

DOI: [10.18372/2410-7840.22.14702](https://doi.org/10.18372/2410-7840.22.14702)  
УДК 004.56

## МЕТОД КОМП'ЮТЕРИЗОВАНОГО ФОРМУВАННЯ АНГЛОМОВНОГО ТЕКСТУ ВІДПОВІДНО ДО ПСИХОЛІНГВІСТИЧНОГО ПОРТРЕТУ ПРОПАГАНДИСТА

*Ярослав Тарасенко*

*Поточна ситуація в державі та світі вимагає розробки надійних методів виявлення та нейтралізації кіберзагроз, однією з яких є інформаційна пропаганда в медіа джерелах. Для протидії подібній загрозі необхідно здійснювати зворотний вплив на пропагандиста, що може бути реалізовано на основі комп'ютерно згенерованого осмисленого тексту за психолінгвістичним портретом пропагандиста. Переважна більшість сучасних методів комп'ютеризованої генерації текстів базується на технологіях машинного навчання, що обмежує можливість ефективної побудови психолінгвістичного профілю за рахунок потреби великої кількості досліджуваної інформації, що обумовлює ризик виявлення протидії пропаганді. Інші ж методи не враховують категоріальність та квантову природу сприйняття, або занадто вимогливі до кількості вхідних даних для побудови контекстуальних зв'язків чи не враховують пропагандний дискурс. Розроблено метод комп'ютеризованого формування англomовного тексту відповідно до психолінгвістичного портрету пропагандиста на основі багаторівневих шаблонів та вдосконаленої методики лісу даних природної мови, шляхом використання методів квантової семантики та семантики Монтегю, що дозволяє підвищити ефективність протидії негативному впливу інформаційної пропаганди в англomовних медіа джерелах завдяки врахуванню індивідуальної суб'єктивної семантичної лінії пропагандиста та імовірних квантових станів лексико-семантичних одиниць, ґрунтуючись на класифікації семантичної частки. Метод дає змогу генерації англomовного тексту відповідно до психолінгвістичного портрету пропагандиста для подальшого впровадження в нього засобів зворотного цільового впливу на пропагандиста. Використання методу важливе при дослідженні проблеми зсуву семантичної лінії тексту в напрямку суб'єктивної семантичної лінії пропагандиста.*

**Ключові слова:** генерація тексту, комп'ютеризована компоновка тексту, протидія пропаганді, категорії семантичної частки, психолінгвістичний портрет пропагандиста, семантична частка в англomовному тексті, категоріальні рівні, квантові стани лексико-семантичної частки.

### ВСТУП

На поточному етапі розвитку держави гостро постає проблема проведення інформаційного протиборства, що зумовлено низкою факторів, до яких, перш за все, відноситься спроба злочинних угруповань впливати на вільне волевиявлення громадян, обмежувати конституційні права і свободи, породжувати ворожнечу і примушування до протиправних дій як на національному так і на міжнародному рівні. Згідно із Стратегією кібербезпеки України, затвердженої указом Президента України від 15 березня 2016 року №96/2016 [18], забезпечення безпеки кіберпростору і його використання в інтересах суспільства та держави є головною задачею. Таким чином, інформаційний вплив на свідомість громадян за допомогою текстової інформації, розміщеної в мережі Інтернет чи будь-яких ЗМІ чи медіа [6] в тому числі міжнародних з ознаками інформаційної пропаганди повинен бути знешкоджений. Це корелює із необхідністю забезпечення життєво важливих інтересів громадянина та держави, зазначеної у Стратегії, де також визначається один з пріоритетних напрямків забезпечення кібербезпеки України, що полягає у створенні системи виявлення та нейтралізації кіберзагроз [18]. Звідси слідує актуальність дослідження,

направленого на підвищення ефективності протидії маніпулятивному впливу ворожої пропаганди в умовах загрози інформаційної війни [1]. А той факт, що англійська мова є основною в світовій практиці використання медіа [2] змушує розглядати саме англomовні тексти як основу.

Для отримання можливості здійснення зворотного впливу на пропагандиста, необхідно вирішити проблему комп'ютеризованого формування англomовного тексту враховуючи особливості суб'єктивного сприйняття семантики тексту пропагандистом. Існуючі сучасні методи та підходи комп'ютеризованої генерації осмислених текстів [3, 8-10, 12] володіють своїми особливостями та слабкими сторонами, а отже не кожен підхід може бути використано, як підґрунтя для виконання поставленої в роботі задачі. Так, багато методик ґрунтуються на використанні нейромереж для генерації тексту, як в роботах [8-10], однак, хоча такі засоби та інформаційні системи на їх основі і можуть створити текст схожий за своєю структурою на авторський, проте врахувати його психолінгвістичний портрет для них проблематично. Методи, які не використовують нейромереж [3, 12] слід значно вдосконалювати в напрямку врахування категоріальності мислення, що вимагає значно більше часу та зумовлює непередбачуваний

результат. Підходи семантичного пошуку та семантичні мережі [17] занадто вимогливі до кількості вхідних даних для побудови контекстуальних зв'язків та потребують велику кількість вхідної інформації про пропагандиста, що неприпустимо в даній ситуації, оскільки може викликати підозру.

