

Н.П. ДАРЧУК, І.В. ВАСИЛЬЄВА, О.М. ВАСИЛЬЄВ

Київський національний університет імені Тараса Шевченка
(Вул. Володимирська 60, Київ 01601; e-mail: vasilev@univ.kiev.ua)

УДК 53.01

**ВЕКТОРНА МОДЕЛЬ
АНАЛІЗУ СТИЛІСТИКИ ТЕКСТІВ**

Стаття присвячена застосуванню фізичних підходів до аналізу авторських стилів українських письменників. Пропонується модель, у якій літературні стилі описуються векторами одичної довжини в багатовимірному просторі. Числовою характеристикою стилю є результат скалярного добутку відповідного вектора на вектор, який визначає загальний стиль для групи авторів. Показано, що цей параметр лінійним чином залежить від рангу автора. Така залежність підтверджує гіпотезу приєднання до більшості, відповідно до якої автори, вибираючи стиль, орієнтуються на стиль своїх успішних колег.

Ключові слова: фізика складних систем, квантитативна лінгвістика, вектор, енергія, розподіл.

Література кожного народу – це найкраще дзеркало його життя.

Іван ФРАНКО

1. Вступ

Фізика постійно розширює межі свого застосування. Яскравим прикладом є напрямок, який називають *фізикою складних систем* (див., наприклад, [1–5] і посилання, що містяться там). У контексті цього підходу з одних і тих самих позицій аналізуються системи різної природи. Список задач, які успішно розв'язувались і розв'язуються методами фізики складних систем, містить (але не обмежується цим) аналіз транспортних систем [6–8], соціально-економічних зв'язків [9–16], структури та характеристик текстів [17–27] та багато іншого. Важливими є декілька обставин. По-перше, як зазначалося вище, досліджувані системи є абсолютно різними за своєю природою. Іноді навіть досить важко визначити, до якої галузі знань належить відповідна проблематика. По-друге, досліджуються такі системи методами, притаманними саме фізиці (в основному статистичній), оскільки їм властиве використання відповідного математичного апарату, а фізичні підходи ще й мають у своєму розпорядженні потужний методологічний арсенал, який дозволяє створювати нові моделі. Остання обставина є вкрай важливою, оскільки

саме методологічні аспекти виходять на перший план у випадках, коли мова йде про міждисциплінарні дослідження. Перевага такого підходу пов'язана з тим, що наявність оригінальних моделей дозволяє не тільки аналізувати поведінку й характеристики досліджуваної системи, а й дає можливість ідентифікувати ті характеристики, які слід визначати на основі статистичних даних.

У статті методи фізики застосовуються для аналізу авторських стилів українських письменників. У *кількісній (квантитативній) лінгвістиці* [28–30] такі задачі є найбільш пріоритетними (див., наприклад, [31–36] та посилання, що містяться там). Один із способів їх розв'язання базується на використанні фізичних методів, включаючи відповідний категоріальний апарат.

2. Формалізація проблеми

Почнемо з того, що формалізуємо проблему. Нехай у літературній мові використовується певна кількість слів (*базова множина*). Будь-який літературний твір можна розглядати як послідовність слів (із відповідними граматичними змінами) із базової множини. Кожен автор характеризується набором текстів, на основі яких можна розраховувати параметри, що дають уявлення про притаманний автору *стиль*. Зокрема, якщо впорядкувати слова з базової множини (наприклад, в алфавітному порядку), то тоді творчість кожного автора можна описати вектором, компоненти якого визна-

© Н.П. ДАРЧУК, І.В. ВАСИЛЬЄВА,
О.М. ВАСИЛЬЄВ, 2021

чають кількість використань відповідного слова у авторських текстах. Разом з тим, розмірність такого вектора була б досить значною, що унеможливило б використання цього підходу на практиці. Більш продуктивним видається підхід, у рамках якого базова множина слів значно обмежується. Наприклад, до неї можна включати лише слова, котрі належать до певної категорії чи поняття. Саме такий підхід використано у статті.

