

Damjan Popič

Filozofska fakulteta Univerze v Ljubljani,
Fakulteta za računalništvo Univerze v Ljubljani
damjan.popic@ff.uni-lj.si
ORCID: 0000-0003-0229-5343

Lyudmila Aleksandrovna Zadorozhnaya

Evrazijska narodna univerza L. N. Gumiljova,
Astana, Kazahstan
mila.alex.zadorozhnaya@gmail.com
ORCID: 0000-0002-8819-5329

Izzivi »internacionalizacije« jezika v digitalni dobi: korpusnojezikovna analiza variantnosti pri prečrkovanju in morfološkem podomačevanju kazahstanskih osebnih imen v angleščini

V prispevku so predstavljeni izzivi transliteracije kazaških lastnih imen, zapisanih v cirilici, v angleščino, zlasti v kontekstu prehoda kazahstanskega pisnega sistema iz cirilice v latinico, ki poteka od leta 2017, z namenom, da bi kazaščina postala mednarodno bolj prepoznavna. Prizadevanje Kazahstana za internacionalizacijo je privedlo do še plodnejših prevodnih izmenjav med kazaščino in angleščino, kar je povzročilo precejšnja razhajanja pri zapisu lastnih imen v angleščini, tudi pri znanih in ključnih osebnostih in imenih iz kazahstanske kulture in zgodovine. Namen prispevka je analizirati stopnjo variantnosti pri prečrkovanju z uporabo korpusnojezikovne metodologije, v ta namen pa smo sestavili korpus angleških podstrani več uradnih/institucionalnih kazahstanskih spletnih strani (vladnih, visokošolskih itd.). S tem korpusom prikažemo imena z najvišjo stopnjo variantnosti v angleščini v njihovih angleških ustreznih in oblikujemo kategorije potencialno problematičnih imen z vidika prečrkovanja, obenem pa tudi z vidika morfološkega podomačevanja oziroma konkurenčnosti med rusificiranimi in kazaškimi oblikami.¹

Ključne besede: prečrkovanje, kazaščina, cirilica, latinica, pisava, jezikovna variantnost

The Challenges of “Internationalizing” a Language in the Digital Age: A Corpus-Linguistic Study of Transliteration and Morphological Adaptation in Rendering Kazakh Cyrillic Names in English

This article presents the challenges of transliterating Kazakh proper names written in the Cyrillic script into English, especially against the backdrop of the government-led transformation of the Kazakh writing system from Cyrillic to Latin script, which has been underway since 2017 and is poised to pass its first milestone in 2029. The aim of the transformation is to make the Kazakh language more easily understandable and international. This turn towards internationalization has also led to more prolific translation from Kazakh into English, which in turn has brought about a significant degree of variation in spelling in English, even when pertaining to key figures and names from the Kazakh culture and history (cf. Abai/Abay, etc. for the poet, philosopher and reformer Abai Qunanbaiuly). The article analyzes the degree of variation by using corpus-lin-

¹ Raziskavo sta podprla projekta »Consistency of Kazakh Onyms in English-Language Texts: Translatory Research« na univerzi Gumilyov Eurasian National University (ENU), Astana, Kazahstan, in Razvojni sklad Filozofske fakultete Univerze v Ljubljani.

This research was supported by the project “Consistency of Kazakh Onyms in English-Language Texts: Translatory Research” at Gumilyov Eurasian National University, and by the Research Development Fund at the Faculty of Arts, University of Slovenia.

guistic methodology. To this end, we constructed a corpus of English subpages of several official/institutional Kazakhstani websites (the government, higher education, etc.). We use this corpus to identify the Kazakh names with the highest degree of variation in their English equivalents and create categories of potentially problematic names in terms of transliteration, as well as in terms of morphological adaption of competitive Russified and Kazakh forms.

Keywords: transliteration, Kazakh, Cyrillic, Latin, script, language variation, reform

1 Uvod

Razvoj pisnih sistemov v različnih jezikih skozi zgodovino je zaznamovalo dinamično prepletanje kulturnih, političnih in tehnoloških vplivov. Od antičnih civilizacij do sodobnih družb so prehodi iz ene pisave v drugo odražali zapleteno prepletenost človeške komunikacije in njenega pisnega izraza. Prehod kazaškega jezika s cirilice na latinico predstavlja pomemben kulturni in jezikovni premik, ki prinaša globoke in daljnosežne posledice za položaj Kazahstana v globalni skupnosti, pri čemer položaj države močno zaznamuje jezikovna pestrost (Smagulova 2008; Aksholakova, Ismailova 2013) ter tudi geopolitični položaj (Pavlenko 2008). Ta pomembna sprememba, ki je bila uvedena leta 2017 z odlokom predsednika republike,² je bila predstavljena kot sredstvo za modernizacijo jezika in države, krepitev njene mednarodne prisotnosti in lažjo integracijo v globalno gospodarstvo ter znanost (Bekzhanova, Makoelle 2022). Latinizacija kazaščine tako ni zgolj sprememba pisave, temveč strateški korak za pozicioniranje jezika (v smeri zahoda) in s tem tudi države v širšem kontekstu globalizacije, tehnološkega napredka in geopolitike.

Ta pomembna in nenadna sprememba, ki jo je predsednik Nazarbaev sicer večkrat omenjal, tudi v pogovorih z drugimi voditelji postsovjetskih držav (Bekzhanova, Makoelle 2022), je terjala in seveda tudi pritegnila pozornost jezikoslovne skupnosti v Kazahstanu, pri čemer je osrednja jezikoslovna institucija, Inštitut za jezikoslovje Akhmeta Baitursynulya,³ prevzela vodilno vlogo in po začetku procesa spreminjanja pisave predstavila več alternativnih abeced in smernic za transliteracijo kazaščine iz cirilice v latinico. Priročniki vključujejo različne referenčne materiale, kot so splošni in terminološki slovarji ter številni drugi viri, vendar pa ti viri podajajo zgolj prečrkovane smernice in variante znotraj enega jezika v dveh različnih pisnih sistemih. To pomeni, da ti priročniki ne ponujajo rešitev za prevajanje kazahstanskih lastnih imen v angleščino, kar pomeni, da to vprašanje ostaja nerešeno – tudi pri segmentu osebnih lastnih imen, ki bi z vidika normativistike morala biti zapisana čim bolj enoznačno. Kot pravi Ermolovich (2001: 26), je ena od temeljnih funkcij lastnih imen prav to, da

² Predsedniški odlok št. 569 je 26. oktobra 2017 podpisal nekdanji (in prvi) predsednik Republike Kazahstan Nursultan Nazarbaev. Odlok je dostopen na spletni strani <https://academy-gp.kz/?p=9661&lang=en>.

³ <https://tbi.kz/eng>.

služijo kot pravni identifikator posameznika in njegovih povezanih lastnosti, prav zato pa je bistveno, da se v vseh virih ohrani dosledna pisna oblika antroponima, zlasti v kontekstu digitalizacije, v okviru katere besedila na spletu niso zgolj besedila, temveč tudi viri za strojno učenje umetno-inteligenčnih sistemov, tako da ima vsaka posamezna variantna raba močan in potencialno dolgotrajen vpliv.⁴

Ta prispevek obravnava posledice kazahstanske pisne reforme, s posebnim poudarkom na prečrkovanju kazahstanskih osebnih lastnih imen, zapisanih v cirilici, v angleščino. Angleščina kot *lingua franca* mednarodnega diskurza igra ključno vlogo pri tem, kako države komunicirajo svojo identiteto in sodelujejo s svetom. Sprejetje latinice naj bi kazaščino približalo mednarodni javnosti, s tem pa bi se spodbudila kulturna izmenjava in pospešilo gospodarsko sodelovanje (Smagulova 2016). Dejstvo, da standardiziranih smernic za prečrkovanje (še) ni, povzroča zmedo pri zapisovanju kazahstanskih imen na angleških spletiščih, kar poskušamo prikazati s korpusnojezikoslovno analizo uradnih spletišč v angleščini. Ta variantnost je izziv tako za jezikovno identiteto kot tudi za mednarodno prepoznavnost Kazahstana, saj je vpliv transliteracije kazahstanskih lastnih imen v angleščino takojšen in globalno viden. Tako se, čeprav se reforma običajno obravnava kot notranje vprašanje jezikovnega načrtovanja, njene posledice raztezajo daleč onkraj zapisovalnih praks. Ker angleščina že služi kot *lingua franca* diplomacije, znanosti in digitalne komunikacije, se vsaka sprememba v referenčnem zapisu osebnih imen odraža v potnih listih, akademskih repozitorijih in drugih storitvah, pravnih dokumentih in indeksih spletnih iskalnikov. Ta prispevek zato združuje dva vidika, ki se v znanstveni literaturi pogosto obravnavata ločeno:

1. reformo pisave kot jezikovnopolitično vprašanje;
2. variantnost pri prečrkovanju kot živi jezikovni praksi.