Для поставленої задачі зручно підходять методи [14, 15], хоча вони і потребують вдосконалення в напрямку врахування індивідуальної семантики пропагандиста. Фундаментальні методи породжувальної граматики [5] та граматики складання дерев [11] можуть застосовуватись в процесі генерації тексту.

### ПОСТАНОВКА ЗАДАЧІ

Метою роботи є підвищення ефективності нейтралізації маніпулятивного впливу інформаційної пропаганди шляхом забезпечення підґрунтя для здійснення зворотного впливу на пропагандиста на основі методу комп'ютеризованого формування англомовного тексту відповідно до психолінгвістичного портрету пропагандиста.

Для вирішення задачі були поставлені такі завдання:

1. Удосконалення існуючих методів генерації осмислених текстів для врахування індивідуальної семантики пропагандиста.

2. Впровадження у формований текст понятійного простору синтаксичних одиниць психолінгвістичного портрету пропагандиста з урахуванням їх характерних квантових станів.

3. Адаптація підходу генерації семантичної складової при комп'ютеризованому формуванні англомовного тексту в напрямку використання ядра семантики психолінгвістичного портрету пропагандиста.

### МЕТОДИКА

Метод квантово-семантичного психолінгвістичного аналізу англомовного тексту пропагандного дискурсу [20] та контент-аналізу семантичної частки в текстах з ознаками психолінгвістичного впливу [21] використовуються для отримання базових вхідних даних при генерації тексту. Методи квантової лінгвістики в інформаційному протиборстві [19] застосовуються для вдосконалення існуючої методики генерації осмислених англомовних текстів. Для генерації тексту використовуються методи породжувальної граматики [5], граматики складання дерев [11], методи варіаційних шаблонів [7], дворівневих шаблонів [15] та методи генерації тексту на основі лісу даних природної мови [14]. Формальна семантика та інтенціональна логіка [16] застосовується для забезпечення семантичної єдності тексту відповідно до структурних

особливостей англійської мови [4]. Методи композиції в журналістському тексті [13] використовуються для забезпечення стильової приналежності до медіа тексту. Використана термінологія ґрунтується на дослідженнях [19-21] та наукових джерелах, на які вони опираються.

### ВИРІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

Як уже згадувалось, існуючі сучасні методи генерації осмисленого тексту природної мови в більшості випадків ґрунтуються на нейромережах. Однак, особливості використання штучного інтелекту вносить свої обмеження, до яких відноситься час навчання нейромережі та необхідний матеріал, що не відповідає обґрунтованому в [19] використанню методу опитування інформанта. Згаданий метод передбачає отримання обмеженої кількості відповідей від пропагандиста, щоб не привернути до процедури протидії увагу, а це, в свою чергу, критично знижує ефективність використання нейромереж. Однак, англійська мова є формалізована, тому потреби в навчанні та побудові моделі мовлення конкретного індивіда немає, як це вимагається при комп'ютеризованій генерації тексту, використовуючи, як основу флективні мови. Подібний стан речей дозволяє використання методів генерації тексту на основі шаблонів, на кшталт [7]. При цьому, виникає потреба удосконалення подібних підходів на основі методів квантової лінгвістики для отримання можливості в якості основи для шаблонів використовувати категоризаційні лексико-семантичні особливості психолінгвістичного портрету пропагандиста, змодельованого за методом [20].

Таким чином, метод комп'ютеризованого формування англомовного тексту відповідно до психолінгвістичного портрету пропагандиста складається з чотирьох етапів: підбір лексико-семантичної основи; формування синтаксичних одиниць; забезпечення семантичної єдності та осмисленості; стилістичне коригування.

#### Підбір лексико-семантичної основи

Для побудови матриці шаблонних лексико-семантичних одиниць необхідно врахувати окремі структурні одиниці семантичної частки психолінгвістичного портрету пропагандиста і сформувати таблицю лексико-семантичних одиниць, згрупованих за граматичними ознаками відповідної частини мови у суворій послідовності частин речення, у зв'язку із сталим порядком слів в англійській мові.

