

АСОЦІАТИВНА МЕРЕЖА «КРИЗА» В ПРОМОВАХ ДЖО БАЙДЕНА

Дружб'як Світлана Володимирівна

доцент кафедри прикладної лінгвістики
Національного університету «Львівська політехніка»
вул. Степана Бандери, 30, Львів, Україна

Рій Андріана Василівна

студентка 4 курсу ОС «Бакалавр»
кафедри прикладної лінгвістики
Національного університету «Львівська політехніка»
вул. Степана Бандери, 30, Львів, Україна

Статтю присвячено дослідженню теоретичних аспектів поняття асоціативна мережа та практичному аналізу асоціативних мереж поняття «криза» на основі корпусу промов Джо Байдена. Необхідність аналізу високочастотних колокацій пов'язана із затребуваністю інформації про стійкі словосполучення для прикладних лінгвістичних завдань. В роботі проаналізовано наукові праці, присвячені проблемі сполучуваності, колокацій, асоціацій та асоціативних мереж. Розглянуто сучасні підходи до визначення та дослідження асоціативних мереж. Підтверджено, що використання корпусу текстів та різних лінгвістичних та статистичних інструментів уможливує проведення ґрунтовного аналізу асоціативної мережі заданого слова. Виявлено, що поняття «асоціація» є психосемантичним зв'язком у свідомості людини між двома й більше психічними явищами, а також їхніми позначеннями в мові. Встановлено, що синонімічним до цього поняття є термін «колокації» – словосполучки, які трапляються в тексті частіше поряд, ніж за випадковою ймовірністю окремо. Проведено практичне опрацювання методів створення асоціативних мереж за допомогою корпусу текстів, створених на матеріалі промов Джо Байдена за період з лютого 2022 – по лютий 2023 років. Для візуальної презентації колокаційних відношень між вузлом та його колокатами використано колокаційний граф. Співставлено та проаналізовано результати, отримані за допомогою кожного з підходів.

Ключові слова: асоціативна мережа, колокація, конкорданс, корпус текстів, показники асоціації, колокаційний граф.

Постановка проблеми в загальному вигляді та її обґрунтування її актуальності. Однією з особливостей мови є повторюваність її сполук та одиниць. Цей аспект є причиною формування стійких колокацій, що в подальшому спричинює автоматичне повторення сполук. Останнім часом спостерігаємо зростання зацікавленості науковців щодо ролі частотності в лінгвістичних дослідженнях та використання статистичних підрахунків для її вираховування (Бобкова, 2014; Ділай, 2022; Лотоцька, 2019). Саме поява корпусів уможливила такі статистичні розвідки. В межах корпусної лінгвістики досить поширеними є питання дослідження асоціативних мереж, колокацій та інших статистичних мір сполучуваності. Визначення та дослідження асоціативних мереж є актуальним та необхідним, оскільки така інформація допомагає у вирішенні різноманітних прикладних завдань у лінгвістиці. Зацікавленість аналізом високочастотних колокацій пояснюємо затребуваністю інформації про стійкі словосполучення, оскільки їх виявлення у певних текстах

допомагає зрозуміти їхній зміст та основну мету, а в подальшому – укласти словники, тезауруси, виконувати автоматичний аналіз тексту чи машинний переклад (Бобкова, 2014).

Аналіз останніх досліджень і публікацій. Теоретико-методологічну базу дослідження асоціативних мереж за допомогою корпусів текстів становлять праці вітчизняних та зарубіжних вчених-лінгвістів: J. Sinclair, Т. В. Бобкової, І. П. Ділай, М. П. Ділай, В. В. Жуковської, О. П. Левченко, Н. Я. Лотоцької, Р. К. Махачашвілі, О. Селіванової, Г. Ситар. Зокрема, праця J. Sinclair розкриває питання колокацій та конкордансів у корпусах текстів (Sinclair, 1991). Наукова стаття І. П. Ділай, М. П. Ділай розкриває поняття та роль частотності у корпусній лінгвістиці (Ділай, 2022). Н. Лотоцька запропонувала поняття колокації, як конструкції, що має структуру вільної словосполучки, підпорядкованої цілісності значення вислову, їй властива регулярність уживання у контексті, ця словосполучка трапляється в тексті частіше

поряд, ніж за випадковою вірогідністю окремо (Лотоцька, 2019). Однак ще не вирішеними залишаються питання побудови асоціативної мережі та її дослідження за допомогою корпусного підходу, що свідчить про актуальність пропонованого дослідження.

