

Шифр «Нейромережа»

**Лексичні засоби увиразнення авторського стилю
журналістів (аналіз із використанням методів корпусної
лінгвістики й штучних нейронних мереж)**

Зміст

Вступ	3
Розділ 1. Переваги використання методів корпусної лінгвістики й штучних нейронних мереж у дослідженні мови медіатекстів	6
Розділ 2. Лексичні особливості публіцистичних текстів Ю. Макарова та О. Покальчука: корпусний аналіз	10
Розділ 3. Порівняльний аналіз лексики текстів Ю. Макарова та О. Покальчука за допомогою штучних нейронних мереж	18
Висновки	26
Список використаної літератури	28

Вступ

У сучасній українській журналістиці є чимало яскравих постатей, які активно працюють для розвитку улюбленої справи. Талановиті журналісти мають індивідуальний стиль – неповторний, який вирізняє їх з-поміж інших колег. Одним із найбільш дієвих елементів увиразнення авторського стилю є лексика, адже публіцисти вдаються до вживання специфічної лексики, часто стилістично обмеженої, аби увиразнити матеріал.

Мова сучасних мас-медіа формується під впливом суспільно-політичних подій, які провокують появу нових лексем. Мовлення виступає засобом інтерпретації власної думки щодо конкретних явищ та подій. Тож на мові постійно позначаються різноманітні зміни в державі. У мовленні сучасних мас-медіа помітними є тенденції, пов'язані з демократизацією та лібералізацією суспільства [1].

Цікавим є дослідження авторської стилістики – однієї із новаторських дисциплін у фаховій підготовці журналістів, яка є частиною загальної стилістики української мови. Професійним спікерам варто усвідомлювати, чим для них є мова, бо це значна частина діяльності.

Найбільш цікавим, на наш погляд, емпіричним матеріалом для дослідження є мовлення відомих публіцистів, адже в їхній творчості найбільш виразно виявляються не тільки загальні мовностильові тенденції, а й простежуються індивідуальні особливості стилю. З огляду на це ми обрали для дослідження авторських стилів, їх порівняння авторів одних із найкращих українських періодичних видань – «Українського тижня» (Юрія Макарова) і «Дзеркала тижня» (Олега Покальчука).

Актуальність нашого дослідження зумовлена необхідністю виявлення індивідуальних мовних рис (ідіостилю) у статтях відомих сучасних публіцистів Юрія Макарова та Олега Покальчука. Актуальність нашої роботи зростає ще й з огляду на те, що дотепер нам не відомі дослідження їхнього ідіостилю, тим

більше проведені з використанням методів корпусної лінгвістики та штучного інтелекту.

Досліджуючи особливості лексики в публікаціях зазначених авторів, можемо виділити такі **аспекти**: функціональний, якісний, індикатор авторського осмислення, формально-змістовий.

Функціональний аспект полягає в особливостях зв'язку у тексті. Якісний аспект указує на теоретичний матеріал, що використовуватиметься нами під час дослідження і є допоміжним у розкритті теми. Індикатор авторського осмислення полягає у стандартності висловлювань, використання специфічних лексем для вираження авторської думки. Формально-змістовий аспект дає змогу дослідити лексико-граматичну структуру корпусу.

Новизна нашого дослідження зумовлена використанням новітніх підходів до вивчення ідіостилю журналістів (методів корпусної лінгвістики та штучних нейронних мереж) і полягає у виявленні й аналізі основних маркерів ідіостилю Юрія Макарова та Олега Покальчука.

Об'єктом нашого дослідження є лексика публікацій Олега Покальчука та Юрія Макарова.

Предметом нашого дослідження є лексичні засоби увиразнення авторського стилю публікацій Олега Покальчука та Юрія Макарова.

Джерелами фактичного матеріалу є статті Юрія Макарова у виданні «Український тиждень» та Олега Покальчука у виданні «Дзеркало тижня. Україна».

Мета – проаналізувати лексичні особливості як виразник авторського стилю публікацій Юрія Макарова, Олега Покальчука.

Для реалізації мети необхідно виконати такі **завдання**:

- окреслити можливості використання методів корпусної лінгвістики та штучних нейронних мереж для вивчення ідіостилю журналіста;
- проаналізувати лексичне наповнення публікацій Юрія Макарова, Олега Покальчука;

- порівняти результати аналізу, реалізованого за допомогою програми Tlex та за допомогою штучного інтелекту;
- зробити порівняльну характеристику ідіостилів досліджуваних журналістів.

Методи, за допомогою яких здійснено наукове дослідження:

- метод корпусної лінгвістики, який допоможе виявити найчастіше уживані лексеми;
- метод аналізу, який допоможе виокремити суттєве і несуттєве у досліджуваних журналістських текстах;
- метод моделювання, що дасть змогу більш поглиблено дослідити об'єкт наукового інтересу;
- історичний метод допоможе вивчити розвиток об'єкта, процеси його формування.

Практичне значення бакалаврської роботи полягає у можливостях наукового використання результатів дослідження або у рекомендаціях щодо їх використання.

Розділ 1. Переваги використання методів корпусної лінгвістики й штучних нейронних мереж у дослідженні мови медіатекстів

Сучасний світ рухається під впливом безперервних потоків інформації. Це зумовлює необхідність її автоматичного опрацювання. Високий рівень комп'ютеризації має вплив не лише на технічні науки, але й зокрема на лінгвістику. Сьогодні є доступними способи опрацювання, збереження та відбору мовного матеріалу. Можливість роботи з лінгвістичними корпусами вимагає великої кількості емпіричного матеріалу та залучення комп'ютерних технологій для отримання об'єктивних висновків. Результат корпусного аналізу провокує появу нових тверджень у галузі мовознавства.