Метод квантово-семантичного психолінгвістичного аналізу англомовного тексту пропагандного дискурсу [20] дає можливість проаналізувати

текст, визначити маніпулятивні ознаки та виявити лексико-семантичні одиниці, на які найбільш виразна реакція зловмисника з огляду на його психологічні особливості. Таким чином, нехай  $r_n$  – лексико-семантична одиниця, на яку був відгук при прихованому опитуванні пропагандиста. При цьому, згідно [20, 21] кожен елемент множини  $r$  буде відноситись до одного з категоріальних рівнів психолінгвістичного портрету. Це означає, що будується категоріальне дерево для лексико-семантичної одиниці. Для поточної задачі, на основі побудованого дерева необхідно знайти його крону, яка в даному випадку виступає множиною вершин найнижчої категорії для кожної лексико-семантичної одиниці, в якій простежується відбиток психолінгвістичного портрету пропагандиста. Пошук крони можна описати формулою 1:

$$B = \{ \forall B_n \in B_n^m \mid B_n^m \in r \}, \quad (1)$$

де  $B_n^m$  – множина найнижчих категорій ядра семантики,  $m$  – глибина дерева,  $n$  – кількість вершин при всіх множинах синонімів  $B_n$  кожної категорії  $m$ . При цьому, формули в статті пов'язані одна з одною та витікають з попередніх, тому відповідні змінні мають сталі пояснення в рамках статті.

Отримана множина повинна бути розсортована по частинам мови та частинам речення та сформована таблиця, яка власне і буде лексико-семантичним матеріалом для формування тексту із окремих речень. Однак, перш за все, необхідно виділити синонімічні конструкції найнижчої категорії ядра семантики за індивідуальною семантичною лінією пропагандиста за принципом видалення синонімічних структур, які не володіють відбитком психолінгвістичного портрету (2):

$$T = \frac{\prod_{i=0}^d \sum_{j=0}^k b^j_i}{\prod_{i=0}^d \sum_{j=0}^k (b^j \in B \mid G^{b^j} < 0,5)_i}, \quad (2)$$

де  $\prod_{i=0}^d \sum_{j=0}^k (b^j \in B \mid G^{b^j} < 0,5)_i$  – синонімічні конструкції, які не володіють відбитком психолінгвістичного портрету та потребують видалення;  $b^k$  – підмножина множини найнижчих категорій з відбитком психолінгвістичного портрету пропагандиста  $B$ , елементами якої є синонімічні конструкції підкатегорії найнижчого порядку,  $d$  – кількість самостійних частин мови, що можуть бути використані в англomовному тексті,  $k$  – кількість сино-

німічних лексико-семантичних одиниць підкатегорії найнижчого порядку,  $G^{b^k}$  – модифікований коефіцієнт Джині, розрахований за (3).

Коефіцієнт Джині повинен бути модифікований в напрямку об'єднання економічних властивостей та особливостей машинного навчання шляхом застосування принципів квантової лінгвістики в інформаційному протистові [19], оскільки в даному конкретному випадку мова йдеться про генерацію тексту природної мови, що змушує враховувати одночасно як соціальну понятійну систему, так і можливості комп'ютеризованої генерації тексту. Однак на відміну від традиційного використання коефіцієнту, де прийнято вважати, що коефіцієнт описує ефективність обернено пропорційно його значенню, то ефективність в даному випадку має прямо пропорційний ефект, що зумовлений віддаленістю від ідеальної системи, яка в даному випадку виступає об'єктивною семантичною лінією, і наближеністю до суб'єктивної семантичної лінії саме пропагандиста. Формалізація процесу визначення функцій, що описують об'єктивну та суб'єктивну семантичну лінію потребує подальшого дослідження, проте в загальному випадку спосіб їх визначення не впливає на роботу спроможність методу:

$$G^{b^k} = 1 - \frac{\int_0^h f(x'')dx - \int_0^h f(x')dx - \int_0^h f(x)dx}{\int_0^h f(x')dx}, \quad (3)$$

де  $f(x)$  – функція, що описує об'єктивну семантичну лінію,  $f(x')$  – описує суб'єктивну семантичну лінію, властиву пропагандисту,  $f(x'')$  – функція, отримана моделюванням семантичної лінії з використанням множини категоріальних лексико-семантичних одиниць. При тому  $h$  – межа завершення тексту, а елементи  $x, x', x''$  множин  $X, X', X''$ , що є областями семантичної частки для об'єктивного, суб'єктивного та пропагандистського тексту відповідно

При цьому, враховуючи, що  $f(x) \rightarrow 0$ , а  $\int_0^h f(x')dx > \int_0^h f(x)dx$ , тоді  $f(x'') \rightarrow f(x')$ , а звідси слідує логічність твердження, що за умови  $G^{b^k} < 0,5$  така множина повинна бути видалена.

