

Андрій РОВЕНЧАК  
**Кількісний аналіз систем письма: алфавіт нко**

*Андрій РОВЕНЧАК – доктор фізико-математичних наук, професор катедри теоретичної фізики Львівського національного університету імені Івана Франка. Наукові інтереси: статистична фізика; конденсація Бозе–Айнштайна; системи з дробовою статистикою; кількісні методи в суспільних і гуманітарних науках; вивчення систем письма; історія науки. Електронна адреса: andrij.rovenchak@gmail.com*

**Р**егіон Західної Африки для вивчення походження й розвитку писемності цікавий тим, що протягом XIX–XX століть тут створили низку писемностей для місцевих мов, насамперед – із родини манде (силабарії вай, менде, кпелле, лоома, бамана). У 1949 р. гвінейський просвітник Соломана Канте розробив алфавіт нко для мов манден, що належать до родини манде і включають, зокрема, мови бамана, джула та манінка. Назва «нко» походить від фрази *N ko* «я кажу» мовами манден. Переважно цю писемність використовують для мови манінка в Гвінеї (манінка-морі), яка є рідною для понад 3,5 млн осіб у Гвінеї, Малі та Сьєрра-Леоне. Алфавіт нко також поширений у Ліберії, Кот-д’Івуарі й у діаспорі (переважно в Нігерії та Єгипті), а кількість користувачів може сягати від сотні тисяч до мільйона.

У статті наведено інформацію про різні аспекти дослідження алфавіту нко, які можна відобразити кількісними параметрами. Насамперед за певними принципами обчислено складність графічних форм кожної з 27 літер, наприклад, крапці відповідає 1, прямому відрізку – 2, а дузі – 3; також певні ваги присвоюють різним типам з’єднання і перетину ліній. Далі зроблено частотний аналіз розподілу літер у корпусі текстів мовою манінка, записаних алфавітом нко, обсягом понад 3,1 млн слів. Це дало, зокрема, можливість простежити, наскільки складність графічної форми знаків корелює з їхньою частотністю. Виявилось, що така кореляція не надто істотна: коефіцієнт кореляції становить  $-0,38$ , тоді як, наприклад, для азбуки Морзе в англійських текстах він сягає  $-0,82$ . Повній зворотній кореляції, коли завжди простіші знаки використовують для частотніших літер, відповідає значення  $-1$ . Також показано, що частотний аналіз може слугувати додатковим обґрунтуванням деяких орфографічних принципів у нко, зокрема в позначенні тонів.

Наступної задачею був розрахунок так званої орфографічної невизначеності: в ідеальному алфавіті, де є однозначна відповідність між фонемами (звуками) графемами (знаками) ця невизначеність дорівнює нулеві. В алфавіті якої її значення досить малі: 0,37 без урахування позначення тонів та 0,22 з їх урахуванням. Для порівняння можна вказати значення, що відповідають «старими» писемностям: в українському алфавіті – 1,12, а дещо простіша італійська ортографія забезпечує невизначеність на рівні 0,56.

Одержані результати можуть бути корисні для вивчення фонотактики, просодичних елементів, досліджень з історії письма та лексикографії, а також для порівняльних і контрастивних студій.

**Ключові слова:** алфавіт; мови манден; манінка; ортографія; складність графічної форми знаків.

## 1. Вступ

Писемність є важливим елементом культури. Завдяки їй ми можемо обмінюватися інформацією на віддалі, причому не лише розділені у просторі, але й відокремлені в часі. До винайдення сучасних засобів комунікації саме письмо – у різних його формах – було єдиним способом передати якісь відомості від людини до людини, не контактуючи безпосередньо. Відкритим залишається питання про те, з'явилася писемність один раз і потім поступово поширилася світом (моногенез) чи виникнення відбулося в кількох різних місцях незалежно (полігенез). В історії писемності натрапляємо на так зване предметне письмо, прикладами якого є *ароко* народу йоруба в Західній Африці чи *кіну* з регіону Анд у Південній Америці<sup>1</sup>. Графічні системи, які походять, найвірогідніше, від піктограм, розвинулися в різних народів до ідеографічного (ранні клинописи, китайські ієрогліфи), складового (кіпрське, японські складові азбуки) та алфавітного письма.

Найдавнішими відомими осередками розвитку писемності зараз вважають Межиріччя і Єгипет. Клинопис із долини Тигру і Євфрату та ієрогліфи з долини Нілу почергово випереджують одне одного в боротьбі за пріоритет у цьому питанні мірою того, як археологи виявляють нові знахідки<sup>2</sup>. Не виключено, що поява однієї з писемностей стала імпульсом до виникнення іншої внаслідок так званої «дифузії стимулу». У будь-якому разі єгипетські ієрогліфи зберігають першість за тяглістю писемної традиції, оскільки через ієратичну та демотичну форми збереглися обриси окремих літер у таких алфавітах, як коптський чи нубійський.

Творення писемностей не є справою лише далекої історії. З цього погляду особливо цікавим є регіон Західної Африки, де протягом XIX–XX століть виникла низка писемностей для місцевих мов<sup>3</sup>. Найбільше тут відзначилися

<sup>1</sup> Hans Jensen, *Die Schrift in Vergangenheit und Gegenwart* (Berlin: Deutscher Verlag der Wissenschaften, 1969), 21–23.