Методи корпусної лінгвістики посідають вагоме місце у сучасному мовознавстві. Вони дають змогу визначити принципи побудови, обробки та використання тексту за допомогою сучасних комп'ютерних технологій.

Корпусний аналіз має цілу низку особливостей:

- емпіричний підхід до аналізу мовних даних (дослідження реальних моделей мовної реалізації);
- використання великих за обсягом структурних корпусів (текстів);
- широке залучення комп'ютерних технологій;
- залучення квалікативних і квантитативних аналітичних методик (визначення найбільш уживаних лінгвістичних одиниць та інше).[2]

Сучасні корпусні дослідження проводяться у два головні способи.

Перший напрям базується на розробці проблем, що стосуються аналізу корпусу, тобто концептуалізації: типологія корпусу, його об'єм, параметри, структурування та ін.

Другий напрям засновано на власне лінгвістичних дослідженнях корпусу, за допомогою яких можливе вивчення мови.[2]

Завдання корпусної лінгвістики полягає у повному й системному відображенні змістовного спілкування мовою. Важливим є також підхід до розгляду проблем лінгвістики у процесі спілкування.

Мета корпусної лінгвістики полягає в об'єктивному описі мовної системи у процесі певної комунікативної ситуації. [2]

Корпусна лінгвістика є лише емпіричною дисципліною, яка дає змогу проводити дослідницьку діяльність. Ця наука водночас є невід'ємною з аналізом лінгвального матеріалу, що дозволяє більш поглиблено вивчати особливості мови. Завдяки застосуванню новітніх технологій, а саме комп'ютера, величезний обсяг мовного матеріалу обробляється з надзвичайною швидкістю. Лінгвіст має змогу отримати більш точні та ґрунтовні висновки, на основі яких може висувати тези, припущення. [2]

Незважаючи на високий рівень популярності та довіри, статус корпусної лінгвістики досі залишається неоднозначним. Дехто з лінгвістів, зокрема МакЕнрі, Е. Вільсон, Р. Ксіао та інші вважають її лише особливою методологією дослідження мовного матеріалу, на противагу фонетики, синтаксису, семантики, граматики. Усі ці галузі вивчення мови досліджують лише окремий її аспект, натомість у корпусній лінгвістиці такого обмеження немає.

Проте більшість вітчизняних та зарубіжних лінгвістів не дотримуються такої точки зору. Вони вважають, що корпусна лінгвістика є самостійною наукою у прикладному мовознавстві. Російські лінгвісти, а саме В.П. Захаров та С.Ю. Богданова, стверджують, що корпусна лінгвістика володіє двома особливостями, які роблять її відмінною з-поміж усіх інших наук: характер аналізованого словесного матеріалу; специфіка інструментарію.

Оскільки корпусні дослідження проводяться із застосуванням комп'ютерів, то цю науку часто утотожнюють зі ще однією дисципліною, а саме комп'ютерною лінгвістикою. Незважаючи на схожість, ці поняття не є ідентичними.

Корпусна лінгвістика оптимізує епістимічну функцію, що пов'язана зі збереженням і передачею інформації. Для цієї науки використання комп'ютера не є визначальним аспектом. Корпусна лінгвістика не моделює функціонування мови в різних ситуаціях, а лише фіксує всі аспекти використання мови.

Робота з методами корпусної лінгвістики передбачає такі кроки:

- ідентифікація мовних даних за допомогою аналізу;
- співвідношення мовних даних за допомогою статистичних методів;
- інтелектуальна інтерпритація результатів.

Безперечно, перші два кроки потребують використання комп'ютера. Натомість третій можливий лише завдяки залученню людських розумових здібностей [Жуковська].

Медіалінгвісти використовують різні програми для роботи з корпусами текстів, серед яких найбільш відомими є LEXA, MonoConc, MicroConcord, TACT, WordSmith, WordCruncher, Manatee [Bonito], IMS Corpus Workbench [CQP], XAIRA, LEXA, Virtual Corpus Manager[VMC], EXMARaLDA Corpus-Manager[Co-Ma].

Для нашого дослідження ми обрали програму «**TLex**». Розробник цієї програми – компанія TshwaneDJe Human Language Technology. Як зазначено на офіційному сайті цієї компанії (він доступний за адресою <https://tshwanedje.com>), вона була створена 2002 року двома співзасновниками – професором-лексикографом Гіллес-Моріс де Шривер (в оригінальному написанні – Gilles-Maurice de Schryver) та програмістом Девідом Джоффе (David Joffe) у Південно-Африканській Республіці.

Спочатку офіс компанії був розташований у місті Преторія (столиця Південно-Африканської Республіки), традиційна африканська назва якого – *Tshwane*, саме від неї й походить назва компанії.

«**TLex**» – це професійна програма, що використовується для укладання термінологічних словників. Програма складається з чотирьох основних модулів: «**TLex Suite**», «**tlTerm**», «**tlCorpus**» та «**tl 2013**». У нашому дослідженні ми використовуємо модуль програми – «**tlCorpus**».

Використання методів корпусної лінгвістики та штучних нейронних мереж для вивчення мови медіатекстів дозволяє дослідникам суттєво зекономити час і зусилля, які використовуються на реалізацію рутинних

процедур: розписування текстів на окремі лексичні елементи (з контекстами й паспортизацією), сортування лексем тощо.

Розділ 2. Лексичні особливості публіцистичних текстів Ю. Макарова та О. Покальчука: корпусний аналіз

Проаналізувавши за допомогою методів корпусної лінгвістики 50 статей **Юрія Макарова**, ми виявили найбільш уживані лексеми. Це дало змогу більш поглиблено дослідити особливості стилю автора. Ми зауважили, що найчастіше Юрій Макаров вживає службові частини мови, і це цілком логічно: службових слів є порівняно небагато, вони часто повторюються в текстах. Ці лексеми самі собою зазвичай не є виразниками авторського стилю, однак статистичні дані щодо їх вживання (на матеріалі великих корпусів текстів) у поєднанні з аналізом повнозначних слів зазвичай дають змогу окреслити індивідуальні особливості стилю журналіста.

