

КОМП'ЮТЕРНА ЛІНГВІСТИКА

УДК 811.161.2'42:004

[https://doi.org/10.17721/um/52\(2022\).93-115](https://doi.org/10.17721/um/52(2022).93-115)

Наталія Дарчук, д-р філол. наук, проф.

КНУ імені Тараса Шевченка, Київ

<https://orcid.org/0000-0001-8932-9301>

КОРЕФЕРЕНТНІСТЬ У ПОЕТИЧНИХ ТВОРАХ ЛЕСІ УКРАЇНКИ

Метою цієї статті є дослідження категорії кореферентності в поетичному мовленні Лесі Українки на основі корпусу текстів української мови, розміщеного на порталі tova.info. Матеріалом слугували прижиттєві видання Лесі Українки. Схарактеризовано категорійну сутність кореферентності, визначено види репрезентантів об'єктів дійсності, їхні функційно-стилістичні різновиди і типові об'єкти репрезентації в поетичному тексті, розглянуто найчастотніші моделі їх представлень.

***Ключові слова:** референт, кореферент, кореферентність, дерево залежностей, кореферентна пара.*

Постановка наукової проблеми та її актуальність. Категорія кореферентності відіграє винятково важливу роль у текстотворенні, оскільки вона зумовлена фундаментальними закономірностями організації тексту. Лінійність тексту і нелінійність описуваної в тексті ситуації неминуче пов'язані зі згадуванням елементів ситуації з іншою номінацією цього об'єкта, яка базується на тому, що було сказано про цей об'єкт, і на тих знаннях, які в тексті не вербалізовані. Тому в тексті виникають ланцюжки кореферентних імен, які, маючи бінарний зв'язок, почасти спонукають до появи вторинної номінації, яка є анафоричним зв'язком із первинною номінацією.

Актуальність обраної проблеми зумовлена тим, що категорія кореферентності є засадничою в текстотворенні, здійснюючи пізнавальну й комунікативну функцію, визначаючи навіть структуру поетичного твору. Попереднє дослідження

поетичної спадщини Лесі Українки продемонструвало у використанні форм граматичної особи індивідуальний стиль поетки, значну роль в утворенні смислу та загальної композиції ліричних віршів [Дарчук].

Метою цієї статті є дослідження категорії кореферентності в поетичному мовленні Лесі Українки на основі корпусу текстів української мови, розміщеного на порталі mova.info. Матеріалом були прижиттєві видання Лесі Українки.

Аналіз досліджень проблеми. Когерентність як загальна, а кореферентність як часткова категорії перебувають у фокусі уваги кількох дисциплін: лінгвістики, кібернетики і комп'ютерної інженерії. І якщо представників точних наук цікавлять методи оцінки зв'язності тексту, пошуки причин порушення зв'язності 1) для генерування тексту, автоматичного реферування, редагування, дидактики [Гогерчак та ін.; Cui et al.; Guinaudeau, Strube; Komen; Le, Mikolov; Li, Novy]; 2) у медичних цілях (для виявлення симптомів ментального захворювання тощо) [Погорілий, Крамов 2018 а; Погорілий, Крамов 2018 б], то в лінгвістиці досліджується цілісність тексту як логіко-семантична, граматична і стилістична співвіднесеність і взаємопов'язаність складових елементів; одна з визначальних характеристик тексту/дискурсу та необхідних умов текстуальності [Кибрик; Кобзарєва; Леднік; Максим'юк; Barzilay, Lapata; Putra, Tokunaga]. Насправді, зацікавлення цією проблематикою принаймні трьох наук свідчить про глибину порушеної проблеми, кінцевою метою досліджень є саме пошук смислу в аналізованому тексті. Що об'єднує в постановці проблем когерентності лінгвістику, кібернетику і комп'ютерну інженерію? Це обов'язкова структуризація фактів, пошук різних формальних моделей їхнього опису і, нарешті, нині чи не основне – пошук найадекватніших комп'ютерних форм представлення даних та оцінка отриманих результатів.

Для лінгвістики категорія кореферентності має значення як ідіостилістична характеристика, про що свідчать наукові розвідки (див. [Куриляк, Максим'юк]). Водночас кореферентність як текстова категорія, яку активно вивчають на матеріалі різних

дискурсів, не була предметом системного аналізу. Мовотворча роль Лесі Українки є всеосяжною, що завжди привертало увагу і літературознавців, і мовознавців. Цікаво було дослідити корелюваність як ідіостилістичну рису творчості поетеси.

Оскільки сучасну наукову добу характеризує активне впровадження новітніх комп'ютерних засобів, закономірним кроком є створення відповідного інструментарію для роботи з текстами поетеси. У цьому полягає водночас і **актуальність**, і **новизна** проекту електронного дослідницького корпусу мови Лесі Українки в межах Корпусу української мови (КУМ), створеного в лабораторії комп'ютерної лінгвістики Навчально-наукового інституту філології Київського національного університету імені Тараса Шевченка. Мета корпусу текстів Лесі Українки – створення такого лінгвістичного і програмного продукту, який би, з одного боку, надавав широкі відомості про мову поетеси, а з іншого – був зручним для використання в різноманітних дослідженнях. Така мета вимагала виконання багатьох **завдань**, серед яких: лінгвістичний аналіз текстів Лесі Українки; укладання бази даних зафіксованих у них мовних одиниць із їхніми граматичними й кількісними характеристиками; створення зручного користувацького інтерфейсу, за яким можна було б здійснювати пошук, сортувати та статистично опрацьовувати всю зібрану в базі інформацію відповідно до потреб дослідників [Дарчук]. Лінгвістичне опрацювання здійснювалося двома способами: 1) автоматично – за допомогою автоматичного морфологічного аналізу (для визначення частин мови та їхніх граматичних форм); 2) автоматизовано – унаслідок роботи лінгвіста, який контролює правильність результатів аналізу, редагує й усуває можливі помилки.