<sup>2</sup> Див. до прикладу: Larkin Mitchell, "Earliest Egyptian glyphs", *Archeology*, 52, 2 (1999): 28–29.

<sup>3</sup> Андрій Ровенчак, Джейсон Глейві, *Африканські системи письма нового часу: Регіон Суб-Сахари* (Нью Гейвен–Львів–Абіджан: Athinkra, 2009); Andriy Rovenchak and Jason M. Glavy, *African writing systems of the modern age: the Sub-Saharan region: First English Edition (revised and expanded from the original Ukrainian edition)* (New Haven–Buena Park–New Rochelle–London–Lviv–Abidjan: Athinkra LLC, 2011).

представники родини мов манде (Mandé, див. рис. 1). Ще у першій чверті XIX ст. силабарій (складове письмо) було створене для мови народу вай (Vai) в Ліберії – можливо, внаслідок «дифузії стимулу» від письма черокі (Cherokee) через американську місію<sup>4</sup>. Століттям пізніше, вже у першій половині XX ст., з'явилися силабарії менде (Mende), кпелле (Kpelle), лоома (Looma, Loma)<sup>5</sup> та бамана (Bamana, Bambara)<sup>6</sup>. У цій статті зосередимо увагу на алфавітній писемності нко з цього ж регіону.



**Рис. 1.** Родини африканських мов. Найбільші об'єднання, які класифікують як сім'ї і навіть макросім'ї: афразійські мови, ніло-сахарські мови (сюди також відносять сонгайські) та нігеро-конголезькі мови. У південній частині континенту виділяють різномірну групу койсанських мов, якими розмовляє автохтонне населення (на відміну від мов банту, що належать до нігеро-конголезьких). Мови манде належать до нігеро-конголезької макросім'ї.  
Джерело: <https://tracingafricanroots.files.wordpress.com/2015/09/mande-2.jpg>

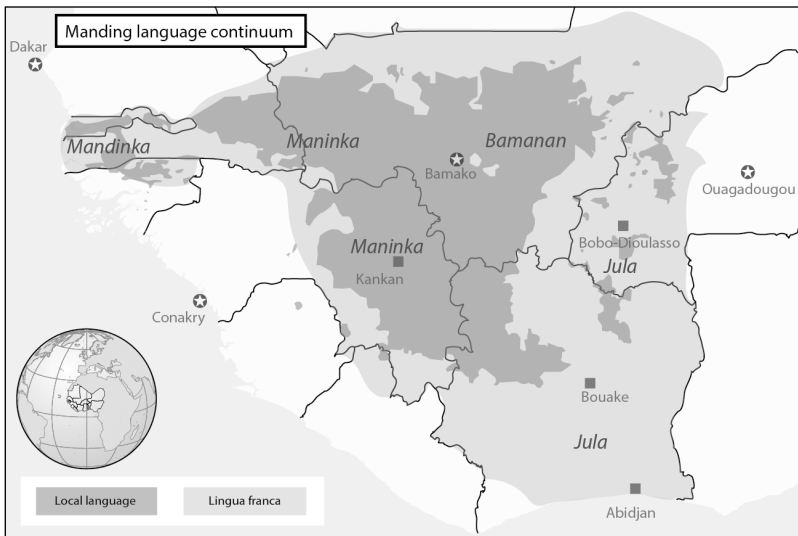
<sup>4</sup> Konrad Tuchscherer and P. E. H. Hair, "Cherokee and West Africa: Examining the origins of the Vai script", *History in Africa*, 29 (2002): 427–486.

<sup>5</sup> David Dalby, "A survey of the indigenous scripts of Liberia and Sierra Leone: Vai, Mende, Loma, Kpelle and Bassa", *African Language Studies*, 8 (1967): 1–51.

<sup>6</sup> Gérard Galtier, "Un exemple d'écriture traditionnelle mandingue : le "masaba" des Bambara-Masasi du Mali", *Journal des africanistes*, 57, 1 (1987): 255–266.

## 2. Алфавіт нко

Алфавітну писемність *нко* (Nко, N'ко, **𞤎𞤵**) використовують переважно для мові манінка в Гвінеї (Maninka-Mori). День писемності відзначають 14 квітня: саме тоді 1949 року гвінейський просвітник Соломана Канте (Sòlomáana Kántè, 1922–1987) завершив роботу над алфавітом нко, який планувався як писемність для мов манден (Manding, Mándén), до яких, крім манінка, також належать споріднені бамана та джула (дьюла; Jula, Dyula, Dioula)<sup>7</sup>, див. рис. 2. Назва писемності походить від фрази *N'ko* «я кажу» мовами манден (складовий носовий /N/ позначає займенник першої особи однини). Нко є одним із трьох основних способів запису мов манден<sup>8</sup>, разом із графікою на латинській основі та адаптаціями арабського письма, відомими під загальною назвою *аджамі*<sup>9</sup> (арабське **عجمي**).



**Рис. 2.** Мови манден. Різні варіанти (бамана, манінка, джула, мандінка) використовують як *lingua franca* в західноафриканському регіоні<sup>10</sup>.