Лексема, до якої журналіст найчастіше звертається – частка **«не»**. Лише у 50 статтях Юрій Макаров ужив її 1605 разів. Це вказує на часті сумніви автора, які він висловлює у текстах. Очевидно, журналіст нерідко стикається з невпевненістю, сумнівами. Юрій Макаров також протиставляє певні тези, факти. Прикладом цього може слугувати речення зі статті *«Електронний вимикач довіри»*, у якому частку «не» вжито саме для протиставлення: *«У таких серйозних речах, як правило, визначаються не з допомогою дилетантів, а звертаються до так званих think tanks (мозкових центрів – зазвичай це якісь фонди, інститути, консалтингові компанії) ...»*.

Прийменник **«в»** вжито 1142 рази, а прийменник **«на»** - 1023 рази. Вживання цієї частини мови вказує на різноманітні чинники, зокрема часте використання іменників. Прийменник **«у»** вжито 585 разів. Отже, Юрій Макаров часто починає речення лексемою **«У»**.

Не менш активно журналіст використовує сполучники, найчастіше сурядні, а саме єднальні. До прикладу: **«й»** вжито 843 рази, **«і»** - 815. Не зважаючи на відсутність лексичного значення, на те, що сполучники не є членами речення та не відповідають на запитання, саме вони слугують для послідовного з'єднання слів та милозвучності у мові. Отже, відіграють вагоме значення у побудові речень.

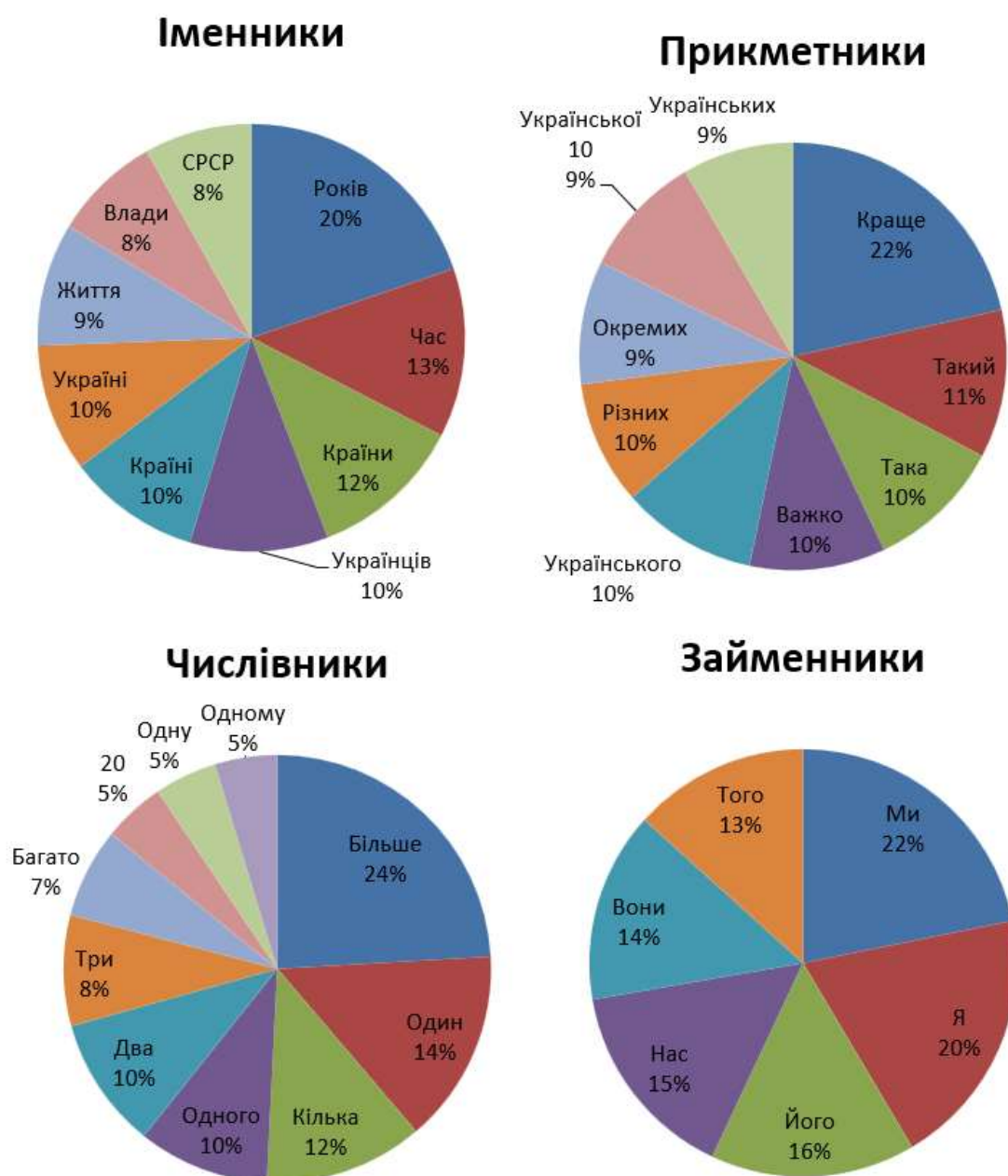
Сполучний засіб «що» вжито 826 разів. Він вказує на уточнення автором нюансів, роз'яснення певних тез. До прикладу: у статті «Камінь на шийі» ця лексема використана саме для деталізації: *«А сентиментальні дрібниці, як-от виконання обов'язку перед громадою та майбутнє країни, визнаються такими, що заздалегідь позбавлені сенсу»*.

