

Автоматизація вибірки антиметаболи за допомогою мови корпусних запитів CQL

Abstract. *The article discusses the automation procedure for extracting stylistic devices, specifically, antimetaboles (a type of chiasmus). Antimetabole is characterized by a set of formal parameters, such as the repetition of lexemes in the sequence A – B – B – A. However, its context and semantic features are necessary for it to be fully functional. One approach for automating the extraction of antimetaboles from textual corpora is to use a specific corpus query language. In the absence of an annotated corpus, regular expressions are a powerful tool for the automatic extraction of antimetaboles. The strengths and weaknesses of both approaches are analyzed in the article.*

Keywords: *chiasmus, antimetabole, corpus query language, regular expression, automatic extraction, data mining, stylistic device.*

Автоматизація та швидке добування корисних даних стає сучасною повсякденністю. Необхідність отримати швидко, докладну відповідь на питання або ж виокремити з великого масиву інформації лише ті частини, що відповідають певним умовам, є запорукою успішної раціоналізації праці, інтелектуального поступу галузі, а також творчої реалізації фахівця і його професійного зростання. Одним з викликів для дослідника мови стає й автоматична вибірка стилістичних фігур. Окрім того, що стилістичні фігури мають певні формальні параметри, що відповідають певним умовам. Ці формальні маркери виконують свої функції лише завдяки тісному зв'язку з семантикою компонентів, з широким і вузьким контекстом, і саме тому їх важко безпомилково й точно описати для опрацювання штучним інтелектом. Тому, вочевидь, автоматизація може стати лише допоміжною ланкою у процесі вибору матеріалу. Але цією допомогою однозначно варто скористатися, якщо вона суттєво заощаджує зусилля дослідника.

Однією з таких дивовижних стилістичних фігур, що характеризується доволі однозначними формальними параметрами, і яка попри свою колосальну привабливість і колорит, отримує достатньо мало уваги серед дослідників, є хіазм. Це й не дивно: читач від нього дістає естетичну насолоду; вдала гра синтаксисом і словами вривається в пам'ять і часто навіть стає крилатою, в той час як дослідник подекуди змушений перечитати сотні сторінок, аби зібрати заледве кілька прикладів. Отже, власне барвистість і глибока естетичне навантаження цієї стилістичної фігури на тлі вражаючого провалля у її дослідженнях спонукали підійти впритул до цього аспекту з сучасної точки зору й за допомогою

використання сучасних технологій добування даних, аби розкрити потенціал комп'ютера у полегшенні вибірки емпіричного матеріалу для дослідження хіазму, зокрема, одного з його найвиразніших різновидів – антиметаболи.

Таким чином, мета цієї розвідки – оцінити потенціал використання двох способів добування даних: 1) шляхом використання мови корпусних запитів (CQL, Corpus Query Language) та подібних; 2) за допомогою використання регулярних виразів для автоматизації вибірки стилістичних фігур.

Хіазм – це «стилістична структурно-синтаксична фігура мовлення, що ґрунтується на перехрещенні, семантичній трансформації синтаксичних структур двох суміжних відрізків тексту: «Не для того ми живемо, щоб їсти, а для того щоб жити» [1, с. 654]. Через поширеність прикладів, подібних до наведеного може скластись хибне враження, ніби у хіазмі має спостерігатися кільцевий повтор лексем за схемою А – В – В – А. Тим не менш, як зазначають дослідники М. Дубрєметц і Дж. Нівре, лексемний повтор у хіазмі спостерігається не завжди, якщо хіазм заснований власне на кільцевому повторі лексем, то він називається антиметаболою [2]. О. О. Тележкіна заострює увагу на синтаксичному аспекті хіазму і наводить приклади з різними лексемами, хоча близькими за значенням [3]. Л. І. Лонська зауважує, як і більшість дослідників, що в основі хіазму лежить обмін синтаксичних позицій між окремими компонентами речення та компонентами кількох предикативних частин [4, с. 102]. Дослідниця аналізує більше десятка визначень і думок вчених щодо цієї фігури. З глибокого огляду дослідниці теж видно, що провідні характеристики хіазму все ж пов'язані з синтаксисом, а не з лексикою. І, тим не менш, в літературі зустрічаються згадки щодо лексичного хіазму (див., приміром, статтю Макса Ненні і Ольги Фішер [5, с. 69]). У той самий час, цей «лексичний хіазм» має свою технічну назву, яку, з огляду на раритетність терміна, рідко згадують. Таким чином, очевидно, що під хіазмом часто мають на увазі антиметаболу.

