



Zakaj latinica ni dovolj: o izgubi informacij pri latinizaciji vzhodnoazijskih imen v knjižničnih katalogih

Why latin script is not enough: information loss through the romanisation of East Asian names in library catalogues

**Kristina Hmeljak Sangawa,
Hyeonsook Ryu, Mateja Petrovčič**

Oddano: 14. 10. 2019 – Sprejeto: 15. 8. 2020

1.01 Izvirni znanstveni članek

1.01 *Original scientific article*

UDK 025.355:025.31

UDK 017:019

Izvleček

Namen: Predstavljamo problem izgube informacij pri pretvarjanju imen iz kitajske, korejske in japonske pisave v latinico ter težave, ki jih nepovratna pretvorba prinaša tako knjižničarjem pri katalogizaciji kot tudi uporabnikom pri iskanju gradiva v vzhodnoazijskih jezikih v izključno latiničnem katalogu.

Metodologija/pristop: Po uvodnem opisu vzhodnoazijskih pisav, ki se uporabljajo za zapis kitajskih, korejskih in japonskih imen, standardov za latinični zapis ter ustanov, ki urejajo normativne podatke gradiva v teh jezikih, analiziramo postopke pretvarjanja zapisov lastnih imen iz vzhodnoazijskih pisav v latinico in obratno, ter ugotovljamo, kdaj pride do izgube informacij in posledično nepovratnosti pretvorbe. Na primeru knjižnice Oddelka za azijske študije Filozofske fakultete Univerze v Ljubljani (FF UL), ki je del Osrednje humanistične knjižnice in vključena v sistem COBISS, opišemo težave, ki jih povzroča izključno latinično zapisovanje podatkov pri upravljanju in rabi knjižnične zbirke v vzhodnoazijskih jezikih.

Rezultati: Iz prikaza različnih postopkov latinizacije kitajskih, korejskih in japonskih lastnih imen je razvidno, da pride v vseh primerih do izgube informacij. V primeru knjižnice Oddelka za azijske študije FF UL to povzroča težave tako upravljalcem kot uporabnikom knjižnične zbirke.

Omejitve raziskave: Raziskava analizira problem nepovratnosti pretvorbe vzhodnoazijskih pisav v latinico na teoretični ravni in na primeru ene knjižnice. V bodočih raziskavah pa bi bilo zanimivo podrobneje analizirati še druge konkretne primere, v katerih izključno latinični katalog otežuje ali onemogoča iskanje in upravljanje knjižničnih virov v nelatiničnih pisavah.

Izvirnost/uporabnost raziskave: Prva raziskava, ki osvetljuje problem latinizacije pri upravljanju knjižničnih zbirk v vzhodnoazijskih pisavah v slovenskem prostoru in predlaga možno rešitev tega problema z uvedbo standarda Unicode.

Ključne besede: latinizacija, kitajska pisava, korejska pisava, japonska pisava, knjižnični katalogi

Abstract

Purpose: The problem of information loss inherent in the romanisation of Chinese, Japanese and Korean names is examined together with the problems it poses both to librarians who record book information and to library users who search for resources written in East Asian languages in a catalogue containing only romanised transcriptions.

Methodology/approach: Firstly, we describe the writing systems used to write Chinese, Korean and Japanese, the romanisation standards for these languages, and the institutions providing authority data in Latin script. We then analyse the procedures for romanising East Asian proper names and then retrieving their original written form, and identify cases of information loss and irreversible transformations. As a case study, we present the difficulties encountered by librarians and users in the Library of the Department of Asian Studies of the Faculty of Arts at the University of Ljubljana (part of the Central Humanities Library and as such using the COBISS system) when using only romanised data to organise and search for library holdings in East Asian languages.

Results: Information loss occurs in all methods of romanisation of Chinese, Korean and Japanese proper names presented in the previous sections. In the case of the Library of the Department of Asian Studies, this is a burden to both librarians and library users.

Research limitations: We analysed the problem of the irreversibility of East Asian script romanisation in theory and in a case study involving one library; a more detailed analysis of other cases of Latin script catalogues hindering the management and use of library holdings in non-Latin scripts could be the object of future research.