Ključna povezava med obema točkama je, da latinizacija kazaščine ponovno načena vprašanje »privzetega črkovanja« za vsa kazahstanska lastna imena, ki se v angleščini že uporabljajo. Za razliko od ruske cirilice, pri kateri imamo enotno prevladujočo prečrkovalno normo (BGN/PCGN), je kazaščina na razpotju: naj uradna komunikacija v angleškem jeziku prevzame neposredno latinščino iz nove pisave (npr. *Qunanbaiuly*), ohrani uveljavljene rusificirane oblike (npr. *Kunanbayev*) ali dopusti nadaljnjo vzporedno rabo? V prispevku poskušamo z uporabo korpusnojezikoslovne metodologije kvantitativno opredeliti distribucijo končnic pri različnih imenih in tako izmeriti entropijo med njimi, hkrati pa želimo čim bolj nazorno prikazati tudi prevlado posameznih oblik.

⁴ To je opazno denimo pri slovenskih poimenovanjih za zemljepisna imena zunaj meja Slovenije, npr. v Italiji, kjer pomanjkljive informacije v slovenščini na Wikipediji vodijo v slabšo zastopnost slovenskih poimenovanj pri Googlovih storitvah, npr. Zemljevidih. Tudi zaradi tega se poraja vprašanje o nujnosti in celo strateški vlogi zagotavljanja verodostojnih informacij v odprtih repozitorijih za (digitalno) preživetje lokalnih in regionalnih poimenovalnih variant.

(Jezikovna) politika – prehod na novo pisavo – torej določa *makrokontekst* naše raziskave, medtem ko korpusnojezikoslovna analiza prečrkovalnih in zapisovalnih praks razkriva *mikroizide*, ki so vidni danes. S povezavo obeh lahko utemeljeno trdimo, da:

1. bo pravopisna reforma brez enotnih smernic variantnost še bolj utrdila, namesto da bi jo zmanjšala;
2. lahko obstoječi vzorci rabe vplivajo na samo izvajanje reforme: če bi denimo za posamezno ime, pri katerem povsem prevladuje določena oblika, predpisali drugo, bi nenadna sprememba lahko naletela na večji odpor pišočih kot pri drugih, bolj enakomerno rabljenih končnicah in zapisovalnih oblikah;
3. je tovrstna variantnost za globalno prepoznavnost države lahko škodljiva, hkrati pa je implementacija tudi močno odvisna od geopolitičnih nazorov aktualne politike.

2 Pisava: kaj izraža in kaj se skriva med vrsticami

Prečkovanje je proces pretvorbe besedila iz enega pisnega sistema v drugega s pomočjo preslikave znakov iz izvorne pisave v znake ciljne pisave na podlagi fonetične ali ortografske enakovrednosti (Coulmas 2003: 36). Ta praksa omogoča, da se leksemi in imena iz jezikov z različnimi pisavami predstavijo na način, ki je razumljiv bralcem, ki z izvorno pisavo niso seznanjeni.

Zgodovina transliteracije je dolga in sega v čase, ko so kulturne izmenjave in trgovina terjale ciljno usmerjene spremembe izrazne oblike. Grki so abecedo prevzeli od Feničanov in jo prilagodili nesemitskemu jeziku, pri čemer so uvedli samoglasnike (Robinson 1995: 166–167), med širjenjem verstev, kot sta krščanstvo in islam, pa je postalo prečkovanje ključno za prevajanje posvetnih besedil in znanstvenih del, kar je pripeljalo do razvoja različnih prečrkovalnih konvencij (Coulmas 2003: 202–203; Diringer 1948: 290, 553).

Prečkovanje je zahteven proces, saj pogosto vključuje dileme o jezikovni standardizaciji, kulturni avtentičnosti in sami izvedbi, zato so v tovrstnih primerih še bolj pomembne jasne smernice, zlasti v sodobnem času, saj sta globalizacija in tehnološki napredek povečala pomen standardiziranih sistemov prečkovanja, kar je v spletnem okolju še toliko bolj pomembno. Mednarodni organi, kot je Mednarodna organizacija za standardizacijo (ISO), so oblikovali smernice, na primer standard ISO 9 za prečkovanje cirilice, da bi spodbudili doslednost in odpravili dvomnost v mednarodni komunikaciji. Kljub tem prizadevanjem pogosto obstaja več standardov transliteracije za isto pisavo, kar povzroča variantnost v zapisu, to pa je opazno tudi pri prečkovanju

imen in toponimov cirilice v latinico, pri katerem različne države in organizacije uporabljajo različne sisteme (Fierman 2009: 1220–1221).

2.1 Menjave pisnih sistemov

Skozi zgodovino so številne države in skupnosti sprejemale korenite spremembe pisave v okviru širših prizadevanj za modernizacijo, uveljavljanje nacionalne identitete ali usklajevanje s političnimi spremembami, eden najbolj znanih primerov pa je reforma pisave v Turčiji leta 1928. Pod vodstvom Mustafe Kemala Atatürka je Turčija arabsko pisavo nadomestila z latinico. To radikalno spremembo, katere cilj je bil povečati stopnjo pismenosti in modernizirati državo, so vodila sovjetska prizadevanja za latinizacijo (*korenizatsiya*), kar je olajšalo integracijo Turčije v zahodni svet (Lewis 1999: 27, 38–39). V zadnjem stoletju so se tako za spremembo pisave odločile številne države, zlasti na podlagi geopolitičnega položaja v razmerju do Sovjetske zveze/Ruske federacije, kot prikazuje tabela 1.

Tabela 1: Kratek zgodovinski pregled glavnih sprememb pisave.

| Država/narod | Leto | Izvorna pisava | Ciljna pisava |
|---------------------|-----------|--------------------------------|---------------|
| Uzbekistan | 1926 | turška | latinica |
| Turkmenistan | 1926 | turška | latinica |
| Azerbajdžan | 1926 | turška | latinica |
| Kazahstan | 1926 | turška | latinica |
| Uzbekistan | 1930 | latinica | cirilica |
| Turkmenistan | 1930 | latinica | cirilica |
| Azerbajdžan | 1930 | latinica | cirilica |
| Kazahstan | 1930 | latinica | cirilica |
| Mongolija | 1941 | tradicionalna mongolska pisava | cirilica |
| Uzbekistan | 1993 | cirilica | latinica |
| Turkmenistan | 1993 | cirilica | latinica |
| Azerbajdžan | 1991–1992 | cirilica | latinica |
| Kazahstan | 2017– | cirilica | latinica |

Kot je razvidno iz tabele 1, je bil v preteklem stoletju v Srednji in Vzhodni Aziji razvoj pisav zelo dinamičen (Grenoble 2003: 49–50). Mongolija je leta 1941 zamenjala pisavo in pod vplivom Sovjetske zveze tradicionalno mongolsko pisavo nadomestila s cirilico. Namen te spremembe je bil izboljšati stopnjo pismenosti in Mongolijo politično in kulturno približati Sovjetski zvezi. Po razpadu Sovjetske zveze je več srednjeazijskih republik – poleg Kazahstana – začelo reformo pisave, da bi s tem okrepile svojo nacionalno identiteto. Azerbajdžan je med letoma 1991 in 1992 prešel s cirilice na latinico (Grenoble 2003: 122), kar je bil del prizadevanj za uveljavitev neodvisnosti in vzpostavitve tesnejših vezi z zahodnimi državami. Podobno je Uzbekistan leta 1993 sprejel latinico, ki je nadomestila cirilico, da bi okrepil svojo kulturno dediščino in

spodbudil uzbeški jezik (Pavlenko 2008). Turkmenistan je prav tako leta 1993 prešel s cirilice na latinico, da bi s tem oživil turkmensko nacionalno identiteto (Landau, Kellner-Heinkele 2001).

2.2 Latinizacija kazaščine

Zgodovino kazaške pisave zaznamujejo korenite spremembe pravopisa, ki so bile pogosto posledica zunanjih geopolitičnih sil in notranjih kulturnih dinamik in reform. Začetki Kazahstana kot suverene države segajo v 15. stoletje, ko so se tri glavna plemena (horde) združila in ustanovila Kazaški kanat, sam pisni sistem pa se je začel oblikovati že veliko prej. Med 5. in 12. stoletjem so uporabljali staro turško runsko pisavo, ki ji je sledila arabska pisava, ki se je postopoma začela uporabljati od 8. stoletja (Ayagan 2005: 40) in se je aktivno uporabljala vse do 20. stoletja, danes pa jo uporabljajo kazahstanski prebivalci v avtonomnih regijah Kitajske (Zhou 2003: 149), v 20. in 21. stoletju pa je Kazahstan doživel in še doživlja dinamičen razvoj svoje pisave, celotno dinamiko pa ponazarja tabela 2.