Формулювання мети і завдань статті.

Метою цієї статті є дослідження асоціативної мережі «криза» на матеріалі дослідницького корпусу промов Джо Байдена за лютий 2022 – лютий 2023 років. Досягнення поставленої мети передбачає виконання наступних завдань: 1) аналіз теоретичних аспектів асоціативних мереж як інструмента лінгвістичних досліджень; 2) дослідження особливостей асоціативних мереж; 3) характеристика методів та інструментів виявлення і відображення асоціативних мереж; 4) створення корпусу промов Джо Байдена, його аналіз.

Виклад основного матеріалу дослідження.

Дослідження колокацій за допомогою корпусного підходу розпочалося після появи праць Дж. Синклера, який сформулював теоретичні засади вивчення колокацій на базі корпусів текстів (Sinclair, 1991). Важливим є саме розуміння поняття «колокація». Отож, у своїй роботі ми дотримуємося визначення, що колокації – це словосполучення, які трапляються в тексті частіше поряд, ніж за випадковою вірогідністю окремо (Левченко, Лотоцька, 2019).

Для здійснення цього корпусного дослідження асоціативних показників в першу чергу було використано систему *Sketch Engine* – інструмент для дослідження найбільш частотних та стійких словосполучень, які класифікуються

за типами відповідно до лексикосинтаксичних шаблонів граматики (Махачашвілі, 2021). Основою дослідження стали промови, що виголошував Джо Байден протягом останнього року. У системі *Sketch Engine* було створено корпус, що містить 610998 токенів та 498000 слів. За допомогою запиту на вкладці *Wordlist*, виявлено, що лема "crisis" зустрічається в корпусі 299 разів. Після цього, у розділі *Word Sketch*, сформовано список найбільш стійких співвіднесенень. На рис. 1 зображено слова, які найчастіше вживаються з досліджуваним словом. На візуалізації можна помітити, що найбільш частотними є словосполучення *climate crisis* та *economic crisis*.

Частотність саме таких колокацій спричинена екстралінгвістичними факторами. Адже останніми роками ми все частіше чуємо про насування світової економічної кризи, яку ще називають «банкопадом» та порівнюють з «Великою Депресією». Окрім цього, початок повномасштабної війни в Україні приблизив світову економіку до початку фінансової кризи. Тому не дивно, що це поняття часто зустрічається в промовах Президента США. Щодо екологічної кризи, то про неї Джо Байден найчастіше згадує на зустрічах з міжнародними політиками, висвітлюючи проблеми глобального потепління, водної кризи чи екологічні проблеми, які спричинилися після нападу росії на Україну.

Одним з основних інструментів для вивчення колокацій на основні корпусних даних є конкорданс. Цим терміном називають рядки корпусу текстів заданої довжини, які представляють