Слід відзначити декілька моментів, важливих для розуміння особливостей запропонованого підходу. По-перше, це стосується означення авторського стилю. Взагалі, поняття стилю досить широке і включає в себе сукупність ознак, що характеризують манеру письменника і вирізняють його твори серед творів інших авторів [37]. Тому, безумовно, не можна стверджувати, що авторський стиль цілком описується певним вектором. Такий вектор слід розглядати лише як одну з характеристик авторського стилю. По-друге, при побудові вектора для характеристики авторського стилю, як зазначалося вище, використовується обмежена група слів. Ця група слів є підмножиною базової множини. Побудований на її основі вектор можна інтерпретувати як проєкцію на відповідну підмножину вектора, побудованого на основі базової множини. Фактично, це означає, що вибираючи різні підмножини базової множини, ми отримуємо різні характеристики авторського стилю, які стосуються тої чи іншої тематики у творчості авторів.

Будемо виходити з того, що базова множина складається з N слів, а кількість авторів становить величину M . Позначимо через $a_n^{(m)}$ кількість використання n -го слова m -м автором ($n = 1, 2, \dots, N$ та $m = 1, 2, \dots, M$). Вектори \mathbf{e}_m визначимо як такі, що мають компоненти

$$e_n^{(m)} = \frac{a_n^{(m)}}{\sqrt{\sum_{n=1}^N (a_n^{(m)})^2}}. \quad (1)$$

Якщо розраховувати скалярний добуток двох довільних векторів \mathbf{e}_m та \mathbf{e}_k як

$$\mathbf{e}_m \cdot \mathbf{e}_k = \sum_{n=1}^N e_n^{(m)} e_n^{(k)}, \quad (2)$$

то вектори будуть мати одиничну норму:

$$|\mathbf{e}_m|^2 \equiv \mathbf{e}_m \cdot \mathbf{e}_m = 1. \quad (3)$$

Вектори \mathbf{e}_m ($m = 1, 2, \dots, M$), побудовані таким способом, будемо інтерпретувати як такі, що описують стиль m -го автора.

Також визначимо вектор \mathbf{a} з компонентами a_n ($n = 1, 2, \dots, N$), котрі розраховуються таким чином:

$$a_n = \frac{A_n}{\sqrt{\sum_{n=1}^N A_n^2}}. \quad (4)$$

При цьому параметри A_m визначаються згідно з співвідношенням

$$A_n = \sum_{m=1}^M a_n^{(m)}. \quad (5)$$

Отже, вектор \mathbf{a} описує загальний стиль, розрахований на основі всіх текстів у сукупності.

Слід зазначити, що скалярний добуток двох векторів (у даному випадку це значення від 0 до 1) можна інтерпретувати як характеристику схожості відповідних стилів (аналог "енергії взаємодії" спінів у моделі, подібній до моделі Ізінга [38, 39]). Фактично, мова йде про адаптацію методики, яка широко використовується у лінгвістичних дослідженнях і базується на розрахунку коефіцієнта подібності (див., наприклад, [40]). Зокрема, якщо скалярний добуток дорівнює 1, то стилі збігаються. Якщо добуток дорівнює 0, то ненульовим компонентам першого вектора відповідають нульові компоненти другого вектора, і навпаки. Це означає, що автори, чий стиль описується такими векторами, використовують абсолютно різні набори слів (із базової множини), серед яких немає "спільних". Зокрема, скалярний добуток векторів \mathbf{a} та \mathbf{e}_m визначає, наскільки стиль m -го автора відрізняється від загального стилю. Величину

$$\varepsilon_m = \mathbf{a} \cdot \mathbf{e}_m \quad (6)$$

можна інтерпретувати як числову характеристику, яка ранжує стиль автора по відношенню до загального стилю. Важливим може видаватися розподіл параметрів ε_m серед групи письменників. Однак для вирішення цього питання необхідно задати спосіб поведінки агентів (письменників) при визначенні власного стилю. Тут простежується аналогія з тим, як елементи статистичної системи розподіляються за енергетичними рівнями в зовнішньому полі, роль якого відіграє вектор \mathbf{a} . Відпо-

відно, спосіб поведінки агентів є аналогом додаткових обмежень (на кшталт *принципу Паулі* [41, 42]), які накладаються на систему.