Originality/practical implications: This is the first analysis of the problem of information loss in romanised catalogues of East Asian library holdings in Slovenia, with a proposal for a possible solution using the Unicode standard.

Keywords: romanisation, Chinese writing system, Korean writing system, Japanese writing system, library catalogues

1 Uvod

V prispevku predstavljamo problem izgube informacij pri pretvarjanju imen iz kitajske, korejske in japonske pisave v latinico in težavah, ki jih nepovratna pretvorba prinaša tako knjižničarjem kot uporabnikom. Težavo občutimo zlasti v knjižnici Oddelka za azijske študije FF UL, ki je del Osrednje humanistične knjižnice in kot taka vključena v sistem COBISS. Od ustanovitve oddelka leta 1995 se z aktivnim povezovanjem s tujimi institucijami in vedno večjo mednarodno prepoznavnostjo oddelka stalno povečuje zbirka gradiv v azijskih jezikih, ki je zdaj verjetno najobsežnejša zbirka knjig in periodike v vzhodnoazijskih jezikih v Sloveniji: v zadnjih dvajsetih letih smo s prijavami na razpise in donacijami prejeli več kot 20.000 knjig in drugih enot v vzhodnoazijskih jezikih (kitajščini, korejščini in japonščini).

Za uspešno evidentiranje gradiva in možnost informiranja uporabnikov o razpoložljivih virih je treba vse enote tudi katalogizirati. V knjižnici smo do zdaj v COBISS zavedli več kot 14.000 enot. Od tega je (poleg knjig v latiničnih pisavah, večinoma v angleščini in slovenščini) nekaj tisoč enot v japonščini, korejščini in kitajščini. Dober del oddelčne knjižnične zbirke še ni zaveden v COBISS, v prvi vrsti zaradi pomanjkanja kadra, a tudi zaradi nepraktičnosti in neučinkovitosti vnašanja podatkov v latinico. V COBISS tako ni zavedenih več tisoč knjig, ki so sicer popisane v inventarni knjigi (do leta 2013 smo jo vodili v obliki tabele excel, ki vsebuje tudi originalne zapise v kitajski, korejski oziroma japonski pisavi), 4500 enot donacije Hamaguchi (ki so bile leta 2014 popisane v ločeni interni datoteki v japonski pisavi in v latinici) ter donacije okoli 8000 kitajskih ter več sto japonskih in korejskih knjig, ki še niso popisane.

Tujejezično gradivo ne prihaja zgolj v knjižnico Oddelka za azijske študije, temveč tudi v Raziskovalno središče za Vzhodno Azijo (EARL), ki sta ga ustanovili Filozofska fakulteta in Fakulteta za družbene vede Univerze v Ljubljani kot osrednji repozitorij za tiskane in digitalne vire z Vzhodne Azije v Sloveniji in ima sedež v Osrednji družboslovni knjižnici Jožeta Goričarja (ODKJG) na Fakulteti za družbene vede Univerze v Ljubljani. Temeljna naloga EARL-a je zagotavljati dostop do virov in storitev ter povezovati Slovenijo z mrežo institucij iz Vzhodne Azije, ki prispevajo znanstveno in strokovno literaturo, elektronske baze podatkov in drugo gradivo, ter s tem spodbujati razumevanje med narodi teh regij, strokovno javnostjo, študenti in fakultetami. Zato da so viri dejansko dostopni, morajo biti seveda ustrezno obdelani.

Gradiva v vzhodnoazijskih jezikih ne prihajajo zgolj v fakultetne knjižnice, temveč tudi v splošne knjižnice. Na straneh Mestne knjižnice Ljubljana preberemo, da je uporabnikom na razpolago zbirka elektronskih knjig Okno Kitajske

v pretežno kitajskem jeziku, ki ga je knjižnica prejela od Šanghajske knjižnice v okviru njihovega mednarodnega projekta Window of Shanghai. Zbirka naj bi bila namenjena tako povezovanju kitajskih izseljencev z matično domovino kakor tudi temu, da mednarodni javnosti približa kitajski jezik in kulturo, sodobno Kitajsko in predvsem Šanghaj skozi najnovejše kitajske publikacije.