Tabela 2: Pregled sprememb pisave v kazaščini (Kassenova, Zhumashev 2025; Riekkinen idr. 2021; Bazarbayeva, Chukayeva 2021).

| Obdobje | Pisni sistem | Ključni dogodek/ ljudje | Značilnosti |
|--------------------------|--|---------------------------------|--|
| 5.–12. stoletje | Runska pisava | Stara turška plemena | Uporabljala so jo turška plemena; zapisi kažejo na razširjeno pismenost v 7. in 8. stoletju. |
| 11.–začetek 20. stoletja | Arabska pisava | Širitev islama | Nadomestila je runske pisave; prevladovala v književnosti in znanosti; povezovala Kazahce z muslimansko identiteto in tradicionalno preteklostjo. |
| 1924 | Prilagajone arabščina (tote jazu/tote zhazu) | Akhmet Baitursynuly | »Varčna pisava«; odstranitev diakritičnih znakov zaradi prilagoditve kazaški fonologiji. Še vedno v uporabi pri govorcih kazaščine na Kitajskem. |
| 1929–1940 | Latinica (<i>janalif</i>) | Sovjetska »kulturna revolucija« | »Enotna turška abeceda«, sprejeta po kongresu v Bakuju leta 1926. Politični cilj: oddaljitev prebivalstva od islama in izogibanje videzu rusifikacije. |
| 1940– | Cirilica | Sovjetska oblast | Politični preobrat k rusifikaciji. Uporaba 42 črk za 28 fonemov. |
| 2017 | Latinica (1. osnutek) | Predsedniški odlok | »Različica z apostrofi«. Močno kritizirana zaradi vizualne neprijetnosti in težav pri digitalni obdelavi besedil. |
| 2018 | Latinica (2. osnutek) | Predsedniški odlok | »Različica z ostrimi strešicami« (akut). Kritizirana zaradi uporabe digrafov (npr. <i>sh</i> , <i>ch</i>) in neupoštevanja načela »en zvok – ena črka«. |
| 2019– | Latinica (nova različica) | Inštitut za jezikoslovje | Uporaba diakritičnih znakov. Sklada se s turškimi/azerbajdžanskimi modeli; cilj je fonetična ekonomičnost in natančen zapis fonemov. |

Po vzpostavitvi sovjetske oblasti v 20. letih 20. stoletja se je med kazahstansko inteligenco vnela ostra razprava o prihodnosti pisave, ki je družbo razdelila na »zahodnjake«, ki so podpirali latinizacijo, in »turkofile«, ki so zagovarjali ohranitev arabske pisave. Sovjetska vlada je aktivno spodbujala latinizacijo kot del širše »kulturne revolucije« s političnim ciljem, da bi turška muslimanska prebivalstva oddaljila od vpliva duhovščine in islama, hkrati pa se izognila takojšnji uvedbi cirilice, ki bi jo lahko razumeli kot nadaljevanje carske rusifikacije.

Po mnenju Mamynove (2022) in Mamynove idr. (2023) je za enega prvih poskusov preoblikovanja kazahstanske pisave zaslužen sloviti kazahstanski jezikoslovec in kulturnik Akhmet Baitursynuly (1872–1937), ki je ustvaril prilagojeno različico abecede, temelječe na arabski pisavi, in jo poimenoval »tote zhazu«. Njegova različica pisave je bila prva, ki je upoštevala fonološke značilnosti kazaškega jezika, hkrati pa je prava pisava v veliki meri prispevala k razvoju izobraževanja in ozaveščanja ljudi v tem obdobju (Mamynova 2022).

Med letoma 1920 in 1930 se je v republikah Sovjetske zveze aktivno izvajala politika latinizacije turških jezikov. Ena od rešitev, predlaganih na kongresu turkologov v Bakuju leta 1926, je bilo priporočilo, da se vse za vse pisave vseh turških jezikov uporabi latinico. To je leta 1927 privedlo do nastanka pisave, imenovane janalif (*yañalif*, tudi *yangalif*, cir. Яңалиф, 'nova abeceda'), latinične pisave za turške jezike.

Slika 1: Končna oblika pisave janalif iz leta 1929.

A B C Ç D E É Ë F G H I J K L M N O Ó P R S T U V X Y Z Z III W

Ta abeceda ni bila povsem enaka angleški in je vsebovala črke, značilne za turške jezike, kar vse do digitalne dobe za mednarodno pojavnost ni bilo problematično. 7. avgusta 1929 so najvišji kazahstanski organi uradno odobrili sprejetje te abecede (Nurgaziev, Kyrykbaev 2024: 53). Janalif se je kot pisni sistem za kazaščino uporabljal v obdobju med letoma 1929 in 1940, sama implementacija pa se je izkazala za izjemno težavno: ljudje so se morali učiti dveh pisav in so pogosto zamenjevali posamezne znake, tako da je bila nova pisava odpravljena s sprejetjem Zakona o prehodu kazahstanske pisave z latinice na cirilico, 13. novembra 1940 (Tilleshov idr. 2020: 195).

Tudi po osamosvojitvi Kazahstana leta 1991 se je za kazaški jezik uporabljala cirilica, je pa bila na državni ravni predlagana izjemno pomembna in daljnosežna sprememba – uporaba angleščine kot *lingue france* (Bekzhanova, Makoelle 2022: 1), kar je

naznanilo uvedbo »trojice jezikov«: kazaščine kot državnega jezika, ruščine kot jezika medetničnega sporazumevanja in angleščine kot jezika integracije v mednarodno okolje in svetovno gospodarstvo. Po mnenju Zharkynbekove in Aimoldine (2016) »trojico kazaščine, ruščine in angleščine podpira večina državljanov«.

Sprejetje te ureditve je imelo daljnosežne posledice tudi na področju izobraževanja in znanosti, saj se ti jeziki zdaj uporabljajo za izobraževanje in znanost v skoraj enakomernem obsegu, tako da je angleščina postala pomemben jezik v podiplomskem izobraževanju (Sulkarnayeva 2017: 76). To je močno povečalo potrebo po predstavitvi kazahstanske kulture ter znanstvenih in gospodarskih dosežkov v angleščini. Vse lokalne konference morajo denimo biti opremljene s povzetki v vseh treh jezikih, kar ponovno odpira vprašanje enotne transliteracije imen znanstvenikov in javnih osebnosti za njihovo ustrezno reprezentacijo v mednarodnem (znanstvenem) okolju – če ne zaradi drugega, vsaj zaradi citiranja, bibliometrije in sledljivosti v digitalnih virih. Pri tem pa ne gre zgolj za vprašanje prečrkovanja iz cirilice, temveč tudi za izbiro morfološke oblike imena (npr. rusificirane končnice v primerjavi s kazaškimi).

Pomembno je upoštevati, da lahko morfološka zgradba osebnih imen v kazaščini deluje kot kazalnik zgodovinskih obdobj (Mnaidarova idr. 2025: 223). Dogodki v prvi polovici 20. stoletja so prispevali k uveljavitvi rusificiranih končnic, ki so nato v kazahstanskih antroponimih povsem prevladale. Po osamosvojitvi pa so konec 20. in v začetku 21. stoletja sledile pomembne spremembe pri zapisovanju in normiranju lastnih imen (Bazarbayeva idr. 2024: 66). Preoblikovanje antroponimične formule je bilo tudi uradno potrjeno v dokumentu Državne onomastične komisije;⁵ dokument med drugim predlaga rabo kazaških obrazil -uly in -kyzy namesto ruskih -ov/-ev ter -ova/-eva (Madieva idr. 2014).

V angleških (digitalnih) virih se zato pogosto prepletata dve različni ravni variabilnosti. Prva je morfološka variabilnost, tj. konkurenca med rusificiranimi in kazaškimi, pogosto patronimičnimi oblikami priimkov, ki v rabi lahko prevzamejo funkcijo »priimka« v latiničnih zapisih. Tako se npr. pojavlja rusificirana oblika *Kunanbayev*, vzporedno pa tudi kazaška oblika *Qunanbayuly*. Druga raven je variabilnost prečrkovanja oziroma latinizacije, tj. več možnih latiničnih zapisov iste izhodiščne oblike (npr. *Kunanbayev/Kunanbaiev* ali *Qunanbayuly/Qunanbaiuly*; vključno z omahovanjem tipa q : k v korenu). Prav ta kombinacija morfološke in prečrkovalne variantnosti povečuje verjetnost, da bo v rabi več različic istega imena, kar posledično vpliva na število na-

⁵ *Conception of Regulating Personal Names, Middle Names and Family Names for the Kazakhs in the Republic of Kazakhstan* (1994).

činov, kako je ime zapisano v angleških, posledično pa tudi drugih latiničnih okoljih (npr. v bibliografijah, citatnih bazah in na uradnih spletnih straneh).

Razmah tovrstnih različic je dodatno povezan z nestandardizirano rabo digitalnih tehnologij, ki pogosto temeljijo na različnih transliteracijskih sistemih, kar potrjujejo tudi nedavne raziskave (Zadorozhnaya 2025: 307). Kot opozarjata Rsaliyeva in Rysbergen (2014: 161), je zato vprašanje znanstveno utemeljenih splošnih načel za prečrkovanje kazaških imen v angleščino postalo posebej pereče. V pričujoči raziskavi zato variantnost obravnavamo na dveh ločenih oseh – (i) izbira morfološke oblike (rusificirana : kazaška) in (ii) variantnost latiničnega zapisa znotraj izbrane oblike – ter v empirični analizi sistematično popišemo in kvantificiramo obe, pri čemer je osrednji predmet analize prav porazdelitev in obseg latiničnih variant v angleških virih.