3. Ефект приєднання до більшості

Важливим є питання щодо того, чи існують якісь закони чи принципи розподілу характеристик для авторських стилів. Ми будемо виходити з гіпотези, що у соціумі існують запити на певні літературні стилі чи теми, і популярність автора залежить від того, наскільки його стиль відповідає очікуванням суспільства. Тому розумно припустити, що деякі (не всі) автори, свідомо чи підсвідомо, вибирають власний стиль орієнтуючись на стиль інших популярних авторів. З іншого боку, природним видається бажання авторів мати власний унікальний стиль. Якщо авторів розглядати як соціальних агентів, то кожен такий агент має зробити певний вибір стосовно власного стилю. Різні автори можуть керуватися різними принципами чи підходами при виборі стилю, що не дозволяє говорити про існування універсальних законів поведінки соціальних агентів. Разом з тим, мова може йти про виявлення груп чи спільнот авторів, які мають спільні механізми реалізації соціального вибору.

Щодо питання про те, як агенти в соціально-економічних системах роблять вибір, то існує декілька теорій та підходів. Один з алгоритмів поведінки називається “ефектом приєднання до більшості” [43]. Суть ефекту полягає у тому, що агент поводить себе аналогічно до більшості інших агентів. У контексті досліджуваної моделі таку поведінку можна інтерпретувати як намагання m -го письменника ($m = 1, 2, \dots, M$) максимізувати значення параметра ε_m . Це фактично означає, що вектор \mathbf{e}_m має бути орієнтований у напрямку вектора \mathbf{a} . Разом з тим, абсолютний збіг стилів зробить авторів нерозрізнюваними для сприйняття читачами. Тому відмінність між різними стилями повинна бути відчутною. Саме тому, що стилі мають відрізнитися, даний ефект не можна звести до принципу переважного приєднання у добре відомій моделі Саймона [44]. Отже, ми можемо очікувати на нові результати.

Позначимо через $\Delta\varepsilon$ мінімальну відмінність між значеннями параметрів ε_m , на якій читачі відчують різницю між стилями різних авторів. З використанням фізичної термінології ситуацію можна

інтерпретувати таким чином. Є набір “енергетичних” рівнів ε_m , і різниця між цими рівнями становить величину $\Delta\varepsilon$. Частинки, роль яких відіграють агенти (письменники), мають перебувати на різних рівнях. Фактично, мова йде про правило, подібне до принципу Паулі у фізиці.

Не обмежуючи загальності, можемо вважати, що автори перенумеровані так, що $\varepsilon_1 \geq \varepsilon_2 \geq \dots \geq \varepsilon_M$. Тоді, враховуючи все вищезазначене, для всіх $m = 1, 2, \dots, M$, маємо

$$\varepsilon_m = \varepsilon_0 - m\Delta\varepsilon, \quad (7)$$

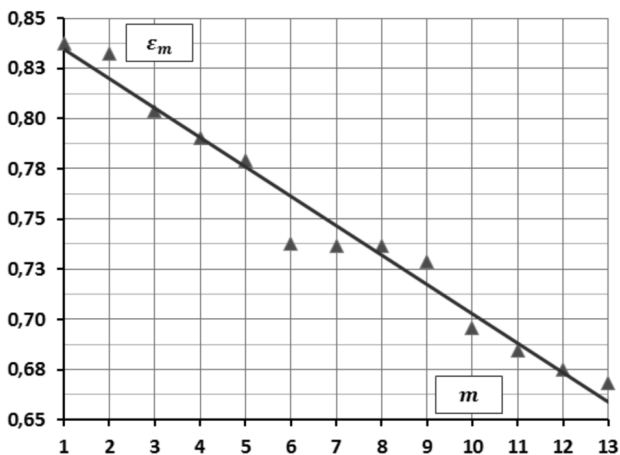
де ε_0 є параметром моделі. Тут ми врахували, що вектор \mathbf{e}_1 , який відповідає найвищому “енергетичному” рівню ε_1 , може не збігатися з вектором \mathbf{a} , оскільки останній не описує реальний стиль, а є лише результатом усереднення.