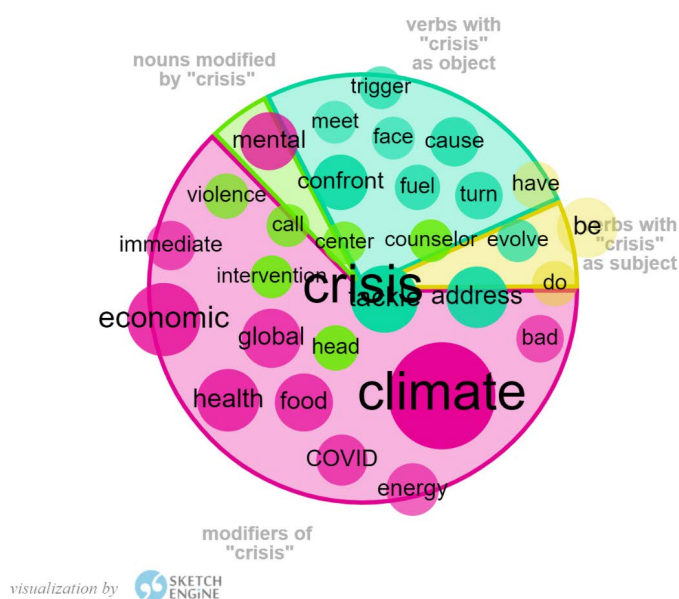


Рис. 1. Знайдені в корпусі словосполучення зі словом crisis

досліджуване слово у контексті (Жуковська, 2013). Таким чином, лінгвіст може отримати інформацію про закономірності вживання досліджуваного слова (нода) та його зв'язки між іншими словами (колокатами). Безумовно, колокації можуть мати кілька варіантів колокатів, які належать ноду (Бобкова, 2015). Зробивши запит на пошук кондордансів зі словом *crisis*, виявлено, що найчастотнішими є такі лексеми: *climate, economic, this, every, health, food, energy, COVID, healthcare, opioid, Missile, immediate, humanitarian, ongoing, future, evolving, financial, major, security*.

Наші результати чітко показують, які саме кризи є найбільш обговорюваними у політичному середовищі. Окрім згадуваних раніше економічної/фінансової та екологічної кризи, у промовах Президента США досить часто згадуються кризи у сфері охорони здоров'я та криза, пов'язана з пандемією COVID-19. Це спричинено тим, що близько 3 років у США були застосовані обмежувальні заходи для протидії COVID-19, які були скасовані лише у квітні 2023 року. В загальному, пандемія стала найбільшою кризою у сфері охорони здоров'я.

Крім цього, виявлено, що Джо Байден часто вживає термін «гуманітарна криза» та «енергетична криза». Обидва терміни вживаються певною мірою в контексті повномасштабної російсько-української війни, в першому випадку – криза біженців, в другому – щодо енергетичного терору росії. Тим не менш, часто згадується також про енергетичну кризу в Європі та США.

Досліджуючи асоціативні зв'язки між певними словами, необхідно оцінити силу такого зв'язку. Для цього використовують статистичні показники асоціації, серед яких *MI-score, t-score log-likelihood, Dice, gmean* та ін. Оскільки в нашому дослідженні застосовано з коефіцієнт *MI-score*, хочемо детальніше про нього розповісти.

Показник *MI-score* дає інформацію про притягання між двома словами, або ймовірність того, що якщо одне слово з'являється в рядку, слово, яке складає з ним колокацію, з'явиться в найближчому оточенні досліджуваного слова (Ситар, 2018). За допомогою цього показника можемо виокремити стійкі словосполучення, власні назви чи спеціальні терміни.

Середовище *Sketch Engine* автоматично вираховує показники асоціації. Відповідно до результатів, найвищі показники мають лекси *Cuban* (9,87) та *climate* (9,13). Крім цього, досить високим є значення *MI-score* для *mental* (8,02),

immediate (8,19), *opioid* (8,28), *economic* (7,44), *health* (6,99), *food* (6,85), *global* (6,70), *COVID* (7,20), *energy* (5,48), *bad* (6,18), *confront* (7,99).

Також варто звернути увагу на показник *log Dice*. Цей статистичний показник використовується для виявлення спільної появи певних слів. Він виражає силу словосполучення. В нашому випадку найвищі показники *log Dice* мають словосполучення зі словами: *climate* (12,2063), *economic* (10,4834), *health* (9,81149), *mental* (9,7914), *food* (9,55982), *every* (9,43974), *global* (9,42326), *COVID* (8,93033), *immediate* (8,83409), *intro* (8,81738), *energy* (8,53123), *bad* (8,25181), *Cuban* (8,20558), *opioid* (8,15451), *confront* (8,13788), *rapidly* (8,13375), *healthcare* (8,08114), *humanitarian* (8,03422).

Отримані результати певною мірою підтверджують попередні дослідження за допомогою конкордансів та аналізу стійких словосполучень. Не дивно, що найвищий показник мають словосполучення "*climate crisis*" та "*economic crisis*". Також цікавим є високий показник появи словосполучення "*mental crisis*" та "*global crisis*". Вживаючи перше словосполучення Президент хоче розповісти про ментальну кризу, як наслідок пандемії *COVID*, а також звертає увагу суспільства на ментальне здоров'я.

Наступним методом визначення асоціативних зв'язків за допомогою корпусів є дослідження N-грам з даним словом. Нагадаємо, що N-грама – це послідовність з n елементів. Тобто з семантичної точки зору, це може бути послідовність звуків, складів, слів або букв. В нашому випадку ми будемо запит на пошук N-грам розміром від 2 до 3 слів, які містять слово *crisis*.

