introduction. Cambridge University Press, 2004. – 168 p.
8. Oxford Thesaurus – An A-Z Dictionary of Synonyms. – 2062 p.
9. Renouf Antoinette, Kehoe Andrew (Editors). Corpus Linguistics: Refinements and Reassessments : Rodopi, 2009. – vii, 462 p.
10. [Електронний ресурс]. – Режим доступу : <http://www.mova.info/corpus.aspx>.
11. [Електронний ресурс]. – Режим доступу : <http://www.cnrtl.fr/>
12. [Електронний ресурс]. – Режим доступу : <http://corpus.byu.edu/bnc/>

**Яницкая Е. Н. Корпусный анализ для определения ключевых лексем-номинантов концепта ВЕРА на материале английского, украинского и французского языков**

**Аннотация.** В статье описаны этапы корпусного анализа для определения ключевых лексем-номинантов концепта ВЕРА в английском, украинском и французском языках. Проанализирован концепт ВЕРА в Британском национальном корпусе, Национальном центре текстовых и лексических ресурсов французского языка и корпусе текстов украинского языка, лаборатории компьютерной лингвистики Института филологии Киевского национального университета имени Тараса Шевченко.

**Ключевые слова:** корпус, корпусная лингвистика, текст, синоним.

**Ianytska O. Corpus analysis to identify key lexemes that nominate the concept of FAITH in the English, Ukrainian and French languages**

**Summary.** The article describes the stages of corpus analysis to identify key lexemes of the nominees of concept FAITH in the English, Ukrainian and French languages. The concept FAITH is analyzed in the British national corpus, the National Center for Textual and Lexical Resources of the French language and the corpus of the Ukrainian language of the laboratory of computer linguistics of the Institute of Philology of Kyiv national University named after Taras Shevchenko.

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