Таким чином, за наявності “ефекту приєднання”, якщо ранжувати письменників за значенням параметрів ε_m , то маємо отримати лінійну залежність ε_m від рангу m .

4. Емпіричні результати

Використаємо наявні “експериментальні” результати для перевірки валідності запропонованої вище моделі. Зокрема, скористаємося даними про використання слів, які належать до категорії “людина/особа”, групою українських поетів (*В. Стус, І. Драч, С. Йовенко, Л. Костенко, М. Вінграновський, М. Воробйов, І. Жиленко, О. Забужко, В. Коломієць, Т. Коломієць, Б. Олійник, Д. Павличко та Л. Скірда*, значення параметрів $M = 13$ та $N = 115$). Аналіз виконувався на основі поетичних творів, написаних у 60-ті, 70-ті та 90-ті роки минулого століття. Для кожного поета об’єм вибірки з текстів становив 20000 слововживань. В таблиці наведено інформацію про кількість лексем з категорії “людина/особа”, які використовують різні автори, вказана кількість лексико-семантичних вживань (ЛСВ) та загальна кількість лексико-семантичних форм у текстах (словник) [45–47].

Вибір саме такої групи поетів зумовлений декількома причинами. Зокрема, необхідно було визначити авторів, котрі належать до певної компактної групи, поведінка якої могла би бути описана у термінах запропонованої вище моделі приєднання до більшості. Ця група не обов’язково має бути локалізована у часі. Навпаки, якщо розглядати



Залежність параметра ε_m від рангу m . Зв'язок між рангом та автором визначається з таблиці

процес приєднання до групи як зайняття певної вакансії у “енергетичному спектрі”, а також ту обставину, що вакансії можуть займатися у довільному порядку, то для повного заповнення “спектра” необхідний значний проміжок часу, який може включати у себе декілька поколінь авторів. Нарешті, наявність зазначеної групи і належність конкретного автора до неї визначається безпосередньою перевіркою. Іншими словами, мета полягає не у тому, аби довести наявність певної закономірності у розподілі характеристик авторських стилів для довільної групи авторів. Мета полягає у тому, аби знайти групу авторів, для яких закономір-

Дані про авторів

Ранг	Автор	ЛСВ	Лексеми	Словник
1	В. Коломієць	341	31	4188
2	Т. Коломієць	181	16	3536
3	Л. Скирда	252	19	3338
4	І. Драч	286	25	3818
5	Б. Олійник	466	42	3463
6	М. Вінграновський	197	24	3252
7	С. Йовенко	272	25	3186
8	Л. Костенко	281	26	4040
9	М. Воробйов	196	15	3193
10	Д. Павличко	194	19	3830
11	О. Забужко	128	13	4181
12	І. Жиленко	317	29	3802
13	В. Стус	210	21	3664

ність існує. Емпіричним шляхом встановлено, що перераховані вище поети належать до однієї групи, яку можна описати у рамках моделі приєднання до більшості.

Важливо також відзначити такі дві обставини. По-перше, вибір групи “людина/особа” визначає підмножину для вектора стилю, яка досліджується. В цьому відношенні цілком припустимо вибрати іншу групу і провести дослідження для відповідної підмножини. Однак при цьому група має бути досить загальною і спільною для багатьох авторів. По-друге, після того, як групу для визначення підмножини вибрано, автори вибираються відповідно до того, наскільки актуальною є у їхній творчості відповідна тематика. Критерієм відбору може бути кількість вживань лексем з вибраної групи.

