

# Ізоморфна латинізація староукраїнських і новоукраїнських текстів: від історичної писемної спадщини до сучасних лінгвістичних технологій

Maksym O. Vakulenko

Institute of Problems of Artificial Intelligence of MES and NAS of Ukraine;  
State Scientific and Technical Library of Ukraine, Kyiv, Ukraine

**Abstract** This article analyses the Old Ukrainian and New Ukrainian alphabets and proposes a system of their isomorphic (simple-correspondent) transliteration into the Latin script, which is necessary for the effective inclusion of Ukraine in international cooperation in the field of information and linguistic technologies, preservation of historical written monuments, development of multilingual corpus linguistics and computational lexicography. Particular attention is paid to the letters that, during the historical development of the Ukrainian language, have had different readings, particularly <є>, <и>, <ъ>, <ь>.

**Keywords** Ukrainian alphabets. Cyrillic-Latinic transliteration. Isomorphic latinisation. Old Ukrainian language. New Ukrainian language. Historical written monuments. Linguistic technologies.

**Summary** 1 Вступні зауваги. – 2 Вимоги до транслітераційних систем. – 3 Транслітераційні надбання для України на початок 2022 року. – 4 Українські історичні алфавіти. – 5 Висновки.



## Peer review

Submitted 2022-10-10  
Accepted 2022-11-25  
Published 2022-12-15

## Open access

© 2022 Vakulenko | 4.0



**Citation** Vakulenko, M.O. (2022). "Ізоморфна латинізація староукраїнських і новоукраїнських текстів: від історичної писемної спадщини до сучасних лінгвістичних технологій". *Balcania et Slavia*, 2(1), 73-90.

**DOI** 10.30687/BES/2785-3187/2022/01/004

Розвиток інформаційних і лінгвістичних інструментів опрацювання природної мови породжує, зокрема, потребу в автоматизованих системах взаємнооднозначної романізації нелатинських алфавітів, які б забезпечували повну еквівалентність вихідного й транслітерованого текстів. З метою збереження історичної писемної спадщини, такі системи повинні бути придатними також і для алфавітів, які були вживані протягом історичного розвитку мови.

У статті проаналізовано староукраїнські та новоукраїнські алфавіти і запропоновано систему їхньої взаємнооднозначної транслітерації латиницею, що необхідно для ефективного включення України в міжнародну співпрацю в сфері інформаційних і лінгвістичних технологій, збереження історичних писемних пам'яток, багатомовної корпусної лінгвістики та комп'ютерної лексикографії тощо.

Особливу увагу приділено літерам, які протягом історичного розвитку української мови мали різне прочитання, насамперед <г>, <є>, <и>, <ь>, <ъ>. Таблиці транслітерації і ретранслітерації охоплюють українські алфавіти від середини XIV століття по теперішній час.

Ключові слова: українські алфавіти, кирилично-латинична транслітерація, ізоморфна латинізація, староукраїнська мова, новоукраїнська мова, історичні писемні пам'ятки, лінгвістичні технології.

## 1 Вступні зауваги

Світові глобалізаційні процеси зумовлюють формування єдиного інформаційного простору, в якому текстова інформація подається переважно за допомогою латинського письма. Тому країни, мови яких використовують нелатинську графіку, виробили системи транслітерації свого алфавіту на латиницю, якими послуговуються в міжнародному спілкуванні: *Hiroshima* (яп.), *Pyeongyang* (кор.), *Xinhua* (кит.), *Hurghada* (ар.), *Iraq* (ар.), *Haifa* (ід.) і под.

На сьогодні використання латиниці вийшло далеко за рамки відтворення окремих назв (власних і загальних) та екзотизмів. Останнім часом питання латинізації алфавітів мов світу досліджують переважно в контексті застосування автоматичних інструментів для транслітерації старовинних і сучасних санскритських (Cai, Wang 2018), яванських (Razak et al. 2018), арабських та інших нелатинських текстів (Bogacz, Klingmann, Mara 2017), де машинну транслітерацію розглядають як необхідний складник інформаційного пошуку (Naji, Allan 2016). Транслітерація на латиницю необхідна для скриптів програмного забезпечення в процедурах опрацювання природної

мови (англ. *Natural Language Processing - NLP*) - це семантичний аналіз повідомлень, пошук термінів і ключових слів, видобуток інформації, машинний переклад, оптичне зчитування текстової інформації, озвучування тексту, розпізнавання мовлення тощо (Vakulenko 2022).

Уніфікована стандартизована система передачі українських назв латиницею (українська латиниця, УЛ) необхідна не тільки для міжнародного спілкування України, а й для ефективного розвитку її лінгвістичних та інформаційних технологій. Тому впровадження відповідної транслітераційної системи, яка повністю відповідає сучасним вимогам, є актуальним завданням.

Наприклад, недавно розроблена національна електронна науково-інформаційна система (URIS) потребує адекватного латинізованого представлення наукометричної інформації, яка включає в себе імена українських науковців, назви їхніх публікацій, назви видань і міст видання тощо. Усі ці тексти стосуються переважно сучасної української мови, але мусять передбачати й коректну передачу літер, уживаних протягом її історичного розвитку - зокрема, алфавітів Лаврентія Кукіля (Зізанія), Мелетія Смотрицького, Олексія Павловського, Пантелеймона Куліша, Михайла Максимовича та інших (див. Русанівський 2004; ІВП).