<sup>7</sup> David Dalby, "Further indigenous scripts of West Africa: Manding, Wolof and Fula alphabets and Yoruba 'holy' writing", *African Language Studies*, 10 (1969): 161–181; Valentin Vydrine, *Manding-English Dictionary (Maninka, Bamana)*, Vol. 1: A, B, D-DAD, supplemented by some entries from subsequent volumes (St. Petersburg: Dimitry Bulanin Publishing House, 1999); Christopher Wyrod, "A social orthography of identity: the N'ko literacy movement in West Africa", *International Journal of the Sociology of Language*, 192 (2008): 27–44.

<sup>8</sup> Coleman Donaldson, "Orthography, Standardization, and Register: The Case of Manding", in *Standardizing Minority Languages: Competing Ideologies of Authority and Authenticity in the Global Periphery*, eds. Pia Lane, James Costa, Haley De Korne (New York: Routledge, 2017), 175–199

<sup>9</sup> Valentin Vydrin, "Ajami script for the Mande languages", in *The Arabic script in Africa: Studies in the use of a writing system*, eds. Meikal Mumin, Kees Versteegh. (Leiden–Boston: Brill, 2014), 199–224.

<sup>10</sup> Coleman Donaldson, "Map of the Manding language continuum". *Zenodo* (2018, December 30), <https://doi.org/10.5281/zenodo.2528947>

Нко є, мабуть, найуспішнішою з нових писемностей, і не лише африканських. Мова манінка є рідною для понад 3,5 млн осіб<sup>11</sup> у Гвінеї, Малі та Сьєрра-Леоне. Завдяки активній діяльності спільноти нко цей алфавіт поширений також у Ліберії та Кот-д'Івуарі й у діяспорі (переважно в Нігерії та Єгипті), а кількість користувачів може сягати від сотні тисяч до мільйона<sup>12</sup>.

Алфавіт нко налічує 27 літер (див. табл. 1), із яких 7 на позначення голосних, 1 для складового носового /N/, 19 для приголосних. Окремими знаками позначають носові приголосні, що виникають внаслідок асиміляції. За допомогою діакритичних знаків набір літер можна істотно розширити для позначення іншомовних звуків. У письмі нко є своїх 10 знаків для цифр, а також низка пунктуаційних знаків, які вживають разом із арабською пунктуацією. Напрямок письма – справа наліво.

Таблиця 1. Літери алфавіту нко.

a	l	b	F	s	□	h	h
e	o	p	ɸ	gb	∇	w	ɖ
i	ʏ	t	ɓ	f	ɖ	y	ɸ
ε	ʌ	j	ɣ	k	ɥ		
u	ɥ	c	ɿ	l	ɸ	<i>dental</i> → n	ɿ
o	ɔ	d	ɓ	m	Δ	y→ɿ	ɸ
ɔ	ɿ	r	ɸ	ɿ	ɸ		
N	ɸ	rr	ɸ	n	ɿ	g	ɿ

Позначення «*dental* → n» відповідає асиміляції зубних після носових, а «y → ɿ» – аналогічній асиміляції фонемі /y/, що відповідає традиційному в африканістиці написанню для звука [j].

Вивчення писемностей зазвичай відбувається на межі кількох наукових дисциплін, скажімо, палеографії<sup>13</sup>, що досліджує рукописні пам'ятки, істо-

<sup>11</sup> “Maninkakan, Eastern”, in *Ethnologue: Languages of the World, Twentieth edition*, eds. Gary F. Simons and Charles D. Fennig (Dallas, Texas: SIL International, 2017), online version: <https://www.ethnologue.com/language/emk>.

<sup>12</sup> Andrij Rovenchak, “Quantitative studies in the corpus of Nko periodicals”, in *Recent Contributions to Quantitative Linguistics*, eds. Arjuna Tuzzi, Martina Benešová, Ján Mačutek (Berlin–Boston: Mouton de Gruyter, 2015). 125–138.

<sup>13</sup> Ірина Нінелівна Войцехівська, «Палеографія», у кн. *Енциклопедія історії України: у 10 т. Т. 8: Па–Прик*, редкол.: В. А. Смолій (голова) та ін.; Інститут історії України НАН України (Київ: Наукова думка, 2011), 34.

рію й еволюцію письма, послуговується методами історії та мовознавства. Ігнац Гельб<sup>14</sup> увів 1952 р. в обіг термін «граматологія» на позначення науки про письмо. Книга Жака Дерріди «Про граматологію»<sup>15</sup>, що вийшла друком 1967 р., містить, зокрема, філософський аналіз поняття «знак». Однак, попри давню традицію, можна сказати, що дослідження систем письма й сьогодні перебуває в зародковому стані. Так, наприклад, уважає Амалія Гнанадесікан<sup>16</sup>. У наступних розділах розглянемо підходи до аналізу писемності, що ґрунтуються на застосуванні кількісних методів. Маючи на меті не переобтяжувати виклад математичними виразами, все ж проілюструємо міркування числовими даними, даючи змогу читачам відчутти справедливість зроблених висновків на підставі порівнянь.