Отже, «антиметабола – це повторення слів у зворотному порядку, антиметатеза – двочленна стилістична фігура, в якій симетрично повторюються ті ж слова, але зі зміною їхніх синтаксичних функцій і змісту» [6]. Однак Т.В.Ф.Броган і А. В. Голсолл вважають, усе ж, що вживання терміна "антиметабола" не виправдане, і ліпше називати хіазмом інверсійну перестановку чи то літер, синтаксису, чи то значень [7, с.58].

Прикметно, що з 62 прикладів, що їх наводить Л. І. Лонська у своїй праці, абсолютно в усіх є бодай один лексичний кільцевий повтор однієї з ключових лексем хіазму, а в 58 з них, тобто, в абсолютній більшості – по два кільцевих повтори. Хоча це й не свідчить, що хіазм на основі лексичного повтору трапляється частіше, ніж без повторів, все ж, зазначене спостереження промовисто показує, наскільки помітнішим, колоритнішим стає хіазм з лексичним повтором, завдяки якому набуває афористичного характеру.

Саме тому ми вирішили зосередитися лише на хіазмах зі стилістичним повтором (антиметаболах) у цій розвідці, тим паче, що повтор як додатковий формальний маркер, попервах має полегшити завдання: здійснити автоматичну вибірку хіазмів з таких джерел: 1) з розмічених текстових корпусів; 2) з будь-якого електронного тексту.

1. Автоматична вибірка хіазму з розмічених текстових корпусів:

З метою вибірки хіазму скористаємося корпусом ГРАК (Генеральний регіонально анотований корпус українською мови[8]). Для автоматичної вибірки хіазмів з кільцевим лексичним повтором необхідно описати мовою корпусних запитів його формальні умови, зокрема, такі: 1) крайні лексеми конструкції однакові; 2) серединні лексеми однакові; 3) крайні й серединні лексеми різні. Якщо знехтувати останньою, на перший погляд, не надто важливою умовою, під час вибірки буде забагато зайвих прикладів, у яких будуть лексеми, повторені чотири рази підряд. У синтаксисі CQL ці умови можна формалізувати так:

```
[1]2:[3]:[4]:[& 1.lemma=4.lemma & 2.lemma=3.lemma&
1.lemma!=2.lemma
```

Зазначені умови хіазму пишуться після знаку ампераснда "&".

Врахуємо, що в хіазмі між першим, другим, третім і четвертим токеном може бути перерва в кілька слів. Перерву від 0 до 3 будь-яких токенів можна оформити так: `[] {0,3}` . Отже, запит буде такий:

```
[1][{0,3}2:[[] {0,3}3:[[] {0,3}4:[& 1.lemma=4.lemma & 2.lemma=3.lemma&
1.lemma!=2.lemma
```

Аби позбутися прикладів, що рясніють повторами мало змістовних пунктуаційних знаків (які теж у корпусі вважаються токенами), та допоміжних частин мови, пропишем явно, які частини мови можуть входити до складу хіазму, за допомогою атрибута "tag" (іменник, дієслово, прикметник, прислівник, займенник) :

```
tag="noun.*|verb.*|adj.*|adv.*|. *pron.*"
```

Тоді тоді запит матиме такий вигляд:

```
1:[tag="noun.*|verb.*|adj.*|adv.*|. *pron.*"][] {0,3}2:[tag="noun.*|verb.*|adj.*|ad
v.*|. *pron.*"][] {0,3}3:[[] {0,3}4:[& 1.lemma=4.lemma & 2.lemma=3.lemma&
1.lemma!=2.lemma within <s/>
```

За цим запитом у корпусі ГРАК, з підкорпусу художніх текстів, знаходиться 1436 прикладів. Не всі з них є хіазмами, трапляються й суто випадкові повтори лексем. Наприклад:

І от, нарешті, вона сама каже мені: " Як дивно: я люблю світ, я люблю людей так, як ніколи не любила. (Володимир Винниченко)

Тим не менш, на прикладі корпусу ГРАК помічаємо, що з 1436 прикладів приблизно третина справді містять хіазм. Наведемо деякі приклади:

Горішній аркуш помальовано стрілами й решітками, решітками і стрілами (Марина Меднікова).