Своєю чергою, відображення старовинних українських текстів латиницею необхідне у зв'язку з потребою збереження культурної та мовної спадщини нашої країни, а також забезпечення її "видимості" та доступності для фахівців і користувачів. Включення текстової інформації щодо українських історичних документів у міжнародні бази та інформаційні ресурси потрібне для проведення досліджень у рамках міжнародних наукових ініціатив - таких, як Європейська дослідницька інфраструктура спадщинної науки (E-RIH) програм ЄС "Horizon 2020" і "Horizon Europa". Зокрема, одним із пріоритетних завдань E-RIH на наступні роки планується зробити захист і збереження друкованої історичної спадщини (Pezzati 2019). При цьому опрацювання інформації щодо цієї спадщини відбувається з використанням інформаційних технологій, які застосовують, зокрема, транслітерацію нелатинських текстів (Striova 2019).

Крім того, очікуване приєднання української мови до кола мов Євросоюзу відкриває раніше недоступні можливості включення української мови в багатомовні сценарії дослідження природної мови, деякі з яких передбачають наявність відповідного взаємнооднозначного транслітераційного механізму.

Розвиток корпусної лінгвістики та комп'ютерної лексикографії базується на використанні відповідних лінгвістичних технологій (Широков 2005; Широков 2011; Вакуленко 2017). Це також передбачає застосування належного механізму латинізації

кириличних текстів, що необхідно для забезпечення вимоги сумісності словника з багатомовними текстовими корпусами та лінгвістичними інструментами.

У цій роботі зосередимося на українських алфавітах, які існували починаючи з XIV століття. Це пов'язано з тим, що саме на той час українська мова набула тих характерних рис, які дають змогу класифікувати її як окрему (Шевельов 2002, 451; Мозер 2018, 9, 12). Таким чином, староукраїнський (середньоукраїнський) період – від середини XIV століття до XVIII століття – є питомих етапом історії власне української мови, який має істотний вплив на її сучасний стан, в той час як давньоукраїнська доба (від IX століття до середини XIV століття) характеризується значною кількістю спільнослов'янських рис і тому потребує окремого розгляду.

## 2 Вимоги до транслітераційних систем

Як відомо, транслітерація – це “заміна літер однієї писемності літерами іншої писемності, незалежно від їхньої вимови” (СУМ-11, т. X: 230). Це формулювання недостатньо чітке, оскільки не конкретизує, про яку саме вимову йдеться – про вимову в межах мови оригінального тексту чи про вимову в якійсь іншій мові, яка використовує транслітеровані знаки. Тому важливим є уточнення О. Реформатського, який відзначав, що прагнути до адекватного читання власних назв у інших мовах не варто (Реформатський 1960, 96). Враховуючи зв'язок між конкретним звуком мовлення, відповідною фонемою і її зображенням на письмі, можна підсумувати, що транслітерація не передбачає максимально “зручного” чи “легкого” читання конвертованого тексту якоюсь іншою мовою, але враховує правила вимови графем у мові оригінального тексту.

Таким чином, транслітерація являє собою механізм алфавітного перетворення тексту в межах одної мови (пор. Вакулєнко 1998, 335; Vakulenko 2004, 213-16; Vakulenko 2015, 41, 49-51) – тому передбачає лише зміну алфавіту, а не зміну самої мови. Прикладом транслітераційного перетворення є співвідношення між гаєвицею та вуковицею в мовах країн колишньої Югославії, які забезпечують еквівалентне відтворення того самого тексту різними алфавітними системами. Юрій Маслов відзначав, що наукова транслітерація базується на принципі взаємної однозначності між вихідними графемами і транслітерованими знаками (Маслов 2007, 284), що важливо для зворотної транслітерації і запобігає втраті інформації.

Більш конкретно вимоги до транслітераційного перетворення сформульовані в преамбулах до Міжнародного

транслітераційного стандарту ISO 9:1995 (ISO 9:1995) і Міждержавного транслітераційного стандарту ГОСТ 7.79-2000 (ГОСТ). До них належать:

- однозначність, що забезпечує стабільність представлення елементів вихідної писемності (літер, слів, виразів) засобами іншої (конвертувальної) писемності;
- зворотливість, що забезпечує можливість однозначного точного відновлення оригінального тексту;
- простота, що забезпечує автоматичне виконання процедури переходу від вихідного тексту до конвертованого і навпаки на основі простих алгоритмів, які переважно зводяться до застосування таблиць заміни знаків однієї системи письма знаками іншої системи письма.

Вважають, що при застосуванні правил конверсії вимоги звукової відповідності конвертованих знаків, естетичні міркування і традиційні норми можуть не завжди бути дотримані, хоча бажано, щоб у кожному окремому випадку порушення традиційних, фонетичних і естетичних норм було мінімальним.

Таким чином, ефективне включення України в міжнародні проекти передбачає, зокрема, можливість ізоморфного транслітераційного перетворення українських кирилических текстів у латиничні, яке забезпечує повне збереження інформації, що міститься в цих текстах. Крім того, для сумісності з програмами автоматичного опрацювання мови необхідно мати транслітераційні правила, які використовуватимуть набір основних літер латинського алфавіту – з кодами ASCII від 33 до 127 (Юнікод: від 0021 до 00A0). Оскільки кількість літер кирилиці та латиниці різна, то в цьому випадку необхідно використовувати комбінації латинських літер на позначення однієї кирилическої. Щоб уникнути неоднозначностей, ці комбінації не повинні бути тотожними транслітерованим комбінаціям кирилических літер. Для цього необхідні модифікатори – символи, які використовують лише в комбінаціях латинських літер для позначення однієї кирилическої. Поширеним транслітераційним модифікатором є знак “**h**” (Якобсон 1965, 111-13).