### 3. Складність графічних форм

Складність певного об'єкта можна визначити як властивість, пов'язану зі сукупністю його багатьох взаємопов'язаних ознак, частин, якостей тощо<sup>17</sup>. Це поняття досить широке<sup>18</sup>, й у різних галузях науки використовують дуже різні способи оцінки – від простого підсумовування кількості компонентів<sup>19</sup> до складних алгоритмів<sup>20</sup>, що можуть включати нетривіальний математичний апарат<sup>21</sup>. Для кількісної характеристики складності графічних знаків Габріель Альтман<sup>22</sup> запропонував оцінити кожен графічний елемент за такими принципами:

- крапці відповідає 1;
- прямій лінії ( $/$ ,  $|$ ) – 2;
- дузі  $\leq 180^\circ$  ( $\oslash$ ,  $\smile$ ,  $\smile$ ) – 3;
- неперервне з'єднання (як у літерах S, C) оцінюється 1;
- ламане з'єднання (L, V, T) – 2;
- перетин (X) – 3.

<sup>14</sup> Ignace Gelb, *A Study of Writing* (Chicago: University of Chicago Press, 1952).

<sup>15</sup> Jacques Derrida, *De la grammatologie* (Paris: Les Éditions de Minuit, 1967).

<sup>16</sup> Amalia E. Gnanadesikan, "Towards a typology of phonemic scripts", *Writing Systems Research* 9 (2017): 14–35.

<sup>17</sup> *Словник української мови: в 11 томах. Том 9* (Київ: Наукова думка, 1978), 277–278.

<sup>18</sup> Про поняття складності див. також у статті Юрія Головача в цьому номері.

<sup>19</sup> Johann Nichols, "Linguistic complexity: a comprehensive definition and survey", in *Language Complexity as an Evolving Variable*, eds. Geoffrey Sampson, David Gil, and Peter Trudgill (New York: Oxford University Press, 2009), 110–125.

<sup>20</sup> Damián Reyes, Alan L. Erera, and Martin W. P. Savelsbergh, "Complexity of routing problems with release dates and deadlines", *European Journal of Operational Research*, 266, 1 (2018): 29–34; Chao Li and Pengjian Shang, "Complexity analysis based on generalized deviation for financial markets", *Physica A*, 494 (2018): 118–128.

<sup>21</sup> Bruce J. West, "Colloquium: Fractional calculus view of complexity: A tutorial", *Reviews of Modern Physics*, 86, 4 (2014): 1169–1186.

<sup>22</sup> Gabriel Altmann, "Script complexity", *Glottometrics*, 8 (2004): 68–74.

Певні модифікації методу щодо врахування кількості з'єднань запропонував Ян Мачутек<sup>23</sup>, а надалі також було враховано суцільне заповнення деяких елементів, що трапляється, зокрема, у вайському письмі<sup>24</sup>. Складність кожної літери обчислюють як суму відповідних складників.

Можна сказати, що розраховане за вказаними принципами значення складності буде приблизно відповідати інтуїтивному уявленню про це поняття, скажімо, літера «Ж» складніша за «Х», яка своєю чергою складніша за «І». Природно, отже, очікувати, що винахідник письма буде орієнтуватися на своє інтуїтивне відчуття і не створюватиме знаків складної форми на позначення часто вживаних звуків (а також складів чи інших комбінацій). Таку гіпотезу зручно перевірити на нових системах письма, створених для конкретних мов, оскільки адаптовані писемності обтяжені спадком тих мов, для яких їх застосовували початково.

У таблиці 2 наведено результати, одержані на підставі корпусу текстів мовою манінка<sup>25</sup>, записаних алфавітом нко, обсяг якого понад 3,1 млн слів<sup>26</sup>. Як видно з перших шести найчастотніших літер, що становлять понад половину всіх літер у текстах, складність лише незначно корелює з частотою вживання. Для числової оцінки можна розрахувати так званий коефіцієнт кореляції Пірсона  $r$ , який змінюється від  $-1$  (зворотна кореляція) до  $+1$  (пряма кореляція). Для літер нко отримаємо  $r = -0,38$ . Це значення як таке не є дуже інформативним, однак його можна порівняти, наприклад, зі складністю азбуки Морзе, яку було створено власне з таких міркувань, щоб найуживаніші літери мали найпростіші позначення. Коефіцієнт кореляції Пірсона для знаків азбуки Морзе в англійських текстах становить  $r = -0,82$ , тобто очікувано близький до  $-1$ , на відміну від значення для алфавіту нко, де така кореляція виявилася дуже незначною. Навіть більше, форма найчастотнішої (і найпростішої) літери на позначення /a/ безпосередньо пов'язана з арабською літерою 'аліф < ʾ >.

<sup>23</sup> Ján Mačutek, "Runes: complexity and distinctivity", *Glottometrics*, 16 (2008): 1–16.

<sup>24</sup> Andrij Rovenchak, Ján Mačutek, and Charles Riley, "Distribution of complexities in the Vai script", *Glottometrics*, 18 (2009): 1–12.

<sup>25</sup> Valentin Vydrin, Kirill Maslinsky, Andrij Rovenchak, Ibrahimia Sory II Condé et al., *Corpus Maninka de Référence et Corpus N'ko* (2016–2017), mode d'accès: <http://cormand.humanum.fr/cormani/projet.html>.

<sup>26</sup> Valentin Vydrin, Andrij Rovenchak, and Kirill Maslinsky, "Maninka Reference Corpus: A Presentation", in *Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 11: TALAF* (Paris: Association Francophone pour la Communication Parlée (AFCP) et Association pour le Traitement Automatique des Langues (ATALA), 2016), 87–94.

Таблиця 2. Частоти і складність літер нко.