На основі наявних емпіричних даних (див. [45–47]) для кожного з авторів було сформовано вектор (у просторі розмірності $N = 115$), який описує стиль автора, а також розраховано вектор \mathbf{a} , який описує узагальнений стиль. На основі даного набору векторів розраховано параметри ε_m ($m = 1, 2, \dots, M$, де $M = 13$), автори впорядковані за значеннями параметрів ε_m . Результати розрахунків проілюстровано на рисунку. Використовуючи лінійну регресійну модель, нами отримано залежність

$$\varepsilon_m \approx 0,849 - 0,015 m, \tag{8}$$

тобто $\varepsilon_0 \approx 0,849$ та $\Delta\varepsilon \approx 0,015$, коефіцієнт детермінації $R^2 \approx 0,972$. Отже, ґрунтуючись на результатах розрахунків, можна стверджувати, що параметр ε_m залежить від рангу m лінійним чином. Однак тут слід зробити важливе зауваження: йдеться про лінійну залежність від рангу для вибраної групи письменників. Іншими словами, даний емпіричний закон не є універсальним і його не можна поширювати на довільні групи письменників. Це означає, що результат не є стійким відносно змін щодо складу групи авторів. Вилучення автора із групи (або включення нового автора у групу) може мати наслідком порушення залежності (7). Більше того, саме належність чи неналежність автора до групи можна було би характеризувати по тому, чи виконується співвідношення (7).

Ще одне зауваження стосується самого принципу приєднання до більшості. При його формулю-

ванні ми вважали вектор \mathbf{a} незмінним, що насправді не так. Фактично, ми виходили з припущення, що кожен з векторів \mathbf{e}_m не впливає на вектор \mathbf{a} . Це можливо, якщо існує певне “ядро” письменників, котрі формують вектор загального стилю, а всі інші приєднуються до цієї групи. У граничному випадку дане “ядро” можна розглядати як таке, що складається з одного, найбільш успішного письменника. Тут можна було б навести аналогію з теорією середнього поля, яка використовується у фізиці критичних явищ і фазових переходів [48, 49]. Зрозуміло, що це є наближення, однак, воно не видається принциповим (хоча й накладає обмеження на застосовність результатів).

5. Висновки

Запропонована нами модель формалізує процес ідентифікації авторського стилю. А саме, модель дозволяє визначати схожість стилів різних авторів за різними групами чи категоріями лексем. На основі моделі можна отримати кількісні оцінки для відповідних характеристик та виконувати не тільки якісне, а й кількісне порівняння. Застосований у роботі підхід може бути корисним для загальної класифікації стилістики текстів, у тому числі і в автоматичному режимі.