3 Analiza

V tem delu predstavljamo korpusnojezikoslovno analizo različic transliteracije kazaških lastnih imen v angleščini in njihovega morfološkega podomačevanja. Preden smo se lotili te korpusnojezikoslovne raziskave, smo med jezikoslovci opravili predhodno anketo, da bi potrdili smiselnost poglobljene študije. Tako smo na mednarodni konferenci »Prevajalske tehnologije in primerjalnoknjiževnostne raziskave v znanstveni zapuščini Akhmeta Baitursynulya« (univerza Gumilyov ENU, Astana, Kazahstan, 28. oktobra 2022) izvedli anketo, s katero smo udeležence prosili, naj ime Abai Qunanbaiuly zapišejo v angleščini. Čeprav smo skupno zabeležili 20 anketirancev, je različnih variant skupno 21, ker je ena anketa (ali en anketiranec/anketiranka) navedel/navedla več variant. Če vse navedene variante obravnavamo kot izbire (skupaj = 21), dobimo naslednje približne odstotke:

1. Abai Qunanbaiuly: $7/21 \approx 33,3 \%$
2. Abai Kunanbaiuly: $4/21 \approx 19,0 \%$
3. Abay Kunanbayuly: $4/21 \approx 19,0 \%$
4. Abai Qunanbayev: $3/21 \approx 14,3 \%$
5. Abay Kunanbayev: $1/21 \approx 4,8 \%$
6. Abay Kunanbaiuly: $1/21 \approx 4,8 \%$
7. Abay Kunanbaev: $1/21 \approx 4,8 \%$

Kljub razdrobljenosti lahko opazimo nekaj izraženih smernic: razlike med zapisom s Q in K, konkurenčne rusificirane in tradicionalne končnice (-baev/bayev proti -baiuly/bayuly/baiyuly) ter končnico -ai/ay. To v praksi pomeni številne potencialne kombinacije in močno razdrobljenost zapisovalnih možnosti, še posebej če upoštevamo

dejstvi, da gre za najpomembnejšo literarno in zgodovinsko osebnost države ter da so bili anketiranci jezikoslovci in prevodoslovci.

3.1 Korpusnojezikoslovna analiza variantnosti v zapisu

V tem odseku poskušamo ugotoviti dejansko variantnost pri prečrkovanju in zapisovanju kazahstanskih priimkov, to pa poskušamo napraviti z obsežno korpusnojezikoslovno analizo.

3.1.1 Podatkovna baza

Za analizo smo sestavili korpus iz najbolj relevantnih kazahstanskih spletnih virov, ki so na voljo v angleščini. Korpus obsega raznolike vire, vključno z novicami, akademskimi članki, vladnimi dokumenti in spletnimi publikacijami, ki izhajajo iz uglednih kazahstanskih spletnih strani in medijev. Ti viri so bili izbrani na podlagi njihove relevantnosti za jezikovno krajino Kazahstana, dostopnosti za angleško govoreče občinstvo in obsežnosti, saj gre za izključno uradne in/ali vladne institucije. Seznam spletnih strani v angleščini, s katerih smo zbrali besedilni korpus,⁶ je naslednji:

1. Urad predsednika Republike Kazahstan (<https://www.akorda.kz/en>)
2. Literarni portal, ki je nastal s podporo Ministrstva za kulturo in informacije ter Centralne komunikacijske službe pri službe predsednika Republike Kazahstan (<https://adebiportal.kz/en>)
3. Portal e-History, državni portal za zgodovino Kazahstana (<https://e-history.kz/en>)
4. Urad predsednika vlade Republike Kazahstan (<https://primeminister.kz/en>)
5. Spletne novice Kazinform (<https://www.inform.kz/en>)
6. Evrazijska narodna univerza ENU (<https://enu.kz/en>)
7. Univerza Abai (Abay) (<https://www.kaznpu.kz/en/>)

Podatkovna zbirka torej zajema široko paleto tematik – politiko, kulturo, gospodarstvo, izobraževanje, zgodovino in družbo, s čimer smo želeli v čim večji meri zajeti večplastnost diskurza. Končna velikost korpusa je navedena v tabeli 3.

Tabela 3: Korpus v številkah.

| | |
|-----------|------------|
| Pojavnice | 30.479.361 |
| Besede | 25.819.605 |
| Povedi | 1.252.219 |

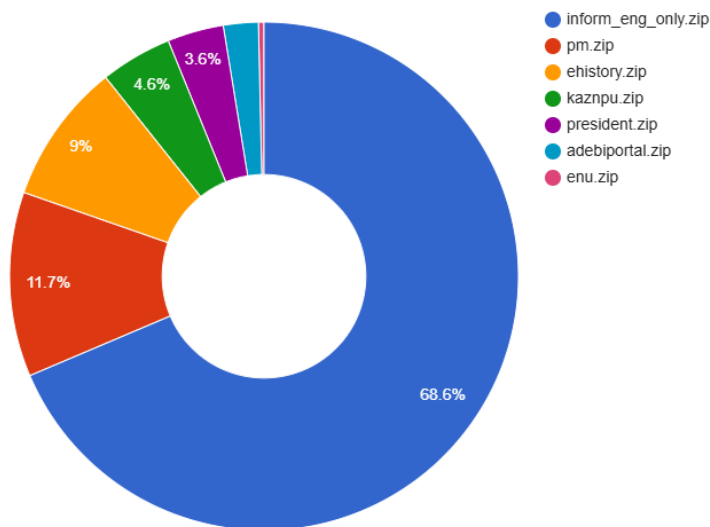
⁶ Za luščenje besedil smo uporabili knjižnico BeautifulSoup (<https://pypi.org/project/beautifulsoup4/>).

Med sestavljanjem korpusa smo si prizadevali, da bi bil čim bolj uravnotežen, vendar smo bili omejeni z velikostjo in razpoložljivostjo besedil v angleščini na posameznih spletnih straneh. Tabela 4 in slika 2 prikazujeta distribucijo pojavnic v različnih delih celotnega korpusa.

Tabela 4: Porazdelitev pojavnic med spletnimi viri.

| Element | Pojavnice |
|--------------------------------------|------------|
| Kazinform | 20.910.476 |
| Urad predsednika vlade | 3.577.638 |
| e-History | 2.729.306 |
| Abayeva (Abaieva) univerza | 1.389.286 |
| Urad predsednika Republike Kazahstan | 1.089.204 |
| Portal Adebi | 675.890 |
| Univerza ENU | 107.561 |

Slika 2: Distribucija besedilnih virov v korpusu.



Kot prikazujeta tabela 4 in slika 2, je porazdelitev pojavnic med temi viri izrazito neenakomerna, saj znaten del izvira iz ene same (novičarske) spletne strani. Kazinform zajema največji segment besedil v korpusu, saj predstavlja približno 20,91 milijona pojavnic, kar je okoli 68,6 % celotnega korpusa. Ta prevlada kaže tudi, da Kazinform kot glavna novičarska agencija igra ključno vlogo pri širjenju angleških vsebin znotraj Kazahstana in zlasti zunaj njega. Glede na njegov širok doseg in bralstvo ima obsežna produkcija zelo močan vpliv na prevladujoče transliteracijske prakse.

Urad predsednika vlade je drugi največji vir, ki zajema približno 3,58 milijona pojavnic (približno 11,7 % celotnega korpusa). Ta segment odraža prizadevanja vlade za sporočanje uradnih informacij in politik v angleščini, namenjenih domačemu, predvsem pa mednarodnemu občinstvu. Takšna uradna komunikacija je ključnega pomena za oblikovanje standardizirane rabe jezika, vključno z normami transliteracije.

Portal e-History zajema približno 2,73 milijona pojavnic (9 %) in ponuja zgodovinsko in kulturno vsebino, ki bogati raznolikost korpusa. Ta vir dodaja globino z vključitvijo besedil, ki lahko zaradi zgodovinskega konteksta vsebine zajemajo različne transliteracijske prakse. Ostali viri zajemajo okoli 10 % vseh besedil.

Na splošno distribucija kaže močno odvisnost od novic in vladnih spletnih strani, kar lahko znatno vpliva na prečkvalne vzorce. Prevladujoča vloga Kazinforma kaže, da bi lahko transliteracijske prakse, ki jih uporablja ta agencija, nesorazmerno vplivale na splošne ugotovitve in potencialno izkrivile podatke, vendar smo se zaradi pomanjkanja razpoložljivih (in ustreznih) virov v angleščini odločili, da v analizo vključimo vse razpoložljive podatke, pri čemer se v celoti zavedamo, da porazdelitev med viri korpusa ni idealna.

V naslednjih odsekih predstavimo analizo neskladij pri transliteraciji, opazovanih v korpusu, s sistematičnim pregledom lastnih imen in izrazov pa želimo pojasniti zapletenost prečkvalnih praks in njihove posledice za medkulturno komunikacijo in jezikovno standardizacijo.

3.1.2 Metodologija

Najprej smo identificirali seznam pogostih kazahstanskih končnic, ki se pojavljajo v moških in ženskih osebnih imenih. Nato smo za vsako končnico iz korpusa izluščili vse pojavitve, zabeležili absolutno frekvenco vsakega korena z vsako od končnic in izvedli statistično analizo z orodjem JupyterLab.⁷

Da bi upoštevali razlike v velikosti korpusa in olajšali primerjavo, smo izračunali relativno frekvenco na milijon besed za vsako kombinacijo korena in končnice. Ta normalizacija nam je omogočila, da smo ocenili razširjenost vsake oblike znotraj korpusa. Nato smo preučili razlike v frekvencah med posameznimi končnicami za vsak koren. Z izračunom absolutnih razlik in odstotkovnih distribucij želimo identificirati preferenčne vzorce v prečkvalnih praksah.