Таким чином, отримана множина  $T$  буде складатися з підмножин  $T_d$ , які відповідають за суми синонімічних конструкцій, відповідно до частин мови, кожна з яких, в свою чергу складається з підмножин рівноправних лексико-семантичних

синонімів найнижчого порядку. Але в даному випадку інтерес представляють підмножини  $T_d$ , як такі, що повинні бути пов'язані за допомогою відповідності (4):

$$T' = (T_1 \times T_2) \cup (T_2 \times T_3) \cup \dots \cup (T_{d-1} \times T_d). \quad (4)$$

Після цього, отриману множину необхідно розташувати в таблиці де лексико-семантичні одиниці знаходяться попарно виду (a, b). Таким чином, описана в [7] формула граничної імовірності стосовно відношення однієї частини таблиці з даними до іншої з огляду на вдосконалення в напрямку врахування індивідуальної суб'єктивної семантичної лінії пропагандистського тексту буде виглядати наступним чином (5):

$$\log p_\theta(b | a) = \log \int_{T'} p_\theta(b | a, T', B_n^m = f_{enc}(a)) p(T') dT, (a, b) \in T, \quad (5)$$

де  $a$  – лексико-семантична одиниця, що виступає в ролі вказівника,  $b$  – ядро семантики.

Таким чином, при врахуванні індивідуальної семантичної лінії для автоматизованої генерації тексту на етапі побудови шаблону застосовується правило єдності вказівника та ядра семантики, а також лексико-синтаксичні особливості англomовного тексту.

#### Формування синтаксичних одиниць

В основі етапу формування синтаксичної структури при комп'ютеризованій генерації англomовного тексту лежать фундаментальні методи породжувальної граматики [5] та граматики складання дерев [11]. Оскільки було прийнято рішення використовувати метод шаблонів, то для виконання поточної задачі найкращим чином підійде належно модифікований підхід використання дворівневих шаблонів [15], який може реалізовуватись на основі згаданих формальних граматик. У зв'язку з тим фактом, що отримана на першому етапі методу таблиця шаблону може виступати лише основним вихідним матеріалом для формування синтаксичних структур, то можна стверджувати, що одна і та ж частина мови зможе виступати різними частинами речення, а отже від її розміщення у визначеній структурі англomовного речення залежить сприйняття речення пропагандистом. Звідси слідує необхідність використання багаторівневого шаблону замість дворівневого для вирішення проблеми квантових станів лексико-семантичної частки для сприйняття цільовою особою.

Маючи дані класифікації семантичних категорій для психолінгвістичного портрету пропагандиста, отриманого за допомогою методів [20, 21],

а також сформовану таблицю лексико-семантичних одиниць, можливо побудувати вектор класифікації (6):

$$\bar{V}(v_1^d, v_2^n, v_3^B), \quad (6)$$

де  $v_1^d$  – координата, що відповідає за приналежність до самостійної частини мови,  $v_2^n$  – за члени речення, а  $v_3^B$  – за категорію лексико-семантичної одиниці.

Таким чином, враховуючи визначений порядок слів англійської мови для розповідного речення можна побудувати узагальнену модель синтаксичної одиниці (7):

$$S = \bar{V}(v_1^d, v_2^1, v_3^B) + \bar{V}(v_1^d, v_2^2, v_3^B) + \bar{V}(v_1^d, v_2^3, v_3^B) + \bar{V}(v_1^d, v_2^4, v_3^B) + \bar{V}(v_1^d, v_2^5, v_3^B) + \bar{V}(v_1^d, v_2^1, v_3^B), \quad (7)$$

де  $v_2^1$  – обставина,  $v_2^2$  – означення,  $v_2^3$  – підмет,  $v_2^4$  – присудок,  $v_2^5$  – додаток.

Для питальних та заперечних моделей така ж, за винятком, що вводяться допоміжні дієслова відповідно до використаного часу. В питальному ж відсутня обставина на початку речення. В реплі випадків, допоміжне дієслово не важливе для загальної лексико-семантичної структури і опосередковано впливає на семантичні стани відношення до тексту.

До того ж, використана модель дворівневих шаблонів дозволяє групувати семи англійської мови в матриці шаблону лексико-семантичної основи. Ця властивість необхідна за умови заплутаності квантових станів сприйняття ядра семантики зловмисником. Оскільки отримуються відповідні пари вказівника та ядра семантики, тому при побудові речення необхідно враховувати, що при використанні категоріальної одиниці, що належить до однієї і тієї ж частини речення, наприклад іменника при використанні в ролі додатка чи підмета, у ядрі семантики буде утворюватись певна кількість станів сприйняття, а, відповідно, може змінюватись і стан сприйняття синтаксичної системи взагалі. Це зумовлює необхідність використання принципу тотожності однакових часток (8), відповідно до якого хвилева функція  $f(x'')dx$ , отримана моделюванням семантичної лінії з використанням множини категоріальних лексико-семантичних одиниць повинна описувати той же самий стан, що і функція  $f(x')dx$ , що описує суб'єктивну семантичну лінію, властиву пропагандисту. Теоретична основа для застосування подібного

підходу описується в [19, 20]. А отже, згідно принципу тотожності однакових часток, ці функції (8) відрізняються множителем  $\lambda$ :

$$f(x')dx = \lambda f(x'')dx. \quad (8)$$

Звідси слідує, що за умов перестановки частин мови в рамках одного речення, необхідно знайти такий стан системи синтаксичної одиниці, за якого реалізуються стани, які не змінюються при перестановці ядер семантики місцями, тому застосовується рівняння (9):

$$\hat{P}f(x')dx = \lambda f(x')dx, \quad (9)$$

де  $\hat{P}$  – оператор перестановки.