V slovenskem prostoru se torej povečuje količina gradiva v vzhodnoazijskih jezikih, s tem pa tudi potreba po bibliografskem sistemu, ki omogoča vnos in iskanje gradiv v nelatiničnih pisavah. COBISS trenutno tega ne podpira, zato v prispevku preverjamo, ali latinični zapis vendarle zadostuje za opis gradiva v vzhodnoazijskih pisavah, ali pa prihaja do izgube informacij pri pretvarjanju iz kitajske, korejske in japonske pisave v latinico.

V prispevku predstavljamo teoretični okvir in terminologijo za opis različnih sistemov pisave in pretvorb med njimi. Opišemo kitajski, korejski in japonski sistem pisave ter različne sisteme za latinični zapis teh jezikov. Predstavimo analizo nekaj primerov zapisov kitajskih in japonskih imen in probleme, ki nastanejo pri izključno latiničnem zapisovanju, reševanje tega problema v drugih sistemih in zaključne misli ter predlog uporabe standarda Unicode za vnašanje in hranjenje zapisov v COBISS-u.

2 Teoretična izhodišča in terminologija: prečrkovanje, transliteracija, transkripcija

V slovenščini se izraz *prečrkovanje* uporablja kot sinonim za dva termina: *transliteracijo* in *transkripcijo*, kar lahko privede do nesporazumov, zato v tem razdelku pojasnjujemo razliko med dvema vrstama pretvorbe.

*Transliteracija*¹ (ali *prečrkovanje* v ožjem smislu) je preslikava zapisa v eni pisavi (npr. v cirilici) v zapis v drugi pisavi (npr. v latinici). To je uporabno takrat, ko moramo zaradi fizičnih omejitev pisalne tehnike (starejši računalniki, telefoni ipd.) ali omejitev ciljnega bralca (npr. ko bralec ne pozna izvorne pisave) uporabiti drugo pisavo, da lahko prenesemo sporočilo (informacijo) prvotnega zapisa. Idealni sistem transliteracije je obojestransko enoznačen: vsakemu znaku izvorne pisave odgovarja točno en znak ciljne pisave, kot npr. pri pretvorbi iz srbske cirilice v srbsko latinico. Nepopoln sistem transliteracije, v katerem znaku ene pisave lahko odgovarja več znakov druge pisave, ni reverzibilen. Tako

¹ Dobesedni prevod besede *trans-literacija* je *pre-črkovanje*.

transliteracijo včasih uporabimo, ko ni drugih izbir, npr. za vnašanje slovenščine v okolju, ki ne podpira šumnikov, zaradi česar uporabimo črke brez strešic namesto šumnikov. Pri taki transliteraciji se izgubi del informacij: ko npr. vidimo zapis Sesev v besedilu, kjer nismo mogli uporabiti šumnikov, ne moremo vedeti, ali je v originalnem zapisu to priimek Sešek ali Šešek.

*Transkripcija*² (ali *zapis*) lahko pomeni dvoje: ali zapis sporočila, ki je bilo samo izgovorjeno (npr. zapis telefonskega pogovora, torej pretvorba govora v pisavo), ali zapis glasov, ki odgovarjajo napisu v neki izvorni pisavi, z uporabo druge pisave, npr. zapis glasov /tʃɑ:rlz/, ki odgovarjajo zapisu angleškega imena Charles, z znaki cirilice Чарлз. Pri transkripciji torej ne gre za preslikavo vsakega znaka ene pisave v dogovorjen znak druge pisave, ampak za bolj kompleksen proces v dveh fazah: najprej moramo po pravopisnih in pravorečnih pravilih izvornega jezika pretvoriti prvotni zapis v glasove, nato pa te glasove zapisati z znaki ciljne pisave, ki so najbližji temu izgovoru.

V Preglednici 1 navajamo nekaj primerov razlik med transliteracijo in transkripcijo.

