

## АЛГОРИТМ АВТОМАТИЧНОЇ ІДЕНТИФІКАЦІЇ МЕТАФОРИ

(на матеріалі творів О. Забужко)

doi: 10.12958/2227-2631-2021-1-43-88-96

**Олена Левченко,**

*доктор філологічних наук, професор  
кафедри прикладної лінгвістики  
Національного університету «Львівська політехніка»  
ORCID ID <https://orcid.org/0000-0002-7395-3772>*

**Віталій Карасьов,**

*магістрант кафедри прикладної лінгвістики  
Національного університету «Львівська політехніка»  
ORCID ID <https://orcid.org/0000-0002-8039-2811>*

У цій праці представлено уточнення методики автоматичної ідентифікації метафори на матеріалі прозових творів О. Забужко. Аналіз моделей метафори, одним із компонентів якої є лексема *думка*, дав змогу виокремити та класифікувати конструкції, які потенційно можуть бути метафорою.

Наголошено, що дослідження мовної картини і окремого письменника, і етносу загалом, зорієнтовано на дослідження метафори як механізму пізнання та відтворення навколишнього світу, як ключ до розуміння мисленневих процесів людини.

Акцентовано на важливості вивчення метафори завдяки залученню корпусних інструментів та застосуванню математичних методів дослідження, адже такий підхід дає змогу отримувати результати, точність та об'єктивність яких суттєво покращиться. Використання корпусного підходу відкриває перспективу створити перелік моделей метафори тієї чи тієї мови, що суттєво полегшить опрацювання природномовної інформації, зокрема автоматичну ідентифікацію метафор.

Установлено, що ідентифікація метафор ментальної сфери передбачає два обов'язкові етапи: автоматизований пошук у корпусі та інтерпретація результатів, що зумовлено специфічними особливостями метафоризації мисленневих процесів і тим, що в сучасних українських корпусах текстів не розв'язано проблему розрізнення омонімії та багатозначності.

Проведено пошук у ГРАК та укладено перелік колокацій двох структурних моделей: іменник *думка* + складник та складник + іменник *думка*. Загальна кількість результатів для двох цих моделей – 234.

Виявлено, що однією з найбільш продуктивних моделей метафоризації концепту ДУМКА є моделі *думка* + дієслово / дієслово + *думка* та *думка* + прикметник / прикметник + *думка*.

У досліджених текстах найчастотнішими є антропоморфні метафори, найчастіше пов'язані із стрімким рухом: *думка набігає, думка доганяє, думка догнала, думка петляла, думка звернула*, що можна одночасно трактувати як особливу рису ідіостилю О. Забужко і як авторське переосмислення ідеї про думку як процес, рух, дію.

*Ключові слова:* автоматична ідентифікація метафори, когнітивна лінгвістика, концептуальна метафора, корпусна лінгвістика, мова контекстних запитів.

An Algorithm of Automated Identification of the Metaphors (based on the novels written by Oksana Zabuzhko) The thesis is a complex study in which the creation of a model of automatic identification of the metaphor is represented. Analysis of models of metaphor, one of the elements of which is the token «думка» made it possible to identify and classify constructions that could potentially be a metaphor.

It is emphasized that the study of the linguistic picture of both the individual writer and the ethnic group as a whole, focused on the study of metaphor as a mechanism of cognition and reproduction of the surrounding world, as a key to understanding human thought processes.

A search was conducted in GRAC, and a list of collocations of two structural models was compiled: noun думка + component and component + noun думка. The total number of results for these two models is 234. It is revealed that one of the most productive models of metaphorization of the concept ДУМКА is the models думка + verb / verb + думка and думка + adjective / adjective + думка. Personalized properties of «думка» are often associated with the rapid movement: *думка набігає, думка доганяє, думка догнала, думка петляла, думка звернула*, which can be interpreted as a special feature of idiostyle O. Zabuzhko and as the author's rethinking of the idea of thought as a process, movement, action.

The process of identifying a metaphor will have features and specific stages depending on the model of the metaphor under consideration. An important step is to create the appropriate tags and variants to get the most accurate results possible. Attempts to concretize abstract concepts using metaphors is a natural process but complex and involves many results. Since this type of research in Ukrainian linguistics is at an early stage, the «manual» analysis and screening of results by the researcher himself occupy a prominent place and require a lot of time.