Літера	Фонетичне значення	Абсолютна частота	Відносна частота, $f_i$	Складність, $C_i$
l	/a/	1707136	18,9 %	2
q	/l/	744519	8,2 %	10
ч	/k/	644025	7,1 %	10
у	/i/	625359	6,9 %	12
ш	/d/	538824	6,0 %	18
^	/ε/	482959	5,3 %	6
		<b>Разом:</b>	<b>52,4 %</b>	

Отже, можемо зробити висновок, що міркування простоти накреслення літер не варто вважати основними для розробників писемності. Це твердження вимагає ще перевірки на інших системах письма, створених «під конкретну мову», серед яких і тубільні африканські, наприклад, вайський силабарій у Ліберії чи бамумське (Bamum) письмо в Камеруні, а також силабарій для мови черокі на південному сході США.

#### 4. Частотний аналіз і ортографічні принципи

Здавалось би, що алфавіт, створений для конкретної мови (чи групи мов) повинен мати просту ортографію, а в ідеалі – однозначну відповідність літер і фонем (чи спрощено – звуків). Цікаво з цього погляду проаналізувати нко, особливо з урахуванням того, що мова манінка є тональною.

Кількісну оцінку, яку ми застосуємо, називають *ортографічною невизначеністю* (англ. *orthographic uncertainty*). Цю величину розраховують за формулою, що походить з теорії інформації:


$$U = \frac{1}{N} \sum_k f_k \log_2 k,$$

де  $N$  – загальна кількість фонем, а  $f_k$  – кількість фонем, що мають  $k$  способів запису. Для ідеального алфавіту – з однозначною відповідністю «фонема–літера» – матимемо  $U = 0$ .

В алфавіті нко є два відхилення від такої однозначної відповідності<sup>27</sup>. Перше пов'язане з тим, що деякі приголосні в певному фонетичному оточенні назалізуються (див. таблицю 1), унаслідок чого тверде /n/ і м'яке /ɲ/ матимуть по два варіанти написання. Крім того, можна вважати, що фонема /r/ реалізується також як /rr/, що позначається окремою рідкісною літерою,

<sup>27</sup> Andrij Rovenchak and Valentin Vydrin, "Quantitative properties of the Nko writing system", in *Text and Language: Structures – Functions – Interrelations. Quantitative perspectives*, eds. Peter Grzybek, Emmerich Kelih, Ján Mačutek (Wien: Praesens, 2010), 171–181.



яка трапляється в записі ідеофонів. Друге відхилення є наслідком правила «економного написання» *сбаралі* (*gbàrali*): у двох складах із однаковими короткими голосними, що мають однаковий тон, першу не пишуть (за деякими винятками). Тобто короткі голосні матимуть по два варіанти написання, один із яких – це нульовий (порожній) символ. Класичним прикладом тут може бути одне з найуживаніших повнозначних слів, <  > *bólo* ‘рука’, у записі якого (читаємо справа наліво!) відсутнє перше /o/.

Розрахунок орфографічної невизначеності в такому разі дає  $U = 0,37$ . Якщо подібні міркування повторити з урахуванням позначення тонів, внаслідок чого кількість голосних «фонем» значно збільшиться, то отримаємо менше значення  $U = 0,22$ . Отже, навіть у новому алфавітному письмі необов’язково досягають однозначного написання, якому відповідає  $U = 0$ , хоча й значення орфографічної невизначеності залишається малим порівняно зі «старими» писемностями.

Для порівняння український алфавіт<sup>28</sup> характеризується невизначеністю  $U = 1,12$ . Внесок у це значення дають неоднозначності відображення голосних (наприклад, <a> і <я> на позначення /a/), м’яких приголосних (як-от <c> перед /i, e, ю, я/ і /ь/), п’ять варіантів позначення /й/ – <й, е, і, ю, я>, врахування асиміляції приголосних (наприклад, <к> у словах ‘якби’, ‘вокзал’ вимовляється як /г/) тощо. Аналогічний розрахунок в італійській орфографії<sup>29</sup> дає  $U = 0,56$  – таке низьке значення можна пояснити насамперед простотою запису голосних і меншою кількістю скупчень приголосних. У нко ж орфографія ще простіша.

Цікаво, що особливості правопису нко можна обґрунтувати на підставі саме кількісних показників. Продемонструємо це на двох прикладах, перший із яких пов’язаний із правилом *сбаралі*, а другий – із принципом позначення тонів.

Виявляється, що пари складів із однаковими короткими голосними досить часто трапляються в манінка, і в багатьох випадках саме вони є частотнішими за інші можливі комбінації<sup>30</sup>, див. таблицю 3. Тому правило *сбаралі* справді має кількісне підґрунтя, хоч і більш докладний аналіз вимагає врахуванням тонів голосних.

Орфографія нко характеризується послідовним позначенням тонів. Їх у манінка виділяють чотири: високий, низький, висхідний і плаваючий. Оскільки голосні в манінка можуть бути короткі й довгі, то діакритичні знаки для тону враховують також довготу голосних, що дозволило обмежи-

<sup>28</sup> Solomija Buk, Ján Mačutek, and Andrij Rovenchak, “Some properties of the Ukrainian writing system”, *Glottometrics*, 16 (2008): 63–79.