Preglednica 1: Primeri transliteracije in transkripcije

Izvorni jezik	Izvorni zapis	Transliteracija (prečrkovanje)	Transkripcija (zapis, ki govorcju jezika ciljne pisave pokaže izgovarjavo)	Jezik ciljnega bralca
srbski	Београд	Beograd	Beograd	slovenščina
ruski	Пётр	Pjotr	Pjotr (prim. prevod: Peter)	slovenščina
japonski	おとうさん	otousan	otoosan	slovenščina
japonski	えいご	eigo	eego	slovenščina
grški	Άγγελος	Aggelos	Angelos	slovenščina
grški	Μπουμπουλίνα	Μpoupoulína	Bubulína	slovenščina
angleški	Charles	Цхарлєс	Чарлз	srbsčina

Če gre tako pri izvorni kot pri ciljni pisavi za sistem s plitkim pravopisom³ (tj. sistem, pri katerem praviloma vsakemu glasu odgovarja en znak in vsakemu znaku odgovarja en glas, blizu idealu Vuka Karadžića »Piši kot govoriš«), je transliteracija velikokrat obenem že transkripcija, kot v primeru imena Beograd. Če pa je izvorni zapis v pisavi z globokim pravopisom (kot je npr. angleški, pri katerem so pravila o razmerjih med glasovi in znaki vezana tudi na morfologijo in

² Dobesedni prevod besede *transkripcija* je *pre-pis*, ki pa ima v slovenščini več pomenov.

³ Izraza *plitki* in *globoki* pravopis uporabljamo kot prevoda angleških izrazov *shallow* in *deep orthography*, kot ju uporabljajo Coulmas (1989, str. 169) in drugi.

leksiko ter zato zelo kompleksna), potem transkripcija ni enaka transliteraciji, kot je razvidno iz zapisa angleškega imena Charles v cirilici v Preglednici 1. Cirilična *transliteracija* Цхарлес se običajno ne uporablja, je pa zvesta preslikava izvirnega latiničnega zapisa; *transkripcija* Чарлз, ki je običajni zapis tega imena v srbski cirilici, je srbskemu bralcu berljiv zapis, ni pa iz tega mogoče rekonstruirati izvirnega zapisa brez dodatnih zunanjih informacij.

Transliteracija in transkripcija imata torej drugačni funkciji. Če je pri pretvorbi iz ene pisave v drugo pomembno, da lahko informacijo kasneje tudi zanesljivo pretvorimo nazaj v izvorno obliko (kot npr. pri bibliografskih podatkih), je nujno, da je pretvorba enoznačna v obe smeri. To je mogoče, če lahko sestavimo prečrkovalno razpredelnico, v kateri vsakemu znaku ene pisave (npr. cirilice) odgovarja točno en znak ali skupek znakov druge pisave (npr. latinice). Z uporabo prečrkovalne razpredelnice je pretvorba enoznačna in jo lahko mehansko izvede tako stroj kot tudi človek, ki ne pozna ne enega ne drugega jezika. Če pa je glavni cilj pretvorbe iz ene pisave v drugo ta, da ciljni bralec pravilno izgovori zapisano sporočilo, je transkripcija bolj smiselna izbira, tudi če iz transkribiranega zapisa ni več mogoče rekonstruirati originalnega. Pri knjižničnih katalogih, ki primarno služijo hranjenju in iskanju pisnih podatkov, je seveda bolj pomembna povratnost pretvorbe kot pa izgovorljivost zapisa.

Prečkovanje iz srbske cirilice v srbsko latinico, ki se uporablja tudi pri zapisu srbskih imen v slovenskih besedilih, je enoznačno, zato prečkovanj zapis v latinici ni samo berljiv, ampak ga lahko tudi brez napak pretvorimo nazaj v cirilico. Zato uporaba latinice v COBISS-u za vnašanje podatkov o knjigah, ki so zapisane v srbski cirilici, ni problematična. Uporabniki, ki iščejo podatke o knjigah v srbščini, praviloma poznajo tako cirilico kot latinico in znajo pretvarjati med obema, pri pretvorbi iz ene pisave v drugo pa ne pride do izgube informacij.