Результат виконання запиту можна побачити на рис. 2, на якому зображено безпосередню N-граму на частотність її вживання в корпусі. У цьому випадку ми отримали 69 N-грам з загальною частотністю вживання, що становить 654. Їх подано в таблиці, в порядку зменшення частотності в корпусі, тобто на першому місці в нас *climate crisis* з частотністю 90, а на останньому – *an economic crisis* з частотністю 3.

Тим не менш, найбільш цікавим для нас є поняття асоціативних (колокаційних) мереж, під яким розуміють розширені колокаційні графи, що показують більші асоціативні патерни; це мережі пов'язаних колокацій, які починається з певного вузла, навколо якого визначаються колокати (Ділай, 2022).

За допомогою корпусів ми можемо побудувати асоціативні (колокаційні) графи та мережі. Колокаційний граф є візуальною презентацією колокаційних відношень між вузлом та

2–3-grams, word (terms: 69, total frequency: 654)

N-gram	Frequency ?	N-gram	Frequency ?	N-gram	Frequency ?
1 climate crisis	90 ...	24 climate crisis and	9 ...	47 crisis caused	4 ...
2 the climate crisis	63 ...	25 into the crisis	8 ...	48 the opioid crisis	3 ...
3 economic crisis	27 ...	26 economic crisis to	8 ...	49 the food crisis	3 ...
4 every crisis	24 ...	27 energy crisis	7 ...	50 turned every crisis	3 ...
5 this crisis	24 ...	28 from economic crisis	7 ...	51 economic crisis caused	3 ...
6 crisis we	23 ...	29 the economic crisis	7 ...	52 crisis we ever	3 ...
7 the crisis	21 ...	30 COVID crisis	6 ...	53 economic crisis since	3 ...
8 of every crisis	16 ...	31 crisis to economic	6 ...	54 crisis in Ukraine	3 ...
9 crisis and	15 ...	32 of crisis	6 ...	55 crisis head on	3 ...
10 a crisis	14 ...	33 economic crisis that	5 ...	56 global energy crisis	3 ...
11 crisis stronger	14 ...	34 climate crisis ever	5 ...	57 crisis since	3 ...
12 every crisis we	13 ...	35 global food crisis	5 ...	58 crisis in this	3 ...
13 crisis stronger than	12 ...	36 the COVID crisis	5 ...	59 crisis head	3 ...
14 crisis in	11 ...	37 crisis ever	5 ...	60 immediate crisis	3 ...
15 health crisis	11 ...	38 The climate crisis	4 ...	61 humanitarian crisis	3 ...
16 of this crisis	11 ...	39 climate crisis is	4 ...	62 health crisis in	3 ...
17 crisis to	11 ...	40 climate crisis with	4 ...	63 of crisis	3 ...

Рис. 2. Знайдені в корпусі N-грами з словом crisis

його колокатами. Для його створення ми використовували інструмент #LancsBox: Lancaster University corpus toolbox. Для побудови графа потрібно використати вкладку GraphColl. Основою для його створення став статистичний показник MI-score (рис. 3).

Використовуючи колокаційний граф, ми проаналізували три виміри колокаційних відношень, а саме силу асоціацій, частотність колокатів та позицію колоката у тексті. Як правило, на зображеннях сила асоціацій виражається як довжина зв'язків між вузлом та колокатом, тобто що ближче колокат до вузла, то сильніші зв'язки. Відповідно, найсильніші асоціації слово crisis має у цьому корпусі зі словами climate, mental, caused, faced, stronger, health.

Частотність відображена відтінком кольору колоката, чим інтенсивніший колір, тим частотніший колокат. В нашому випадку це слова the, every, climate, this, economic. Аналізуючи цей граф, бачимо, що позиція колоката в тексті відображена як позиція колоката на графі, зокрема зліва від crisis: climate, immediate, mental, economic, COVID, food, global, а справа – caused, faced, stronger, ever, which.