Знайшовши  $\lambda$ , шляхом впливу на обидві частини рівняння оператора підстановки, можна побудувати модель синтаксичної одиниці (10) з урахуванням характерних квантових станів сприйняття ядер семантики пропагандистом:

$$\begin{aligned} S = & \lambda \bar{V}(v_1^d, v_2^1, v_3^B) + \lambda \bar{V}(v_1^d, v_2^2, v_3^B) + \\ & \lambda \bar{V}(v_1^d, v_2^3, v_3^B) + \lambda \bar{V}(v_1^d, v_2^4, v_3^B) + \\ & \lambda \bar{V}(v_1^d, v_2^5, v_3^B) + \lambda \bar{V}(v_1^d, v_2^6, v_3^B) + \\ & \lambda \bar{V}(v_1^d, v_2^7, v_3^B). \end{aligned} \quad (10)$$

При цьому, не слід забувати, що координата вектору, яка відповідає за члени речення для обраної частини мови може набувати різних значень при кожному стані системи, тому одночасно без прив'язки до пропагандиста може виникати одночасно  $l$  вимірів сприйняття стану, а отже буде отримано багатовимірний вектор виду (11):

$$\lambda \bar{V}(v_1^d, v_2^n, \dots, v_l^n, v_{l+1}^B). \quad (11)$$

Звідси слідує використання замість дворівневого шаблону  $l$  – рівневого.

### Забезпечення семантичної єдності та осмисленості

Синтаксичні структури є основою для формування семантичної структури тексту. При тому, що самі структури несуть семантичне значення, вони повинні формувати єдину цілісну семантичну систему для забезпечення осмисленості тексту.

Дослідження в роботі [4], доводять, що структура семантики англійської мови ієрархічна, при тому, що співвідноситься із синтаксичною структурою та природою розуму людини. Звідси слідує, що необхідно ґрунтуватись на синтаксичних структурах, згенерованих при виконанні другого етапу методу, беручи до уваги структуру ядра семантики психолінгвістичного портрету пропагандиста для врахування природи розуму та сприйняття зловмисником, а також розглядати ієрархічну класифікацію ядер семантики.

Найбільш ефективний підхід, який може бути застосований при поточній задачі є використання генерації природної мови на основі лісу даних природної мови [14]. Це зумовлено можливістю проєкції семантичної та формальної класифікації. Семантичні ядра психолінгвістичного портрету будуть знаходитись на другій знизу позиції в ієрархії семантичних одиниць, тобто виступати в ролі сем. В процесі застосування підходу генерації тексту на основі лісу даних природної мови необхідно виконання умови (12):

$$\begin{cases} B^{m-1} \supset B_n^m \subset B_{f'(x)}^{m-1}; \\ \exists! B^m \in B : f(x'') \rightarrow f(x'). \end{cases} \quad (12)$$

Відповідно до законів проєкції семантичної і формальної класифікації одна на одну, одному вузлу семантичної класифікації (що в даному випадку виступають вершини, відмінні від вершин найнижчої категорії  $B^{m-1}$ ) відповідає декілька вузлів формальної класифікації, яким є вершини найнижчої категорії ядра семантики  $B^m$ , а це підтверджується фактом що подібні вершини є синонімами. При цьому лише один із синонімів може належати ядру семантики пропагандиста за умови, що змодельована семантична лінія наближається до суб'єктивної семантичної лінії пропагандиста. Крім того вузлу формальної класифікації може відповідати декілька вузлів семантичної класифікації  $B_{f(x)}^{m-1}$ , таких, що визначають відношення пропагандиста до семантичної одиниці (наприклад одиниця з позитивним значенням може трактуватися з негативної точки зору), а, отже,  $\hat{P}f(x')dx \Leftrightarrow B_{f(x)}^{m-1}$ .

Як відомо, виходячи із законів психолінгвістики, семантична єдність та осмисленість забезпечується єдиною семантичною направленістю тексту. Мається на увазі розкриття певної семантичної складової в одному реченні, часткове повторення цієї складової в наступному та насичення додатковою складовою. Загальний принцип може бути представлений формулою (13):

$$T = \sum_{i=1}^e cB_{i-1}^r B_i^r \bar{S}, \quad (13)$$

де  $c$  – коефіцієнт, який відповідає за відсоток збереження ядра семантики  $B^r$ , що характерне пропагандисту.