Слід мати на увазі, що йдеться про великі та надвеликі масиви текстових даних, які підлягатимуть опрацюванню в автоматичному режимі. Це означає, що до транслітераційних таблиць необхідно мати відповідні автоматизовані засоби транслітерації та ретранслітерації.

### **3 Транслітераційні надбання для України на початок 2022 року**

На початок 2022 року для української мови набули більшого чи меншого поширення такі системи латинізації.

Міжнародний транслітераційний стандарт ISO 9:1995. Інформація і документація - Транслітерація кирилических літер у латинські - Слов'янські та неслов'янські мови.

Переваги: майже повне забезпечення вимоги взаємної однозначності (за винятком <ь>); наявність таблиці з базовими літерами латинської абетки (без діакритичних знаків); часткове врахування слов'янської латиніграфічної традиції (транслітерація <й>, <ц>).

Недоліки: непослідовне врахування слов'янської латиніграфічної традиції (транслітерація йотованих літер); відсутність більшості літер, уживаних протягом історичного розвитку східнослов'янських мов (<Ѣ>, <ѣ>, <ѥ>, <Ѧ>, <ѧ>, <Ѩ> та інших); некоректна транслітерація <ь> апострофом.

Українська латиниця, рекомендована Транслітераційною комісією 16 листопада 2000 року (Вакуленко 2015, 234-54, 37-359).

Переваги: відповідність затвердженням Держстандартом України принципам транслітерації, зокрема повне забезпечення вимоги взаємної однозначності; наявність таблиці з базовими літерами латинської абетки (без діакритичних знаків); послідовне врахування слов'янської латиніграфічної традиції; наявність транслітераційної програми для прямої і зворотної транслітерації.

Недолік: відсутність літер, уживаних протягом історичного розвитку української мови (<Ѣ>, <ѣ>, <ѥ>, <Ѧ>, <ѧ>, <Ѩ> та інших).

Постанова № 55 Кабінету Міністрів України від 27.01.2010 "Про впорядкування транслітерації українського алфавіту латиницею" (ВТУАЛ).

Перевага: використання базових літер латинської абетки (без діакритичних знаків).

Недоліки: порушення вимоги взаємної однозначності вихідного і транслітерованого текстів (<г>, йотовані, <ь>); ігнорування слов'янської латиніграфічної традиції, яка використовує <ј> і діакритичні знаки; орієнтація на англійську мову як на посередника; вживання "h" як модифікатора і як окремої літери; відсутність літер, уживаних протягом історичного розвитку української мови (<Ѣ>, <ѣ>, <ѥ>, <Ѧ>, <ѧ>, <Ѩ> та інших); відсутність транслітерації <ь>.

Істотні порушення вимоги взаємної однозначності між вихідним і транслітерованим текстом унаслідок використання англійської мови як посередника призводять до систематичного спотворення і хибної ідентифікації назв: Ярун - Ярунь (Yarun),

Гальченко - Галченко (Halchenko), Яцків - Яцьків (Yatskiv), Тронко - Тронько (Tronko), Воронко - Воронько (Voronko), Банкова - Банькова (Bankova), Паньківська - Панківська (Pankivska), Польова - Полова (Polova), Лялько - Ліалко (Lialko), Ліліана - Ліляна (Liliana), Маріан - Мар'ян (Marian), медіана - медяна (mediana), Возіанов - Возянов (Voziyanov), Гундеріан - Гундерян (Hunderian), Клаузіус - Клаузюс (Klausius), Пії - Пій (Pii), Лар'їн - Ларін (Larin), Левитський - Левицький (Levytskyi), Тотька - Тотська (Totska), Чернятський - Черняцький (Cherniatskyi) тощо. Це виключає можливість відновити вихідну форму після зворотної транслітерації.

Із назви цієї постанови випливає, що транскодувальна таблиця, яка лежить у її основі, зорієнована виключно на алфавіт - тобто на впорядковану послідовність окремих літер <a>, <b>, <v>, ..., <y>, <z>. Тому не дивно, що ця таблиця непридатна для транслітерації текстів, які є комбінаціями літер.

Таблиця романізації українського алфавіту Бібліотеки Конгресу США (URLC).

Переваги: часткове забезпечення вимоги взаємної однозначності (за винятком <г> і <ь>); наявність транслітераційної відповідності <i> - <і>.

Недоліки: безальтернативне вживання надрядкових знаків (лігатур і коротки); порушення вимоги взаємної однозначності вихідного і транслітерованого текстів у складених словах та абрєвіатурах; вживання "h" як модифікатора і як окремої літери; ігнорування слов'янської латинографічної традиції; орієнтація на англійську мову як на посередника; відсутність літер, уживаних протягом історичного розвитку української мови (<8>, <Ѡ>, <Ѳ>, <Ѣ> та інших); некоректна транслітерація <ь> апострофом.