*Key words:* automatic metaphor identification, cognitive linguistics, conceptual metaphor, contextual query language, corpus linguistics.

**Вступ.** Метафора – це складний мовленнєвий та мисленнєвий субстрат, адекватне та вичерпне розшифрування якого потребує застосування найрізноманітніших підходів. Корпусний підхід має особливо важливе значення з огляду на її природу, наближає нас до розуміння метафори як явища ментальної сфери загалом, як механізму

концептуалізації та сприйняття світу, а також способів її функціонування та творення.

У процесі дослідження мовної картини світу того чи того автора особливе місце посідає аналіз метафор. І в мовленні, і в художньому тексті метафори виконують ключову функцію, оформлюючи й аргументуючи структуру авторського світосприйняття. Саме метафора є тим механізмом, завдяки якому письменник може створювати ті чи ті образи й нові значення, водночас «регулюючи» ставлення реципієнта до них.

**Аналіз останніх досліджень і публікацій.** Автоматична ідентифікація метафори є предметом дослідження таких учених, як Н. Аллан, Л. Генді (Gandy, Allan, Atallah, Frieder, Howard, Kanareykin, ... Argamon, 2013), К. Шутова (Shutova, 2013), М. Ласт, Я. Нуман (Neuman, Assaf, Last, 2013).

Зокрема, автори праці «Robust Extraction of Metaphors from Novel Data» (Strzalkowski, Broadwell, Taylor, 2013) формулюють етапи ідентифікації метафори. Запропонований загальний алгоритм складається з п'яти основних кроків: від отримання вхідного тексту до класифікації вхідних даних – метафоричних або дослівних. У вітчизняній науці про мову цей напрям дослідження порівняно новий: створюється база, яка буде точкою відліку для більш глибоких досліджень, а також єдиний універсальний алгоритм автоматичної ідентифікації метафори. Серед дослідників, які активно працюють у цій галузі, належить виокремити О. Левченко, М. Ділай, Н. Романишин та ін. (Петрук, Левченко, 2019; Levchenko, Tyshchenko, Dilai, 2021; Levchenko, Romanyshyn, Dosyn, 2019; Levchenko, Romanyshyn, 2019).

**Метою праці** є корпусний аналіз метафор ментальної концептосфери, представлених у романах О. Забужко, а також визначення структурних моделей цієї групи метафор у текстах письменниці. Реалізація поставленої мети передбачає розв'язання таких **завдань**: дослідити метафору та механізми її утворення з погляду когнітивної лінгвістики; проаналізувати домінантні моделі метафоризації, наявні в текстах О. Забужко. Вибір об'єкта дослідження зумовлено тим, що Оксані Забужко, одній із найбільш знакових вітчизняних авторок сьогодення, притаманний унікальний авторський стиль.

Матеріалом дослідження слугували прозові твори письменниці, які було отримано за допомогою Генерального регіонально анотованого корпусу української мови (ГРАК) (Шведова, 2019), з яких було сформовано підкорпус текстів для подальшого аналізу.

Інтерпретація метафори від звичайного тропу, художнього засобу до категорії когнітивного феномену стала можлива завдяки її переосмисленню в річищі когнітивної лінгвістики. Постійний розвиток науки про мову, неперервне науково-дослідницьке заглиблення в людську діяльність зрештою зумовили те, що класичне трактування

метафори як тропу, засобу «прикрашання» мови стало нефункційним і неефективним. Відтак, на пограниччі наукових дисциплін, які раніше лише віддалено або й зовсім ніяк не стосувались метафори, з'являються нові напрями дослідження. У сучасній науці про мову метафора є об'єктом вивчення кількох напрямів. Особливий статус вона отримала в межах когнітивної лінгвістики, яка розглядає її як ключ до розуміння мисленневих процесів людини. З огляду на складну природу метафори, такий підхід є виправданим. Проте лише комплексне вивчення та застосування різноманітних підходів дасть змогу визначити або максимальньо наблизитися до визначення мисленневих процесів і моделей метафоротворення.