<sup>29</sup> Gerald Bernhard and Gabriel Altmann, “The phoneme–grapheme relationship in Italian”, in *Analysis of Script. Properties of Characters and Writing Systems*, eds. Gabriel Altmann and Fan Fengxiang (Berlin: de Gruyter), 11–21.

<sup>30</sup> Andrij Rovenchak, “Phoneme distribution, syllabic structure, and tonal patterns in Nko texts”, *Mandenkan. Bulletin semestriel d’études linguistiques mandé*, 47 (2011): P. 77–96.

Таблиця 3. Пари складів із короткими голосними\*.

	Ca	Ce	Cε	Ci	Co	Cɔ	Cu
Ca	<b>1967</b>	53	239	961	197	130	421
Ce	128	<b>189</b>	1	20	42	5	18
Cε	183	16	<b>887</b>	85	18	65	37
Ci	<b>695</b>	63	94	386	72	81	30
Co	177	9	11	7	<b>460</b>	12	2
Cɔ	241	16	28	21	12	<b>683</b>	3
Cu	<b>307</b>	13	14	8	3	17	220

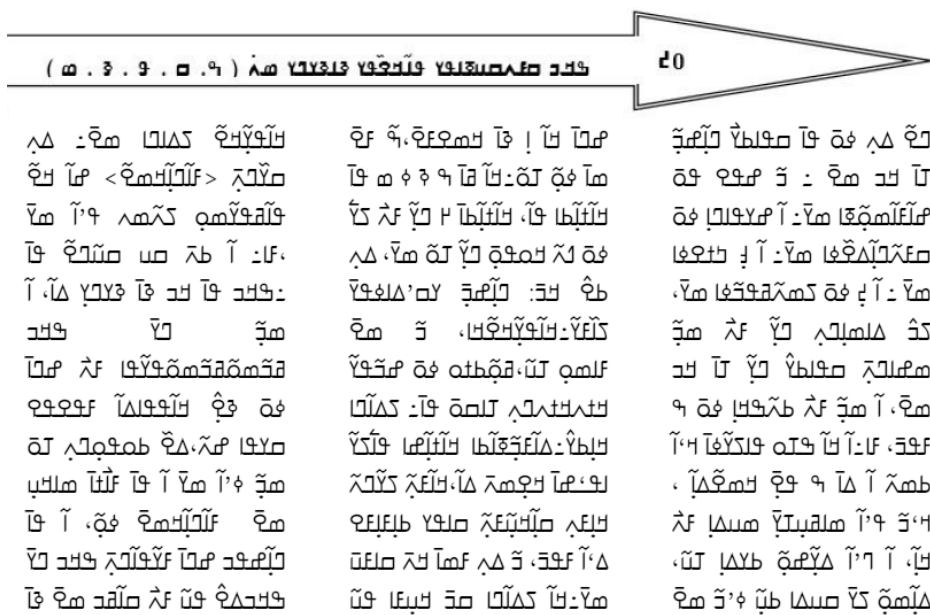
\* "C" відповідає приголосним. Напівгрубим шрифтом і підкресленням показано найбільші значення. Як бачимо, переважно вони лежать на діагоналі, обгрунтовуючи певною мірою правило *сбаралі*.

ти кількість відповідних літер. Переважно трапляються високий і низький тони, причому перший – удвічі частіше за другий<sup>31</sup>. Тому цілком логічним виглядає вибір не позначати на письмі саме високий тон над літерами для коротких голосних, оскільки вони трапляються значно частіше за довгі.

Для ілюстрації на рис. 3 наведено фрагмент сторінки з гвінейського періодичного видання "Wentere". Звернемо увагу на те, що крапки під голосними позначають назалізацію. Водночас треба сказати, що *відсутність* діакритичного знака над останньою голосною слова відповідає *наявності* тонального артикля, який реалізується як плаваючий тон – причину такого орфографічного рішення також можна спробувати дослідити за допомогою кількісного аналізу.

Як бачимо з двох проаналізованих випадків, кількісні міркування добре обгрунтовують особливості ортографії nko. Цей висновок є очікуваним, оскільки правила правопису не повинні бути наслідком довільних рішень. Зважаючи на вказані кількісні характеристики, можемо також припустити, що для впровадження цих принципів було достатньо якісних оцінок – того, як людина відчуває рідну мову: пари складів з однаковими голосними в манінка справді трапляються *істотно частіше* за інші, а високий тон *удвічі* переважає низький.

<sup>31</sup> Andrij Rovenchak, "Quantitative studies in the corpus of Nko periodicals", in *Recent Contributions to Quantitative Linguistics*, eds. Arjuna Tuzzi, Martina Benešová, Ján Mačutek (Berlin–Boston: Mouton de Gruyter, 2015). 125–138.



**Рис. 3.** Фрагмент сторінки 4 (записаної як «04» на вістрі стрілки у верхній правій частині сторінки: 40, тобто справа наліво, на відміну від того, як числа записують в арабському письмі) місячника “Wentere” (грудень 2015 р.), джерело: *Bibliothèque électronique N’ko*, <http://cormand.huma-num.fr/maninkabiblio/periodiques/wentere12.pdf>

## 5. Висновки

На кількох простих прикладах було продемонстровано низку кількісних підходів до аналізу писемности на прикладі алфавіту нко, який використовують як одну із систем письма для мови манінка та кількох споріднених мов у Західній Африці. Показано, до якої міри кількісні дані можуть прояснити функціонування писемности, зокрема з погляду форми літер і правил правопису.