Об'єктом дослідження є ідіостиль поетичного доробку Лесі Українки, який потребує глибокого й різнорівневого наукового аналізу, **предметом** – функціонування категорії корелюваності.

Опис фактичного матеріалу та методів, застосованих у процесі аналізу. Матеріалом були прижиттєві видання Лесі Українки: “Відгуки (із циклу “Невольницькі пісні”, “Легенди”, “Ритми”, “Хвилини”); “Думи і мрії” (“Відгуки”, “Мелодії”

“Невільничі пісні”, “Поєми”); “На крилах пісень” (I та II розділи). Проаналізовано 153 тексти. Загальний обсяг матеріалу – 38479 слововживань. Кількість речень – 3218. Корпус розміщено на мовно-інформаційному порталі mova.info. Застосовано методи дистрибутивного, компонентного, кількісного та описового аналізів.

Наукова новизна, теоретична і практична цінність отриманих результатів полягає в тому, що ідіостиль Лесі Українки з погляду ідіостилістичної організації тексту потребує докладних, виважених, об’єктивно цілісних студій глибинного характеру, а розроблений лінгвістичний і програмний продукт Корпусу текстів Лесі Українки надає широкі відомості про авторську мову і є зручним для використання в різноманітних дослідженнях, зокрема текстотворенні. Схарактеризовано категорійну сутність кореферентності, визначено види репрезентантів об’єктів дійсності, їхні функційно-стилістичні різновиди й типові об’єкти репрезентації в поетичному тексті, розглянуто найчастотніші моделі їх представлень.

Виклад основного матеріалу й обґрунтування результатів дослідження. *Референт* – об’єкт позамовної дійсності у конкретній мовній ситуації; предмет референції. *Кореферент* – мовний знак, співвідносний з іншим мовним знаком за одним і тим самим референтом. *Кореферентність* – це відношення між референтом і кореферентом, тобто між словами або словосполученнями, які позначають той самий об’єкт, маючи один і той самий референт. Кореферентність як важлива текстова категорія забезпечує тематичну неперервність, інформаційну завершеність, когерентність, тобто цілісність тексту.

Установити кореферентність у цьому проекті означає виявити в тексті всі згадування, які належать до тієї самої сутності. Способом вираження референта може бути слово, словосполучення, а кореферента – слово, словосполучення або підрядне речення.

Розмічування тексту, тобто встановлення кореферентних пар, здійснювалося автоматизовано, за участі лінгвіста, з урахуванням специфіки поетичного тексту Лесі Українки, у якому між членами пари майже завжди дистантний зв’язок і почасти від-

сутнє узгодження в категорійних граматичних формах, про що йтиметься нижче¹. Кожен текст вірша дублювався на екрані: зліва блакитним кольором маркувався кореферент (К), справа жовтим кольором – референт (Р), що давало змогу візуально їх розрізнити. Натисканням клавіші “запам’ятати” у викидному списку обирався тип відношень між К і Р, після чого при К і Р приписувався відповідно парний елемент: кореференту – Р, а референту – К. У разі, коли одному референту відповідали кілька кореферентів, тобто було кілька вербальних репрезентацій об’єктів художньо інтерпретованої дійсності, утворювався ланцюжок.

Ми дотримувалися таких загальних правил:

1. Позначати бінарну пару якомога довшим висловом, що характеризує сутність описуваної ситуації. Наприклад:

В дворі старого замчища-руїни зібрались ми поважні учені голови гладенькі й кучеряві (ідентичність) на віче, все поважніми усіх (ідентичність), учені головими усіх (ідентичність), гладенькі й кучерявими усіх (ідентичність), і віком не малі – якби зложити літа усіхповажні учені голови гладенькі й кучеряві (ідентичність), либонь, століття вийшло б!

У наведеному реченні кореференту *ми* приписується згадка про референт *поважні учені голови, гладенькі й кучеряві*; кореференту *усіх* (якби зложити літа усіх, либонь, століття вийшло б) також приписується той самий референт.

2. У разі, якщо є складнопідрядне речення зі співвідносним словом, останнє кваліфікують як референт, а підрядне речення – як кореферент, напр.:

садочка спогадів що що (ідентичність) моїх, що садочка спогадів садочка спогадів (ідентичність) міниться барвистими квітками при світлі мрій, мов при західнім сонці.

У цьому фрагменті референтом буде *садочка спогадів*, а кореферентом – *що міниться барвистими квітками при світлі мрій*.

3. Іменникові вислови, які є предикатами, вважаються кореферентними лише тоді, коли вони позначають еквівалентні сут-

¹ Для більш структурованих текстів, напр., наукового чи ділового дискурсу, розроблено програму автоматичного виявлення кореферентності в українськомовному тексті.

ності, а не включення однієї сутності в іншу, як, напр., у реченні: *Макрон є президентом Франції*, у якому слово *президентом* вжито в предикатному значенні.

При завершенні етапу автоматизованої розмітки текстові дані автоматично інвертувалися в табличну форму, де представлено загальні кількісні дані про тип кореферентних відношень як для конкретного тексту, так і для цілого вузла текстів, напр., збірки віршів, або для всіх творів поетки (табл. 1).

Таблиця 1. Кількісні дані типів зв'язку кореферентного зв'язку

Назва кореферентного зв'язку	Загальна кількість	Кількість у відсотках
ідентичність	1445	95,1
частина-ціле	27	1,8
предикативна ідентичність	6	0,4
розрив сукупності	29	2,0
асоціація	10	0,7
метонімія	3	0,2

На матеріалі вказаних збірок автоматично укладено таблицю, де вміщено 1520 референційно тотожних одиниць, які утворювали бінарну пару. Серед проаналізованих 153 текстів у 23 не трапилися випадки кореферентності. Кількісно оцінити насиченість тексту категорією кореферентності допоможе пропонований нами індекс кореферентності:

$$I_{co} = K / N,$$

де I_{co} – індекс кореферентності; K – кількість бінарних пар; N – кількість речень у тексті.