Pri vnašanju podatkov, ki so izvorno zapisani s kitajskimi pismenkami, pa transliteracija (tj. prečkovanje v ožjem smislu) sploh ni možna, saj je pogoj za transliteracijo oziroma preslikavo ene pisave v drugo ta, da je v ciljni pisavi vsaj toliko znakov kot v izvorni pisavi. Različnih kitajskih znakov je več deset tisoč, zato ne moremo sestaviti prečrkovalne razpredelnice, v kateri bi vsakemu kitajskemu znaku odgovarjal en znak latinice. Pri takem številu znakov je edina rešitev za enoznačno pretvorbo kodirna tabela, v kateri vsaki kitajski pismenki odgovarja unikatni skupek vsaj petih latiničnih znakov ali števil, kar je za računalniško obdelavo smiselno in izvedljivo, za človeka pa zapis v tako pretvorjenih kodah seveda ni praktično berljiv. Ko hočemo človeku, ki ne pozna kitajske pisave, omogočiti branje kitajskega zapisa, zato uporabimo transkripcijo (in ne transliteracijo), tj. zapis izgovorjave kitajskih besed v latinici. Taka transkripcija je potem berljiva, ni pa povratna, saj s transkripcijo zapišemo samo izgovor besede, ne pa

tudi informacije o tem, katera od možnih pismenk se za to besedo uporablja v izvornem zapisu. Za iskanje bibliografskih podatkov je to podobno nerodno, kot če bi za angleško ime Jean Wright v knjižnični katalog vnesli samo transkripcijo (tj. zapis izgovorjave) Džiin Rajt (ali kar Džin Rajt, saj v slovenskem pravopisu ne zapisujemo razlike med dolgimi in kratkimi samoglasniki), potem pa bi, ko bi želeli poiskati avtorico v drugem katalogu ali jo citirati, morali najprej poiskati, ali se originalno zapiše Jean Wright ali Geene Right ali Gean Rite ali Djeen Rhyt ipd. Pri angleščini tega ne počenjamo, ker je jasno, da je za obdelavo, hranjenje in rabo bibliografskih podatkov bolj pomemben zapis kot izgovorjava. Tudi praktično ni problemov, ker so računalniški sistemi nastali najprej za angleščino, zato vsi računalniški nabori znakov vključujejo tudi vse znake angleščine, tudi nabori, ki so se razvili za slovenščino ali druge jezike, ki ne uporabljajo npr. črk w, x in y. Ko pa samo z latiničnim naborom znakov vnašamo podatke, ki so izvorno zapisani v kitajskih pismenkah, pride do nepovratne izgube informacij.

V naslednjem razdelku podrobneje pojasnimo ustroj kitajskega, korejskega in japonskega sistema pisave ter razlog, zakaj se besede, ki se izgovorijo (in v latinici zapišejo) enako, lahko v teh jezikih zapišejo na več različnih načinov.

3 Sistemi pisav vzhodnoazijskih jezikov

Črkovne pisave so le eden od možnih sistemov zapisovanja jezikov. Latinica spada med pisave, pri katerih posamezni znaki zapisujejo glasovne enote jezika,⁴ to so abecede (kot so latinica, cirilica in grška pisava), pri katerih vsak znak primarno zapisuje posamezni fonem, in zlogovnice (kot sta japonski hiragana in katakana), pri katerih en znak zapisuje en zlog. Obstajajo pa tudi pisave, ki primarno zapisujejo pomenske enote jezika, tj. morfeme in besede: poleg egipčanskih hieroglifov in drugih izumrlih pisav v to skupino spada kitajska pisava. Korejski in japonski pa sta sistema pisave, ki uporabljata znake obeh tipov.

Za prvi računalniški sistem katalogizacije knjig je bila latinica sprva tudi optimalni način zapisovanja, saj je bil ustroj in razvoj informacijske tehnologije tesno povezan z angleščino in angleško pisavo. Prvi format za izmenjavo informacij ASCII namreč odraža ravno to – obsega zgolj elemente (črke), ki so potrebni za zapis angleščine. Za zapis vzhodnoazijskih jezikov pa se uporabljajo sistemi

⁴ Pisave, ki primarno zapisujejo glasovne enote, po Bekešu (1999) in Coulmasu (1989) imenujemo *kenemske*, medtem ko tiste, ki primarno zapisujejo pomenske enote, imenujemo *pleremske*. Podrobneje o tem Hmeljak Sangawa (2019).

Kot vidimo iz Preglednice 2, zlog 'li' zapisujemo s približno 120 različnimi pismenkami, pri čemer vsako pismenko povezujemo z vsaj enim pomenom.