Висновки та перспективи подальших досліджень. Дослідження асоціативної мережі «криза» в промовах Джо Байдена дає можливість зробити висновки про те, що найбільш частотними є словосполучення climate crisis та economic crisis. Аналіз вживання слова crisis в контекстах показало, що найчастіше з досліджуваним слово

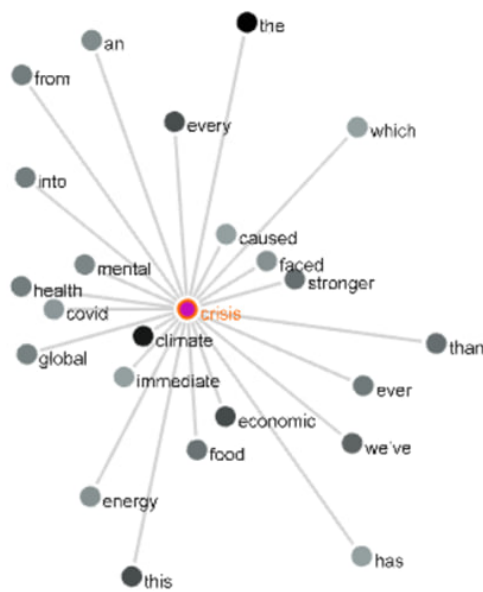


Рис. 3. Колокаційний граф слова crisis

вживаються такі: *climate, economic, this, every, health, food, energy, COVID, healthcare, opioid, Missile, immediate, humanitarian, ongoing, future, evolving, financial, major, security*. Цей перелік висвітлює основні проблеми, які є актуальними для сучасного американського суспільства. Крім згадуваних раніше економічної та екологічної кризи, це і криза у сфері охорони здоров'я, спричинена пандемією COVID-19, і гуманітарна криза, в контексті повномасштабної російсько-української війни та напливом біженців з України, і енергетична криза, однією з причин якої став енергетичний терор росії. Вивчення статистичних показників *MI-score* та *log Dice* лише підтвердили результати, отримані в попередніх дослідженнях. Для візуального відображення отриманої інформації було створено колокаційний

граф, за допомогою якого показано, що найсильніші асоціації слово *crisis* має у цьому корпусі зі словами *climate, mental, caused, faced, stronger, health*; найчастотнішими колокатами є: *the, every, climate, this, economic*; а також отримали інформацію про позицію колоката в тексті.

Отже, використання корпусу текстів та різних лінгвістичних та статистичних інструментів дало змогу провести ґрунтовний аналіз асоціативної мережі «криза». Оскільки основою нашого дослідження стали промови за лютий 2022 – лютий 2023 років, перспективи подальшого дослідження полягають у збільшенні матеріалу дослідження, проведення зіставного дослідження в англійській та українській мовах та порівняння результатів для виявлення особливостей вживання колокацій зі словом «криза».

ЛІТЕРАТУРА

1. Бобкова Т. В. Корпусний словник колокацій: методика укладання. *Наукові записки Національного університету «Острозька академія»*. Серія: Філологічна. 2015. № 52. С. 39–41.
2. Бобкова Т. В. Теоретико-методологічні підходи до вивчення колокацій. *Вісник Київського національного лінгвістичного університету*. Серія: Філологія. 2014. № 17 (2). С. 14–22.
3. Жуковська В. В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир : Вид-во ЖДУ ім. І. Франка, 2013. 142 с.
4. Ділай І. П., Ділай М. П. Роль частотності у лінгвістичних дослідженнях. *Закарпатські філологічні студії*. 2022. № 21. С. 146–151.
5. Левченко О., Лотоцька Н. Розмежування понять фразеологізм та колокація в сучасному мовознавстві (на матеріалі корпусу текстів Романа Іваничука). *Науковий журнал*. 2019. № 10. С. 193–199.
6. Махачашвілі Р. К., Білик К. М. Дослідження функціонально-лінгвістичних аспектів заголовків текстів новин у сучасному французькому медіадискурсі. *Науковий вісник Міжнародного гуманітарного університету*. 2021. № 48. С. 139–143.
7. Ситар Г. Статистичний аналіз прислів'їв і приказок: показник асоціації mutual information (на матеріалі українського національного лінгвістичного корпусу). *Лінгвістичні студії*. 2018. № 35. С. 170–176.
8. #LancsBox: Lancaster University corpus toolbox. URL: <http://corpora.lancs.ac.uk/lancsbox/> (дата звернення: 20.04.2023).
9. Sinclair, J. *Corpus. Concordance and Collocation*. Oxford : Oxford University Press. 1991.
10. Sketch Engine. URL: <http://www.sketchengine.co.uk> (дата звернення: 20.04.2023).