Цікавим є використання сполучника «але», який вживається у текстах 106 разів. Така лексема вказує на протиставлення або ж зміну думки автора у процесі написання журналістського твору чи зіставлення суперечливих думок: «То українці все-таки ксенофоби чи ні? Кажу відразу: так. Але не всі». Використання сполучника у такому контексті є прикладом наполегливої та рішучої натури автора. Отже, Юрій Макаров – особистість із яскраво вираженою позицією, яку він не боїться розкривати у статтях.

Цікавим є використання автором займенника «я», загалом – 160 разів. Це є яскравим прикладом того, що Юрій Макаров не завжди дотримується правила «золотої середини» у статтях, але доволі часто висловлює власну думку, іноді надто рішуче. *«Я маю один інструмент вимірювання градусу напруження в суспільстві, яким не володіють жодні соціологи: просто він дуже суб'єктивний і не квантифікується. Це поведінка водіїв у щільному потоці машин»* (стаття «Коротке дихання»). Юрій Макаров – один із провідних журналістів України, тому його думку щодо певної теми зазвичай цікаво дізнатися. Безперечно, у його статтях простежується яскраво виражена авторська позиція.

Часто Юрій Макаров уводить у контекст займенник «я» для того, щоб «підживити» статтю фактами з власного життя. «Я досі так і не був у «Межигір'ї», якось не склалося» (стаття «Знайомство по горизонталі»). Цікаво, що журналіст 75 разів починає статті саме цим займенником. Можна зробити висновок, що Юрій Макаров іноді підкреслює увагу читача на власній персоні. Така практика є притаманною для аналітичної журналістики, в дусі якої пише Юрій Макаров.

Безперечно, не рідше журналіст використовує самостійні частини мови, оскільки саме вони складають вагому частину речень. Аналізуючи іменники, ми зауважили, що в публікаціях Юрія Макарова переважає суспільно-політична лексика. Лексема «років» вжита 76 разів, а до прикладу «країни» – 64 рази. «Нові праві» різного ступеня провокативності й відв'язаності потроху пробують визначати порядок денний і в цитаделях демократії на кшталт Франції з її «Національним фронтом», і в посткомуністичних країнах включно з Угорщиною, Чехією, Польщею». (стаття «Тінь Шпенглера» №31 (455)).



Діагр. 2.1. Частотний аналіз самостійних слів у текстах Ю. Макарова: іменні частини мови

Часте використання в текстах прикметників може свідчити про тяжіння автора до порівнянь та граційних описів. Зокрема найчастіше, а саме 200 разів, вживається лексема «які». Використання цього слова найчастіше вказує на уточнення. «Журналістці тицяють під ніс дозвільні документи, яких нібито немає, на «скандальний» об'єкт, а камера й не дивиться в той бік» (стаття «Хвороба червоних очей» №30 (454)).

Цікавим є вживання Юрієм Макаровим у статтях дієслів. 178 разів журналіст використав лексему «є». Часте вживання цієї лексеми вказує на впевненість автора, доведення певних аргументів. Окрім цього «є» дієсловом, що використовується для опису сучасних подій, явищ (Юрій Макаров найчастіше вживає саме в цьому контексті) . Наприклад: «А вона взагалі є чи її немає? Це я знову про сучасну українську культуру — як у вузькому сенсі, як кажуть, «культурку», себто сучасні художні практики («ВВ» та ансамбль Вірського, Ройтбурд і Гапчинська, Іздрік та Шкляр), так і в загальному, як артикульовану ієрархію цінностей, втілену в корпусі текстів та артефактів.(стаття «Чорний квадрат у чорній скринці №48 (420)).

Вагоме місце посідають також прислівники. Автор нерідко вдається до їх вживання. Зазвичай з метою уточнення, де або ж як відбулася подія. До прикладу: прислівник «коли» вжито 125 разів. «Є такі безвихідні ситуації, коли, наприклад, ти опиняєшся в зоні агресії не дуже врівноваженої, натренованої на людей вівчарки» (стаття «Терпіти не можна наступати» №31 (403)).

Для дослідження ми обрали 50 медіатекстів **Олега Покальчука**, опублікованих в інтернет-ресурсі «Дзеркало тижня». Нами виявлено 21349 лексем. Слід зазначити, що у текстах Олег Покальчук близькі народу теми розкриває крізь призму психології та філософії, що дещо ускладнює розуміння. Разом із тим саме ця особливість вирізняє автора з-поміж інших, зацікавлює читача.

Найчастіше вживаною лексемою з-поміж проаналізованих нами 50 публікацій є сполучник «і» – 1458 разів. *«Америка в її очах хоче захопити весь*

світ, задля наживи готова на все, нав'язує всім свою бездуховність, безкультур'я, примітивізм, усіх зневажає і нічого про них не знає, занепадає інші культури (потрібне підкреслити), бачить світ дуже примітивно — обміняй Америку на Росію, і, о диво!» («Про братерство по розуму», 27.12.2017).

Олег Покальчук часто завдяки одному реченню немовби дає перепочити читачеві, проаналізувати отриману інформацію, дати відповіді на певні запитання. Наприклад: *«І Україна за це проголосувала» («Бібі та стрічечка», 18.05.2018).* У наведеному реченні автор за допомогою сполучника «І» наче підсумовує попереднє висловлення.