Як показує дослідження, лише у 20 текстах індекс перевищує 1, а максимальне значення (1,42) має вірш “Забута тінь”, де всього 24 речення, але вірш буквально пронизаний кореферентними зв'язками, тобто в середньому в кожному реченні є посилання на дійових осіб. У першому реченні згадується ім'я Данте, *флорентійського вигнанця, співця*, а потім формами особового займенника *він* заміщується його ім'я у 12 випадках і лише в прикінцевому реченні повторюється. Друга дійова особа, Беатріче

Портінарі, також згадується двічі – у середині вірша й наприкінці, у решті речень її ім'я також заміщено формами особового займенника *вона*. За рахунок таких зв'язків вірш стає ніби зцементованим єдиним цілим, композиційно підтверджуючи силу кохання: *Безсмертна пара Данте й Беатріче, Потужна смерть не розлучила їх*.

На другому місці вірш “До природи” (1,25), який за обсягом невеликий (4 речення), але насичений парадигматичними формами особового займенника *ти*, референтом якого є природа, до якої поетка звертається як до живої істоти: *Натуро-матінко! Я на твоєму лоні дитячі радощі і горе виливала і матір'ю тебе я щиро звала*.

На третьому місці вірш “Віче” (1,14), у якому 29 речень, які також зв'язані займенниками, що заміщують дійових осіб вірша і самого оповідача, який був учасником таємного віче та свідком подій часів Жанни д'Арк.

Іншим коефіцієнтом є індекс покриття тексту явищами кореферентності:

$$I_{ct} = K / L \times 100 \%,$$

де I_{ct} – індекс кореферентності для тексту; L – довжина тексту у словоформах; K – кількість бінарних пар.

Цей індекс указує на те, який відсоток тексту охоплений явищем кореферентності. Як показало дослідження, на першому місці драматична поема “Одержима” (8,85 %). Це зрозуміло, оскільки вона побудована у формі філософського діалогу Міріам (одержима) з Месією. На другому місці вже згадуваний вірш “Забута тінь” (8,5 %), а на третьому – прозовий вірш “Жертва” (8,4 %), який невеликий за розміром (155 словоформ, 9 речень), але цілісний саме за рахунок текстотвірної категорії кореферентності.

Нами укладено таку класифікацію відношень кореферентності:

1. **Ідентичність** – позначає вислови, що стосуються однієї сутності реального світу. Приклад: *Де поділися ви, голосні слова, що без вас моя туга німа?* (“Де поділися ви, голосні слова” / “Ритми”).

На рис. 1 представлено дерево залежностей речення, у якому кожен вузол – словоформа, а різнокольорові дуги – типи синтаксичних зв'язків між словоформами, які розшифровано у крайньому правому стовпчику. Референт-згадка *голоснії слова* та кореферент-згадка *ви* стоять поруч у головному реченні, а другий кореферент-згадка (*без вас*) – у складнопідрядному реченні, тобто референт між двома кореферентами пов'язаний із ними дугою “ідентичність”, що ілюструє симетричність цього відношення.

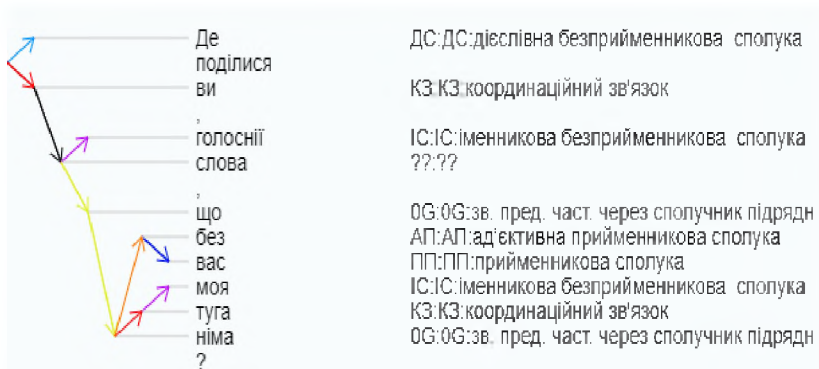


Рис. 1. Дерево залежностей речення *Де поділися ви, голоснії слова*

Стрілки від анафор *ви, вас* спрямовані на їхнього антецедента *голоснії слова*: *ви* → *голоснії слова* ← *вас*.

2. **Розрив сукупності** – пов'язує вислів з іншими висловами, які у сукупності стосуються однієї сутності реального світу. Приклад: *Сиділи ми в садку, там саме зацвітав, і сипався з каштанів білий цвіт. Мені сім літ минуло, а їй, либонь, минуло двадцять літ* (“Забуті слова”).

У цьому реченні стрілка спрямована від вислову, який розривається, до висловів, які описують елементи сукупності: *ми* → *мені, їй*. Цікаво, що в цьому вірші не згадуються імена дійових осіб.

