

КВАНТИТАТИВНИЙ ПІДХІД ДО АНАЛІЗУ СТРУКТУРИ КРЕОЛІЗОВАНОГО ВІРШОВАНОГО ТЕКСТУ

THE QUANTITATIVE APPROACH TO THE ANALYSIS OF THE STRUCTURE OF THE CREOLIZED VERSE TEXT

Ключюкіплікчі О.О.,

orcid.org/0000-0001-5475-4543

кандидат філологічних наук, доцент,

доцент кафедри англійської філології і перекладу

Київського національного лінгвістичного університету

У статті показано переваги залучення методів квантитативної лінгвістики для формального моделювання креолізованого віршованого тексту (далі – КВТ). Під останнім розуміємо візуальну поезію у всіх її проявах: тексти американської та української конкретної поезії та фігурні вірші. Структура КВТ складається з вербальних та невербальних елементів (графічні виражальні засоби та графічні стилістичні прийоми), які виокремлені за допомогою лінгвoseміотичного аналізу. Квантитативний підхід використовується для підрахунків кількості та якості складників тексту у спеціально розробленому програмному забезпеченні у мережі Інтернет. Результати підрахунків використовуються у створенні формальної моделі КВТ у вигляді математичної формули.

Ключові слова: моделювання тексту, креолізований віршований текст, візуальна поезія, квантитативна лінгвістика, вербальні та невербальні елементи, лінгвoseміотичний аналіз.

В статті показані переваги використання методів квантитативної лінгвістики для формального моделювання креолізованого стихотворного тексту (далі – КСТ). Под последним понимается визуальная поэзия во всех её проявлениях: тексты американской и украинской конкретной поэзии и фигурные стихи. Структура КСТ состоит из вербальных и невербальных элементов (графические выразительные средства и графические стилистические приёмы), которые выделены при помощи лингвoseміотического анализа. Квантитативный подход используется для подсчёта количества и качества составляющих текста в специально разработанном программном обеспечении в сети Интернет. Результаты подсчётов используются для составления формальной модели КСТ в виде математической формулы.

Ключевые слова: моделирование текста, креолізованный стихотворный текст, визуальная поэзия, квантитативная лінгвістика, вербальные и невербальные элементы, лінгвoseміотический анализ.

The article shows the advantages of the engagement of the quantitative linguistic methods for the formal modelling of the creolized (multimodal) verse text (CVT). The latter one is understood as the visual poetry in all its manifestations: the texts of the American and Ukrainian concrete poetry and shape poems. The structure of the CVT consists of the verbal and nonverbal elements (graphic expressive means and graphic stylistic devices), singled out with the help of the linguosemiotic analysis. The quantitative approach is used for the counting of the number and quality of the text constituents in the specially designed software in the Internet. The counting results are used for the compilation of the formal model of the CVT as a mathematical formula.

Key words: text modeling, creolized (multimodal) verse text, visual poetry, quantitative linguistics, verbal and nonverbal elements, linguosemiotic analysis.

Постановка проблеми. Лінгвісти все більше використовують кількісні дані у своїх дослідженнях. Фонетисти, соціолінгвісти, психолінгвісти та комп'ютерні лінгвісти користуються такими розрахунками протягом десятиріч, але сьогодні й дослідники фонології, синтаксису та історії мови також звертаються до квантитативних методів. Наприклад, Келлер [1] вимірював прийнятність речення через психофізичну техніку під назвою оцінка величини (magnitude estimation). Також Боерсма та Хейс [2] застосували ймовірнісні міркування в обмеженому перерозподілі алгоритму для теорії оптимальності. Отже, освоєння квантитативних методів стає все більш невіддільним компонентом лінгвістичних досліджень [3].

Оскільки креолізація тексту – це відношення вербальної складової до невербальної, у ході дослідження було вироблено формальну модель КВТ як математичну формулу відношення невербальних елементів (графічних виражальних засобів + графічних стилістичних прийомів + зображувальних елементів) до вербальних (літера, морфема, слово, вираз\речення, текст). Для представлення формальної моделі кожного тексту необхідно підрахувати кількість елементів в кожному КВТ у їх відношенні один до одного. Для цього необхідно звернутися до квантитативної лінгвістики.