Poleg tega smo za vsak koren izračunali entropijo in variacijsko vrzel. Entropija je statistično merilo nepredvidljivosti ali raznolikosti končnic (pripon) imen. Če se sko-

⁷<https://jupyter.org/>.

raj vsa imena končajo enako (npr. vsa se končajo na »-ova«), je predvidljivost visoka, entropija pa nizka. Nasprotno pa je, če so različne končnice enakomerno razporejene in se pojavljajo s podobno pogostostjo, predvidljivost nizka, entropija pa visoka, saj to kaže na bogat in enakomerno porazdeljen nabor variantnosti. Variacijska vrzel pa je po drugi strani bolj neposredno merilo razlike med dejanskim številom edinstvenih končnic, ki smo jih opazili v svojih podatkih, in največjim možnim številom možnih končnic, ki jih dopušča slovnica jezika ali omejitve določene podatkovne baze. Ti dve merili se dopolnjujeta, saj entropija količinsko opredeljuje stopnjo naključnosti in enakomernosti v porazdelitvi končnic imen (kvaliteta variantnosti), medtem ko variacijska vrzel izpostavlja stopnjo izkoriščenosti celotnega potencialnega obsega variacije (kvantiteta variantnosti).

V vsaki tabeli uporabimo vzorec 30 oblik zaradi omejenega prostora, vendar vključujemo tudi povezave do obeh glavnih tabel. Celotna zbirka podatkov vsebuje 1217 moških osnov z petimi možnimi končnicami in 633 različnih ženskih osnov s štirimi možnimi končnicami.

3.1.3 Moška imena

Tabela 5 prikazuje moške oblike, razvrščene po najvišji pogostosti.⁸ Osredotočili smo se na končnice -baev, -baiev, -baiuly, -bayev in -bayuly.

Tabela 5: Moške oblike, razvrščene po pogostnosti.

| osnova | baev | baiev | baiuly | bayev | bayuly | št. vseh pojavitev | entropija | št. različnih končnic | variacijska vrzel |
|--------|------|-------|--------|-------|--------|--------------------|-----------|-----------------------|-------------------|
| Nazar | 45 | 0 | 0 | 15120 | 0 | 15165 | 0,029 | 2 | 0,003 |
| Sapar | 6 | 0 | 0 | 648 | 0 | 654 | 0,075 | 2 | 0,009 |
| Balyk | 82 | 0 | 0 | 547 | 0 | 629 | 0,558 | 2 | 0,13 |
| Bozum | 11 | 0 | 0 | 464 | 0 | 475 | 0,159 | 2 | 0,023 |
| A | 9 | 0 | 0 | 449 | 3 | 461 | 0,195 | 3 | 0,026 |
| Usken | 25 | 0 | 0 | 425 | 0 | 450 | 0,31 | 2 | 0,056 |
| Ashim | 9 | 0 | 0 | 409 | 0 | 418 | 0,15 | 2 | 0,022 |
| Kunan | 29 | 1 | 33 | 266 | 31 | 360 | 1,26 | 5 | 0,261 |
| Kozy | 77 | 0 | 0 | 231 | 2 | 310 | 0,862 | 3 | 0,255 |
| Zhamau | 11 | 0 | 0 | 297 | 0 | 308 | 0,222 | 2 | 0,036 |
| Kusher | 7 | 0 | 0 | 276 | 1 | 284 | 0,2 | 3 | 0,028 |
| Zhuma | 39 | 0 | 6 | 220 | 5 | 270 | 0,873 | 4 | 0,185 |
| Buri | 21 | 0 | 0 | 230 | 0 | 251 | 0,415 | 2 | 0,084 |

⁸ Celotna tabela je na voljo tukaj: https://www.dropbox.com/scl/fi/icj4ym64fdvayct5fovi/stem_by_ending_with_entropy_masculine.xlsx?rlkey=bxdbvonlo20t07ip4ilzdr781&dl=1.

| osnova | baev | baiev | baiuly | bayev | bayuly | št. vseh pojavitev | entropija | št. različnih končnic | variacijska vrzel |
|--------|------|-------|--------|-------|--------|--------------------|-----------|-----------------------|-------------------|
| Tuime | 4 | 0 | 0 | 244 | 0 | 248 | 0,119 | 2 | 0,016 |
| Yermek | 1 | 0 | 0 | 217 | 0 | 218 | 0,042 | 2 | 0,005 |
| Atam | 4 | 0 | 0 | 211 | 0 | 215 | 0,134 | 2 | 0,019 |
| Kira | 38 | 0 | 0 | 155 | 0 | 193 | 0,716 | 2 | 0,197 |
| Sarsen | 2 | 0 | 1 | 179 | 0 | 182 | 0,136 | 3 | 0,016 |
| Zhoshy | 4 | 0 | 0 | 173 | 0 | 177 | 0,156 | 2 | 0,023 |
| Zha | 18 | 0 | 1 | 149 | 0 | 168 | 0,543 | 3 | 0,113 |
| Mussa | 4 | 0 | 0 | 164 | 0 | 168 | 0,162 | 2 | 0,024 |
| Myn | 10 | 1 | 0 | 124 | 1 | 136 | 0,503 | 4 | 0,088 |
| Ashik | 1 | 0 | 0 | 128 | 0 | 129 | 0,065 | 2 | 0,008 |
| Kara | 2 | 0 | 0 | 122 | 0 | 124 | 0,119 | 2 | 0,016 |
| Bishim | 11 | 0 | 0 | 109 | 0 | 120 | 0,442 | 2 | 0,092 |
| Turgum | 1 | 0 | 0 | 114 | 0 | 115 | 0,072 | 2 | 0,009 |
| Alpys | 11 | 0 | 7 | 88 | 0 | 106 | 0,821 | 3 | 0,17 |
| Kuli | 4 | 0 | 0 | 100 | 0 | 104 | 0,235 | 2 | 0,038 |
| Sauda | 0 | 0 | 0 | 101 | 0 | 101 | 0 | 1 | 0 |
| Pshem | 12 | 0 | 0 | 87 | 0 | 99 | 0,533 | 2 | 0,121 |

Tabela 5 kaže, da rusificirana končnica *-bayev* še vedno prevladuje: v 25 od 30 osnovah predstavlja vsaj 80 % vseh pojavitev, pogosto pa izpodriva vse konkurenčne oblike. Najbolj nazoren primer je *Nazar-*, kjer oblika *Nazarbayev* sama zajema 15.120 od 15.165 pojavitev, kar entropijo zniža na skoraj nič (0,03), medtem ko variacijska vrzel znaša zgolj 0,003. Takšne skoraj neobstoječe vrzeli nam povedo, da pri javnih imenih z največjo prepoznavnostjo (predvsem pri aktualnih imenih iz sveta politike) variantnosti skorajda ni: končnica *-bayev* je utrjena norma.

Kot kaže tabela 5, so najpogostejše oblike torej najbolj stabilne. Tabela 6 pa prikazuje moške oblike z najvišjo stopnjo variantnosti (in minimalno skupno frekvenco 10).

Tabela 6: Moške oblike, razvrščene po entropiji.

| osnova | baev | baiev | baiuly | bayev | bayuly | št. vseh pojavitev | entropija | št. različnih končnic | variacijska vrzel |
|---------|------|-------|--------|-------|--------|--------------------|-----------|-----------------------|-------------------|
| Sagyr | 5 | 0 | 3 | 4 | 6 | 18 | 1,955 | 4 | 0,667 |
| Kurman | 2 | 0 | 3 | 6 | 2 | 13 | 1,834 | 4 | 0,538 |
| Kulzha | 6 | 0 | 0 | 5 | 2 | 13 | 1,46 | 3 | 0,538 |
| Zholdy | 2 | 0 | 1 | 12 | 4 | 19 | 1,457 | 4 | 0,368 |
| Berkim | 7 | 0 | 4 | 13 | 0 | 24 | 1,428 | 3 | 0,458 |
| Naiman | 5 | 0 | 1 | 17 | 2 | 25 | 1,32 | 4 | 0,32 |
| Sandy | 3 | 0 | 1 | 6 | 0 | 10 | 1,295 | 3 | 0,4 |
| Ak | 3 | 0 | 0 | 9 | 2 | 14 | 1,287 | 3 | 0,357 |
| Qunan | 0 | 0 | 26 | 13 | 4 | 43 | 1,279 | 3 | 0,395 |
| Kunan | 29 | 1 | 33 | 266 | 31 | 360 | 1,26 | 5 | 0,261 |
| Nur | 7 | 0 | 0 | 9 | 1 | 17 | 1,253 | 3 | 0,471 |
| Kekil | 17 | 0 | 4 | 55 | 2 | 78 | 1,19 | 4 | 0,295 |
| Duisen | 2 | 0 | 1 | 7 | 0 | 10 | 1,157 | 3 | 0,3 |
| Nurly | 5 | 0 | 0 | 12 | 1 | 18 | 1,135 | 3 | 0,333 |
| Kenes | 6 | 0 | 0 | 14 | 1 | 21 | 1,116 | 3 | 0,333 |
| Kenzhe | 22 | 0 | 0 | 43 | 2 | 67 | 1,089 | 3 | 0,358 |
| Ait | 9 | 0 | 2 | 31 | 0 | 42 | 1,009 | 3 | 0,262 |
| Argyn | 6 | 0 | 0 | 6 | 0 | 12 | 1 | 2 | 0,5 |
| Tele | 5 | 0 | 0 | 5 | 0 | 10 | 1 | 2 | 0,5 |
| Daulet | 12 | 0 | 0 | 13 | 0 | 25 | 0,999 | 2 | 0,48 |
| Kily | 7 | 0 | 0 | 8 | 0 | 15 | 0,997 | 2 | 0,467 |
| Kydyr | 5 | 0 | 0 | 6 | 0 | 11 | 0,994 | 2 | 0,455 |
| Shangyt | 5 | 0 | 0 | 6 | 0 | 11 | 0,994 | 2 | 0,455 |
| Elu | 8 | 0 | 0 | 6 | 0 | 14 | 0,985 | 2 | 0,429 |
| Mamyr | 7 | 0 | 0 | 10 | 0 | 17 | 0,977 | 2 | 0,412 |
| Bazar | 0 | 0 | 14 | 21 | 0 | 35 | 0,971 | 2 | 0,4 |
| Asem | 0 | 0 | 4 | 0 | 6 | 10 | 0,971 | 2 | 0,4 |
| Konys | 5 | 0 | 0 | 8 | 0 | 13 | 0,961 | 2 | 0,385 |
| Sokpak | 15 | 0 | 0 | 9 | 0 | 24 | 0,954 | 2 | 0,375 |
| Tazhi | 19 | 0 | 0 | 32 | 0 | 51 | 0,953 | 2 | 0,373 |