Таким чином, жодна з цих систем латинізації не містить літер історичних українських алфавітів, а більшість суперечать прийнятним принципам транслітерації - насамперед вимозі ізоморфної відповідності між вихідним і транслітерованим текстами, що необхідно для ідентифікації назв. Це робить такі системи неефективними для сучасних інформаційних і лінгвістичних ресурсів. Разом з тим, розробка української латиниці, схвалена Транслітераційною комісією 16 листопада 2000 року, цілком узгоджується з необхідними вимогами до транслітераційних стандартів і тому може бути покладена в основу історичної української латиниці.

## 4 Українські історичні алфавіти

Реєстр староукраїнських літер наведено (з певними варіаціями) в “Слов’янській граматиці” Лаврентія Кукіля, або Лаврентія Зізанія (Кукіль), “Граматиці” Максима Герасимовича, або Мелетія Смотрицького (Смотрицький), “Лексиконі слав’яноросському і імен толкованії” Памва Беринди (Беринда). Реєстр новоукраїнських літер включає в себе алфавіти О. Павловського, М. Лучкая, М. Шашкевича, Я. Головацького, І. Вагилевича, М. Максимовича, П. Куліша, М. Драгоманова, М. Гатцука, Є. Желехівського та сучасний український алфавіт (Півторак 2004) (див. також ІУП). Через велику кількість літер як модифікатори будемо використовувати не тільки “h”, а й “w”, іноді прямиий апостроф (’).

Як відомо, розвиток української мови супроводжувався характерними фонологічними змінами (Мейє 1951, 15-125; Безпалько та ін. 1957, 27-171; див. також Соболевській 1907, 20-151; Истрина 1915, 8-35), які варто відобразити і в латинізованому алфавіті. З одного боку, походження староукраїнських літер зі старослов’янських передбачає наявність графічної подібності між їхніми романізованими відповідниками, а з другого, постання притаманних українській мові специфічних фонетичних рис зумовлює необхідність відповідних графічних модифікацій. Таким чином, ми розглядаємо давньоукраїнську, староукраїнську та новоукраїнську як три різні, хоч і близькі одна до одної мови. Відповідні латинізовані алфавіти мусять бути подібними, але разом з тим відображати характерні особливості історичної вимови. Це насамперед стосується латинізації літер <г>, <є>, <и> в різні періоди існування української мови.

На території сучасної України елементи латиниці на основі польської, чеської, німецької, угорської, латинської, румунської, англійської та інших мов застосовувалися починаючи з XIV століття (Шевельов 2002, 451). У XIX столітті у Західній Україні деякі праці друкувалися латинською графікою, у Львові латинським алфавітом видавалася газета “Robotnyk” (Матвіяс 2011). Автор галицького букваря та граматики Йосип Лозинський активно поширював ідею використання латинської абетки в українському письмі, виклавши 1834 року свою позицію в статті “Про запровадження польської абетки в руську писемність” (ХМУЛМ, 32; ІУП, 64). У правилах орфографії Михайла Драгоманова замість кириличної літери <й> уживалася латинська <j>, а знак <j> був уживаний і в графіці М. Тулова 1879 року (ІУП, 343). 1859 року Йосиф Іречек намагався був запровадити для української мови латинську правописну систему на основі чеської мови, але Іван Франко та інші українські культурні діячі гостро її розкритикували (ІУП, с.



219). Антін Кобилянський писав латиницею, близькою до графіки сербохорватської та чеської мов (ХМУЛМ, 32). Подібний варіант української латиниці запропонував і Сергій Пилипенко у статті “Odvertyj lyst do vsix, xto cikavyt’sja sijeju spravoju” (Пилипенко 1923; Москаленко 1968, 28; див. також ІУП, 348-51).

Розглянемо найпроблемніші літери українських алфавітів.

Літера <г>. Ця літера існувала ще в протокирилиці, про що свідчать писемні пам’ятки IX століття (Брайчевський 2009, 29-30, 142). На еллінське коріння протокирилиці, яка вже існувала на час просвітницької діяльності Кирила та Мефодія і якою користувалися тоді в Києві, вказував Ю.В. Павленко (Брайчевський 2009, 15). Зокрема, українська літера <г> походить із давньогрецької гами <γ>, якій у латинографічних мовах відповідає <g>: *Gregory* (англ.), *geology* (англ.), *giro* (іт.), *gusto* (ісп.), *grafolog* (чес.). Таке походження передбачає наявність графічного елемента “g” і в транслітерації давньоукраїнської <г>. Збереження цього елемента в транслітерації <г> в пізніші періоди відображає тяглість української писемної спадщини, наступність алфавітів, які існували в історії української мови.

У старослов’янській (а отже, і в давньоукраїнській) мові літері <г> відповідав проривний (зімкнений) звук [g] принаймні до XII століття, а між 1157 і 1215 роками відбулася його спірантизація (Шевельов 2002, 450). – зміна проривної артикуляції на щілинну. Принаймні з 1334 року українську <г> почали передавати латинською <h>: *Hryczkone Kossaczowicz, Behohosc, haliciensis, Torhowycze, Drohobicz, Halicz, Hrubeschow, Halicie, Horbach* тощо (Шевельов 2002, 451). Тимчасом звук [g] стали позначати двознаком кг (Соболевський 1907, 43). Разом з тим, на місці <г> трапляється й <g>: *Bogdano-, mogilla, Bogdanowicz, Golosa, Galich, Gallicia, Mogula, Bereg, Igrischtya* та інші – а також **gh**: *Beregh, Ugh* (Шевельов 2002, 451). З іншого боку, за свідченням Агатангела Кримського, послідовники галичанина Якуба Гаватовича, який ще в 1619 році послуговувався польською графікою для української мови, транслітерували через <h> літеру <x>, а не <г>: хрін – *hrin* (ІУП, 398-9).