Аналіз останніх досліджень і публікацій. Для моделювання інформації про тексти корпусу, об'єднаного проблематикою лінгвокогнітології,

В.Г. Шкляревський обирає метод когнітивного мапування, який має на меті моделювання тексту в термінах структуризації його концептуального простору [4, с. 168]. В. І. Перебийніс під моделюванням тексту розуміє дослідження текстів «шляхом побудови й вивчення їх моделей, використання моделей для визначення чи уточнення характеристик і раціоналізації способів побудови нових за конструкцією об'єктів» [5, с. 92]. Вчені активно застосовують жорсткий різновид когнітивного мапування, що ґрунтується лише на даних текстів [6, с. 254].

Сучасний етап у розвитку систем обробки інформації взагалі характеризується як етап формалізації та застосування знань [7]. На сьогодні практично сформувалася й активно розвивається нова науково-технічна дисципліна – інженерія знань, яка об'єднала зусилля фахівців у галузі прикладної інформатики, математики, лінгвістики, психосемантики. Одним з основних завдань інженерії знань, як і раніше, є пошук мови логічної *систематизації знань* – важливої ланки між мовами опису знань, які використовуються у сучасних системах штучного інтелекту, і знаннями фахівця, що вербалізуються на професійному діалекті прикладної математики [8, с. 2-4].

Для точнішого опрацювання креолізованих віршованих текстів в роботі пропонується використання програми лінгвосеміотичного аналізу креолізованого віршованого тексту, яку розміщену в мережі Інтернет. Дана програма показує використання вербальних та невербальних елементів у КВТ, їх співвідношення та допомагає вивести лінгвістичну формулу КВТ.

Постановка завдання. На завершальному етапі формального моделювання КВТ після лінгвосеміотичного аналізу й виділення складових тексту необхідно звернутися до квантитативної лінгвістики задля підрахування кількісно-якісного складу структурних елементів. Статистичні дослідження використовуються у комунікативній лінгвістиці та при дослідженні прагматики тексту. Наприклад, контент-аналіз як методика частоти появи в тексті певних одиниць змісту – контентів, що дозволяє робити висновки про наміри творця тексту або можливі реакції адресата. Контент-аналіз – це дослідницька техніка для об'єктивного, системного й кількісного опису наявного змісту комунікації, яка відповідає цілям дослідника [9]. Деякі дослідники, наприклад В.А. Ядов, наголошують на кількісному аспекті контент-аналізу, пропонуючи таке визначення: «Контент-аналіз – це переведення в кількісні показники великої кількості текстової (чи записа-

ної на плівку) інформації» [10]. Зосереджуючись на кількісно-якісній стороні даного методу, В. С. Бондар і М. А. Допіра формулюють найкоротше визначення, що адекватно відбиває сутність контент-аналізу: «Контент-аналіз є вимірюванням тексту чи іншого символічного матеріалу, що має методологічне підґрунтя й відповідає завданням соціальних наук» [11, с. 18].

Квантитативна лінгвістика – це сфера міждисциплінарних прикладних пошуків, де основним інструментом виступає вивчення мови та мовлення, є кількісні та статистичні методи аналізу [12, с. 307-308]. У квантитативній лінгвістиці широко використовується кількісний математичний апарат – теорія множин, математична логіка, теорія алгоритмів тощо. А. Загнітко зауважує, що створення структурних моделей цілком ґрунтується на застосування статистики в мовознавстві, оскільки використання статистичних методів у мовознавстві дозволяє доповнити структурну модель мови зі ймовірнісним компонентом, тобто створити структурно-ймовірнісну модель з властивим для неї потужним пояснювальним потенціалом [12, с. 307-308]. Суттєвим є те, що завдання побудови структурно-ймовірнісної моделі функціонування мови належить до теоретичних проблем лінгвістики та складає один із виявів теорії мови. У прикладній сфері квантитативна лінгвістика виявлювана тими фрагментами цієї моделі, що використовуються для лінгвістичного моніторингу функціонування мови, дешифрування кодованого тексту, авторизації/атрибуції тексту та тому подібного [13; 14].