Teoretično se lahko vse te pismenke pojavijo kot del lastnega imena (podobno kot v slovenščini iz pridevnika *jasen* izhaja lastno ime *Jasna*), trinajst od zgoraj navedenih pismenk, s katerimi se lahko zapiše zlog *li*, pa se uporablja tudi kot priimek (podobno kot v slovenščini iz imena poklica *kovač* izhaja priimek *Kovač*, ali iz imena živali *čuk* priimek *Čuk*).

Ko torej pretvorimo pismenke v polni pinjin s toni, se kar 76 različnih pismenk združi v zapis *Lì* v četrtem tonu, 35 v *Lí* v drugem tonu in 16 v *Lǐ* v tretjem tonu. Ob naslednjem koraku, ko zaradi dodatne poenostavitve opustimo naglase, ki označujejo tone, nastane še več enakozvočnic. Popolnoma jasno je, da je rekonstrukcija v obratno smer nemogoča. Iz črkovnega zapisa »Li« ne moremo vedeti, za katero od 118 pismenk gre.

3.2 Korejski sistem pisave

Na Korejskem polotoku so v 2. stoletju n. št. pod kitajsko nadvlado začeli uporabljati kitajsko pisavo, sprva samo za pisanje besedil v klasični kitajščini. V naslednjih stoletjih so razvili sistem *idu* ali *imun* za zapisovanje korejščine s kitajskimi pismenkami. Ker je bila kitajščina sprva izolacijski jezik z večinoma enozložnimi besedami, korejščina pa aglutinacijski z večinoma večzložnimi besedami, pre-gibnimi priponami in drugačnim vrstnim redom besed, je bila raba kitajskih pismenk za zapis korejščine zelo kompleksen sistem, ki je zahteval dobro poznavanje velikega števila kitajskih pismenk in zapletenih pravil za njihovo rabo za zapis korejščine. Zato je leta 1446 kralj Sejong izumil in razglasil bolj praktičen sistem pisave, v katerem vsak znak zapisuje glasovno enoto. To je sistem *hangeul* (poslovenjeno *hangul*), ki se tudi danes uporablja za zapis korejščine (Coulmas, 1989, str. 120; DeFrancis, 1989, str. 186; Hannas, 1997, str. 54–55).

Za vsak glas korejskega glasovnega sistema se uporablja en znak (14 osnovnih znakov za soglasnike in 10 osnovnih znakov za samoglasnike), ki tudi grafično ponazarja mesto izgovarjave v človeškem govornem aparatu, tako npr. znak ㄱ za glas /k/ nakazuje jezik, ki se dotika mehkega neba ob izgovarjavi tega glasu, znak ㄴ za glas /n/ pa nakazuje jezik, ki se dotika dlesni (Taylor in Taylor, 2014, str. 183). Znaki za posamezne glasove so potem grafično združeni v sestavljene znake približno kvadratne oblike (ki je podobna obliki kitajskih pismenk), od katerih vsaka oblika zapisuje en zlog in je kombinacija dveh znakov, če je zlog odprt (tj. sestoji iz enega znaka za začetni soglasnik in enega za samoglasnik ali samoglasniški sklop, npr. 가, 너, 도, 루), ali kombinacija treh znakov, če je zlog zaprt (tj. sestoji iz

enega znaka za začetni soglasnik, enega za samoglasnik ali samoglasniški sklop ter enega znaka za zaključni soglasnik, npr. 간, 널, 돔, 룡). Zlogi so za človeka bolj dojemljiva enota kot posamezni glasovi in zato se lažje nauči brati sosledje grafičnih enot, ki zapisujejo zloge, kot pa sosledje znakov, ki zapisujejo posamezne glasove (Daniels, 2001, str. 68; Taylor in Taylor, 2014, str. 200).

Sistem je izredno racionalen in nadvse prilagojen korejskemu jeziku, kljub temu pa se je skozi zgodovino ohranila tudi raba kitajskih pismenk zaradi njihovega kulturnega ugleda. V Severni Koreji so sicer rabo kitajskih pismenk popolnoma opustili, v Južni Koreji pa se še občasno uporabljajo zlasti v strokovnih besedilih kot dodatni zapis za ločevanje kitajskih tujk, ki imajo veliko enakozvočnic, in za zapis lastnih imen. Pri pretvorbi teh kitajskih pismenk v latinico pride do podobne izgube informacij kot pri pretvorbi kitajskih imen, ki je opisana v razdelku 3.1. Poleg tega je pri zapisu z znaki hangula grafično vidna meja med zlogi, pri pretvorbi v niz latiničnih črk pa se informacija o meji med zlogi izgubi. Tako se npr. imeni Bom-i (v pisavi hangul 봄이) in Bo-mi (v pisavi hangul 보미) običajno obe pretvorita v latinični zapis Bomi brez oznake za mejo med zlogi, ki je torej dvoumno. Podrobneje o tem v razdelku 4.2.