Усі ці приклади латинізації українських назв навряд чи можна вважати серйозними науковими спробами створити цілісну і всеохопну транслітераційну систему для українського алфавіту, яка б забезпечувала еквівалентний латиничний запис вихідного слова – адже на той час українська мова ще не мала навіть усталеного кириличного алфавіту. Це були радше намагання записати українські назви таким чином, щоб їх могли прочитати більш-менш близько до оригінального звучання носії польської, німецької, угорської, латинської та інших мов – на що вказують характерні закінчення та буквосполуки в латинізованих формах. Різної у відтворенні української <г>, очевидно, пояснюється

тим, що ні <g>, ні <h> не забезпечували точного прочитання: перша давала проривний звук, якого вже не було в оригіналі, а друга відповідала в перелічених мовах не дзвінкому, а глухому звукові, а на кінці слова як окрема графема не вживалася.

Не виключено, що вживання двознака **gh** у словах *Beregh, Ugh* саме і було спробою запровадити транслітераційну відповідність для літери <g>, яка враховує її походження з грецької гами і факт спірантизації відповідного звука. У сучасній українській мові кінцева <g> читається як частково оглушений дзвінкий задньоязиковий щілинний приголосний [χ] (Vakulenko 2019, 42-3), тому таке позначення фонетично мотивоване.

Юрій Шевельов наголошує на важливості морфологічного чинника в зміні [g] > [χ], адже альтернантами <g> були щілинні. При цьому проривний [g] утратив статус фонемі. Дальше перетворення [χ] на глибше артикульований глотковий звук не мало фонологічних або морфологічних наслідків. Приводом до цієї зміни, зазначає дослідник, було виникнення протетичного [г] з природно глотковою артикуляцією (Шевельов 2002, 453-4).

Експериментальні фонетичні дані (Тоцька 1981, 83, 90) свідчать, що пом'якшений варіант української фонемі /г/ реалізується задньоязиковим щілинним звуком [χ<sup>l</sup>]. У світлі того, що палаталізація, веляризація і фарингалізація взаємно виключають одна одну (Chomsky, Halle 1968, 307), навряд чи слід очікувати регулярного постання гортанного та глоткового алофонів [h<sup>l</sup>] і [χ<sup>l</sup>]. Подібна ситуація склалася в литовській мові, де фонема /χ/ іноді реалізується як гортанний [h]: оскільки палаталізований варіант у ній завжди задньоязиковий [χ<sup>l</sup>], то [χ] має перевагу перед [h] (Pakegys 1995). Отже, врахування пом'якшених алофонів фонемі /г/ також призводить до необхідності вживати елемент “g”.

Варто відзначити, що висновок Юрія Шевельова про відсутність фонологічних наслідків глибшої артикуляції /г/ надзвичайно важливий. З одного боку, він відображає тенденцію української мови до перетворення проривного звука [g] (і відповідної фонемі) на щілинний. Попри інтенсивні контакти з мовами, де фонема /g/ є поширеною, українська фонема /г/ перебуває внаслідок цієї тенденції на периферії української фонологічної системи. Про це свідчать мовні факти: у словнику Бориса Грінченка слів із літерою <g> відносно небагато, з яких переважна більшість мають паралельні форми з <г> (Грінченко), і активність цієї фонемі найнижча з усіх українських фонем, складаючи менше 0,1% (Жовтобрюх 2004).

Лаврентій Кукіль (Зізаній) називає свій твір “Грама́тіка”, вживаючи літеру <g>. Далі в тексті трапляється словосполучення “наѡкъ кграма́тїкѡ” (Кукіль, 4). Це вказує на прогресивну асиміляцію щілинного звука, який відповідає вимові <g>, у

зімкнений під впливом попереднього зімкненого [k]. Крім того, таке постання проривного алофона вказує на те, що в староукраїнську добу також не було окремих фонем /г/ і /гʲ/, а друга лише виступала периферійним варіантом першої. Це є додатковою підставою включити елемент “g” у транслітерацію літери <г>.

З іншого боку, такий висновок Юрія Шевельова надає більшої ваги задньоязиковому алофону [x] порівняно з глотковим [ʒ] і гортанним [h]. Існування в українській мові паралельного чергування дзвінких і глухих фонем /г/ - /ж/ - /зʲ/ (нога - ніжка - на носі) і /х/ - /ш/ - /сʲ/ (вухо - вушко - у вусі) вказує на фонематичний характер саме цього алофона (Vakulenko 2019, 44). Наявність акустичної пари /х/ - /г/ підтверджує і поява <г> у слові “гарний”, яке походить від гр. χάρη ‘краса, чеснота’ (Шевельов 2002, 447).

Така фонематична симетрія виникла в українській мові саме внаслідок спірантизації /г/. У сучасній російській мові, наприклад, де ця тенденція має нехтовно малий вплив, цієї симетрії немає: третій елемент у фонемних тріадах занепав, а чергування проривної /г/ в /ж/ (нога - ножка) відповідає чергуванню щілинної /х/ в /ш/ (ухо - ушко).