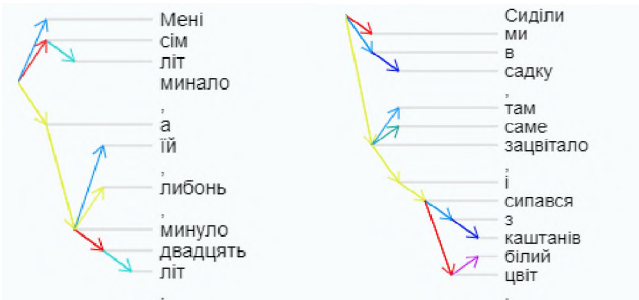


Рис. 2. Дерево залежностей речень із вірша “Забуті слова”

3. **Частина-ціле** – вислів позначає сутність, що є частиною сутності, позначеної іншим висловом. Приклад: *І відхилили розколоне віко того саркофага ... Що ж там було? Ра-Менеїс була горда цариця, дочка фараонів, Гарна й страшна, мов Урея, змія золотая, Що обвивала подвійний вінець двох Єгиптів, Мала чоло діамантове й темні рубінові очі (“Ра-Менеїс”):*



Рис. 3. Дерево залежностей речення з поезії “Ра-Менеїс”

У цьому прикладі стрілка спрямована від слова, що позначає частину (*чоло, очі*), до слова, що позначає ціле (*змія*): *змія* ← *чоло й очі*.

4. **Предикативна ідентичність** – поєднує предикативний вислів із номінативним, що стосуються однієї сутності реального світу. Приклад: *Я з ласки бога цариця, бо, гляньте, – сиджу на престолі* (“Завжди терновий вінець”):

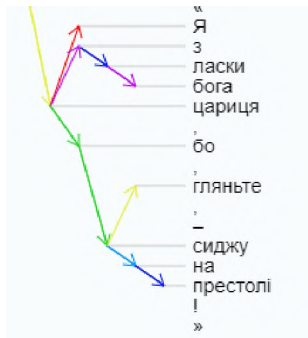


Рис. 4. Дерево залежностей речення з поезії “Завжди терновий вінець”

У цьому прикладі стрілка спрямована від іменникового вислову до предикативного: *цариця* → *сиджу*.

5. **Метонімія** – при семантичних переносах можлива корелюваність між іменними групами, які позначають різні об’єкти, пов’язані між собою лише метонімічно. Приклад: *А далі слово змінилося у спів, і вся громада до дзвінкої промови прилучилась* (“Віче”):

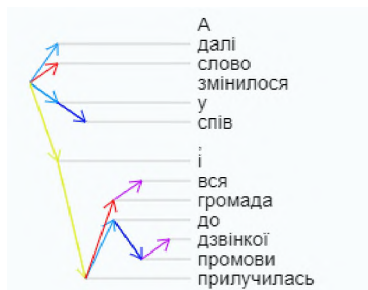


Рис. 5. Дерево залежностей речення з поезії “Віче”

У цьому реченні слово → *спів* кореферентні, стрілка показує напрям зв'язку.

6. **Асоціація** – часткова пов'язаність. Приклад: *Годівлю дав юрбі, тілам і душам, всім дав спокій, а сам у сій пустелі пасе думок отари незчисленні* (“Одержима”):

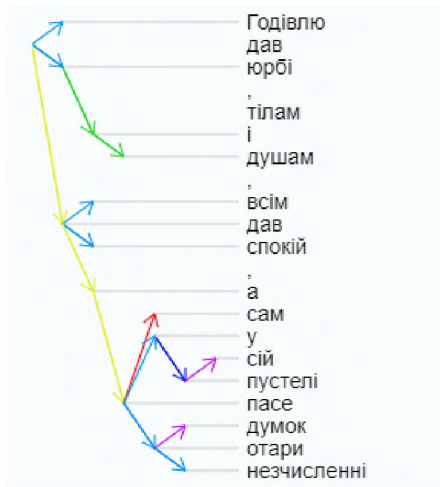


Рис. 6. Дерево залежностей речення з поезії “Одержима”

У цьому прикладі стрілка спрямована від вислову-згадки до вислову, на який здійснено посилання: ...пасе *отари* → *думок*.

Таке унаочнене представлення на прикладі ДЗ дає змогу встановити місцерозташування пари Р-К у реченні, позицію відносно одна одної. Як бачимо з табл. 1, 95,1 % усіх кореферентних пар – це ідентичні зв'язки, на другому і третьому місцях перебувають “розрив сукупності” та відношення “частина-ціле”. Проте серед шести типів кореферентності лише перший, ідентичність, може бути визначений формально, за алгоритмічними правилами. Синтаксичний аналізатор дає змогу автоматично встановити кореференцію для особових (*він, вона, воно, вони*), присвійних (*його, її, їх*), зворотного (*себе*), відносних (*який*) займенників на підставі правил пошуку конгруентного слова, тобто слова, яке узгоджується із займенником

у роді та числі. Для формулювання правил використовуємо таку інформацію:

1. Позиція К і Р у ланцюжку речення;
2. Позиція К і Р у дереві залежностей;
3. Належність сегмента з К або Р до ускладнювачів речення – підрядного речення або звороту;
4. Тип синтаксичного зв'язку К і Р.
5. Тип синтактико-семантичного відношення К і Р.
6. Наявність у референта залежних слів.
7. Перетин у роді та числі К і Р.

Автоматичне вилучення дасть змогу побудувати семантичне дерево кореферентних пар для цілого тексту (рис. 7).



Рис. 7. Міжреченнєва кореферентність

Проілюструємо його на матеріалі вірша “Віче” (із циклу “Невольницькі пісні”: Відгуки). Спершу подаємо фрагменти речень:

1-е: *...мені не прийдеться побачить того садочка спогадів моїх, що міниться барвистими квітками...*

2-е: *...повстав дитячий спогад і кличе, й вабить: глянь на мене...*

3-е: *...зібрались ми на віче, все поважні, учені голови, гладенькі й кучеряві, і віком не малі – якби зложити літа усіх.*

4-е: *Ми всі були на зборах, всі дванадцять.*

5-е: *Обачні люди, тямилі ми добре, що ми живем у небезпечний час...*

На рис. 7 представлено кореферентні пари з наведених речень, які автоматично були вилучені з дерев залежностей п'яти речень: 1К–1Р – мітки кореферента і референта 1-го речення, 2К–2Р – другого, 3К–3Р – третього, 4К–4Р – четвертого і 5К–5Р, 6К–6Р, 7К–7Р – п'ятого, шостого та сьомого речень. Корреферентні пари розміщені в овальних фігурах, пунктирні дуги спрямовані від кореферента-згадки до референта-згадки. До членів пари проведено суцільні дуги від вузлів дерева залежностей, які уточнюють контекст (напрямок дуг у фрагментах дерева залежностей зберігається).