Другою найчастіше вживаною лексемою є частка «не». Її Олег Покальчук у аналізованих нами матеріалах вжив 1369 разів. Використовує автор «не» найчастіше для заперечення чогось або протиставлення. Наприклад: *«Неможливо взяти на них один-єдиний правильний акорд для всіх, якщо самі люди цього не захочуть» («На «с» починається, на «ч» закінчується», 19.05.2017), «Не в сенсі розміру е-декларацій, а як особистість» («Не будіть сову», 21.04.2017).*

Доволі часто Олег Покальчук безпосередньо звертається до своїх читачів. Автор встановлює умови: якщо ви не хочете, щоб було так, тоді матимете певні наслідки. Наприклад: *«Не хочете до "злочинного" Мінздор'я — Гугл вам у поміч!» («Синдром навченої втоми», 25.05.2018), «Не шукайте після цього маніяків із потроєною пильністю, краще просто помийте посуд» («Слабка версія», 26.10.2018).* Олег Покальчук ставить у певні умови свого читача дещо різко, проте такий прийом є дієвим. У першому реченні автор схиляє читача звертатися до лікарів, якщо є скарги на здоров'я: *«Про соматичні причини: тут не можна бити в психологічний бубон, а треба йти і здавати аналізи. Не хочете до "злочинного" Мінздор'я — Гугл вам у поміч! Але показники мають бути об'єктивними».*

У тому ж матеріалі Олег Покальчук нібито радить читачам: *«Не слід випускати з уваги дивні погодні рекорди. Вони, як і преса, не демонічні, однак*

теж трішки напружують». Це виражене конструкцією *«Не слід випускати»*, у складі якої є аналізована частка.

Цікавим є використання Покальчуком конструкції *якщо ти не (відома постать)*. Із-поміж проаналізованих нами текстів автор вдається до цього прийому двічі: *«Але що з того, коли добре кальцинований череп скалиться тобі з історичної могили, якщо ти не Гамлет?»* («Кому потрібен інформаційний захист», 05.11.2018), *«Всерйоз критикувати особливості медікомунікації, особливо з боку споживача, — марна річ, якщо ти не Лютер»* («Індульгенція для колаборанта», 31.08.2018). Зрозуміти ці речення неможливо, якщо читач не володіє інформацією про згаданих постатей. Такі речення можуть ускладнити розуміння тексту. З іншого боку це навпаки може змотивувати людину поцікавитися, хто такі Гамлет і Лютер.

Третя найчастіше вживана лексема – прийменник «в» - 1163 рази. Слід зазначити, що Олег Покальчук послідовно дотримується правил евфонії (милозучності), чергуючи у-в. Наприклад: *«Людина представляє певну корпоративну точку зору, наче вона її особиста, отримує гроші від корпорації за роль марнославного лоха, в результаті всі задоволені, всі сміються»* («Індульгенція для колаборанта», 31.08.2018), *«У 1920-х роках позитив ненадовго повернувся»* («Передчуття великого», 21.04.2018).

Іноді Олег Покальчук у своїх медіатекстах згадує Радянський Союз, часом із метою порівняння. Аби вказати час, місце, автор використовує прийменник в (у): *«Православ'я Київського патріархату — це така "демократична платформа в КПРС" (було таке явище в останні роки СРСР)»* («Анатомія священного скелета», 27.04.2018), *«У 20-х роках минулого століття в СРСР ввели таку статтю в Кримінальний кодекс, але тоді було цілком достатньо мати благородне походження або офіцерський чин, щоб тебе розстріляли»* («Що таке саботаж?», 17.08.2018).

Часто Олег Покальчук вживає конструкцію «в Україні» (31 раз): *«Якої насправді в Україні немає»* («Напівмертве море», 12.09.2017). Слід зазначити, що виражальною ознакою авторського стилю Олега Покальчука є короткі

прості речення. Це зумовлено насиченістю текстів автора, їх складністю. Тож читачу необхідна умовна пауза між складними для розуміння реченнями.



Діагр. 2.2. Частотний аналіз самостійних слів у текстах О. Покальчука: іменні частини мови

П'ятою найчастіше вживаною лексемою з-поміж аналізованих нами текстів є «що» (950 разів). Вона може бути як сполучником, такі сполучним словом. Це доволі важлива відмінність, яку необхідно ідентифікувати. «Що вище людина на соціальних щаблях, то більше їй є що приховувати» («Чому приховують інформацію», 13.04.2018).

«Що більше людина себе вважає "простою людиною" (тобто з самого початку відмежовується від влади, обраної такими ж, як вона сама), то важливіше для неї отримати підтвердження власної переваги» («Чому приховують інформацію», 13.04.2018). У цьому реченні виражена ще одна особливість ідіостилю Олега Покальчука: автор застосовує елементи психологічного аналізу. Це також вказує на особливість тематичного наповнення публікацій автора: психологія людини як соціальної істоти.

Також для публікацій Олега Покальчука характерні питальні речення, що спонукають читача до роздумів: *«Що з усім цим можна робити?» («Синдром навченої втоми», 25.05.2018).*

Підсумовуючи наведене, можемо зазначити, що використання методів корпусної лінгвістики допомагає досліднику уникнути рутинних процедур, які займають багато часу й зусиль. Однак використання спеціалізованих програм для роботи з корпусами текстів є дієвим тільки на підготовчому етапі, оскільки надалі дослідникові доводиться аналізувати дані програм «вручну». У перспективі цю частину роботи може допомогти полегшити використання штучних нейронних мереж, тому на цьому аспекті дослідження зупинимося в наступному розділі.

Розділ 3. Порівняльний аналіз лексики текстів Ю. Макарова та О. Покальчука за допомогою штучних нейронних мереж

Стрімкий розвиток технологій машинного навчання уможлиблює розгляд традиційних наукових питань під новим кутом зору. Одним із цікавих і перспективних напрямів є аналіз особливостей природної мови, зокрема й української, за допомогою штучних нейронних мереж (ШНМ). Дослідження значних масивів даних, робота з великими корпусами текстів з використанням комп'ютерних технологій може не тільки спростити лінгвістам рутинні процедури (розписування текстів на картки, вибіркового аналізу мовних одиниць), а й допомогти розвинути нові аспекти й напрями лінгвістичного аналізу.