3.3 Japonski sistem pisave

Na Japonskem so pisavo spoznali prek korejskih posrednikov, najverjetneje v 5. stoletju. Sprva so tudi na Japonskem uporabljali kitajske pismenke samo za pisanje v klasični kitajščini, postopoma pa so razvili sistem za zapisovanje japonskega jezika. V 9. stoletju so iz poenostavljenih kitajskih pismenk razvili dve zlogovnici, hiragano in katakano, ki zapisujeta glasovne in ne pomenske enote: vsak znak zapisuje en zlog. Kljub temu so za uradne spise še dalje uporabljali kitajske pismenke, s katerimi so zapisovali tako kitajske tujke (teh je še danes za skoraj polovico japonskega besedišča) kot tudi japonske polnopomenske besede. Kitajske tujke so skupaj s pismenkami, ki jih zapisujejo, prevzeli v različnih obdobjih z različnih območij Kitajske, kjer so se različno izgovarjale, zato se še danes nekatere pismenke izgovarjajo na več načinov. Poleg tega se iste pismenke uporabljajo tudi za zapis japonskih besed z enakim pomenom, tako da lahko ista pismenka služi za zapis več različnih kitajskih tujk (ki so sicer etimološko povezane, a se izgovarjajo različno, ker izhajajo iz različnih kitajskih narečij) in več japonskih sopomenk s podobnim pomenom (tj. prevodi teh kitajskih besed v japonščino), ki se izgovarjajo popolnoma drugače od kitajskih tujk, ker sta kitajščina in japonščina nesorodna jezika.

Dandanes pisanje in branje pismenk za splošno besedišče predpisuje seznam, ki ga je določilo japonsko Ministrstvo za izobrazbo (zadnja reforma je bila leta 2010). Za standardni zapis japonščine predpisuje 2136 pismenk in 4388 njihovih

izgovarjav, torej povprečno približno dve možni izgovarjavi za vsako pismenko, nekatere pa imajo več kot dve izgovarjavi. Tako se npr. pismenka 行 uporablja tako za zapis domačih japonskih glagolov *ikimasu* (ki pomeni 'iti') in *okonaimasu* (ki pomeni 'izvesti') kot tudi za zapis kitajskih tujk *kō* (npr. v besedi *ginkō*, ki pomeni 'banka'), *gyō* (v pomenu 'vrstica') in *an* (npr. v besedi *andon*, ki pomeni 'laterna'). Sodobna standardna japonščina se zapisuje s kombinacijo kitajskih pismenk, zlogovnic hiragane in katakane, za kratice pa se uporablja tudi latinica. S kitajskimi pismenkami se zapisujejo polnopomenske besede (samostalniki, pridevniki, glagoli), z zlogovnico hiragano funkcijske besede in pregibne končnice, z zlogovnico katakano pa tujke.