Тетяна Кулик: чужа серед своїх, своя серед чужих – Чудна річ життя
(Марина Меднікова).

Пойняв – не дурак, дурак би не пойняв (Марина Меднікова).

Я люблю жінок, жінки люблять мене – не бачу тут нічого страшного
(Тетяна Малаярчук).

Утім погода може вибрати собі свято, але свято не може вибрати погоду
("Україна Молода", 2011).

І тільки тоді, коли Перший, різко змінивши інтонацію, ткнув йому пальцем у груди (ми підозрюємо вас, ми вас підозрюємо, вас підозрюємо ми!), Артур Пена остаточно допер, що саме цей тут Перший, отже, немає сенсу крутити замотиличеною головою на всі боки, відбиваючись від їхніх зізнавайтесь [...] (Юрій Андрухович).

– Та, скажу вам, у мене взагалі не виходить гуляти містом: робота – дім, дім – робота (Анатолій Сахно).

Я дивився на них, вони дивилися на мене (Владислав Івченко).

Оскільки мова CQL [9] використовується у численних корпусах, саме вона була обрана для ілюстрації вибірки. При її використанні слід пам'ятати, що значення атрибутів (тобто текст, що пишеться в лапках), зокрема, частиномовна розмітка, залежить від конкретного корпусу, а не від мови запитів. Тому у відкритих корпусах Sketch Engine [10] запит для добування хіазмів з кільцевим лексичним повтором матиме дещо інший вигляд.

Зокрема, для корпусу 'FOR EUROPARL, SAMPLE, SPANISH' адаптований запит буде такий:

```
1:[tag="N.*|V.*|ADJ|ADV|DM|INT|PP.*"][]{}0,3}2:[tag="N.*|V.*|ADJ|ADV|DM|INT|PP.*"][]{}0,3}3:[4:[]& 1.lemma=4.lemma & 2.lemma=3.lemma& 1.lemma!=2.lemma within <s/>
```

Після належних адаптацій запиту для корпусу FOR EUROPARL, SAMPLE, ENGLISH він матиме такий вигляд:

```
1:[tag="N.*|V.*|J.*|RB.?[PP.?""][]{}0,3}2:[tag="N.*|V.*|J.*|RB.?[PP.?""][]{}0,3}3:[4:[]& 1.lemma=4.lemma & 2.lemma=3.lemma& 1.lemma!=2.lemma within <s/>
```

Адаптації в цьому разі підлягали лише номінації частиномовних тегів, які різняться від корпусу до корпусу.

2. Автоматична вибірка хіазму зі звичайного тексту за допомогою регулярних виразів:

Якщо хіазми або інші структурні дані необхідно добути зі звичайного тексту, який не забезпечений спеціальною мовною (лексичною, граматичною) розміткою, на допомогу прийдуть регулярні вирази, завдяки яким можна автоматизувати вибірку суто за формальними параметрами мовних знаків безвідносно до їхніх лексико-граматичних характеристик. Регулярний вираз – це рядок або модель, що описує умови, яким відповідає певна множина різних рядків, послідовностей. За допомогою регулярних виразів зручно автоматично обирати ті фрагменти тексту, які відповідають умовам, що нас цікавлять, і відсіювати ті послідовності, які таким умовам не відповідають.

Очевидним недоліком регулярних виразів для обраного об'єкта дослідження є неможливість використання синтаксичної розмітки (хоча за певних умов це можливо). Натомість величезною перевагою є більша свобода у структуруванні тексту на ті чи інші одиниці, в описі тих чи інших показників.

Для опису кільцевих повторів за допомогою регулярного виразу розбіємо весь вираз на групи; кожна група міститиметься в окремій парі дужок; нумерація груп розставлена у нижньому рядку:

| | | | | | | | | | | |
|-------|---------|------|------|---------|------|---|--------|------|--------|------|
| (>?s) | (\w{4}) | (.*) | s | (\w{4}) | (.*) | s | (\4) | (.*) | s | (\2) |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| | | | Лек1 | | Лек2 | | Лекс.2 | | Лекс.1 | |

У першій групі зазначається, що виразу має передувати пробіл або інший роздільник, у другій групі має бути слово довжиною не менше чотирьох літер, третя група відповідає пропуску між першою і другою лексею хіазму, у четвертій групі має бути слово довжиною не менше чотирьох літер.