У першому реченні кореферентна пара *садочок спогадів* (Р) – сполучне слово *що* (К) розташована у двох частинах речення – головному й підрядному, отже, вона зв'яже ці два речення синтаксично. У другому реченні пара *дитячий спогад* (Р) і *мене* (К) як персоніфікація пов'язана з попередньою парою лексично, завдяки лексемі *спогад* (*садочок спогадів, дитячий спогад*), демонструючи зв'язність цих двох речень. До референта третього речення *все поважні, учені голови, гладенькі й кучеряві* спрямовані дуги кореферентів третього (*ми; усі*) та четвертого (*ми*) речень, причому уточнюється значення референта третього речення: *всі дванадцять* – це *все поважні, учені голови, гладенькі й кучеряві*. До референта п'ятого речення *обачні люди* спрямовані дуги шостої та сьомої кореферентних пар із кореферентом *ми* (*тямилі ми; живемо ми*). Отже, у цьому уривку тексту смисловим центром є три референти-згадки: *садочок спогадів, все по-*

важні, учені голови, гладенькі й кучеряві та обачні люди, які цементують текст через свої кореференти-згадки.

Помітно, що в творах Лесі Українки вербальна репрезентація об'єктів реальної та художньо інтерпретованої дійсності виражається в наявності не лише кореферентних пар, а й рядів, коли біля кореферента накопичуються кілька посилань, утворюючи ряди моделей вербальних репрезентацій. Наприклад, у поезії “Завжди терновий вінець...”:

Завжди терновий вінець буде кращий воно (асоціація), ніж царська корона воно (асоціація).

Завжди величніша путь на Голгофу воно (асоціація), ніж хід тріумфальний воно (асоціація).

Так одвіку було й так воно Завжди терновий вінець буде кращий ніж царська корона Завжди величніша путь на Голгофу ніж хід тріумфальний (асоціація) буде довіку, поки житимуть люди і поки ростимуть терни

– кореферент *воно* заміщує два референти, які зафарбовані червоним кольором і приписані йому маленькими літерами, утворюючи кореферентні ряди, зв'язані асоціативно, оскільки і в першому, і в другому випадку *терновий вінець* асоціюється з *царською короною*, а *величніша путь* – із *ходом тріумфальним*.

У вірші “Єреміє, зловісний пророче в залізнім ярмі!..” після звертання до Єремії (референт) у наступних реченнях згадка замінюється на кореферент *ти* у різних парадигматичних формах зі зв'язком ідентифікації:

Єреміє тобі *ти* *серце* твоє *ти* *ти* *серце* *ти* твої сльози (ідентичність), зловісний пророче в залізнім ярмі!

Певне, *серце* господь *тобі* Єреміє (ідентичність) дав із твердого кришталю: *ти* Єреміє (ідентичність) провидів, що люд буде гнить у ворожій тюрмі, – як же *серце* твоє Єреміє (частина-ціле) не розбилося від лютого жалю?

Як *ти* Єреміє (ідентичність) міг дочекатись, чи справдиться слово твоє?

Роєм стріли ворожі на божее місто летіли, – певне, чарами *ти* Єреміє (ідентичність) гартував тоді *серце* ньому Єреміє (частина-ціле) своє, що на ньому *серце* (ідентичність) ламался навіть ворожій стріли!

Отже, при референті після автоматизованого розмічування накопичується кореферентний ряд. Можна припустити, що художній поетичний текст через специфіку текстотворення за рахунок створюваних образів відзначається не лише контактними, але й дистантними позиціями референт-кореферент. Як показує дослідження, для ідіостилю Лесі Українки характерні саме дистантні зв'язки. Наприклад, поезія “Єреміє, зловісний пророчє в залізнім ярмі!..”: Єврейські мелодії (із циклу “Невольницькі пісні”) починається звертанням до Єремії у першому реченні, а всі наступні речення (з другого по п'яте) містять лише згадку про нього і звертання, виражені різними формами особових і присвійних займенників. Під час кожного згадування того самого об'єкта (Єремія) відбувається нова номінація цього об'єкта, яка базується на тому, що вже було сказано про нього, і на тих знаннях, які не вербалізовані в тексті, тобто на екстралінгвальних знаннях. Це знання про постать Єремії: Єремія – один із чотирьох святих пророків, покликаний Господом бути пророком серед іудеїв, яких він зі слізьми на очах благав не відступати від Бога, не вірити ідолам, передбачав біди і війни, але люди насміхалися з нього, сварили його і навіть намагалися вбити. Напр.: *серце господь тобі дав із твердого кришталю; ти провидів, що люд буде гнить у ворожій тюрмі; серце твоє не розбилось від лютого жалю; як ти міг дочекатись, чи справдиться слово твоє?; чарами ти гартував тоді серце своє, що на ньому ламалися навіть ворожії стріли!*

Вірш закінчується звертанням до Єремії, і в останніх трьох реченнях є аналогічне згадування. Така кореферентна структура вірша сприяє його конкатенації, адже міжреченнєві не лише граматичні (напр., сполучники на початку речення), а й смислові зв'язки посилюють цілісність твору.

До контактних можна віднести складнопідрядні речення зі співвідносним іменником-референтом і підрядним означальним зі сполучним словом, напр.:

...більш мені не прийдеться побачить того садочка спогадів ще що (ідентичність) моїх, що садочка спогадів садочка спогадів (ідентичність) міниться барвистими квітками при світлі мрій (“Віче”).