**Виклад основного матеріалу.** Метафоризаційний потенціал тієї чи тієї мови залежить від багатьох чинників. Так, українська мова має низку особливостей, завдяки яким метафоричні процеси практично не мають обмеження, що підтверджують результати нашого дослідження. Будучи незамінною частиною української народнопісенної спадщини, метафора стала невід'ємним елементом побуту українців. Численні фразеологізми, паремії, порівняння, лайка, прокльони, пісні різних жанрів часто у своїй основі мають метафору. Так, в українській усній словесності можна виокремити поняття сталої метафори, тобто створені okazіонально метафоричні звороти виявилися настільки влучними та промовистими, що стали вельми впізнаваними і за межами, зокрема, народних пісень. Скажімо, *уста завжди солодкі як мед, або й солодші, брати – соколята, сестри – голубоньки* тощо. Також наявність діалектів, можливості інверсії і загальний демократизм мовленневих процесів надають носіям української мови практично невичерпні можливості для творення метафор, а особливо письменникам, які професійно володіють словом.

Якщо раніше вивчення метафори відбувалося в межах традиційної семантики на основі інтерпретації значення слів та здебільшого без залучення фонові інформації та кількісних даних, то сьогодні завдяки корпусним інструментам і застосуванню математичних методів дослідження можна отримувати результати, точність та об'єктивність яких суттєво покращились. Отже, дослідження метафори за допомогою корпусного підходу відкриває перспективу створити перелік моделей метафори тієї чи тієї мови, що суттєво полегшить опрацювання природномовної інформації, зокрема автоматизовану ідентифікацію метафор.

Ідентифікація метафор ментальної сфери передбачає два обов'язкові етапи: автоматизований пошук у корпусі та інтерпретація результатів, що зумовлено специфічними особливостями метафоризації мисленневих процесів і тим, що в сучасних українських корпусах текстів не розв'язано проблему розрізнення омонімії та багатозначності.

З огляду на те, що розвідку присвячено виявленню моделей метафоризації концепту ДУМКА у творах О. Забужко, було укладено перелік колокацій двох структурних моделей: іменник *думка* + складник

та складник + іменник *думка*. Загальна кількість результатів для двох цих моделей – 234. Кількість словосполук із лексемою *думка* в прямому значенні суттєво нижча від непрямих, проте, зважаючи на специфіку деяких текстів письменниці, належить аналізувати кожний конкретний випадок окремо.

У процесі дослідження було сформовано 16 запитів, за допомогою яких можна виявити потенційні метафори в корпусі текстів (див. табл. 1).

Таблиця 1

Запит	Кількість контекстів	Кількість метафор	%
[lemma="думка"] [tag="verb.*"]	29	19	65,5%
[lemma="думка"] [lemma="бути"]	7	6	85,7%
[lemma="думка"] [lemma="бути"] [tag="verb.*"]	1	1	100%
[lemma="думка"] [lemma="бути"] [tag="adj.*"]	1	1	100%
[tag="verb.*"] [lemma="думка"]	18	15	83,3%
[tag="adj.*"] [lemma="думка"]	118	29	24,5%
[lemma="думка"] [tag="adj.*"]	9	0	0%
[tag="verb.*"] [tag="adj.*"&tag!="*pron.*"] [lemma="думка"]	2	2	100%
[tag="*adj.*"&tag!="*pron.*"] [lemma="думка"]	57	43	75,4%
[tag="*noun.*"&tag!="*pron.*"&tag!="*adj.*"] [lemma="думка"]	14	8	57,4%
[lemma="думка"] [tag="adj.*"] [tag="adj.*"]	1	1	100%
[tag="adj.*"] [tag="adj.*"] [lemma="думка"]	22	6	27,2%
[lemma="думка"] [lemma="бути"] [tag="adj.*"]	1	1	100%
[tag="verb.*"] [tag="adj.*"] [lemma="думка"]	13	9	69,2%
[tag="verb.*"] [tag="*pron.*"] [lemma="думка"]	12	10	83,3%
[tag="verb.*"] {2} [tag="*pron.*"] [lemma="думка"]	1	0	0%

Доволі часто натрапляємо на конструкції *подумана думка, думка думається, додумати думку, думку додумати не дадуть* тощо, наявність яких у текстах свідчить про брак у мові лексем, які б передавали властивості думки як ментального явища без застосування переносних

значень. Також, досліджуючи саме метафору, варто одразу відмежувати інші художні засоби та явища. Одним з таких засобів є порівняння. З огляду на зазначене, модель сполучник «мов»/«немов»/«наче»/«неначе» + дієслово + іменник «думка» або «мов»/«немов»/«наче»/«неначе» + іменник «думка» + дієслово в цьому дослідженні не взято до уваги.