Кількість текстів, написаних українською мовою, невпинно зростає. Значна частина текстів представлена онлайн, що спрощує доступ до емпіричної бази досліджень. Ці тексти мають різне функціональне призначення, у них використовуються мовні засоби, що відповідають меті й завданням спілкування в певній комунікативній ситуації.

У дослідженні текстів ми використали методику, яку розробив дослідник Максим Лупей. Відповідно до його методики, відібрані тексти проходять подальшу обробку з допомогою програми, написаної на мові програмування C++. Ця програма видаляє непотрібні відступи та порожні частини тексту та формує файл csv, у якому вся інформація розбита і згрупована згідно з наперед заданими параметрами. Саме з цим і працює в подальшому штучна нейронна мережа. Наступним кроком є векторизація, яка є надзвичайно важливою ланкою, і кінцевий результат роботи дуже чутливий до проведеного етапу векторизації вхідних даних. Далі проводиться розподіл текстів на навчальні та тестові. Після поділу текстів із ними працює штучна нейронна мережа, яка навчається та проводить тестові визначення авторів тексту та видає результати. Це все повторюється визначену кількість разів. У кінці відбувається збір результатів експериментів та визначається точність і похибка проведених досліджень [8].

Наведені приклади з результатами роботи штучної нейронної мережі (на ілюстраціях указано тільки незначну частину отриманих даних) узгоджуються з результатами нашого дослідження за допомогою методів корпусної лінгвістики.

Ми визначили такі риси ідіостилю **Юрія Макарова**:

- тяжіння до порівнянь теперішніх часів із минувиною;
- сумніви у висловах;
- протиставлення;
- яскраво виражене авторське «я».

Однією з особливостей авторського стилю Юрія Макарова є гостре викриття вад українського суспільства та критика сучасності. Тому, аналізуючи статті цього журналіста, не можна не зауважити питому вагу суспільно-політичної лексики.

Ядро суспільно-політичної лексики формують особливості епохи. Саме характерні ознаки розвитку кожного історичного періоду впливають на цей шар лексики.

Ще однією не менш важливою авторською особливістю Юрія Макарова є тяжіння до історизмів. Особливо переважає тема впливу Росії на Україну впродовж століть. Окрім цього, проаналізувавши статті, ми зауважили, що Юрій Макаров небайдужий до історії українського народу ХХ століття (варто зауважити, що за національністю Юрій Макаров – болгарин). Така перевага історичної тематики спонукала нас виокремити окрему категорію для виявлення ідіостилю – «історизми».

1. *Царя* – 2 рази.
2. «*Вісла*» (у значенні політичної спецоперації 1947 року) – вжито 1 раз.
3. пацифікації (жорстокий режим Ю. Пілсудського) - вжито 1раз
4. «*Перебудова*» (політика КПРС 1985-1991 років) - вжито 1 раз.

5. *Індустріалізація* (це комплекс заходів з прискореного розвитку промисловості, вжитих ВКП(б) у період другої половини 20-х до кінця 30-х років) – вжито 1 раз.

6. ГУЛАГ (у СРСР у 1934–1956 – х роках підрозділ НКВС, який керував системою виправничо-трудових (офіційна назва у 1920-х роках — концентраційних) таборів) – вжито 1 раз;

7. *Коліївщини* (козацько-селянське національно-визвольне повстання в Правобережній Україні у 1768-1769 роках) – вжито 1 раз;

8. *Лівобережжя* (історично сформована назва частини України) - вжито 1 раз;

10. *Січ* (адміністративний та військовий центр запорозького козацтва) – вжито 1 раз;

Спостерігається використання Юрієм Макаровим розмовної лексики:

*«Я не бачу, не чую поміж громадськості жодних ознак наявності не те що креслень майбутнього, а бодай **завалючого** ескізу»* («Як тебе полюбити, Києве мій»).

*«То був дивом уцілілий фрагмент попереднього, докомуністичного життя — спотворений, занедбаний, **запаскуджений**, але все ж таки співмірний людині»* («Ламати не можна»).

*«Звідти Альфред Рейнгольдовіч кинув гасло до російської опозиції: **гайда** всі на Україну!»* («Ворог мого ворога»)

Використання стилістично маркованих, лайливих слів:

*«**кляті питання відповідальності**»* – підкреслює важкість сприйняття та неминучість цих питань. Клятий *лайл*. – про людину, яку проклинають, засуджують, ненавидять (Словник української мови: в 11 томах. — Том 4, 1973. — Стор. 195.).

*«Один Кісєльов — Дмитрій — голос тих 89% або негідників, **або дебілів**, або конформістів (або все нараз), які активно чи пасивно беруть участь у війні проти моєї країни»* («Ворог мого ворога»).

Юрію Макарову вдається вдало висловлювати іронію, створюючи незвичні мовні конструкції: *«Саме під цими символами — революція-індустріалізація-дєди воєвалі-космос-олівьє — вони вбивають щодня українців і все ще готуються до визвольного походу на Новоросію»* («Ламати не можна»).

*«...вакханалія навколо комунальних тарифів — цирк за межами пристойності, але якби свого часу всілякі там обленерго не були віддані в концесію **власникам-упирям**, а потрапили в руки справжніх ефективних менеджерів, які насамперед зменили б енергозатрати, то не було б ані нинішньої загрози соціального вибуху...»* («Глитай, або ж павук»). Тут «власники-упирі» – висловлення автором зневаги та несприйняття.