Слід наголосити також, що вживання літери <h> на позначення дзвінкого глоткового (фарингального) щілинного звука не є коректним. У Міжнародному фонетичному алфавіті (IPA) такий звук позначається знаком “ɦ” (який утворений із латинської літери <G>), а знак “h” відповідає глухому гортанному щілинному звуку. Таким чином, правила МФА також указують на необхідність наявності елемента “g” у транслітерації української <г>.

Як уже було зауважено, застосування “h” як модифікатора дає змогу забезпечити ізоморфну відповідність між множинами графем кириличного алфавіту та латиничного, який має менше базових літер. При цьому вживання “h” як окремої літери призводитиме до хибного ототожнення буквосполук і літер зг і <ж>, кг і <х>, сг і <ш> та інших.

Щобільше, фрикативна вимова <г> не є тим показовим маркером, який докорінно відрізняє українську мову від російської. Олексій Соболевський зазначає, що не тільки в окремих говірках, а й у літературній російській мові теж є подібна вимова: благодарить, господина, бога, богатъ, мягок, лёгок, дёготь, Богуславъ; іноді в словах голубъ, горох (Соболевській 1907, 126). Інші форми свідчать про наявність у російській мові оглушеного задньоязикового щілинного звука [x] перед приголосними: мяхкій, дёхтя (Соболевській 1907, 126) – що цілком відповідає вимові українських слів воґко, легко тощо (Білодід 1969, 257, 398).

Таким чином, у системі української латиниці, яка охоплює староукраїнські та новоукраїнські алфавіти, літеру <г> транслітеруємо через <ǵ>, **gh**.

Літера <s> читається як поєднання літер <c> і <d> (Кукіль, 9). Ця вимова літери “зелo” чується в деяких українських діалектах: \*здїсти, \*здло, тим часом як форми \*дзїсти, \*дзло в українській мові невідомі. Отже, транслітеруємо <s> через <ʒ>, **zdh**.

Літера <e>. У староукраїнську добу ця літера читалася як [e] після приголосних: “мене”, “добре”, “ключем” (Кукіль, 3), а також у єврейських, грецьких і латинських словах (Смотрицький, 12; див. також ІУП, 135). На початку слова та після голосного вона читалася як [je]: “оуединенїє”, “изглашенїє”, “естество” (Смотрицький, 12); “потребное”, “ей”, “ѣсть”, “естественной”, “небывает”, “моєи” (Кукіль, 4); “тлькованїС”, “КїевскїА” (Беринда, 3). При цьому було позиційне розрізнення великої та малої літери (Кукіль, 3). У новоукраїнських алфавітах (крім систем М. Гатцука та М. Максимовича) після приголосних стали вживати <e> (ІУП). Тому староукраїнську <e> транслітеруємо в <ê>, **ew**; а новоукраїнську в **je** (крім алфавітів М. Гатцука та М. Максимовича).

Літера <i>. Ця літера походить від еллінської ети (іти) <η>, відповідно в давньоукраїнській мові вона вимовлялася як передній піднесений звук [i] (Жовтобрюх, Кулик 1965, 139). Олексій Соболевський зазначає, що цей “чистий” старослов'янський звук зник у другій половині XI століття. Після нього залишився звук [j], який свого часу стояв між попереднім приголосним та [i] (Соболевській 1907, 46). Це вказує на йотовану вимову давньоукраїнської літери <i> після голосних, що підтверджується фактом виникнення літери <й> із <i> (Кукіль, 4; ІУП, 33).

Випадки вживання <ы> замість <i> та навпаки, що вказує на ствердіння голосних перед <i>, з'являються в галицько-волинських пам'ятках уже в XII століття, але до XIV століття вони дуже рідкісні (Соболевській 1907, 135). Починаючи з XIV століття, українські пам'ятки постійно змішують букви <i> та <ы> (Безпалько та ін. 1957, 154). Приклади поплутання цих літер, які наводить Юрій Шевельов (2002, 483), указують на те, що прочитання <i> наближалось до [i] в позиції після приголосних. Така зміна звучання <i> призвела до того, що в XVI-XVII столітті іншомовний звук [i] стали передавати не через <i>, як раніше, а через <ѣ> (Шевельов 2002, 484).

Проте наявність у староукраїнських текстах форм “в Вилни”, “в дрѣкарни” (Кукіль, 2); “и” (сполучник), “искѣство”, “досконали”, “каторїи” (Кукіль, 3); “всаки”, “чесной філософїи”, “естественной богослови”, “моєи”, “тыи”, “речи” (Кукіль, 4); “имаймо”, “дрѣзи” (Кукіль, 5); “изглашенїє” (Смотрицький, 12); “їменъ”, “КїювїА” (Беринда, 3) і под. вказує на те, що в цей період вимова <i> варіювалася: крім уже згаданого звучання, на початку слова та після деяких приголосних (зокрема, в деяких випадках після <l>, <ч>, <з>) вона вимовлялася близько до сучасної <i>, а після голосних (як і в давньоукраїнську добу) позначала йотований звук.

Оскільки транслітерація не передбачає врахування позиційних особливостей вимови, слід вибрати одне позначення. Враховуючи сучасну практику латинізації, тенденції розвитку відповідної фонетики та її характеристики в сучасній українській мові, віддаємо перевагу відповідності <i> - <y> (Вакуленко 1998; Vakulenko 2004; 2018).