В свою чергу, необхідне застосування правил формальної семантики [16]. В даному випадку правило описується виразом (14):

$$\left\{ \begin{array}{l} \text{якщо в } B_n^m \text{ } n = 1, \text{ то } |B_n^m|^{f(x'), B} = F(B_n^m) \\ \text{якщо в } B_n^m \text{ } n > 1, \text{ то } |B_n^m|^{f(x'), B} = B(B_n^m) \end{array} \right., \quad (14)$$

де  $F$  – функція, що описує усі істинні значення формальної класифікації, характерної для психолінгвістичного портрету пропагандиста.

Таким чином, враховуючи особливості проєкції семантичної та формальної класифікацій, векторну природу синтаксичних структур та імовірні квантові стани лексико-семантичних одиниць, формування семантично однорідного тексту з урахуванням класифікації ядер семантики психолінгвістичного портрету пропагандиста буде відбуватись за принципом (15):

$$T' = \lambda \hat{P}T_n [L(\hat{P}T_n)(f(x'))], \quad (15)$$

де  $L$  – предикат відносно змінного завдяки оператору перестановки  $\hat{P}$  значення  $n$ -го речення тексту  $T$ .

### Стилістичне коригування

Оскільки, розглядається проблема пропаганди в ЗМІ, отже і вплив на пропагандиста повинен здійснюватись за допомогою текстів, які не привертують зайвої уваги та не вибиваються із загальної стилістики публіцистичних журналістських текстів. Згенеровані окремі частини тексту необхідно об'єднувати з використанням деяких елементів методики композиції журналістських текстів [13]. А саме, враховувати подробиці обстановки, поведінки, інтересів, переживань та інші характерні пропагандисту особисті риси, які можливо виявити при застосуванні методу [20]. Такі подробиці входять в рамки дискурсу тексту. Звідси слідує необхідність komponування ядер рівня семантичної класифікації в умовах пропагандного дискурсу за принципом (16):

$$T'' = \lambda \hat{P}T_n^{Attr(Q_n)} [L(\hat{P}T_n^{Attr(Q_n)})(B^{m-1})], \quad (16)$$

де  $Attr(Q_n)$  – синонімічні конструкції з ознаками психолінгвістичного впливу, що визначаються в [20].  $B^{m-1}$  – ядро рівня семантичної класифікації для психолінгвістичного портрету пропагандиста.

### ВИСНОВКИ

Таким чином, в роботі вирішено актуальну наукову задачу підвищення ефективності нейтралізації маніпулятивного впливу інформаційної пропаганди шляхом забезпечення підґрунтя для здійснення зворотного впливу на пропагандиста на основі розробки методу комп'ютеризованого формування англomовного тексту відповідно до психолінгвістичного портрету пропагандиста.

Удосконалено методику генерації осмислених текстів, що ґрунтується на використанні шаблонів на основі методів квантової лінгвістики шляхом застосування правил єдності вказівника та ядра семантики для отримання можливості в якості основи використовувати категоризаційні лексико-семантичні особливості психолінгвістичного портрету пропагандиста і, таким чином, врахування індивідуальної суб'єктивної семантичної лінії пропагандиста.

Впроваджено понятійний простір синтаксичних одиниць психолінгвістичного портрету пропагандиста у формований текст на основі дворівневих шаблонів за рахунок їх вдосконалення в напрямку багаторівневості, шляхом використання принципу тотожності однакових часток для урахування характерних квантових станів сприйняття лексико-семантичної частки пропагандистом та його суб'єктивної семантичної лінії.

Адаптовано підхід генерації осмисленого англomовного тексту на основі лісу даних природньої мови за рахунок проєкції семантичної та формальної класифікації семантичної складової шляхом забезпечення єдності семантичних часток класифікаційної ієрархії семантичних одиниць для використання ядра семантики психолінгвістичного портрету пропагандиста.

Вперше розроблено метод комп'ютеризованого формування англomовного тексту відповідно до психолінгвістичного портрету пропагандиста на основі багаторівневих шаблонів та лісу даних природньої мови, що було вдосконалено методами квантової семантики за рахунок використання семантики Монтеґо та методу квантово-семантичного психолінгвістичного аналізу, що дозволяє підвищити ефективність протидії негативному впливу інформаційної пропаганди в англomовних медіа джерелах завдяки врахуванню індивідуальної семантичної лінії пропагандиста та імовірних квантових станів лексико-семантичних одиниць.

Подальшого дослідження потребує формалізація процесу визначення об'єктивної і суб'єктивної семантичної лінії та виділення функцій, що їх описують, а також дослідження впливу предикативної заплутаності на предикативну групу, так само, як зв'язок предикативної групи з іншими членами речення за умов впровадження зловмисником у текст маніпулятивних засобів.