У шостій і восьмій групах за допомогою наскісної риси відбувається порівняння зі значенням четвертої і другої груп, тобто, другого і першого слова хіазму.

Запит, наведений нижче, виключає крапки всередині псевдохіазму, тобто, дозволяє вести пошук у межах розповідного речення:

(>?s)(\w{4,})([^\.\? \!]*s)(\w{4,})([^\.\? \!]*s)(\4)(.*)s(\2)

Не виключаємо з хіазму знак оклику та знак питання, оскільки в експресивному мовленні ці розділові знаки цілком можуть розривати одне цілісне висловлення. Тим не менш, якби це було потрібне, регулярний вираз мав би такий вигляд:

(>?s)(\w{4,})([^\.\? \!]*s)(\w{4,})([^\.\? \!]*s)(\4)(.*)s(\2)

З урахуванням того, що хіазм може займати не один, а два рядки у файлі, додамо символи переходу на новий рядок:

(>?s)(\w{4,})([^\.\? \!]*n*[^\.\? \!]*s)(\w{4,})([^\.\? \!]*n*[^\.\? \!]*s)(\4)(.*)s(\2)

Оскільки антиметаболи часто мають характер афоризмів, вибір текстового матеріалу зупинився на книзі прислів'їв, доступній у відкритому доступі проекту Gutenberg "Dictionary of English Proverbs and

Proverbial Phrases" [11] Результати з Proverbial Phrases:

9. **ADVICE.** *What every one asks, what every one gives, but what very few take--advice.*

23. **ALE.**

He that buys land buys many stones,

He that buys flesh buys many bones,

He that buys eggs buys many shells,

He that buys ale buys nothing else.

283. **COCK.**

If the cock moult before the hen,

We shall have weather thick and thin;

But if the hen moult before the cock,

We shall have weather hard as a block.

369. **CUCKOO.**

When the cuckoo comes to the bare thorn,

Sell your cow and buy your corn,

But when she comes to a full bit,

Sell your corn, and buy your sheep.

419. **DEVIL.**

The Devil Was Sick, the Devil a Monk Would Be;

The Devil Grew Well, the Devil a Monk Was He.

536. **FAITH.** *Love asks faith, and faith asks firmness.*

574. **FIDDLER.**

Let not the fiddler play the fife,

Nor fifer play the fiddle.

696. **GOOD.** *Do all the good you talk of; but talk not of all the good you do.*

881. **JUST.** *Just is honest, and honest is just.*

976. **LITTLE HOUSE.**

A little house well fill'd,

A little land well till'd,

And a little wife well will'd.

1071. **MERRY.** *Merry meet, merry part, merry meet again.*

1254. **PLACE.** *A place for everything, and everything in its place.*

1398. **SAY WELL.**

"Say well," and "do well," end with one letter;

To say well, it is well, but to do well is better.

1429. **SHOP.** *Keep thy shop, and thy shop will keep thee.*

1547. **STRIFE.** *Haste makes waste, and waste makes want, and want makes strife between the good man and his wife.*

1537. STONE WALL.

*Hard **upon** hard **makes** a bad stone wall;*

*But soft **upon** soft **makes** none at all.*

1547. STRIFE. *Haste **makes** waste, and **waste** makes want, and want makes strife between the good man and his wife.*

1598. THINK.

*Who **thinks** to live must live to **think**,*

Else mind and body lose their link.

1656. TREES. *Set trees **poor**, and they will grow **rich**; set them **rich**, and they will grow **poor**.*

1695. VALENTINE'S DAY.

*On **Valentine's** day will a **good** goose lay,*

*If she be a **good** goose, her dame well to pay,*

*She will lay two eggs before **Valentine's** day.*

1713. WANT OF.

*For **want** of a nail the **shoe** is lost; for want of a **shoe** the horse is lost; for **want** of a horse the rider is lost.*

1858. WOULD.

*"He that **wolde** not **when** he might,*

*He shall not **when** he **wold-a**."*

Прислів'я 9, 23, 419, 976, 1071, 1537 і 1713 не містять хіазму: кільцевий повтор лексем насправді в них є лише випадковим збігом. Однак результат автоматичної вибірки у 15 корисних прикладів з 28 вважаємо задовільним. З автоматичної вибірки видно, що частка корисних прикладів становить 68,1 %. І хоча застосований вираз не працює ідеально, з-поміж знайдених фрагментів значно легше вибрати потрібні послідовності, ніж з усіх 1880, які містяться в книзі.