Літера <ѣ>. На думку Ю. Шевельова, в кінці давньоукраїнського періоду ця літера вже вимовлялася близько до закритого звука [e] (Шевельов 2002, 260). О. Соболевський зазначає, що в галицько-волинських пам'ятках XIV-XV століття літера <ѣ> читалася близько до [i], можливо, як дифтонг [ie] (1907, 74). Тому <ѣ> транслітеруємо в <ě>, **ieh**.

Літери <ь>, <ѣ>. Припускають, що на початку давньоукраїнського періоду (IX століття) голосний звук [ъ] (у прочитанні літери "єр") вимовлявся близько до короткого [u], а голосний [ь] ("єрчик") - близько до короткого [i] або [e]. У мові східних слов'ян зредуковані голосні <ь> та <ѣ> існували до першої половини XII століття (Соболевський 1907, 41; Безпалько та ін. 1957, 79, 117). У процесі розвитку української мови з [ъ] в сильній позиції розвинувся [o] (Безпалько та ін. 1957, 128), що дає підстави транслітерувати <ь> через <ǫ>, **oh**. Своєю чергою, <ь> перебуває у відношенні доповнювальності до <й> (Алпатов и др. 1983, 76), що дає фонетичні підстави транслітерувати знак м'якшення подібним до <й> чином: <ь> - <j> (Якобсон 1965; Тилков 1982, 238), а в ізольованій позиції <ь> - <j>, **hj**. Такий вибір дає змогу уникнути неприродного вживання на місці <ь> апострофа, який в українській мові означає відсутність пом'якшення, а в Міжнародному фонетичному алфавіті - зупинку фонації, що призводить до хибних ототожнень текстів (Вакуленко 2015, 247).

Літера <ѣ>. Ще в старослов'янські часи сформувалася вимова цієї літери як [f] (Мейє 1951, 37), яка в українській мові не зазнала відчутних змін. Про подібність цих звуків у староукраїнську добу і відповідно поплутання <ѣ> із <ф> пише Мелетій Смотрицький (Смотрицький, 13; див. також ІУП, 36). Тому транслітеруємо <ѣ> через <f>, **fh**.

Інші літери. Щоб забезпечити можливість відновлення вихідного тексту, важливо коректно відображати графічні варіації написання окремих літер, які не пов'язані з відмінностями їхньої вимови. Це стосується літер <ѣ> і <з>, <ѣ> та <і>; <ѣ> та <8>; <ѣ> та <ѣ>. Потрібно також зберігати історично зумовлені графічні відмінності між літерами, які в процесі розвитку української мови стали вимовлятися однаково: <ѣ>, <ѣ> та <ѣ>; <ѣ>, <ѣ> та <ѣ>; <ѣ>, <ѣ> та <ѣ> тощо.

## 5 Висновки

Отже, ми розглянули можливі варіанти транслітерації староукраїнських і новоукраїнських текстів і запропонували оптимальні рішення, які забезпечують виконання принципів транслітерації і враховують особливості історичного розвитку української писемності. Викладені міркування лягли в основу нового національного транслітераційного стандарту (ДСТУ 9112:2021), який містить таблиці історичної української латиниці - кирилично-латиничної транслітерації давньоукраїнських, староукраїнських і новоукраїнських алфавітів з використанням діакритичних знаків (система А) і базових літер латинського алфавіту з модифікаторами (система Б). Ці системи слід використовувати в таких випадках.

Сфера використання системи А:

- друкована продукція (книги, журнали, буклети, альманахи, довідники, карти, атласи тощо);
- контент сайтів;
- назви транспортних засобів та інших об'єктів, які належать Україні;
- інформаційні системи;
- бланки організацій;
- вивіски;
- історичні пам'ятки;
- поштові марки;
- адреси на поштовині.

Сфера використання системи Б:

- електронна пошта;
- адреси сайтів і вебсторінок;
- скрипти комп'ютерних програм;
- документи, які засвідчують особу;
- багатомовні текстові корпуси;
- бібліографічні записи;
- дорожні щити;
- назви міст в аеропортах.

Належне впровадження цього стандарту є необхідною передумовою ефективної євроінтеграції України.

Ця робота підтримана грантом Національної стипендіальної програми Словацької республіки на підтримку мобільності студентів, аспірантів, викладачів університетів, дослідників і художників (2020).