У головному реченні *садочок спогадів* набуває додаткової номінації за рахунок підрядного речення, у якому створюються нові образи, а референти, завдяки такій дескрипції, набувають унікальних, одинично маркованих репрезентантів об'єкта-референта.

Наступним методичним кроком є визначення типових об'єктів репрезентації дійсності в поетичному тексті Лесі Українки, найчастотніших граматичних моделей їхнього представлення. Для цього укладена робоча таблиця (табл. 2) для занесення даних, а саме:

- перша колонка – кореферент (К);
- друга колонка – референт (Р);
- третя колонка – морфологічна модель пари (К/Р): через слеш частиномовний код і код категорійних характеристик (рід, число, відмінок);
- четверта колонка – синтаксична модель – функція членів речення кожного з членів пари (К/Р);
- п'ята – позиція кореферента відносно референта в реченні;
- шоста – загальна семантична характеристика кожного з членів пари в термінах чотирьох доменів: ім'я власне; загальне; реальне; ірреальне.
- сьома колонка – назва твору.

Таблиця 2. Скріншот робочої таблиці кореферентності

Кореф	Реф	Морф. мод Р-К	Синт. мод Р-К	Поз. Кореф	Х-ка. реф	Назва тексту
неї	зірка	МЗ/КІ	обст/підм	пост	загальне, реальне	Зберегти «Чи тільки ж блискавцями літати...»
їм	словам	МО/ЛО	дод/дод	пре	загальне, реальне, ірр	Зберегти «Чи тільки ж блискавцями літати...»

Чи тільки зірка тим ясніше сяє , чим
теперва чорніша вколо неї ?

Чому ж би їм не залити угору , мов
жайворонка спів , дзвіночком
срібний ? Чи тільки ж блискавцями
літати словам отим , що з туги
народились ?

Якщо перші дві й остання колонки заповнюються автоматично, то решта колонок вручну за підтримки контексту, який отримуємо натисканням миші на кореференті. Така таблиця формується для кожного тексту після розмічування, набору текстів, напр., цілої збірки або всіх творів конкретного автора.

За даними Частотного словника творів Лесі Українки [10] серед найчастотніших слів друге місце після сполучника *і* посідає особовий займенник *я* (абсолютна частота 866, середня – 22, 8), що свідчить про дві ситуації: або у своїй ліричній поезії поетка веде мову про себе і свої почуття, або від імені персоніфікованого ліричного героя. І саме через співвіднесеність Корреферент – Референт встановлено, що превалює саме перша ситуація, тому що лише у 73 з 866 вдалося автоматизовано приписати референт. У решті випадків він був відсутній, оскільки під *я* розумілася сама Леся Українка, а 73 вживання займенника є лише в поемах “Одержима” та “Саул”. Отже, констатуємо, що більшість поезій еготивні.

У подальшому дослідженні ми обмежилися збіркою прижиттєвих видань “Відгуки” (загальна кількість словоформ – 9115). Розподіл абсолютних частот особових займенників у позиції кореферента такий:

<i>ти</i> – 104	<i>вона</i> – 40	<i>ви</i> – 29
<i>він</i> – 89	<i>вони</i> – 39	<i>воно</i> – 2
<i>я</i> – 73	<i>ми</i> – 23	

Як бачимо, позиції змістилися: займенник *ти* посідає перше місце, що пояснюємо взаємодією поета в момент творчості з предметом почуття, до якого він звертається на *ти*. Серед займенників 3-ої особи найчастотнішим є займенник чол. р. *він*, далі *вона*, *вони*. Серед особових форм у множині майже однаково частоту мають займенники *ми*, далі *ви*. Отже, можна припустити, що категорія особи належить до ядрових категорій поетичної граматики Лесі Українки, визначаючи її специфіку і структуру ліричного вірша і, відповідно, високий рівень кореферентності – зв’язності поетичного тексту. Насправді особові займенники виражають концептуальну тотожність.

У другій колонці вміщено репрезентатори об’єктів художньо інтерпретованої дійсності, серед яких:

1) власні назви, які представляють індивідуального референта – ліричного героя, дійової особи (*Міріам* (108), *Месія* (89), *Саял* (19));

2) загальнономінативні одиниці різного функціонального призначення й семантичного наповнення (*учитель* (8), *люди* (8), *жінка* (7), *пісня* (7), *люд* (6), *людина* (2) тощо), тобто різні об'єкти нарації;

3) описові конструкції “*той, хто...*”, “*такий, що...*”, “І/Г, *що...*”, “І/Г, *які...*”, які функціонують у ролі одинично маркованих репрезентантів об'єкта (19); це складні денотати.

Третя і четверта колонки таблиці вміщують морфологічні (третя колонка) та синтаксичні (четверта колонка) моделі репрезентації К–Р. Дослідження показує, що і на морфологічному, і на синтаксичному рівнях у межах першої групи кореферентних відношень “ідентичність” майже 67,4 % від 486 кореферентних пар зазнали зміни у граматичній морфологічній категорії відмінка, а відтак змінили синтаксичну роль члена речення. Наприклад, додаток/підмет (тут і далі лівобічний кореферент, правобічний – референт) (115 прикладів), підмет/додаток (20), підмет/означення (10), додаток/обставина (5), додаток/означення (7) тощо. Це підтверджує думку про різноманітність способів маніфестування об'єктів інформаційно-комунікативного наповнення поетичних текстів та багатовимірну інтерпретацію референта. Можна припустити, що поетичний дискурс Лесі Українки цьому сприяє.