Результати дослідження можливо використовувати як основу для побудови інформаційної системи генерації англomовного тексту відповідно до психолінгвістичного портрету пропагандиста та

впровадження в нього засобів зворотного цільового впливу на пропагандиста, а також результати вкрай важливі при дослідженні проблеми зсуву семантичної лінії тексту в напрямку суб'єктивної семантичної лінії пропагандиста.

#### ЛІТЕРАТУРА

- [1] A. Krishnan, *Military Neuroscience and the Coming Age of Neuromarketing*, London: Routledge, 2017, 270 p.
- [2] D. Owolabi, C. Nnaji, "The English language and the mass media as tools for sustainable development in multilingual nations", *International Journal of Language and Linguistics*, Vol. 1, Issue 4, pp. 124-130, 2013.
- [3] E. Krahmer, M. Theune, *Empirical Methods in Natural Language Generation: Data-oriented Methods and Empirical Evaluation*, Berlin, Heidelberg: Springer Science & Business Media, 2010, 353 p.
- [4] J. Feist, *Semantic Structure in English*, Amsterdam, Philadelphia: John Benjamins Publishing Company, 2016, 452 p.
- [5] N. Chomsky, Á. J. Gallego, D. Ott, "Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges", *Catalan Journal of Linguistics*, Special Issue: Generative Syntax. Questions, Crossroads, and Challenges, pp. 229-261, 2019.
- [6] N. Fitzpatrick, "Media Manipulation 2.0: The Impact of Social Media on News, Competition, and Accuracy", *Athens Journal of Mass Media and Communications*, Vol. 4, Issue 1, pp. 45-62, 2018.
- [7] R. Ye, W. Shi, H. Zhou, Z. Wei, L. Li, "Variational Template Machine for Data-to-Text Generation", *Proceedings of International Conference on Learning Representations*, Addis Ababa, Ethiopia, April 30, 2020. [Електронний ресурс]. Режим доступу: <https://openreview.net/pdf?id=HkejNgBtPB>.
- [8] S. Chen, J. Wang, X. Feng, F. Jiang, B. Qin, C.-Y. Lin, "Enhancing Neural Data-To-Text Generation Models with External Background Knowledge", *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, China, November 3-7, pp. 3022-3032, 2019.
- [9] S. Welleck, K. Brantley, H. Daume III, K. Cho, "Non-Monotonic Sequential Text Generation", *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, USA, June 10-15, 2019. [Електронний ресурс]. Режим доступу: <http://proceedings.mlr.press/v97/welleck19a/welleck19a.pdf>.
- [10] T. C. Ferreira, C. van der Lee, E. van Miltenburg, E. Krahmer, "Neural data-to-text generation: A comparison between pipeline and end-to-end architectures", *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, China, November 3-7, pp. 552-562, 2019.
- [11] T. Lichte, L. Kallmeyer, "Tree-Adjoining Grammar: A Tree-Based Constructionist Grammar Framework for Natural Language Understanding", *2017 AAAI Spring Symposium Series*, Stanford, USA, March 27-29, Technical Report SS-17-02, pp. 205-212, 2017.
- [12] Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, E. P. Xing, *Toward Controlled Generation of Text*, eprint [arXiv:1703.00955v4](https://arxiv.org/abs/1703.00955v4), 2018. [Електронний ресурс]. Режим доступу: <https://arxiv.org/abs/1703.00955>.
- [13] А. Дмитровский, "Виды композиций в журналистском тексте", *Ученые записки ОГУ. Серия: Гуманитарные и социальные науки*, № 1, С. 189-197, 2009.
- [14] Д. Личаргин, "Методы генерации предложений естественного языка на основе леса данных естественного языка", *Сибирский журнал науки и технологий*, № 3 (43), С. 51-56, 2012.
- [15] Д. Личаргин, Е. Чубарева, Н. Николаева, "Решение проблемы вариативности при генерации фраз естественного языка на основе двухуровневых шаблонов", *Филологические науки. Вопросы теории и практики*, № 11 (29), Ч. I, С. 110-112, 2013.
- [16] И. Герасимова, *Формальная грамматика и интенциональная логика*, М.: ИФ РАН, 2000, 156 с.
- [17] О. Волковский, Е. Ковылин, "Компьютерная система интеллектуального семантического поиска с использованием генерации текстов", *Вестник Херсонского национального технического университета*, № 3-1 (66), С. 238-244, 2018.
- [18] Про рішення Ради національної безпеки і оборони України від 27 січня 2016 року «Про Стратегію кібербезпеки України». [Електронний ресурс]. Режим доступу: <https://zakon.rada.gov.ua/laws/show/96/2016>.
- [19] Я. Тарасенко, "Використання принципів квантової лінгвістики в інформаційному протидіянні", *Безпека інформації*, Том 25, № 2, С. 96-103, 2019.
- [20] Я. Тарасенко, "Метод квантово-семантичного психолінгвістичного аналізу англomовного тексту пропагандного дискурсу", *Сучасні інформаційні системи*, Том 3, № 4, С. 62-68, 2019.
- [21] Я. Тарасенко, "Метод контент-аналізу семантичної частки в текстах з ознаками психолінгвістичного впливу", *Системи управління, навігації та зв'язку*, Том 6, № 58, С. 92-96, 2019.