Однією з найбільш продуктивних моделей метафоризації концепту ДУМКА є модель *думка* + дієслово/ дієслово + *думка*, де компонент *думка* може поставати і об'єктом, і суб'єктом дії. Саме на основі цієї моделі можна виявити, як концептуалізується ДУМКА в досліджуваних текстах. За допомогою тега / [lemma="думка"] [tag="verb.\*"] у корпусі текстів письменниці виявлено 29 словосполук, серед яких є низка фразеологізмів, як-от: *спадати на думку, мати на думці*, наприклад: «Отак і не дадуть чоловікові зосередитися, ні одної пунької *думки* *додумати не дадуть* ... І чого ця коза так кричить?» (Музей покинутих секретів, 2009); «Як оприлюднити, де? — мені трохи прикро, що його *думка* *звернула в інше русло*» (Музей покинутих секретів, 2009); «В неінтерпретованих (читай – нечитаних) творах жива *думка* *завмирає* – пригадаймо класичну логічну загадку» (Філософія і культура: притомність нації, 1994); «Або контузія, контузія якось краще... *Думка* *набігає* за думкою, самовільними розрядами, нестримними, як перейми» (Музей покинутих секретів, 2009).

Зауважимо, що в межах цієї моделі можна виокремити лише одну підгрупу метафор – метафори, де *думка* є суб'єктом дії: *думка працювала, думка доганяє, думка згубила* тощо. Отже, у межах моделі *думка* + дієслово у творах О. Забужко *думка* не є об'єктом дії.

Окрім того, у межах цієї групи виокремлено підгрупу метафор, які побудовані за моделлю *думка* + дієслово *бути* (тег [lemma="думка"] [lemma="бути"]). Для уточнення результатів використано теги [lemma="думка"] [lemma="бути"] [tag="verb.\*"] і [lemma="думка"] [lemma="бути"] [tag="adj.\*"]. Виявлено такі метафоризації: *думка була явно чужа, думка була бликнула*. Зважаючи на специфіку цієї моделі метафори, напрями метафоризації тут обмежені. Так, *думка* зазвичай виконує функції, ролі та дії, притаманні людям: *думка звернула, думка завмерла, думка набігає* тощо.

Іншою структурною моделлю метафори, де одним із складників є дієслово, є модель дієслово + *думка*. За допомогою тега [tag="verb.\*"] [lemma="думка"] знайдено 18 контекстів, з яких 15 – метафори. У межах цієї групи виокремлюємо дві підгрупи метафор: 1) метафори, де лексема *думка* є суб'єктом дії (*глухо скидається думка, блимає думка*); 2) метафори, де лексема *думка* є об'єктом дії (*ткнешся думкою, розігнатися думкою*).

Зважаючи на те, що дослідження проведено на матеріалі художніх текстів, треба враховувати специфіку авторського стилю. Так, авторка в одному контексті використовує різні метафори: *обморочує, блимає*

думка. До прикладу: «*Обморочує, блимає думка, — він же обморочує мене, як професійний спокусник*» (Музей покинутих секретів, 2009).

За допомогою тега [tag="adj.\*"] [lemma="думка"/][lemma="думка"] [tag="adj.\*"] у корпусі текстів письменниці виявлено відповідно 118 та 9 словосполук. Аналіз метафор цієї моделі ускладнено тим, що результати пошуку за тегом [tag="adj.\*"] [lemma="думка"] містять також словосполучки займенник + іменник «думка» (*моя думка, його думка* тощо), оскільки займенник стоїть перед іменником. Отже, щоб отримати максимально точні результати, використано уточнений тег [tag="verb.\*"] [tag="adj.\*"&tag!="\*pron.\*"] [lemma="думка"], завдяки якому знайдено 2 контексти, які є метафорами: *приходить нова думка, стеля власних думок*. Проте завдяки «ручному» аналізу результатів пошуку за тегом [tag="adj.\*"] [lemma="думка"] виявлено лише 29 метафор (із 118 результатів). Похибка на нерозрізнення омонімії є доволі значною, про що також свідчить пошук, здійснений за тегом [tag="adj.\*"] [lemma="думка"], який був ускладнений чималою кількістю контекстів штибу *естетична думка, пострадянська думка, філософічна думка*, які не можна вважати метафорою.