У цій групі щодо позиції К–Р, як показує дослідження, у творах Лесі Українки в 77,2 % кореферент передує референтові, а в 22,8 % є у постпозиції, тобто кореферентні вказівні номінації семантично наповнюються постпозиційним референтом. Додатково проаналізовано текстотвірну сутність кореферентності, а саме: яка семантика номінативних одиниць референтів на рівні лексико-семантичних груп: *індивідуальне, загальне, реальне, ірреальне, складний денотат*. Виявляється, що найбільший відсоток (53,5 %) становить група референтів зі значенням “індивідуальне, реальне”, тобто іменники, які позначають власну назву дійової особи та є реальною живою істотою. На другому місці перебуває іменник зі значенням “загальна назва”, який є реальною істотою. На третьому місці – складний

денотат, який називає комунікативну ситуацію, виражену підрядним означальним.

Таку саму картину спостерігаємо у другій і третій групах “частина – ціле” та “розрив сукупності”: зміна морфологічних категорій відмінка й числа, а також почасти зміна синтаксичних категорій К–Р. Семантика референтів стабільно позначає загальні та реальні сутності. Решта відношень представлена невеликою кількістю прикладів.

Висновки. Завдання визначити кореферентність у повному обсязі на практиці розв’язано частково, оскільки, по-перше, потрібні прагматичні знання про аналізований текст, по-друге, авторські вподобання та поетичний стиль, коли референт при різних кореферентах згаданий лише раз або є уявним, утруднює отримання достовірних даних. Як показав наш досвід, установити кореферентність із високим ступенем достовірності можемо тоді, коли:

- референтом є власна назва дійової особи;
- референтом є загальні назви як назви осіб (*пан лицар, товариші колишні, ученик*), абстрактні назви (*слово, душа, робота*), конкретні назви (*гори, блискавиця, хмарка, квітка*). Референтність такої згадки не зумовлена наявністю в читача екстралінгвальних знань;
- відносні займенники (*що, як-ий, як-а, як-е, як-і*) у будь-якій граматичній формі, які не мають анафоричних референтів і кореферентні останній найближчій іменній групі того самого речення, узгодженій за родом і числом;
- особовий займенник, який, урахувуючи атропоцентричний характер мови і мислення, намагається посісти в реченні позицію суб’єкта – підмета. Непрямі відмінки кореферента ускладнюють установлення референта, оскільки такий займенник може бути кореферентним будь-якому іменникові з попередніх речень. Тому застосування автоматичного синтаксичного аналізу не дає змоги повною мірою розв’язати це завдання.

ДЖЕРЕЛА

Частотний словник творів Лесі Українки. URL : <http://www.mova.info/cfq1.aspx?fdid=lupv> (дата звернення: 21.02.2022).

ЛІТЕРАТУРА

Гозерчак Г., Дарчук Н., Кривий С. Представлення, аналіз та видобування знань з неструктурованих природномовних текстів. *Кибернетика та системний аналіз*. 2021. Т. 57. № 3. С. 164–183.

Дарчук Н. П. Категорія особи у творах Лесі Українки. *Лінгвістичні студії (Linguistic Studies)*. 2020. Вип. 40. С. 19–35.

Кибрик А. А. Референция и рабочая память: о взаимодействии лингвистики с психологией и когнитивной наукой. *Материалы первой российской интернет-конференции по когнитивной науке*. Москва : УМК Психология, 2004. С. 29–43.

Кобзарева Т. Ю. Проблема кореференции в рамках поверхностно-синтаксического анализа русского текста. *Компьютерная лингвистика и интеллектуальные технологии: Труды Международной конференции Диалог'2003*. Москва, Наука, 2003. С. 278–284.

Куриляк Л. П. Кореферентність у сучасному українському текстотворенні: автореф. дис... канд. філол. наук: 10.02.01. Івано-Франківськ, 2004. 24 с.

Леднік О. С. Когезія та когерентність як категорії зв'язного тексту. *Науковий часопис Національного педагогічного університету імені М. П. Драгоманова. Серія 10: Проблеми граматики і лексикології української мови*. 2010. Вип. 6. С. 119–123.

Максим'юк О. В. Кореферентність нерозкладних компонентів у структурі речення: автореф. дис... канд. філол. наук: 10.02.01. Чернівці, 2006. 20 с.

Погорілий С. Д., Крамов А. А. Метод розрахунку когерентності українського тексту. *Реєстрація, зберігання і обробка даних*. 2018 а. Т. 20. № 4. С. 64–75.

Погорілий С. Д., Крамов А. А. Автоматизована екстракція структурованої інформації з множини веб-сторінок. *Проблеми програмування*. 2018 б. № 2–3. С. 149–158.

Barzilay R., Lapata M. Modeling local coherence: An entity-based approach. *Computational Linguistics*. 2008. Vol. 34, No 1. P. 1–34.

Cui B., Li Y., Zhang Y., Zhang Z. Text Coherence Analysis Based on Deep Neural Network. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 2017. P. 2027–2030.

Guinaudeau C., Strube M. Graph-based local coherence modeling. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. 2013. Vol. 1. P. 93–103.

Komen E. R. Coreference Annotation Guidelines. *Coreference coding manual*. 2009. URL : http://erwinkomen.ruhosting.nl/doc/2009_Coref_CodingManual_V2-0.pdf (дата звернення: 16.12.2021).

Le Q., Mikolov T. Distributed representations of sentences and documents. *International Conference on Machine Learning*. 2014. P. 1188-1196.

Li J., Hovy E. A model of coherence based on distributed sentence representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014. P. 2039–2048.

Putra J. W. G., Tokunaga T. Evaluating text coherence based on semantic similarity graph. *Proceedings of TextGraphs-11: the Workshop on Graph-based Methods for Natural Language Processing*. Vancouver, Canada, 2017. P. 76–85.