Олег Покальчук у своїх журналістських текстах вживає різну лексику, спеціалізовану, зокрема, з галузі психології, філософії тощо. При цьому вживані поняття автор не пояснює, таким чином спонукає розібратися людину.

Цікавими є варіації вживання аналізованим автором однієї лексеми. Це привертає увагу під час прочитання. Розгляньмо інтерпритацію слова «зрада»:

*«Отже, український чорно-білий дискурс **"зрада—перемога"** є одночасно синонімом національної наївності, яка ґрунтується на вірі»* («Жертва хіндса»)

*«На невдоволення адептів **"зрадоперемоги"**, зазначу, що в цифровому світі таких понять взагалі існувати не може»* («Буква цифрі не »)

*«Седньовічний жанр цілком добре почувається в українському (та й, утім, у кожному) інформаційному просторі XXI століття, то нам ну ніяк не обійтися без **зрадопереможної** теми інквізиції»* («Вереск і сміх»)

*«Слід зазначити, що рівень персонального негідництва влади сильно перебільшений **"зрадофілами"**»* («Мишобрати і наші»).

*Інтерес у тому, що ця тема прямо впирається в наші власні **"зрадопереможні"** баталії* («Бібі та стрічечка»).

У своїх медіатекстах Олег Покальчук використовує жаргонізми:

*«...в криміналізації якогось етносу загрожує не тільки втратою цього електорату (статистично це може бути незначна величина), а й подальшим медійним **"зашикваром"**, організованим політичними супротивниками»* («Між кибиткою та «Мерседесом»).

*«Колись ця середньовічна **лафа** закінчиться»* («Анатомія священства»)

*«Можна було б, звичайно, укотре познущатися з профільних міністерств, які незворушно звітують перед суспільством про успіхи в цілком радянському **окозамилувальному стилі**» («Мишобрати і наші»).*

Спостерігаємо вживання Олегом Покальчуком частини «квазі»: квазілідерами, квазістатистики, квазірелігійна, квазіпростір

КВАЗІ... Перша частина складних слів, що значенням відповідає прикметникам несправжній, уявний, напр.: квазівчений, квазіспеціаліст і т. ін.

Словник української мови: в 11 томах. — Том 4, 1973. — С. 128.

Олег Покальчук – автор, який вирізняється емоційністю, сміливістю у висловлюваннях, тож йому притаманне використання емоційно забарвлених слів, навіть згрубілих форм.

*«У сучасному жорстокому світі наївність часто є синонімом **нетямущості, йолопства**» («Жертва хіндса»).*

*«Як називали ми **гадів** гадами, так і називатимемо» («Усе, що нас не вбиває»)*

Нерідко речення у матеріалах Покальчука здаються абсурдними, зрозуміти їх непросто. Наприклад, матеріал «Як правильно вистрелити собі в ногу» (зацікавлює і дивує водночас формулювання заголовку) розпочинається так: *«Де ж у всьому цьому рука Путіна, що стріляє нам у ногу?» — запитає **порохобот**»*. Такий незвичний початок «примушує» читати далі, адже постає питання: що має на увазі автор і чи не з'їхав він із глузду. Далі Олег Покальчук пише про «відчуття колективного безумства» українського суспільства, яке насправді безумством не є. Відтак автор пояснює нам про рецептори: *«По-перше, чим чутливіший рецептор, тим далі він від самої конструкції (кістки, які становлять опорно-руховий апарат, самі по собі нервів не мають, та й у місцях тертя використовують природні наркотики організму). І хоча все взаємопов'язане, щирі претензії невротиків на інтуїцію можуть лише викликати жаль, та аж ніяк не стати предметом прогностичного аналізу», а ще далі згадує про істероїдну поведінку — «одну з форм когнітивного самозахисту організму».*

Це скидається на матеріал для спеціалізованого з психології видання. Насправді ж цей приклад демонструє важливу особливість ідіостилу Олега Покальчука – використання психології для пояснення процесів, які відбуваються в суспільстві.

Слід зауважити, що Олег Покальчук часто згадує події, які відбувалися раніше. Наприклад: *«Формально діяло перемир'я, а в реальності жодна зі сторін його не виконувала, аж поки 1969-го Саддам Хусейн (так-так, той самий) "Березневим маніфестом" гарантував право курдів на самовизначення»* («Про братерство»); *«На їхньому рахунку, зокрема, вбивства британського міністра Уолтера Гіннесса в 1944-му, посередника ООН графа Фольке Бернадота у 1948-му»* («Бібі та стрічечка»); *«Історію з сектою "Білого Братства" 1993-го, спробу масового самогубства при пророкуванні кінця світу 24 листопада пам'ятаєте?»* («Біле братство»).

Лексеми, за якими штучний інтелект визначає авторство

Юрія Макарова

Лексема	У матеріалах Макарова	У матеріалах Покальчука
Мене	39	11 (0,01%)
Президент	5	6 (0,01%)
Повідомлення	1	3
Кожні	1	1
Побутові	0	1
Пам'ять	0	8 (0,01%)
Не	861	1438 (1,92%)
Добрий	0	3
Останній	5	3

Як ми вже зазначали, для матеріалів Юрія Макарова характерне використання займанника «я» та похідного від нього, зворотного «себе» (як виділяє програма). Притаманним є і використання частки «не».

*Лексеми, за якими штучний інтелект визначає авторство
Покальчука*

Лексема	У матеріалах Покальчука	У матеріалах Макарова
Дуже	180 (0,25)	20
Політичний	14	2
Етнічний	1	0
Соціальний	9	4
Стрімкий	1	0
Люди	92	19
Можуть	59	10
Суспільство	66	7
Злочинність	14	0

Олег Покальчук – соціальний психолог, тож примітними є виокремлені програмою лексеми: «соціальний», «суспільство», – це абсолютно визначає змістове наповнення журналістських матеріалів автора.