1. M. Tsizh, V. Novosyadlyj, Yu. Holovatch, N.I. Libeskind. Large-scale structures in the Λ CDM Universe: network analysis and machine learning. *Month. Not. R. Astronom. Soc.* **495**, 1311 (2020).
2. Ю. Головач, М. Дудка, В. Блавацька, В. Пальчиков, М. Красницька, О. Мриглод. Статистична фізика складних систем у світі та у Львові. *Журнал фізичних досліджень* **22**, 2801 (2018).
3. Ю. Головач, М. Дудка, В. Блавацька, В. Пальчиков, М. Красницька, О. Мриглод. Статистична фізика складних систем. Препринт I CMP-17-06U (Львів, 2017).
4. Y. Holovatch, V. Palchykov. Complex networks of words in fables. In: *Maths Meets Myths: Complexity-Science Approaches to Folktales, Myths, Sagas, and Histories*. Edited by R. Kenna, M. MacCarron, P. MacCarron (Springer, 2016).
5. Ю. Головач, Р. Кенна, П.М. Керрон, П. Сарканич, Н. Федорак, Дж. Хосе. Математика й міти – кількісний підхід до порівняльної мітології. *Україна модерна* **27**, 108 (2020).
6. R. de Regt, C. von Ferber, Yu. Holovatch, M. Lebovka. Public transportation in UK viewed as a complex network. *Transportmetrica A* **15**, 722, (2019).
7. Yu. Holovatch, R. Kenna, S. Thurner. Complex systems: Physics beyond physics. *Eur. J. Phys.* **38**, 023002 (2017) [arXiv:1610.01002].
8. B. Berche, C. von Ferber, T. Holovatch, Yu. Holovatch. Transportation network stability: A case study of city transit. *Adv. Compl. Syst.* **15**, 1250063 (2012) [arXiv:1201.5532].
9. F. Jovanovic, C. Schinckus. Econophysics: A new challenge for financial economics. *J. Hist. Econom. Thought* **35**, 319 (2012).
10. R. Mantegna, H. Stanley. *An Introduction to Econophysics* (Cambridge University Press, 2000).
11. C. Schinckus, F. Jovanovic. Towards a transdisciplinary econophysics. *J. Econom. Method.* **20**, 164 (2013).
12. D. Stauffer. A biased review of sociophysics. *J. Stat. Phys.* **151**, 9 (2013) [arXiv:1207.6178v1].
13. C. Castellano, S. Fortunato, V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009) [arXiv:0710.3256].
14. S. Galam. Sociophysics: A review of Galam models. *Int. J. Mod. Phys. C* **19**, 409 (2008) [arXiv:0803.1800].
15. О.М. Васильєв. Модель розповсюдження чуток у суспільстві з опортуністичною поведінкою. *Журнал фізичних досліджень* **22**, 3801 (2018).
16. О.М. Васильєв, О.В. Чалий. Моделювання макроекономічної динаміки методами еконофізики. *Журнал фізичних досліджень* **17**, 4801 (2013).
17. А. Ровенчак, С. Бук. Квантові розподіли і дослідження текстів: температура та література. *Україна модерна* **27**, 29 (2020).
18. S.N. Buk, Yu. Krynytskyi, A. Rovenchak. Properties of autosemantic word networks in Ukrainian texts. *Adv. Compl. Syst.* **22**, 1950016 (2019).
19. A. Rovenchak, S. Buk. Part-of-speech sequences in literary text: Evidence from Ukrainian. *J. Quantit. Linguist.* **25**, 1 (2017).
20. A.A. Rovenchak, S. Buk. Defining thermodynamic parameters for texts from word rank-frequency distributions. *J. Phys. Stud.* **15**, 1005 (2011).
21. A.A. Rovenchak, S. Buk. Application of a quantum ensemble model to linguistic analysis. *Physica A* **390**, 1326 (2011) [arXiv:1011.5076].
22. Ю. Головач, В. Пальчиков. Лис Микита і мережі мови. *Журнал фізичних досліджень* **11**, 22 (2007).
23. О. Васильєв, О. Чалий, І. Васильєва. Математичні методи й моделі в лінгвістиці. *Україна модерна* **27**, 9 (2020).
24. О. Васильєв, І. Васильєва. Фізика за межами фізики: фізичні підходи в квантитативній лінгвістиці. *Український фізичний журнал* **65**, 141 (2020).
25. A. Vasilev, I. Vasileva. Text length and vocabulary size: Case of the Ukrainian writer Ivan Franko. *Glottometrics* **43**, 1 (2018).
26. О.М. Васильєв, О.В. Чалий, І.В. Васильєва. Про “екзотичні” задачі фізики, Вінні-Пуха та закон Зіпфа. *Журнал фізичних досліджень* **17**, 1001 (2013).