REFERENCES

Hoherchak, H., Darchuk, N. & Kryvyi, S. (2021). Representation, analysis and extraction of knowledge from unstructured natural language texts. *Cybernetics and Systems Analysis*, 57, № 3, 164–183 (in Ukrainian).

Darchuk, N. P. (2020). The category of person in poetry of Lesia Ukrainka. *Linguistic Studies*, 40, 19–35. <https://doi.org/10.31558/1815-3070.2020.40.1.2> (in Ukrainian).

Kibrik, A. A. (2004). Referenciya i rabochaya pamyat: o vzaimodejstvii lingvistiki s psihologiej i kognitivnoj naukoj. *Proceedings of the first Russian Internet conference on cognitive science* (pp. 29–43), Moskva : UMK Psihologiya (in Russian).

Kobzareva, T. Yu. (2003). Problema koreferentsii v ramkah poverhnostno-sintaksicheskogo analiza russkogo teksta. *Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference Dialogue'2003* (pp. 278–284). Moskva : Nauka (in Russian).

Kuryliak, L. P. (2004). Coreference in modern Ukrainian text creating. *Extended abstract of Candidate's thesis (Linguistics)*. Ivano-Frankivsk (in Ukrainian).

Liednik, O. S. (2010). Koheziia ta koherentnist yak katehorii zviaznoho tekstu. *Naukovyi chasopys Natsionalnoho pedahohichnoho universytetu imeni M. P. Drahomanova. Seriiia 10: Problemy hramatyky i leksykolohii ukraïnskoi movy*, 6, 119–123 (in Ukrainian).

Maksymiuk, O. V. (2006). Coreference of stable components in the structure of the sentence. *Extended abstract of Candidate's thesis (Linguistics)*. Chernivtsi (in Ukrainian).

Pohorilyy, S. D. & Kramov, A. A. (2018 a). Method of the coherence evaluation of Ukrainian text. Reiestratsiia, zberihannia i obrobka danykh, 20, № 4, 64–75. <https://doi.org/10.35681/1560-9189.2018.20.4.178945> (in Ukrainian).

Pohorilyy, S. D. & Kramov, A. A. (2018 b). Avtomatyzovana ekstraktsiia strukturovanoj informatsii z mnozhyhny veb-storinok. Problemy prohramuvannia, 2–3, 149–158 (in Ukrainian).

Barzilay, R. & Lapata, M. (2008). Modeling local coherence: An entity-based approach. Computational Linguistics, 34, No 1, 1–34 (in English).

Cui, B., Li, Y., Zhang, Y. & Zhang, Z. (2017). Text Coherence Analysis Based on Deep Neural Network. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 2027–2030). <https://doi.org/10.1145/3132847.3133047> (in English).

Guinaudeau, C. & Strube, M. (2013). Graph-based local coherence modeling. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, (vol. 1, pp. 93–103) (in English).

Komen, E. R. (2009). Coreference Annotation Guidelines. Coreference coding manual. URL : http://erwinkomen.ruhosting.nl/doc/2009_CorefCodingManual_V2-0.pdf (date of access.: 16.12.2021) (in English).

Le, Q. & Mikolov, T. (2014). Distributed representations of sentences and documents. International Conference on Machine Learning (pp. 1188–1196). <https://doi.org/10.48550/arXiv.1405.4053> (in English).

Li, J. & Hovy, E. (2014). A model of coherence based on distributed sentence representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2039–2048) (in English).

Putra, J. W. G. & Tokunaga, T. (2017). Evaluating text coherence based on semantic similarity graph. Proceedings of TextGraphs-11: the Workshop on Graph-based Methods for Natural Language Processing (pp. 76–85), Vancouver, Canada. <https://doi.org/10.18653/v1/W17-2410> (in English).

Natalia Darchuk, Dr Hab, Prof.
Taras Shevchenko National University of Kyiv

Coreference in the poetic works of Lesia Ukrainka

The category of coreference plays an important role in the creation of texts, as it is based on fundamental principles of text organization. The linearity of a text and the non-linearity of a situation described in the text are closely related

to the mention of elements of a situation with another nomination of this object. Therefore, chains of coreferential names appear in the text and, having a binary relation, partly cause the emergence of a secondary nomination, which is an anaphoric relation with the primary nomination.

The purpose of this article is to study the category of coreference in the poetic speech of Lesia Ukrainka based on the corpus of the Ukrainian language represented on mova.info. The object of the research is the idiosyle of Lesia Ukrainka's poetry, and the subject is the functioning of the category of coreference.

The categorical essence of coreference is characterized, the types of representatives of real objects, their functional and stylistic varieties and typical objects of representation in a poetic text are determined, and the most frequent structures are considered.

Software for working with coreference has been developed and implemented. It has a user-friendly interface, which allows searching, sort, and doing the quantitative processing of the collected information according to the needs of researchers. On the material of 153 texts, 1520 referentially identical pairs are established. The coreferentiality index and the index of coreference coverage of the text proposed in this article helped to quantitatively assess the saturation of the text with the category of coreference. In addition, the classification of coreferential relations is composed: 1) identity; 2) collection gap; 3) part-whole; 4) predicative identity; 5) metonymy; 6) association. A dependency tree of a sentence made it possible to determine the location of coreferential pairs in the sentence, the position relative to each other, the belonging of the segment with the coreferent or the referent to the complicators of the sentence, i.e., a subordinate clause or phrases. This representation also helped to establish the type of syntactic relationship between the coreferent and the referent, the type of syntactic-semantic relationship between them, the presence of referent-dependent words, the intersection of the coreferent and the referent in gender and number. Automatic extraction of coreferential pairs from dependency trees allowed building intersentential coreference.

Keywords: referent, coreferent, coreference, dependency tree, coreferential pair.

Стаття надійшла до редколегії 14.03.22

Рекомендована до друку 14.04.22



Creative Commons Attribution 4.0 International License (CC BY 4.0)