Останніми роками бурхливо розвивається і має суттєвий лінгвістичний потенціал комп'ютерний дизайн тексту та його компонентів, зокрема шрифтів тощо [12, с. 314]. Наявне програмне забезпечення дозволяє виводити на монітор комп'ютера всю смугу друкованого тексту книги, часопису, газети, рекламного буклета і компоувати текст та мальовничі елементи оригінал-макету в єдине ціле. Текст у цьому вимірі постає як елемент зображення, а зображення – як елемент тексту. Технологічним є питання їхнього оптимального поєднання, оптимізація креолізації тексту тощо. Суто практичні принципи побудови оригінал-макету видання повинні підтримуватися дослідженнями у сфері лінгвістики та семіотики, які мають виявляти закони поєднання змісту тексту, зображення (малюнок, рисунок, таблиця, схема, графік тощо), семантики гарнітури (шрифту) та способів виділення компонентів тексту (курсив, відступ, пропуск, розрив і тому подібне). Наявні в лінгвістиці та семіотиці теоретичні напрацювання

(теорія «захистів» Олександра Реформатського тощо) на сьогодні мало «вмонтовані» у проблематику комп'ютерно-прикладних завдань [13; 15].

Досвід показує [16, с. 6-8], що використання математичних методів для специфікацій та моделювання програмного забезпечення може сприяти розв'язанню проблеми побудови надійного програмного забезпечення для формальної обробки креолізованих віршованих текстів. Математичні методи, які б ефективно використовувалися у виробництві програмного забезпечення мають задовольняти деякі основні вимоги. По-перше, записи, які застосовуватимуться в специфікаціях мають бути стандартизовані, а отже виробництво багатьох проектів може провадитись групами розробників без непорозумінь. Це дає можливість використовувати вже розроблені специфікації. По-друге, програмні засоби потребують автоматизації маніпулювання з формалізованим текстом. Обидві ці передумови вимагають від таких нотацій стабільності та ясності. Традиційно виділяють наступні фази життєвого циклу програми:

- аналіз вимог;
- специфікація;
- проектування;
- реалізація;
- тестування та нагодження;
- експлуатація та супроводження.

Однією з важливих теоретичних причин використання формальних методів (далі – ФМ) специфікацій програм, є те, що самі програми є математичними об'єктами, бо вони мають формальну семантику та виражені формальною мовою, а отже, можуть бути оброблені за допомогою математичних засобів. Таким чином, виникає можливість доведення певних властивостей програм, зокрема властивості правильності, за допомогою математичних методів [16, с. 6-8].

Виклад основного матеріалу. При лінгвістичному підході до розпізнавання зорових образів як невербальної складової креолізованого тексту А. Г. Аркад'єв та Е. М. Браверман приділяють увагу належному опису фігури [17, с. 68]. Опис складається з фраз за наступним правилом: фраза складається з назв фрагментів та вказівок про їх взаємне розташування. Вчені визнають, що складні фігури належать до одного образу, якщо складені таким чином їх описи однакові або принаймні близькі між собою. Алгоритм розпізнавання складного образу на основі лінгвістичного підходу вирішується «з вчителем». В такому випадку людина, крім класифікації матеріалу, опрацьовує і вводить до машини інформацію словами (наприклад, перелік простих характер-

них фрагментів й характеристики їх взаємного розташування), визначає правила опрацювання і складає опис кожного з образів. Оперуючи цим матеріалом, машина розшукує у введених зображеннях характерні фрагменти, визначає їх взаємні розташування, складає опис зображень та, порівнюючи їх зі складеними «вчителем» описами класів, виробляє класифікацію [17, с. 68].

Для формального вивчення креолізованих віршованих текстів та виведення формальної моделі КВТ було створено корпус українських та американських креолізованих віршованих текстів. Згідно з М. Мальберг, корпусний аналіз художніх текстів добре підходить до філологічного кола Л. Шпітцера, створюючи «корпусно-стилістичне коло», яке включає власне лінгвістичний аналіз, корпусно-стилістичний аналіз і літературну оцінку твору, тобто його інтерпретацію [18, р. 139-140]. Перевагою корпусних студій художнього дискурсу О.П. Воробйова називає одночасне охоплення значного за обсягом текстового матеріалу, що дозволяє виявити нові, нерідко несподівані тенденції й авторські ідіосинкразії, які не простежуються на більш обмеженому емпіричному матеріалі за умови його ручного опрацювання [19, с. 54].