Če izločimo osnove, ki se pojavijo manj kot 10-krat, in jih razvrstimo po entropiji, dobimo jasno sliko, kje so razlike v zapisu. Dve vodilni osnovi – Sagyr- (entropija 1,96) in Kurman- (1,83) – kažeta skoraj šolsko štiridelnostno distribucijo: nobena od končnic ne doseže 45 % vseh pojavitev. Pri osnovi Sagyr- na primer rusificirana oblika -bayev zajema le 4 od 18 pojavitev, medtem ko kazahstanski patronimik -bayuly prevladuje s 6,

-baev in -baiuly pa sta manjša, a pomembna akterja. Variacijska vrzel (0,667) potrjuje, da dve tretjini pojavitev imena nimata prevladujočega zapisa.

Na splošno ti podatki kažejo, da se visoka variantnost pojavlja predvsem pri srednje pogostih imenih – tistih, ki so preveč pogosta, da bi jih lahko ignorirali, vendar še niso standardizirana v medijih ali uradnih krogih. Kadar osnova doseže več tisoč pojavitev (npr. Nazar-), rusificirana končnica -bayev še vedno prevladuje; kadar pa je pojavitev manj, je praksa pri transliteraciji bistveno bolj neenotna.

3.1.4 Ženska imena

Tabela 7 prikazuje najpogostejše ženske oblike (zgolj s končnicami -baieva, -baikyzy, -bayeva in -baykyzy, saj korpus ne vsebuje nobenega imena, ki bi se končal na -baeva), razvrščene po pogostosti. Treba je opozoriti, da so frekvence pri ženskih oblikah bistveno nižje.⁹

Tabela 7: Ženske oblike, razvrščene po pogostosti.

| osnova | -baieva | -baikyzy | -bayeva | -baykyzy | št. vseh pojavitev | entropija | št. različnih končnic | variacijska vrzel |
|--------|---------|----------|---------|----------|--------------------|-----------|-----------------------|-------------------|
| Nazar | 0 | 0 | 300 | 0 | 300 | 0 | 1 | 0 |
| Tulek | 0 | 0 | 83 | 0 | 83 | 0 | 1 | 0 |
| Kulam | 0 | 0 | 62 | 0 | 62 | 0 | 1 | 0 |
| Assau | 0 | 0 | 59 | 0 | 59 | 0 | 1 | 0 |
| A | 0 | 0 | 58 | 0 | 58 | 0 | 1 | 0 |
| Oraz | 0 | 0 | 57 | 1 | 58 | 0,126 | 2 | 0,017 |
| Zhuma | 0 | 0 | 46 | 0 | 46 | 0 | 1 | 0 |
| Movke | 0 | 0 | 40 | 0 | 40 | 0 | 1 | 0 |
| Otun | 0 | 0 | 39 | 0 | 39 | 0 | 1 | 0 |
| Rym | 0 | 0 | 37 | 0 | 37 | 0 | 1 | 0 |
| Tauke | 0 | 0 | 34 | 0 | 34 | 0 | 1 | 0 |
| Buri | 0 | 0 | 32 | 0 | 32 | 0 | 1 | 0 |
| Zhaksi | 0 | 0 | 31 | 0 | 31 | 0 | 1 | 0 |
| Tursyn | 0 | 0 | 29 | 0 | 29 | 0 | 1 | 0 |
| Košer | 0 | 0 | 28 | 0 | 28 | 0 | 1 | 0 |
| Shirin | 0 | 0 | 27 | 0 | 27 | 0 | 1 | 0 |
| Tau | 0 | 0 | 25 | 1 | 26 | 0,235 | 2 | 0,038 |
| Bas | 0 | 0 | 25 | 0 | 25 | 0 | 1 | 0 |

⁹ Celotna tabela je na voljo tukaj: https://www.dropbox.com/scl/fi/uxmjbbcq0fuxuiywr68h/feminine_stem_by_ending_with_entropy.xlsx?rlkey=l8duqp1n8b8d67sckaus6clsb&st=-9689h1iv&dl=1.

| osnova | -baieva | -baikyzy | -bayeva | -baykyzy | št. vseh pojavitev | entropija | št. različnih končnic | variacijska vrzel |
|--------|---------|----------|---------|----------|--------------------|-----------|-----------------------|-------------------|
| Kungei | 0 | 0 | 24 | 0 | 24 | 0 | 1 | 0 |
| Yendi | 0 | 0 | 23 | 0 | 23 | 0 | 1 | 0 |
| Isa | 0 | 0 | 22 | 0 | 22 | 0 | 1 | 0 |
| Sarsem | 0 | 0 | 22 | 0 | 22 | 0 | 1 | 0 |
| Uner | 0 | 0 | 22 | 0 | 22 | 0 | 1 | 0 |
| Aman | 0 | 0 | 20 | 1 | 21 | 0,276 | 2 | 0,048 |
| Jarkim | 0 | 0 | 20 | 0 | 20 | 0 | 1 | 0 |
| Konka | 0 | 0 | 20 | 0 | 20 | 0 | 1 | 0 |
| Sagim | 0 | 0 | 19 | 0 | 19 | 0 | 1 | 0 |
| Kolum | 0 | 0 | 18 | 0 | 18 | 0 | 1 | 0 |
| Sarsen | 0 | 0 | 18 | 0 | 18 | 0 | 1 | 0 |
| Kar | 0 | 0 | 17 | 0 | 17 | 0 | 1 | 0 |

V nasprotju z moškimi oblikami so ženska imena bistveno bolj enolična. Sedemindvajset od tridesetih najpogostejših osnov se pojavlja izključno z rusificirano končnico -bayeva. Pri zgolj treh osnovah – Ora-, Tau- in Aman- – je tradicionalna končnica -baykyzy konkurenčna. Z drugimi besedami, nobeno žensko ime v pasu visoke frekvence ne kaže pomembnega tekmovanja med končnicami.

Tabela 8 prikazuje najbolj spremenljive ženske oblike, vendar brez zahteve po minimalni frekvenci (10 pojavitev), saj so frekvence nizke.

Tabela 8: Ženske oblike, razvrščene po entropiji.

| osnova | -baieva | -baikyzy | -bayeva | -baykyzy | št. vseh pojavitev | entropija | št. različnih končnic | variacijska vrzel |
|---------|---------|----------|---------|----------|--------------------|-----------|-----------------------|-------------------|
| Beissen | 0 | 1 | 1 | 0 | 2 | 1 | 2 | 0,5 |
| Bisen | 0 | 0 | 1 | 1 | 2 | 1 | 2 | 0,5 |
| Duysen | 0 | 0 | 2 | 3 | 5 | 0,971 | 2 | 0,4 |
| Ora | 0 | 0 | 3 | 2 | 5 | 0,971 | 2 | 0,4 |
| Nurly | 1 | 0 | 2 | 0 | 3 | 0,918 | 2 | 0,333 |
| Sapar | 0 | 0 | 2 | 1 | 3 | 0,918 | 2 | 0,333 |
| Serik | 0 | 0 | 2 | 1 | 3 | 0,918 | 2 | 0,333 |
| Tule | 0 | 1 | 2 | 0 | 3 | 0,918 | 2 | 0,333 |
| Duzel | 0 | 0 | 3 | 1 | 4 | 0,811 | 2 | 0,25 |
| Sultan | 0 | 1 | 3 | 0 | 4 | 0,811 | 2 | 0,25 |
| Mergen | 0 | 0 | 4 | 1 | 5 | 0,722 | 2 | 0,2 |
| Myrza | 0 | 0 | 4 | 1 | 5 | 0,722 | 2 | 0,2 |
| Bazar | 0 | 0 | 10 | 2 | 12 | 0,65 | 2 | 0,167 |
| Sansyz | 0 | 0 | 6 | 1 | 7 | 0,592 | 2 | 0,143 |