На підставі текстів, записаних алфавітом нко, для мови манінка також було одержано деякі висновки, що ґрунтуються на законах квантитативної лінгвістики, таких як закон Ціпфа та закон Менцерата–Альтмана<sup>32</sup>. Зокрема, аналіз наявного електронного корпусу текстів періодики обсягом близько 650 тис. слів дозволив оцінити розмір так званого «ядерного словника» (*core vocabulary*) та уточнити поняття мори як одиниці вимірювання довжини складів.

Результати описаних кількісних досліджень можуть бути корисними для вивчення фонотактики, просодичних елементів, у галузі історії письма та лексикографії, а також для порівняльних і контрастивних студій.

<sup>32</sup> Andrij Rovenchak, “Quantitative studies in the corpus of Nko periodicals”, in *Recent Contributions to Quantitative Linguistics*, eds. Arjuna Tuzzi, Martina Benešová, Ján Mačutek (Berlin–Boston: Mouton de Gruyter, 2015). 125–138.

Andrij ROVENCHAK  
**A Quantitative Analysis of Writing Systems:  
The N'ko Alphabet**

*Andrij ROVENCHAK – Doctor of Sciences (Physics and Mathematics), Professor of the Department for Theoretical Physics, Ivan Franko National University of Lviv. Scientific interests: statistical physics; Bose–Einstein condensation; systems obeying fractional statistics; quantitative methods in social sciences and humanities; studies of writing systems; history of science. Email: andrij.rovenchak@gmail.com*

The region of West Africa is of interest for the study of the origin and development of writing because a number of scripts were created there for several local languages during the 19th and 20th centuries, especially for the Mande family (the Vai, Mende, Kpelle, Looma, and Bamana syllabaries). In 1949 the Guinean enlightener Soulemayne Kanté developed the N'Ko alphabet for the Manding (Manden) languages, which belong to the Mande family and include, in particular, Bamana (Bambara), Jula (Dyula, Dioula), and Maninka. The name "N'Ko" originates from the phrase *N ko* 'I say' in Manding languages. This script is predominantly used in Guinea for Maninka (Maninka-Mori), which is native to more than 3.5 million people in Guinea, Mali and Sierra Leone. The N'Ko alphabet is also widely used in Liberia, the Côte d'Ivoire, and the African diaspora (mainly in Nigeria and Egypt) by a hundred thousand to a million persons.

This article provides information about studies of various aspects of the N'Ko alphabet. First of all, the complexity of the graphic forms of each of the 27 letters is calculated according to certain principles. For example, the point corresponds to 1, the straight line segment is 2, and the arc is 3; also certain weight is given to various types of connections and crossed lines. A frequency analysis of the distribution of letters is undertaken in the corpus of Maninka texts written in N'Ko, with more than 3.1 million words. This made it possible, in particular, to trace the extent to which the complexity of the graphic form of the signs correlates with their frequency. It appears that such a correlation is not very significant: the correlation coefficient is  $-0.38$ , whereas, for example, for the Morse code in English texts it reaches  $-0.82$ . The full inverse correlation, when simpler characters are always used to represent more frequent letters, corresponds to  $-1$ . It has also been shown that frequency analysis can serve as a further justification for certain orthographic principles in N'Ko, particularly of tone notation.

The next task was to calculate orthographic uncertainty: in an ideal alphabet, where there is a one-to-one correspondence between phonemes (sounds) and graphemes (signs), this uncertainty is equal to zero. In the N'Ko alphabet, its values are quite small: 0.37 without taking into account the tone notation, and 0.22 with tone notation. For comparison, the values corresponding to some "old"

writing systems are as follows: in the Ukrainian alphabet, it equals 1.12, while a slightly simpler Italian orthography provides uncertainty at the level of 0.56.

The results obtained in this study can be useful for studying phonotactics, prosodic elements, and the history of writing and lexicography, as well as in comparative and contrastive studies.