Особливістю корпусно-стилістичного методу аналізу є й те, що він сфокусований на розкритті системних кореляцій форми, змісту та функцій [20, р. 9] значущих елементів художнього тексту або текстів з опорою на рекурентність, частотність та комбінаторику його/їх ключових слів, словосполучень та мотивів у розгортанні оповіді, формуванні когнітивних пріоритетів автора як основи для побудови відповідних когнітивних мап, встановленні співвідношень між стилістично маркованими структурами та породжуваними художніми ефектами. Це зближує корпусну стилістику зі стилістикою когнітивною, встановлюючи між ними взаємодоповнювальні зв'язки [19, с. 55].

Теоретичним підґрунтям корпусної лінгвістики найбільше є структуралізм, чи, за словами В. Виноградова: «сукупність поглядів на мову та методи її дослідження, в основі яких лежить розуміння мови як знакової системи, у межах якої виділяються структурні елементи (одиниці мови, їхні класи тощо)». Крім цього, структуралізм виявляє прагнення до чіткого, наближеного до точних наук, формального опису мови. І саме це чи не найбільше вирішило долю структуралізму як теоретичного підґрунтя корпусної лінгвістики [21].

У нашому дослідженні для опрацювання фактичного матеріалу в практичній частині було

Выберите файл

Название рисунка

Автор

Verbal Items	Graphic Expressive Means	Graphic Stylistic Devices	Pictorial Items
Letters <input type="checkbox"/>	Text Segmentation <input type="checkbox"/>	Specific title's form <input type="checkbox"/>	Colour <input type="checkbox"/>
Morphemes <input type="checkbox"/>	Graphon <input type="checkbox"/>	Graphic imagery <input type="checkbox"/>	Background <input type="checkbox"/>
Words <input type="checkbox"/>	Punctuation <input type="checkbox"/>	Graphic pastiche <input type="checkbox"/>	Perspective <input type="checkbox"/>
Sentences <input type="checkbox"/>	Capitalisation <input type="checkbox"/>	Language game <input type="checkbox"/>	Object <input type="checkbox"/>
Text <input type="checkbox"/>	Signs of the Other Semiotic System <input type="checkbox"/>	Graphic logogriph <input type="checkbox"/>	Figuration <input type="checkbox"/>
	Graphic-orthographic views of other langs <input type="checkbox"/>	The Text's graphic model transposition <input type="checkbox"/>	
	Acronyms and Abbreviation <input type="checkbox"/>	Simulacra <input type="checkbox"/>	

Рис. 1. Завантаження нового файлу

розроблено і впроваджено в роботу програмне забезпечення з лінгвосеміотичного аналізу КВТ, розташоване на сайті <http://lingvosemiotic.com.ua/> в мережі Internet.

На першому етапі роботи з програмним забезпеченням з лінгвосеміотичного аналізу КВТ пропонується вибрати текст з вже внесених на сайт або завантажити новий. Програма підтримує графічні формати .jpg, .png, .gif. У вже опрацьованих текстах наявна кругова діаграма співвідношення елементів КВТ у відсотках та кількісно (вербальні елементи та невербальні: зображувальні, графічні виражальні засоби й графічні стилістичні прийоми). На основі цих даних створюються формальні моделі кожного КВТ.

Якщо завантажується новий текст, то на другому етапі вводяться вербальні елементи на основі даних з таблиць лінгвосеміотичного аналізу за методикою когнітивно-семіотичного підходу до аналізу КВТ. В поля для заповнення вводяться цифри. Програма розроблена англійською мовою. Виділяються такі вербальні елементи:

- Літери /letters;
- Морфеми /morphemes;
- Слова /words;
- Речення /sentences;
- Текст /text.

На третьому етапі вносяться невербальні елементи. Класифікація графічних виражальних

засобів та графічних стилістичних прийомів вноситься згідно з попереднім дослідженням [22]. Графічні виражальні засоби /graphic expressive means:

- Сегментація тексту/Text segmentation;
- Типи набору та шрифт /Font and Typeface;
- Графон/Graphon;
- Пунктуація/Punctuation;
- Капіталізація/Capitalisation;
- Знаки інших семіотичних систем/Signs of the other semiotic systems;
- Графічно-орфографічні іншомовності/Graphic-orthographic presentation of other languages words;
- Скорочення та абрєвіатури/Acronyms and abbreviations;
- Еквівалент тексту/Text equivalent.