#### МЕТОД КОМП'ЮТЕРИЗОВАНОГО ФОРМУВАННЯ АНГЛОМОВНИХ ТЕКСТІВ В СООТВЕТСТВИИ С ПСИХОЛІНГВІСТИЧЕСКИМ ПОРТРЕТОМ ПРОПАГАНДИСТА

Текущая ситуация в государстве и мире требует разработки надежных методов выявления и нейтрализации киберугроз, одной из которых является информационная пропаганда в медиа источниках. Для противодействия подобной угрозе необходимо осуществлять обратное влияние на пропагандиста, которое может быть реализовано на основе компьютерно сгенерированного осмысленного текста по психолінгвістическому портрету пропагандиста. Подавляющее большинство

современных методов компьютеризированной генерации текстов базируется на технологиях машинного обучения, что ограничивает возможность эффективного построения психолингвистического профиля за счет необходимости большого количества исследуемой информации, что обуславливает риск обнаружения противодействия пропаганде. Другие же методы не учитывают категориальность и квантовую природу восприятия, или слишком требовательны к количеству входных данных для построения контекстуальных связей или не учитывают пропагандный дискурс. Разработан метод компьютеризированного формирования англоязычного текста в соответствии с психолингвистическим портретом пропагандиста на основе многоуровневых шаблонов и усовершенствованной методики леса данных естественного языка, путем использования методов квантовой семантики и семантики Монтегю, что позволяет повысить эффективность противодействия негативному влиянию информационной пропаганды в англоязычных медиа источниках благодаря учету индивидуальной субъективной семантической линии пропагандиста и возможных квантовых состояний лексико-семантических единиц, основываясь на классификации семантической частицы. Метод позволяет генерировать англоязычный текст в соответствии с психолингвистическим портретом пропагандиста для дальнейшего внедрения в него средств обратного целевого воздействия на пропагандиста. Использование метода важно при исследовании проблемы сдвига семантической линии текста в направлении субъективной семантической линии пропагандиста.

**Ключевые слова:** генерация текста, компьютеризированная компоновка текста, противодействие пропаганде, категории семантической частицы, психолингвистический портрет пропагандиста, семантическая частица в англоязычном тексте, категориальные уровни, квантовые состояния лексико-семантической частицы.

**THE METHOD OF ENGLISH TEXT'S  
COMPUTERIZED FORMATION  
IN ACCORDANCE WITH  
THE PROPAGANDIST'S PSYCHOLINGUISTIC  
PORTRAIT**

The current situation in the country and in the world requires the development of reliable methods for detecting and neutralizing cyber threats, one of which is information propaganda in media sources. To counter such a threat, it

is necessary to carry out the reverse influence on the propagandist, which can be realized on the basis of a computer-generated meaningful text based on the propagandist's psycholinguistic portrait. The vast majority of modern methods for computerized text generation are based on machine learning technologies, which limits the ability of effectively building a psycholinguistic profile due to the need for a large amount of input information, and this leads to the risk of detecting the propaganda counteraction. Other methods do not take into account the categorical and quantum nature of perception, they are too demanding on the input information amount to build the contextual connections, or do not take into account the propaganda discourse. It was developed the method of English text's computerized formation in accordance with the propagandist's psycholinguistic portrait based on multilevel templates and improved methods of natural language data forest, by using the methods of quantum semantics and Montague semantics, which allows to increase the effectiveness of counteracting the negative impact of English-language propaganda, due to the taking into account propagandist's individual subjective semantic line and the probable quantum states of lexical-semantic units, based on the semantic particle's classification. The method allows generating the English-language text in accordance with the propagandist's psycholinguistic portrait for further implementing in it the means of reverse target influence on the propagandist. The method's use is important in studying the problem of shifting the text's semantic line in the direction of the propagandist's subjective semantic line.

**Keywords:** text generation, computerized text composition, counter propaganda, categories of semantic particle, psycholinguistic portrait of propagandist, semantic particle in English-language text, categorical levels, quantum states of lexical-semantic particle.

**Тарасенко Ярослав Володимирович**, кандидат технічних наук, асистент кафедри інформаційних технологій проектування, Черкаський державний технологічний університет.

E-mail: yaroslav.tarasenko93@gmail.com

Orcid ID: 0000-0002-5902-8628

**Тарасенко Ярослав Владимирович**, кандидат технічних наук, асистент кафедри інформаційних технологій проектування, Черкаський державний технологічний університет.

**Tarasenko Yaroslav**, candidate of engineering science, assistant lecturer of the department of information technology of designing, Cherkassy state technological university.