**Keywords:** alphabet; Manding languages; Maninka; orthography; complexity of the graphic form of signs

### Bibliography

- Altmann, Gabriel. "Script complexity". *Glottometrics* 8 (2004): 68–74.
- Bernhard, Gerald and Gabriel Altmann. "The phoneme–grapheme relationship in Italian". In *Analysis of Script. Properties of Characters and Writing Systems*, edited by Gabriel Altmann and Fan Fengxiang, 11–21. Berlin: de Gruyter.
- Buk, Solomija, Ján Mačutek, and Andrij Rovenchak. "Some properties of the Ukrainian writing system". *Glottometrics* 16 (2008): 63–79.
- Dalby, David. "A survey of the indigenous scripts of Liberia and Sierra Leone: Vai, Mende, Loma, Kpelle and Bassa". *African Language Studies* 8 (1967): 1–51.
- "Further indigenous scripts of West Africa: Manding, Wolof and Fula alphabets and Yoruba 'holy' writing". *African Language Studies* 10 (1969): 161–181.
- Derrida, Jacques. *De la grammatologie*. Paris: Les Éditions de Minuit, 1967.
- Donaldson, Coleman. "Orthography, Standardization, and Register: The Case of Manding". In *Standardizing Minority Languages: Competing Ideologies of Authority and Authenticity in the Global Periphery*, edited by Pia Lane, James Costa, and Haley De Korne, 175–199. New York: Routledge, 2017.
- "Map of the Manding language continuum". *Zenodo* (2018, December 30), <https://doi.org/10.5281/zenodo.2528947>
- Galtier, Gérard. "Un exemple d'écriture traditionnelle mandingue : le "masaba" des Bambara-Masasi du Mali". *Journal des africanistes* 57, no. 1 (1987): 255–266.
- Gelb, Ignace. *A Study of Writing*. Chicago: University of Chicago Press, 1952.
- Gnanadesikan, Amalia E. "Towards a typology of phonemic scripts". *Writing Systems Research* 9 (2017): 14–35.
- Jensen, Hans. *Die Schrift in Vergangenheit und Gegenwart*, 21–23. Berlin: Deutscher Verlag der Wissenschaften, 1969.
- Li, Chao and Pengjian Shang. "Complexity analysis based on generalized deviation for financial markets". *Physica A* 494 (2018): 118–128.
- Mačutek, Ján. "Runes: complexity and distinctivity". *Glottometrics* 16 (2008): 1–16.
- "Maninkakan, Eastern". In *Ethnologue: Languages of the World, Twentieth edition*, edited by Gary F. Simons and Charles D. Fennig. Dallas, Texas: SIL International, 2017. Online version: <https://www.ethnologue.com/language/emk>.
- Mitchell, Larkin. "Earliest Egyptian glyphs". *Archeology* 52, no. 2 (1999): 28–29.
- Nichols, Johann. "Linguistic complexity: a comprehensive definition and survey". In *Language Complexity as an Evolving Variable*, edited by Geoffrey Sampson, David Gil, and Peter Trudgill, 110–125. New York: Oxford University Press, 2009.

Reyes, Damián, Alan L. Erera, and Martin W. P. Savelsbergh. "Complexity of routing problems with release dates and deadlines". *European Journal of Operational Research* 266, no. 1 (2018): 29–34.

Rovenchak, Andrij. "Phoneme distribution, syllabic structure, and tonal patterns in Nko texts". *Mandenkan: Bulletin semestriel d'études linguistiques mandé* 47 (2011): 77–96.

– "Quantitative studies in the corpus of Nko periodicals". In *Recent Contributions to Quantitative Linguistics*, edited by Arjuna Tuzzi, Martina Benešová, and Ján Mačutek, 125–138. Berlin–Boston: Mouton de Gruyter, 2015.

Rovenchak, Andrij ta Jason Glavy. *Afrykans'ki systemy pys'ma novoho chasu: Region Sub-Sakhary*. New Haven–Lviv–Abidjan: Athinkra, 2009.

Rovenchak, Andrij and Jason M. Glavy. *African writing systems of the modern age: the Sub-Saharan region: First English Edition (revised and expanded from the original Ukrainian edition)*. New Haven–Buena Park–New Rochelle–London–Lviv–Abidjan: Athinkra LLC, 2011.

Rovenchak, Andrij, Ján Mačutek and Charles Riley. "Distribution of complexities in the Vai script". *Glottometrics*, 18 (2009): 1–12.

Rovenchak, Andrij and Valentin Vydrin. "Quantitative properties of the Nko writing system". In *Text and Language: Structures – Functions – Interrelations. Quantitative perspectives*, edited by Peter Grzybek, Emmerich Kelih, and Ján Mačutek, 171–181 Wien: Praesens, 2010.

Tuchscherer, Konrad and P. E. H. Hair. "Cherokee and West Africa: Examining the origins of the Vai script". *History in Africa* 29 (2002): 427–486.

Voitsekhivs'ka, Iryna Ninelivna. "Paleohrafiia". U kn. *Entsyklopediia istorii Ukrainy: u 10 t. T. 8: Pa–Pryk*, redkol.: V. A. Smolii (holova) ta in.; Instytut istorii Ukrainy NAN Ukrainy, 34. Kyiv: Naukova dumka, 2011.

Vydrin, Valentin. "Ajami script for the Mandé languages". In *The Arabic script in Africa: Studies in the use of a writing system*, edited by Meikal Mumin and Kees Versteegh, 199–224. Leiden–Boston: Brill, 2014.

Vydrin, Valentin, Kirill Maslinsky, Andrij Rovenchak, Ibrahimia Sory II Condé et al. *Corpus Maninka de Référence et Corpus N'ko*. 2016–2017. Mode d'accès: <http://cormand.huma num.fr/cormani/projet.html>.

Vydrin, Valentin, Andrij Rovenchak, and Kirill Maslinsky. "Maninka Reference Corpus: A Presentation". In *Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 11: TALAF*, 87–94. Paris: Association Francophone pour la Communication Parlée (AFCP) et Association pour le Traitement Automatique des Langues (ATALA), 2016.

Vydrine, Valentin. *Manding–English Dictionary (Maninka, Bamana), Vol. 1: A, B, D–DAD, supplemented by some entries from subsequent volumes*. St. Petersburg: Dimitry Bulanin Publishing House, 1999.

West, Bruce J. "Colloquium: Fractional calculus view of complexity: A tutorial". *Reviews of Modern Physics* 86, no. 4 (2014): 1169–1186.

Wyrod, Christopher. "A social orthography of identity: the N'ko literacy movement in West Africa". *International Journal of the Sociology of Language* 192 (2008): 27–44.