Графічні стилістичні прийоми/graphic stylistic devices:

- Специфічне оформлення заголовку/Specific form of the title;
- Графічна образність/Graphic imagery;
- Графічний пастиш/Graphic pastiche;
- Мовна гра/Language game;
- Графічний логогриф/Graphic logogriph;
- Транспозиція графічної моделі тексту/The text's graphic model transposition;
- Симулякр/Simulacra.

Завершальний етап – внесення зображувальних елементів. Дані зображувальних еле-

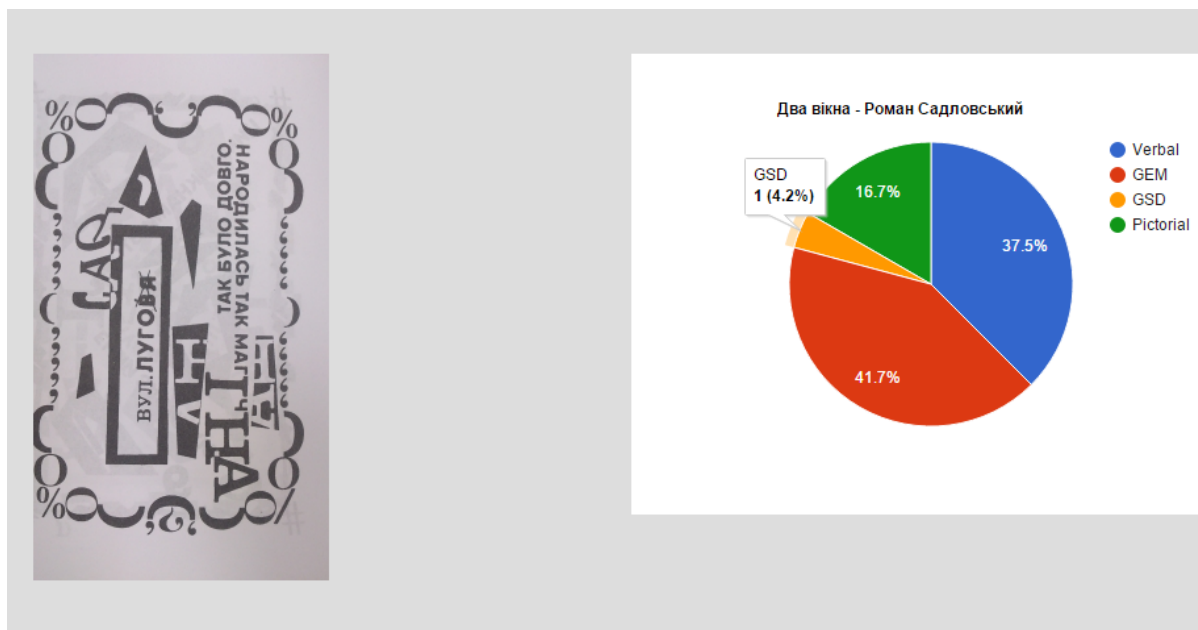


Рис. 2. Опрацьований текст та діаграма його структурних елементів

ментів вносяться згідно з класифікації [23].
Зображувальні елементи / Pictorial elements:

- Колір/Colour;
- Фон/Background;
- Перспектива/Perspective;
- Об'єкт/Object;
- Оформлення/Figuration.

Сторінка завантаження файлу з текстом та виділення структурних елементів виглядає таким чином (Рис. 1).

В результаті програма видає діаграму (Рис. 2), яка показує кількісне співвідношення всіх структурних елементів КВТ для кожного типу тексту.

Отримана відсоткова кількість структурних елементів слугує матеріалом для виведення формул структури КВТ. На основі цих даних будуть американські та українські формальні моделі КВТ.