REFERENCES

1. Bobkova, T. V. (2015). Korpusnyi slovnyk kolokatsii: metodyka ukladannia. [Corpus dictionary of collocations: compilation method]. *Naukovi zapysky Natsionalnoho universytetu "Ostrozka akademiia"*. [in Ukrainian]
2. Bobkova, T. V. (2014). Teoretyko-metodolohichni pidkhody do vyvchennia kolokatsii. [Theoretical and methodological approaches to the study of collocations]. *Visnyk Kyivskoho natsionalnoho linhvistychnoho universytetu*. [in Ukrainian]
3. Zhukovska, V. V. (2013). Vstup do korpusnoi linhvistyky: navchalnyi posibnyk. [Introduction to corpus linguistics: a textbook]. Zhytomyr : Vyd-vo ZhDU im. I. Franka. [in Ukrainian]
4. Dilai, I. P. & Dilai, M. P. (2022). Rol chastotnosti u linhvistychnykh doslidzhenniah. [The role of frequency in linguistic research]. *Zakarpatski philolohichni studii*. [in Ukrainian]
5. Levchenko, O. & Lototska, N. (2019). Rozmezhuvannia poniat phrazeolohizm ta kolokatsiia v suchasnomu movoznavstvi (na materialiv korpusu tekstiv Romana Ivanychuka). [Distinguishing the concepts of phraseologism and collocation in modern linguistics (on the material of Roman Ivanychuk's corpus of texts)]. *Naukovyi zhurnal*. [in Ukrainian]
6. Makachashvili, R. K. & Bilyk, K. M. (2021). Doslidzhennia funksiionalno-linhvistychnykh aspektiv zaholovkiv tekstiv novyn u suchasnomu frantsuzkomu mediadyskursi. [Study of functional-linguistic aspects of headlines of news texts in modern French media discourse]. *Naukovyi visnyk Mizhnarodnoho humanitarnoho universytetu*. [in Ukrainian]

7. Sytar, H. (2018) Statystychnyi analiz prysliviv i prykazok: pokaznyk asotsiatsii mutual information (na materialii ukrainskoho natsionalnoho linhvistychnoho korpusu). [Statistical analysis of proverbs and sayings: the association index mutual information (based on the material of the Ukrainian national linguistic corpus)]. *Linhvistychni studii*. [in Ukrainian]
8. #LancsBox: Lancaster University corpus toolbox. Retrieved from: <http://corpora.lancs.ac.uk/lancsbox/> [in English]
9. Sinclair, J. (1991). *Corpus. Concordance and Collocation*. Oxford : Oxford University Press. [in English]
10. Sketch Engine. Retrieved from: <http://www.sketchengine.co.uk> [in English]

ASSOCIATION NETWORK “CRISIS” IN JOE BIDEN'S SPEECHES

Druzhbiak Svitlana Volodymyrivna

*Associate Professor at the Applied Linguistics Department
Lviv Polytechnic National University
30, Stepana Bandery str., Lviv, Ukraine*

Rii Andriana Vasylivna

*a 4th year Bachelor's student at the Department of Applied Linguistics
Lviv Polytechnic National University
30, Stepana Bandery str., Lviv, Ukraine*

The article is devoted to the study of the theoretical aspects of the concept of associative network and the practical analysis of associative networks of the concept "crisis" based on the corpus of Joe Biden's speeches. The necessity of analysing high-frequency collocations is related to the demand for information about stable word combinations for applied linguistic tasks. The paper analyses scientific works on the problem of compatibility, collocations, associations and associative networks. It is proved that the use of a corpus of texts and various linguistic and statistical tools makes it possible to conduct a thorough analysis of the association network of a given word. It is revealed that the concept of "association" is a psychosemantic connection in the human mind between two or more mental phenomena, as well as their designations in the language. It is established that the term "collocation" is synonymous with this concept – word combinations that occur in the text more often side by side than by random probability separately. The methods of creating associative networks were practically tested using a corpus of texts based on Joe Biden's speeches from February 2022 to February 2023. A collocation graph is used to visually represent the collocation relations between a node and its collocates. The results obtained using each approach are compared and analysed.

Keywords: *associative network, collocation, concordance, corpus of texts, association indicators, collocation graph.*