27. О.М. Васильєв, І.В. Васильєва. Особливості створення математичних моделей у лінгвістиці. *Вісник Херсонського національного технічного університету* **69** 2(3), 99 (2019).
28. Ю. Тулдава. *Проблеми і методи квантитативно-системного дослідження лексики* (Валгус, 1987).
29. Р. Пиотровский, К. Бектаев, А. Пиотровская. *Математическая лингвистика* (Высшая школа, 1977).
30. В.В. Левицкий. *Квантитативные методы в лингвистике* (Рута, 2005).
31. Н. Дарчук, І. Денисенко, О. Сірук, В. Сорокін. Теоретичні питання моделювання ідеографічного тезауруса української мови. *Українське мовознавство* **24**, 107 (2002).
32. Н.П. Дарчук, Л.А. Алексієнко, В.М. Сорокін. Параметризована база даних поетичного мовлення як джерело філологічних студій. *Мова: Проблеми прикладної лінгвістики* **9**, 15 (2004).
33. Н.П. Дарчук. Поетичний словник в аспекті мовної картини світу. *Українське мовознавство* **35**, 55 (2006).
34. Н.П. Дарчук. Дослідницький корпус української мови: основні засади і перспективи. *Вісник Київського національного університету ім. Тараса Шевченка: Літературознавство, мовознавство, фольклористика* **21**, 45 (2010).
35. Н.П. Дарчук. Автоматичний синтаксичний аналіз текстів корпусу української мови. *Українське мовознавство* **43**, 11 (2013).
36. Н.П. Дарчук. Напрямки формалізації семантики. *Мовні і концептуальні картини світу* **46**, 385 (2013).
37. Р.Т. Гром'як, Ю.І. Ковалів, В.І. Теремко. *Літературознавчий словник-довідник* (Академія, 1997).
38. H.S. Green, C.A. Hurst. *Order-Disorder Phenomena* (Interscience, 1964).
39. В.М. McCoy, T.T. Wu. *The Two-Dimensional Ising Model* (Cambridge University Press, 1973).
40. G. Salton, A. Wong, C.S. Yang. A vector space model for information retrieval. *Commun. ACM* **18**, 613 (1975).
41. В. Паулі. *Общие принципы волновой механики* (ГИТТЛ, 1947).
42. А.С. Давыдов. *Квантовая механика* (Наука, 1973).
43. Т. Веблен. *Теория праздного класса* (Прогресс, 1984).
44. H.A. Simon. A Behavioral Model of Rational Choice. *Quart. J. Econom.* **69**, 99 (1955).
45. Н.П. Дарчук. Структурно-статистическая база данных современного украинского языка на основе частотных словарей. *Слово и словарь = Vocabulum et vocabularium: Сб. науч. тр. по лексикографии* (ГрГУ, 2005).
46. Л.А. Алексієнко, Н.П. Дарчук, О.Н. Зубань, В.В. Сорокін. Параметризована база даних поетическої речі як джерело і інструмент філологічних студій. *Матеріали міжнародної конференції "Комп'ютерна лінгвістика без границь"* (Санкт-Петербург, 2004).
47. І. Васильєва. Використання комп'ютерного тезауруса в дослідженні мови поетів. *Лексикографічний бюлетень* **13**, 161 (2006).
48. Ш. Ма. *Современная теория критических явлений* (Мир, 1980).
49. А.З. Паташинский, В.Л. Покровский. *Флуктуационная теория фазовых переходов* (Наука, 1981).

Одержано 02.06.20

N.P. Darchuk, I.V. Vasileva, A.N. Vasilev

VECTOR MODEL FOR THE TEXT STYLE ANALYSIS

The application of physical approaches to the analysis of the authorial styles of Ukrainian writers has been considered. A model, where the literary styles are described in a multidimensional space with the help of unit vectors, is proposed. The numerical characteristic of the style is the scalar product of the corresponding vector and a vector that determines the general style for a group of authors. This parameter is shown to depend linearly on author's rank. This behavior confirms the hypothesis of joining the majority, according to which an author, when selecting his/her literary style, takes the style of his/her successful colleagues into account.

Keywords: physics of complex systems, quantitative linguistics, vector, energy, distribution.