Висновки. Таким чином, у роботі доведена практичність використання квантитативних методів при дослідженні структури креолізованого віршованого тексту. Процедура кількісно-якісного підрахування складових елементів КВТ дає змогу наочного представлення формальної моделі КВТ та подальшого порівняння американських та українських візуалізованих текстів на рівні їх будови.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Keller F. A psychophysical law for linguistic judgments. *Proceedings of the 25th Annual Conference of the Cognitive Science Society*. – Boston, 2003. P. 652–657.
2. Boersma P., Hayes B. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*. 2001. #32(1). P. 45–86.
3. Johnson K. *Quantitative Methods in Linguistics*. Malden, Massachusetts: Wiley-Blackwell, 2008. 296 p.
4. Шкляревський В.Г. Етапи метарозмітки корпусу англомовних текстів з лінгвокогнітології. *Вісник КНЛУ. Серія Філологія*. Том 18. Київ: Вид-во КНЛУ, 2015. №1. С. 165–172.
5. Перебийніс В.І. Математична лінгвістика: навчальний посібник. Київ: Вид. центр КНЛУ, 2014. 125 с.
6. Єсіпович К.П. Феномен когнітивного картування в сучасній лінгвістичній парадигмі. *Studia Linguistica*. 2013. Вип. 7. С. 254–257.
7. Лазаренко О.В., Яковенко А.А. Моделювання процесу узагальнення в системі автоматичного реферування: монографія. Харків: Вид-во НУА, 2007. 123 с.
8. Замаруєва І. Мова–інформація–знання. *Інформація та нові технології*. Київ, 1996. № 1. С. 2–4.
9. Roberts C. *Text analysis for the social sciences: methods for drawing statistical inferences from texts and transcripts*. Mahwah New Jersey: Lawrence Erlbaum Associates, 1997. 316 p.
10. Ядов В.А. Стратегия социологического исследования. Москва: Академкнига; Добросвет, 2003. 596 с.
11. Бондар В.С., Допіра М.А. Розгляд методу контент-аналізу з погляду кількісно-якісних технік проведення. *Наукові записки НАУКМА. Серія «Соціологічні науки»*. 2007. Том 70. С. 17–26.

12. Загнітко А. Статус прикладної лінгвістики в системі лінгвістичних наук. *Донецький вісник Наукового товариства ім. Шевченка*. Т. 28. Донецьк : Український культурологічний центр, Східний видавничий дім, 2010. С. 291–318.
13. Демська-Кульчицька О. Основи національного корпусу української мови. Київ: Інститут української мови, 2005. 220 с.
14. Краснобаєва-Чорна Ж.В. Квантитативна лінгвістика: навч. посібник. Донецьк: ДонНУ, 2007. 79 с.
15. Ситар Г.В. Теоретичні та практичні проблеми перекладу: навчально-методичний посібник. Донецьк: ДонНУ, 2009. 88 с.
16. Омельчук Л.Л. Формальні методи специфікації програм. Київ : УкрІНТЕІ, 2010. 78 с.
17. Аркадьєв А.Г., Браверман Э.М. Обучение машины классификации объектов. Москва: Наука, Главная редакция физико-математической литературы, 1971. 192 с.
18. Mahlberg M. Corpus stylistics. *The Bloomsbury Companion to Stylistics*. London; New York : Bloomsbury, 2016. P. 139–156.
19. Воробйова О.П. Корпусна стилістика: данина моді чи потреба сучасності? *Україна і світ: діалог мов та культур*: матеріали міжнародної науково-практичної конференції, 29 березня – 31 березня 2017 року. Київ: Вид. центр КНЛУ, 2017. С. 54–55.
20. Nørgaard N., Busse B. and Montoro R. Key Terms in Stylistics. London; New York : Continuum, 2010. 269 p.
21. Демська О.М. Українська корпусна лінгвістика: сьогодення і перспективи. *Диво слово*. 2011. № 10. С. 32–34.
22. Вялікова О.О. Система графічних засобів і прийомів в американській та українській постмодерністській поезії (на матеріалі віршованих текстів): дис.... канд. філол. наук : спец. 10.02.15 «Загальне мовознавство». Київ, 2012. 206 с.
23. Вялікова О.О. Формальні підходи до аналізу креолізованого віршованого тексту. *Сучасне літературознавство та прикладна лінгвістика: американські та британські студії*: Матеріали II Міжнародного симпозіуму: Ч. 2, м. Київ, 9-11 квітня 2014 р. Київ : Талком, 2014. С. 331–335.