## Література

- Алпатов, В.М.; Вентцель, А.Д.; Городецкий, Б.Ю. и др. (1983). *Лингвистические задачи*. М.: Просвещение.
- Безпалько, О.П.; Бойчук, М.К.; Жовтобрюх, М.А.; Самійленко, С.П.; Тараненко, І.Й. (1957). *Історична граматики української мови*. Київ: Радянська школа.
- Білодід, І.К. (ред.). (1969). *Сучасна українська літературна мова: фонетика*. К.: Наукова думка.
- Брайчевський, Михайло. (2009). *Походження слов'янської писемності* [4-е вид.]. К.: Вид. дім "Києво-Могилянська академія". 154 с.
- Вакуленко, М.О. (1998). "Восточнославянская латиница в международном контексте". *Slavia*, R 67, 333-9.
- Вакуленко, М.О. (2015). *Українська термінологія: комплексний лінгвістичний аналіз*. Івано-Франківськ: Фоліант.
- Вакуленко, М.О. (2017). "Віртуальна українсько-російсько-англійська термінографічна лабораторія з фізики: сучасні лінгвістичні технології у фаховій мові". Піпер, П.; Йованович, В. (редкол), *Словенська термінологія данас*. Београд: Српска академија науки и уметности: Институт за српски језик САНУ, 679-89.
- Жовтобрюх, М.А.; Кулик, Б.М. (1965). *Курс сучасної української літературної мови: Частина I*. Київ: Радянська школа.
- Жовтобрюх, М.А. (2004). "Фонологічна система". *Русанівський та ін 2004*, 765.
- Истрина, Е.С. (1915). *Руководство по истории русского языка*. Петроградъ: Издание Я. Башмакова и Ко.
- Маслов, Ю.С. (2007). *Введение в языкознание*. 6-е изд., стереотип. М.: Асадемія; СПб.: Филол. фак. СпбГУ.
- Матвіяс, І.Г. (2011). "Особливості фонетичної системи в західноукраїнському варіанті літературної мови". *Мовознавство*, 4, 16-21.
- Мейе, А. (1951). *Общеславянский язык*. М.: Иностранная лит-ра.
- Мозер, Міхаель. (2018). *Історія української мови*. Київ: Ліббез. [http://shron1.chtyvo.org.ua/Michael\\_Moser/Istoriia\\_ukrainskoi\\_movy.pdf](http://shron1.chtyvo.org.ua/Michael_Moser/Istoriia_ukrainskoi_movy.pdf).
- Москаленко, А.А. (1968). *Історія українського правопису (радянський період)*. Одеса: Вид-во Одеського ун-ту.
- Пилипенко, Сергій. (1923). "Odvertyj lyst do vsix, xto cikavyt'sja cijegu spravoju". *Червоний шлях*, 6-7(X), 267-8.
- Півторак, Г.П. (2004). "Український алфавіт". *Русанівський та ін 2004*, 737-8.
- Реформатский, А.А. (1960). "Транслитерация русских текстов латинскими буквами". *Вопросы языкознания*, 5, 96-103.
- Русанівський, В.М. (2004). "Староукраїнська літературна мова". *Русанівський та ін 2004*, 641-4.
- Русанівський, В.М.; Тараненко, О.О.; Зяблюк, М.П. та ін (редкол) (2004). **Українська мова: енциклопедія**. К.: Вид-во "Укр. енцикл". ім. М.П. Бажана.
- Соболевский, А.И. (1907). *Лекции по истории русского языка*. Москва: Университетская типография, Страстной бульварь.
- Тилков, Д. (гл. ред.). (1982). *Грамматика на съвременния български книжовен език. Том 1. Фонетика*. София: Наука и изкуство.
- Тоцька, Н.І. (1981). *Сучасна українська літературна мова: фонетика, орфоепія, графіка, орфографія*. Київ: Вища школа.

- Шевельов, Юрій. (2002). *Історична фонологія української мови*. Харків: Акта.
- Широков, В.А. (2005). *Корпусна лінгвістика*. К.: Довіра.
- Широков, В.А. (2011). *Комп'ютерна лексикографія*. К.: Наукова думка.
- Якобсон, Р.О. (1965). "О латинизации международных телеграмм на русском языке". *Вопросы языкознания*, 1, 111-13.
- Bogacz, B.; Klingmann, M.; Mara, H. (2017). "Automating Transliteration of Cuneiform from Parallel Lines with Sparse Data". *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 1, 615-20.
- Cai, Z.; Wang, W. (2018). "2DLDA-based Compound Distance for Similar Online Handwritten Tibetan Transliteration of the Sanskrit Character Recognition". *Proc. of International Conference on Frontiers in Handwriting Recognition, ICFHR*, 223-8.
- Chomsky, N.; Halle, M. (1968). *The Sound Pattern of English*. New York; Evanston; London: Harper & Row Publishers.
- Naji, N.; Allan, J. (2016). "On Cross-Script Information Retrieval". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9626, 796-802.
- Pakerys, A. (1995). *Bendrinės lietuvių kalbos fonetika*. Vilnius: Enciklopedija, Valiulio Leidykla.
- Pezzati, L. (December 6, 2019). Private communication. Odesa.
- Razak, S.M.A.; Abu Seman, M.S.; Mamat, W.A.W.Y.W.; Mohammad Noor, N.H.N. (2018). "Transliteration Engine for Union Catalogue of Malay Manuscripts in Malaysia: E-JAWI Version 3". *Proc. of International Conference on Information and Communication Technology for the Muslim World, ICT4M*, 58-63.
- Striova, J. (December 6, 2019). Private communication. Odesa.
- Vakulenko, M. (2004). "Simple-Correspondent Transliteration Through a Slavonic Latin Alphabet". *Journal of Language and Linguistics*, 3(2), 213-28.
- Vakulenko, M.O. (2015). "Practical Transcription and Transliteration: Eastern-Slavonic View". *Govor*, 32(1), 35-56.
- Vakulenko, M.O. (2018). "Ukrainian Vowel Phones in the IPA Context". *Govor*, 35(2), 189-214. <https://www.hfi1o1oskod.hr/images/HFD/Govor/Govor-2-2018-web.pdf>.
- Vakulenko, M.O. (2019). "Ukrainian Consonant Phones in the IPA Context with Special Reference to /v/ and /gh/". *Linguistica online*, 22, 36-61. <http://www.phil.muni.cz/linguistica/art/vakulenko/vak-001.pdf>.
- Vakulenko, M.O. (2022). "Normalization of Ukrainian Letters, Numerals, and Measures for Natural Language Processing". *Digital Scholarship in the Humanities* (in press).