Модель метафори, складником якої є дієслово, може уточнюватися за допомогою інших частин мови, зокрема прикметника – *мучила невідчепна й безглузда думка, підбирає впущену думку, важко ворухнулася сонна пресонна думка, видавалася нестерпучою думка*. Отже, модель **дієслово + прикметник + ім.** «думка» є продуктивною щодо творення метафор. Ця модель може також набувати таких варіантів: **дієслово + прикметник + "думка"**, **прикметник + дієслово + "думка"**, **"думка" + прикметник + дієслово**. Використовуючи тег [tag="verb.\*"] [tag="adj.\*"] [lemma="думка"], виявлено 13 контекстів, 9 з яких є метафорами. У цій групі напрями метафоризації дещо різноманітніші, ніж у попередній. Думці притаманні антропоморфні риси, вона може бути, скажімо, *сонною*. Однак тут вона також метафоризується речоморфно – *впущена думка*.

Авторка може надавати низку однорідних означень слову *думка*, які було виявлено в таких інваріантних моделях: **прикметник + думка /думка + прикметник**, як **прикметник + прикметник... + думка /думка + прикметник + прикметник...**, (теги [lemma="думка"] [tag="adj.\*"] [tag="adj.\*"] і [tag="adj.\*"] [tag="adj.\*"] [lemma="думка"]). Знайдено 1 і 22 контексти відповідно. До прикладу: «*А Ленця, відай, саме її й потребувала, – але ця друга думка, наздоганяючи першу, вже тільки ковзає по свідомості*» (Дівчатка, 1998); «*Що знай з дня на день випікала голову, не пропускаючи жодної прохолодної думки, панував виразний дух замкненої спільноти: цікаві позирки*» (Польові дослідження з українського сексу, 1994). Наведені вище приклади відповідають різним типам метафоризації, де ДУМКА – це речовина чи ДУМКА – це впорядкована система тощо.

Окрім моделей метафор, які мають високу частотність, виявлено й такі, що трапляються лише кілька разів, як-от: **іменник + думка/ думка + іменник** ([tag="noun.\*"] [lemma="думка"] / [lemma="думка"] [tag="noun.\* "]), для ефективнішого пошуку використано такий тег [tag=".\*noun.\*"&tag!=".\*pron.\*"&tag!=".\*adj.\*"] [lemma="думка"], за допомогою якого було знайдено 14 словосполук, 8 з них – метафори, наприклад: «*Ця функція блискуче описана Орвеллом – „звукити **обрії думки**, унеможливити ересь”*» (Мова і Влада, 1990); «*Не забуваймо й ще одне: на відміну від питомої мови російського народу, енкратична російська «новомова» призначається до комунікації (так і в академічному мовознавстві речено: **головна функція «русского языка как средства межнационального общения» – комунікативна**), тобто не до **продукування думки**, а до трансляції вже готових думок (згадаймо Славіну!)*» (Мова і Влада, 1990); «*Творення придатної для філософського розмислу мови – це наче синтез алмазів: кожне повновартісне абстрактне поняття має ряхтіти багатьма гранями, відбиваючи в собі цілу історію **сходження думки до предмета***» (Філософія і культура: притомність нації, 1994).