| osnova | -baieva | -baikyzy | -bayeva | -baykyzy | št. vseh pojavitev | entropija | št. različnih končnic | variacijska vrzel |
|--------|---------|----------|---------|----------|--------------------|-----------|-----------------------|-------------------|
| Ait | 0 | 1 | 7 | 0 | 8 | 0,544 | 2 | 0,125 |
| Boran | 0 | 0 | 8 | 1 | 9 | 0,503 | 2 | 0,111 |
| Kenzhe | 0 | 1 | 11 | 0 | 12 | 0,414 | 2 | 0,083 |
| Balta | 0 | 1 | 14 | 0 | 15 | 0,353 | 2 | 0,067 |
| Aman | 0 | 0 | 20 | 1 | 21 | 0,276 | 2 | 0,048 |
| Tau | 0 | 0 | 25 | 1 | 26 | 0,235 | 2 | 0,038 |
| Ali | 0 | 0 | 57 | 1 | 58 | 0,126 | 2 | 0,017 |
| Nazar | 0 | 0 | 300 | 0 | 300 | 0 | 1 | 0 |
| Tulek | 0 | 0 | 83 | 0 | 83 | 0 | 1 | 0 |
| Kulam | 0 | 0 | 62 | 0 | 62 | 0 | 1 | 0 |
| Assau | 0 | 0 | 59 | 0 | 59 | 0 | 1 | 0 |
| A | 0 | 0 | 58 | 0 | 58 | 0 | 1 | 0 |
| Zhuma | 0 | 0 | 46 | 0 | 46 | 0 | 1 | 0 |
| Movke | 0 | 0 | 40 | 0 | 40 | 0 | 1 | 0 |
| Otun | 0 | 0 | 39 | 0 | 39 | 0 | 1 | 0 |
| Rym | 0 | 0 | 37 | 0 | 37 | 0 | 1 | 0 |

Seznam najbolj variantnih ženskih imen razkriva, da se skoraj vsa resnična konkurenca pojavlja v izredno majhnih vzorcih. Dve vodilni osnovi – *Beissen-* in *Bisen-* – dosežeta entropijo = 1, ker sta njuni skupni frekvenci zelo nizki (2): ena oblika z *-bayeva* in ena *-baykyzy* (ali *-baikyzy*). Ena sama dodatna pojavitev v korpusu bi njuno entropijo zmanjšala za polovico. Enak vzorec se ponavlja pri osnovah *duysen-*, *Ora-*, *Nurly-* in *Sapar-*: entropija se giblje okoli 0,9, ker ima vsaka osnova zgolj tri do pet pojavitvev in ena sama pojavitev *-kyzy* zadostuje, da poruši ravnovesje.

Pogostejše osnove – *Bazar-* (12 znakov), *Kenzhe-* (12), *Balta-* (15) in *Aman-* (21) – še vedno kažejo, da ima rusificirana oblika *-bayeva* zgolj eno pravo alternativo (*-baykyzy*), vendar nikoli več kot dve pojavitvi na osnovo. Nad pragom 25 pojavitvev (*Tau-*, *Ora-*) se konkurenčna končnica izravna na statistično opombo, kar potrjuje, da se uveljavljene pisave *-bayeva* hitro stabilizirajo, ko ime pridobi javno prepoznavnost. To strnjeno pomeni naslednje:

1. Edina produktivna alternativa končnice je *-baykyzy*.

Razen ene izjeme (*-baieva*) (sprememba samoglasnika v korenu besede in ne prava sprememba končnice) je vsaka oblika, ki ni *-bayeva*, *-baykyzy*. Različice *-baikyzy* ali *-bayqyzy*, ki jih napoveduje kazahstanska fonologija, so praktično odsotne, kar nakazuje, da angleški mediji privzeto uporabljajo črkovanje »ay« + »k«, kadar uporabljajo končnico *-kyzy*.

2. V vseh primerih sta dve različni končnici.

Nobena ženska osnova se ne pojavlja s tremi ali štirimi končnic, za razliko od moškega sklopa, kjer soobstajajo *-bayev*, *-baev* in *-bayuly*. To poudarja asimetrijo med spoloma: za ženske ostaja izbira transliteracije binarna – tradicionalna *-kzyzy* proti sovjetski *-eva* – in zgolj najmanjši del korpusa ohranja prvo obliko.

Če povzamemo, ženska spremenljivost torej obstaja, vendar je statistično šibka. Le peščica imen srednje pogostosti ohranja končnico *-kzyzy*, vendar vsi štejejo manj kot 30 pojavitev; v trenutku, ko koren preseže ta prag, se utrdi končnica *-bayeva*.

4 Zaključek

Prehod kazaškega jezika s cirilice na latinico je ključni trenutek v jezikovni in kulturni evoluciji Kazahstana. Ta prehod, ki se je začel leta 2017, je del širših prizadevanj za modernizacijo države in integracijo v globalno skupnost. Naša raziskava se je poglobila v posledice tega prehoda, zlasti v transliteracijo kazahstanskih lastnih imen v angleščino – proces, ki bo imel zaradi odsotnosti uradnih smernic za prečrkovanje in zaradi znatne variantnosti pri zapisovanju imen izjemno težo pri določanju statusa in dinamičnosti v digitalni prihodnosti.

Korpusnojezikoslovna analiza angleških spletišč v Kazahstanu je razkrila znatne razlike v zapisovanju kazahstanskih osebnih lastnih imen, kar poudarja izzive, ki jih prinaša sprememba pisave. Kvantitativni podatki so izpostavili prevladujoče oblike, vendar so razkrili tudi opazna odstopanja, ki prispevajo k dvoumnosti in potencialnim nesporazumom.

S to raziskavo smo želeli ugotoviti, kako se kazahstanska imena dejansko zapisujejo v sodobni angleščini, in vprašati, kaj nam ta raba lahko pove o tekočem prehodu Kazahstana s cirilice na latinico. Korpusni podatki so nedvoumni: rusificirane končnice iz sovjetske dobe – *bayev* (moški) in *bayeva* (ženski) – še vedno prevladujejo v javnem diskurzu v angleščini. Med najpogostejšimi imeni variantnosti malodane ni, kar ponazarjata zlasti obliki *Nazarbayev* (15.120/15.165 pojavitev) in *Nazarbayeva* (300/300), tj. priimka nekdanjega kazahstanskega predsednika in njegove žene.

Segment imen, ki so manj pogosta, daje drugačno, bolj dinamično sliko. Približno ena od desetih moških osnov z 10–500 pojavitvami v korpusu se še vedno pojavlja s 3–5 končnicami. To so imena – *Sagyr-*, *Kurman-*, *Kunan-* –, pri katerih so kazahstanske patroni-

mne končnice *-bayuly/-baiuly* dejansko konkurenčne končnici *-bayev*. Njihov obstoj potrjuje, da se je kazahstanska razprava o latinizaciji prenesla v angleško jezikovno prakso, vendar le v kontekstih, v katerih nobeden od zapisov še ni dosegel kritične mase. Ženska imena so bistveno manj variantna; celo »najbolj variantne« osnove redko presegajo eno samo pojavitev končnice *-baykyzy* v primerjavi s končnico *-bayeva*. Z drugimi besedami, variantnost v zapisu je v veliki meri pojav, vezan na moška imena.

To verjetno pomeni, da menjava pisave sama po sebi ne bo spremenila variantnosti v zapisu. Brez izrecnih stilističnih priročnikov in institucionalnih smernic bodo rusificirane končnice svojo statistično prevlado ohranile tudi v obdobju latinice, vprašanje pa ostaja, kakšen bo vpliv geopolitike in razmerja Kazahstana do Rusije. Prečrkovalne prakse torej niso niti naključne niti popolnoma ustaljene: odvisne so od pogostnosti imen, spola in vpliva preteklih in sodobnih jezikovnapolitičnih in jezikovnačrtovalskih praks. Zgolj podrobno – tako kvantitativno kot tudi kvalitativno podkrepjeno – razumevanje, kje v jeziku obstaja prava variantnost – in zakaj do nje prihaja – jezikoslovcem in načrtovalcem omogoča, da reformo pisave v Kazahstanu pretvorijo iz zgolj grafičnega ukrepa v širše preoblikovanje jezikovne identitete.