Отже, антиметабола є поширеним різновидом хіазму, яка заснована на кільцевому повторі двох лексем, і саме вона потрапляє до кола уваги дослідників частіше, ніж суто синтаксичний хіазм. Завдяки формальному повтору антиметабола характеризується однозначними маркерами, що уможливають її автоматичну або, принаймні, напівавтоматичну вибірку з тексту.

Мова корпусних запитів дозволяє здійснювати вибірку різних структур як за лексичними, так і морфосинтаксичними й пунктуаційними показниками, в той час як регулярні вирази мають більш обмежені можливості, також враховуючи той факт, що частина корпусних запитів вже включає до себе регулярний вираз. У той самий час кількість умов, що задаються мовою корпусних запитів, значно обмежена: вони стосуються лише можливості порівняння наявних у корпусі атрибутів і частоти токенів, у той час як при застосуванні регулярних виразів можливо задавати будь-які види умов, і не лише щодо атрибутів, а й також відносно їхніх значень; поза тим, у разі застосування регулярного виразу у програмному коді умову можна описувати складними логічними виразами, у той час як у мові корпусних запитів можлива лише лінійна послідовність умов за

допомогою кон'юнкції і диз'юнкції. Окрім того, регулярні вирази дозволяють формально опрацьовувати будь-які електронні тексти у програмах редагування текстів, оскільки формальне опрацювання мовних знаків відбувається безвідносно до їхніх лексико-граматичних характеристик, а корпусні запити працюють лише у спеціально анотованих корпусах текстів.

Перспективно поширити обрані методи вибірки матеріалу й на суто синтаксичний хіазм. Очевидно, що для його вибірки ліпше підійде використання запитів на матеріалі синтаксично анотованих корпусів.

Література

1. Селіванова О.О. Сучасна лінгвістика: термінологічна енциклопедія. – Полтава: Довкілля-К. – 2006. – 712 с.
2. Dubremetz M., Nivre J. Syntax Matters for Rhetorical Structure: The Case of Chiasmus, 2016. – P. 47-53. – Режим доступу: 10.18653/v1/W16-0206.
3. Тележкіна О.О. Хіазм як мовностилістичний прийом у віршовій оповіді (на матеріалі української поезії II половини – початку ХХІ століття) // Лінгвістичні дослідження. – 2017. – Вип. 46.
4. Лонська Л.І. Структурно-семантичні різновиди хіазму (на матеріалі українських паремій). Мовознавчий вісник. – Вип. 20. Черкаси : Видавець Чабаненко Ю, 2015. С. 99–104.
5. Nänny M., Fischer O. Form Miming Meaning. – Philadelphia, PA, USA: John Benjamins Publishing, 1999. – 443 p.
6. Лесин В.М. Антиметабола // Літературна енциклопедія в п'яти томах – К., 1988-1995. – Тт. 1. Інститут літератури імені Т. Г. Шевченка Академії наук Української РСР. – 536 с.
7. Brogan T.V.F., Halsall A.W. Antimetabole // The Princeton Encyclopedia of Poetry and Poetics. – Princeton: Princeton University Press, 2012. – 440 p.
8. Пошук у корпусі ГРАК. – Режим доступу: <http://uacorpus.org/Kyiv/ua/poshuk-u-grak/poshuk-cherez-interfejs-nosketchengine>
9. Шведова М., Фокін С. Основи мови запитів CQL (на матеріалі корпусу ГРАК). – Режим доступу: <https://drive.google.com/file/d/1Ti3VGpFYbzXRYO0tPVZoyGT343e9jiWl/view>
10. Sketch Engine Open Corpora. – Режим доступу: <https://app.sketchengine.eu/#open>
11. Preston Th. Dictionary of English Proverbs and Proverbial Phrases Proverbial Phrases / Produced by Delphine Lettau, Lisa Reigel, and the Online Distributed Proofreading Team. – 2022. – Режим доступу: <http://www.pgdp.net>