**Висновки.** Отже, з аналізу випливає, що найпродуктивнішим напрямом метафоризації є антропоморфний. Антропоморфність *думки* найчастіше пов'язана зі стрімким рухом: *думка набігає, думка доганяє, думка догнала, думка петляла, думка звернула*, що можна трактувати як особливу рису ідіостилю О. Забужко і як авторське переосмислення ідеї про думку як процес, рух, дію. У досліджених текстах авторки кілька раз повторено метафору *думка думку доганяє* та її варіантів, доповнених іншими складниками. Виокремлюємо таку метафору, як ДУМКА – ЦЕ СТИХІЯ: *думки текли, думки мчать, думка звернула в інше русло, стихійний плин думки, думка пекла* тощо. Структурна модель прикметник + іменник «думка» може відповідати різним напрямам метафоризації. Це підтверджують і використані письменницею прикметники: *безробітна, сонна-пресонна* – персоніфікація, *потайна, палюча* – стихія, *жива* – персоніфікація, *прохолодна* – речоморфізм тощо.

Процес ідентифікації метафори матиме особливості і специфічні етапи залежно від моделі аналізованої метафори. Важливим етапом роботи є створення відповідних тегів та їхніх варіантів для отримання якомога точніших результатів. Намагання вербалізувати абстрактні поняття за допомогою метафор – процес природний, проте складний і передбачає значну кількість результатів. Оскільки цей вид досліджень у вітчизняній лінгвістиці перебуває на початковій стадії, «ручний» аналіз і відсіювання результатів самим дослідником потребує великих затрат часу. Проте поступовий розвиток цієї галузі дасть змогу значно спростити й скоротити етап безпосередньої роботи з корпусом, створивши чіткий алгоритм пошуку метафори.



### Література

- Петрук, О. Ю. & Левченко, О. П. (2019). Ідентифікація метафори моделі білий + іменник за методикою кількісного аналізу словникової дефініції. *Young Scientist*. № 10. Р. 505 – 511.
- Шведова, М. (2019). Генеральний регіонально анотований корпус української мови (ГРАК) як інструмент дослідження лексико-граматичної варіативності. *Людина. Комп'ютер. Комунікація : зб. наук. пр.* / за ред. проф. О. П. Левченко. Львів : Вид-во Львівської політехніки, С. 145 – 148.
- Gandy, L., Allan, N., Atallah, M., Frieder, O., Howard N., Kanareykin, S. ... Argamon, S. (2013). Automatic Identification of Conceptual Metaphors With Limited Knowledge. *Proceedings of the AAAI Conference on Artificial Intelligence*. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/8648>.
- Levchenko, O. & Romanyshyn, N. (2019). Modern Approaches to Automated Identification of Metaphor. *Вісник Львівського університету. Серія філологічна*. № 70. Р. 288 – 298.
- Levchenko, O., Romanyshyn, N. & Dosyn, D. (2019) Method of automated identification of metaphoric meaning in adjective + noun word combinations (based on the Ukrainian language). *CEUR Workshop Proceedings, Workshop proceedings of the 8th International conference on Mathematics*. Information technologies. Education, MoMLeT&DS. P. 370 – 380.
- Levchenko, O., Tyshchenko, O. & Dilai, M. (2021). Automated Identification of Metaphors in Annotated Corpus (Based on Substance Terms). *5th International Conference on Computational Linguistics and Intelligent Systems* (April 22 – 23, 2021, Kharkiv, Ukraine).
- Neuman, Y., Assaf, D. & Last, M. (2013). Metaphor Identification in Large Texts Corpora. *PLoS ONE* 8(4): e62343. URL: <https://doi.org/10.1371/journal.pone.0062343>.
- Shutova, E. (2013). Metaphor Identification as Interpretation. *Second Joint Conference on Lexical and Computational Semantics (\*SEM)*. P. 276 – 285.
- Strzalkowski, T., Broadwell, S. & Taylor, A. (2013). Extraction of Metaphors from Novel Data. *Conference: In Proceedings of the First Workshop on Metaphor in NLP at the North American Association of Computational Linguistics Conference (NAACL-2013)*. Atlanta USA. URL: [https://www.researchgate.net/publication/267928109\\_Robust\\_Extraction\\_of\\_Metaphors\\_from\\_Novel\\_Data](https://www.researchgate.net/publication/267928109_Robust_Extraction_of_Metaphors_from_Novel_Data).

Стаття надійшла до редакції 15.10.2021 р.

Прийнято до друку 19.10.2021 р.

Рецензент – д. філол. наук, проф. Глуховцева К. Д.