Viri in literatura

- AKSHOLAKOVA, Assem, Nurgul ISMAILOVA, 2013: The Language Policy of Kazakhstan and the State Language in Government Service. *Procedia – Social and Behavioral Sciences* 93. 1580–1586.
- ANTROPNOMYIC DICTIONARY [Антропонимдик сөздик], 2018: Almaty: Ministry of Education and Science of the Republic of Kazakhstan Science Committee, Akhmet Baitursunov Institute of Language Education. 250.
- AYAGAN, Burkitbai (ur.), 2005: *National Encyclopedia of Kazakhstan*. Zv. 3. Almaty: Kazakh Encyclopaedia. 560.
- BAZARBAYEVA Zeynep Muslimovna, Talshyn Kanatkyzy CHUKAYEVA, 2021: The New Kazakh Latinized Script. *Bulletin of L. N. Gumilyov Eurasian National University. Philology Series* 3/136. 24–30. DOI: <https://doi.org/10.32523/2616-7174-2021-136-3-24-30>.
- BAZARBAYEVA, Zeynep Muslimovna, Didar SADYK, 2024: Certain Issues in Spelling Onomastic Names. *Bulletin of Sh. Ualikhanov KU* 4. 60–70. DOI: <https://doi.org/10.59102/kufil/2024/iss4pp60-70>.
- BAZARBAYEVA, Zeynep Muslimovna, Talshyn Kanatkyzy CHUKAYEVA, 2023: The Phonematic Principle of the Kazakh Script Based on Latin Graphics. *Bulletin of L. N. Gumilyov Eurasian National University. Philology Series* 1/142. DOI: <https://doi.org/10.32523/2616-678X-2023-142-1-8-14>.
- BEKZHANOVA, Zhazira, Tsediso Michael MAKOELLE, 2022: Latinization of the Kazakh Alphabet: Implications for Education, Inclusion, and Social Cohesion in Kazakhstan. *SAGE Open* 12/4.
- COMRIE, Bernard, 1999: Script and Identity in the Turkic Languages. *Language and National Identity in Central Eurasia*. Ur. William Fierman. Washington: Woodrow Wilson Center. 25–36.
- COMRIE, Bernard, Gerald STONE, Maria POLINSKY, 1996: *The Russian Language in the Twentieth Century*. Oxford: Clarendon Press.
- COULMAS, Florian, 2003: *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge: Cambridge University Press.
- DIRINGER, David, 1948: *The Alphabet: A Key to the History of Mankind*. New York: Funk & Wagnalls.
- ЕРМОЛОВИЧ, Дмитрий, 2001: *Имена собственные на стыке языков и культур*. Москва: Р. Валент.

- [ERMOLOVIČ, Dmitrij, 2001: *Imena sobstvennye na styke jazykov i kul'tur*. Moskva: R. Valent.]
- FIERMAN, William, 2006: Language and Education in Post-Soviet Kazakhstan: Kazakh-Medium Instruction in Urban Schools. *The Russian Review* 65/1. 98–116.
- FIERMAN, William, 2009: Identity, Symbolism, and the Politics of Language in Central Asia. *Europe-Asia Studies* 61/7. 1207–1228.
- GRENOBLE, Lenore A., 2003: *Language Policy in the Soviet Union*. New York idr.: Kluwer Academic Publishers.
- KASSENOVA, Meruyert D., Rymbek M. ZHUMASHEV, 2025: From the History of Translation Kazakh Script into Latin Graphic in the 1920 and 1930s. *Bulletin of the Karaganda University. History. Philosophy Series* 30/1. 139–147.
- LANDAU, Jacob M., Barbara KELLNER-HEINKELE, 2001: *Politics of Language in the Ex-Soviet Muslim States: Azerbaijan, Uzbekistan, Kazakhstan, Kyrgyzstan, Turkmenistan, and Tajikistan*. London: Hurst & Company.
- LEWIS, Geoffrey, 1999: *The Turkish Language Reform: A Catastrophic Success*. Oxford: Oxford University Press.
- MADIEVA, Gulmira B., Roza M. TAYEVA, 2014: The Evolution of the Kazakh Surnames: 'For' and 'Against'. *Procedia – Social and Behavioral Sciences* 116. 4798–4802.
- MAMYNOVA, Bagdan, 2022: The Newspaper »Kazakh« (1913–1918): Spelling Features and Punctuation in the Process of Approbation of the Usul-Sotie (Tote Zhazu) Script. *Philology Series* 2/186. 62–76.
- MAMYNOVA, Bagdan, Garifolla ANES, Umit ANESSOVA, 2023: The Beginning of the 20th Century: Spelling Features and Punctuation in the Process of Approbation of the Letter Usul-Sotie (Tote Zhazu) in the Newspaper «Kazakh» (1913-1918). *Vestnik of Saint Petersburg University. Asian and African Studies* 15. 568–587. DOI: <https://doi.org/10.21638/spbu13.2023.308>.
- MEIRAMBEKOVA, Lyazat K., Gulnaz R. DAUTOVA, 2021: The Historical Role of Kazakh Language in the System of Turkic Languages. *Eurasian Journal of Philology Science and Education*, 181(1). 119–126. DOI: <https://doi.org/10.26577/EJPh.2021.v181.i1.ph13>.
- MNAIDAROVA, Miramgul, Gulnar SARSEKE, Ibrahim SAHIN, 2025: Anthroponyms in the Kazakh and Turkish Languages: A Historical Review of Their Studies and Structural Features. *Treatises and Documents Journal of Ethnic Studies* 94. 208–226.
- NURGAZIEV, T. E., A. B. KYRYKBAEV, 2024: Methodological Aspect of Kazakhstan's Transition to Latin Graphic Writing. *International Journal of Humanities and Natural Sciences* 4–5/91. 53–56. DOI: <https://doi.org/10.24412/2500-1000-2024-4-5-53-56>.
- PAVLENKO, Aneta, 2008: Multilingualism in Post-Soviet Countries: Language Revival, Language Removal, and Sociolinguistic Theory. *International Journal of Bilingual Education and Bilingualism* 11/3–4. 275–314. DOI: <https://doi.org/10.1080/13670050802271517>.
- RIEKKINEN, Mariya, Aizhan ZHATKANBAYEVA, Kanat KOZHABEK, Gulnar AIGARINOVA, 2021: Latinizing the Alphabet in Kazakhstan: Rationale, Legal Foundation, and Possible Impact on the Status of the Russian Language. *European Yearbook of Minority Issues Online* 18. 1–25.
- ROBINSON, Andrew, 1995: *The Story of Writing: Alphabets, Hieroglyphs & Pictograms*. London: Thames & Hudson.
- RSALIYEVA, Nursaula, Kyzdarkhan RYSBERGEN, 2014: Onomastic Space of Kazakhstan: Current Condition and Problems. *Names and Their Environment: Proceedings of the 25th International Congress of Onomastic Sciences*. Zv. 4. Ur. Carole Hough, Daria Izdebska. Glasgow: University of Glasgow. 153–165.
- SMAGULOVA, Juldyz, 2008: Language Policies of Kazakhization and Their Influence on Language Attitudes and Use. *International Journal of Bilingual Education and Bilingualism* 11/3–4. 440–475.
- SULKARNAYEVA, Assiya, 2017: Analysis of PPP Use and Types of Questions at MA TESOL Modules. *Bulletin of KazGUU University JSC. Series Philological Sciences* 1/33. 76–81.
- ZADOROZHNAJA, Lyudmila Aleksandrovna, Aidana Almabekovna KUTPANOVA, Azatbek Manonuly BADALIYEV, 2025: Kazakh Proper Names in English Translation: Variability and Risks of Non-Standardized Usage of Computer-Assisted Translation Tools. *Bulletin of L. N. Gumilyov Eurasian National University. Philology Series* 4/153. 299–310. DOI: <https://doi.org/10.32523/2616-678X-2025-153-4-299-310>.
- ZHARKYNBEKOVA, Sholpan, Aliya AIMOLDINA, 2016: The Role of English Language in the Context of New Language Policy Implementation in Kazakhstan. *Sociolinguistic Transition in Former*

Eastern Bloc Countries: Two Decades after the Regime Change. Ur. Marián Sloboda, Petteri Laihonen, Anastassia Zabrodskaja. Frankfurt am Main: Peter Lang (Prague Papers on Language, Society and Interaction, 4). 399–418.

ZHOU, Minglang, 2003: *Multilingualism in China: The Politics of Writing Reforms for Minority Languages, 1949–2002*. Berlin: Mouton de Gruyter.

Summary

This article examines how Kazakhstan's ongoing shift from Cyrillic to Latin (initiated in 2017) interacts with English-language visibility: when Kazakh personal names move into English institutional and media texts, they often circulate in multiple, competing spellings. We built a 25-million-word corpus from English subpages of major Kazakhstani official/institutional sites (government portals, universities, national history and culture portals, and the Kazinform news agency) and quantified variation in two dimensions: (i) morphological choice between Russified surname endings and Kazakh patronymic forms (-bayev/-bayeva vs -uly/-kyzy variants) and (ii) alternative Latin renderings of the same base (e.g., q~k; ai~ay; -baev/-baiev/-bayev; -baiuly/-bayuly). A small expert elicitation at a 2022 conference already shows severe fragmentation for arguably the most recognizable historic figure of Kazakhstan, Abai Qunanbaiuly, with seven different spellings, whereas corpus results confirm strong frequency effects: the most visible public names are highly stabilized in Russified forms (e.g., Nazarbayev, Nazarbayeva), while mid-frequency male bases display the highest entropy, with 3–5 endings co-occurring and no single form reaching dominance. Female names are far less variable: -bayeva overwhelmingly prevails, and -kyzy forms appear only sporadically and mostly in very small samples. Overall, the study argues that script reform alone will not reduce real-world variation in English without explicit institutional guidelines because digital publication and technology-mediated transliteration amplify and entrench competing forms with consequences for identity, citability, and traceability in global information systems.