Отже, за допомогою штучної нейронної мережі й завдяки статистичному аналізу з використанням програм для роботи з корпусами текстів визначено типові лексичні риси індивідуальних стилів досліджуваних авторів. Схожість результатів, отриманих за допомогою обох методів, може свідчити про значний рівень їх достовірності. Однак обраний напрям дослідження ще потребує подальшого вивчення, розвитку, поглиблення.

Висновки

Мова сучасних мас-медіа формується під впливом суспільно-політичних подій, які зумовлюють появу нових лексем та актуалізацію наявних. Особливо відчутною до змін у державі є мова засобів масової інформації. У мовленні сучасних мас-медіа сильними є тенденції, пов'язані з демократизацією та лібералізацією суспільства. Ці тенденції особливо простежуються в мовленні провідних українських публіцистів, до яких належать і Ю.Макаров та О.Покальчук, медіатексти яких ми проаналізували з погляду увиразнення лексичних елементів.

Корпусне дослідження виконане за допомогою програми «Tlex», яка дозволяє визначити частотність вжитих лексем. Відтак, проаналізувавши контекст, ми зробили наступні висновки щодо особливостей стилів аналізованих авторів.

Однією з особливостей авторського стилю Юрія Макарова є гостре викриття вад українського суспільства та критика сучасності. Тому, аналізуючи статті цього журналіста, не можна не зауважити питому вагу суспільно-політичної лексики.

Ще однією не менш важливою авторською особливістю Юрія Макарова є тяжіння до історизмів. Особливо переважає тема впливу Росії на Україну впродовж століть. Окрім цього, проаналізувавши статті, ми зауважили, що Юрій Макаров небайдужий до історії українського народу ХХ століття (варто зауважити, що за національністю Юрій Макаров – болгарин). Така перевага історичної тематики спонукала нас виокремити окрему категорію для виявлення ідіостилу – «історизми». Спостерігається також і використання Юрієм Макаровим розмовної лексики.

Олег Покальчук у своїх журналістських текстах вживає різну лексику, спеціалізовану, зокрема, з галузі психології, філософії тощо. При цьому вживані поняття автор не пояснює, таким чином спонукає розібратися людину.

Олег Покальчук – автор, який вирізняється емоційністю, сміливістю у висловлюваннях, тож йому притаманне використання емоційно забарвлених слів, навіть згрубілих форм.

Наступним аспектом нашого дослідження стало використання штучного інтелекту: програма демонструє лексеми, за якими визначає авторство того чи іншого журналіста. Такі результати збіглися з отриманими раніше завдяки програмі Tlex.

Серед перспектив дослідження варто виділити доцільність застосування апробованих у нашій роботі технологій (методів корпусної лінгвістики та штучних нейронних мереж) для вивчення мовлення ширшого кола українських публіцистів із використанням більших корпусів текстів.

Список використаної літератури

1. Дубравська Д. М. Внесок корпусної лінгвістики у сучасну лексикографію. *Наукові записки [Національного університету "Острозька академія"]*. Сер. : Філологічна. 2012. Вип. 27. С. 13-16. Режим доступу: http://nbuv.gov.ua/UJRN/Nznuoaf_2012_27_5
2. Жуковська В.В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир: Вид-во ЖДУ ім. І. Франка, 2013. 142 с.
3. Бабоч Н.Д. Практична стилістика і культура української мови. Л.: Світ, 2003. 432 с.
4. Єрмоленко С. Нариси з української словесності: стилістика та культура мови. К.: Довіра, 1999 .
5. Навальна М.І. Динаміка лексикону української періодики початку ХХІ ст. : [монографія]. К. : Інститут української мови ; Видавничий дім Дмитра Бураго, 2011. 328 с.
6. Панченко С. А. Українська мова в засобах масової комунікації: Навчально-методичний посібник. З.: Запорізький національний технічний університет, 2009. 103 с.
7. Переломова О. С. Авторська стилістика. Суми: СумДУ, 2014. 108 с.
8. Lupei Maksym, Mitsa Alexander, Repariuk Volodymyr, Sharkan Vasyl. Identification of authorship of Ukrainian-language texts of journalistic style using neural networks. *Eastern-European Journal of Enterprise Technologies*. ISSN 1729-3774. 2020. № 1/2 (103). P. 30-36.

АНОТАЦІЯ

наукової роботи «Нейромережа»

У роботі проаналізовані особливості лексики публіцистичних текстів Юрія Макарова та Олега Покальчука з використанням методів корпусної лінгвістики й штучних нейронних мереж.

Мета дослідження – проаналізувати лексичні особливості як виразник авторського стилю публікацій Ю. Макарова, О. Покальчука. Для реалізації мети окреслено можливості використання методів корпусної лінгвістики та штучних нейронних мереж для вивчення ідіостилу журналістів; проаналізовано лексичне наповнення публікацій Ю. Макарова, О. Покальчука; зроблено порівняння результатів аналізу, реалізованого за допомогою програми Tlex та за допомогою штучного інтелекту; наведено порівняльну характеристику ідіостилів досліджуваних журналістів.

Новизна дослідження зумовлена використанням новітніх підходів до вивчення ідіостилу журналістів – методів корпусної лінгвістики та штучних нейронних мереж. Крім зазначених методів, у роботі використано також методи аналізу, моделювання, історичний метод.

Загальна характеристика наукової роботи. Робота містить вступ, три розділи, висновки, список використаної літератури. Кількість сторінок – 28, кількість діаграм – 8, ілюстрацій – 2, таблиць – 2.

Ключові слова: ідіостиль, лексика, корпусна лінгвістика, штучна нейронна мережа.