Pri lastnih imenih uradni predpisi določajo nabor znakov, ki se smejo uporabiti za zapis osebnega imena in priimka (hiragana, katakana, standardnih 2136 pismenk in 863 dodatnih pismenk, ki se smejo uporabiti le za zapis lastnih imen), ne določajo pa dovoljenih izgovarjav, tako da lahko starši, ko otroku izberejo ime, določijo npr. arhaično ali celo popolnoma originalno izgovarjavo pismenke. Poleg tega v rojstnem listu, ki je prvi in osnovni dokument, kjer je zapisano uradno ime, ni zapisa izgovarjave ne v latinici ne v eni od japonskih zlogovnic (razen če se ime uradno zapiše samo v eni od zlogovnic, kar je razmeroma redko), zato se lahko izgovarjava spremeni tudi kasneje. Tako je npr. japonski igralec, čigar ime se zapiše 井上陽水, kar se je sprva izgovarjalo Inoue Akemi, spremenil izgovarjavo v Yōsui, ohranil pa isti zapis, saj se znaka 陽水 lahko bereta tako Akemi kot Yōsui. Primarna oblika lastnega imena v uradnem poslovanju je namreč pisna oblika. Tudi v slovenščini izgovarjava uradnega imena ni popolnoma določena v zapisu, saj ne zapisujemo naglasa: tako se lahko npr. ime Aleš ali priimek Kovačič ali Hmeljak v izgovoru naglasi na prvem ali drugem zlogu, kar lahko izbere in tudi spremeni nosilec imena, saj v uradnem zapisu te informacije ni. Razlika med japonščino in slovenščino je v tem, da se japonski zapisi lahko berejo na bistveno več zelo različnih načinov.

V japonščini sta zato v knjižničnih katalogih običajno obe informaciji: standardni zapis (običajno v pismenkah) in zapis izgovarjave (v fonetskem zapisu), saj se isti zapis lahko izgovori na več načinov, obratno pa se lahko vsaka izgovarjava lastnega imena zapiše na več možnih načinov. Tako se npr. japonski priimek prve avtorice, ki se v latinici zapiše Sangawa, v konkretnem primeru avtoričine družine zapiše s pismenkama 寒川, a priimek z glasovno podobo Sangawa se lahko na Japonskem zapiše tudi z drugimi pismenkami, pri tem pa ima vsak zapis drug pomen: 寒川 pomeni »mrzla reka«, 三川 pomeni »tri reke«, 參川 je starejši zapis za »tri reke«, 山川 pa pomeni »gore in reke«. Znak 河 je enakovredna različica znaka 川 za besedo *kawa* »reka«. Tudi pretvorba teh pismenk v latinico ni enoznačna, saj se vsak od teh zapisov lahko prebere na več načinov, kot je prikazano v Preglednici 3.

Preglednica 3: Razmerja med zapisi in izgovarjavami nekaj japonskih priimkov

Izgovarjava	Možni zapisi	Možne izgovarjave
Sangawa	寒川 ali 寒河	kot priimek Sangawa, Samukawa, Samugawa, Sôgawa, Sôkawa, Sogawa, Kangawa, Kankawa, Sagawa, kot krajevno ime tudi Sabukawa, kot moško osebno ime tudi Kansan
	三川 ali 三河	Sangawa, Sankawa, Mikawa ali Mitsukawa
	参川 ali 参河	Sangawa, Sankawa ali Mikawa
	山川 ali 山河	običajno Yamakawa ali Yamagawa, redkeje Sangawa

Kot skrajni primer lahko navedemo priimek Saitô, ki je eden od najbolj pogostih na Japonskem in se lahko zapiše na 52 različnih načinov: 斎藤、齊藤、齋藤、齊藤、西島、西藤、西塔、西東、才藤、濟藤、西頭、佐藤、齋藤、西堂、齊藤、歳桃、斎藤、妻藤、西当、齋当、細藤、濟藤、齊藤、犀藤、斉当、佐困東、再東、齊藤、斎東、崔藤、齊藤、西當、濟藤、齋藤、再藤、濟藤、細東、削藤、柴燈、西塘、斉東、斉當、財藤、佐井藤、佐居東、才頭、采東、裁頭、西藤、齋東、齋藤、齊當

Podobno je tudi pri osebnih imenih, tako da je iskanje vseh možnih zapisov imena in priimka, za katere vemo samo izgovor, kot je zabeležen v latinici v COBISS-u, izredno zamudno.

4 Mednarodni standardi za latinični zapis kitajščine, korejščine in japonsčine

Za vse tri pisave so nastali tudi črkovni sistemi. Da proces pretvorbe ni tako enostaven, pričajo številni sistemi, ki so poskušali te jezike zapisati bodisi v latinico ali cirilico.

4.1 Latinični zapis kitajščine

Danes mednarodno sprejeti črkovni zapis kitajske pisave se imenuje *Hanyu pinyin*. S poslovenjenim izrazom *pinjin* poimenujemo sistem, ki ga je izoblikoval Odbor za prenovo pisnega jezika Ministrstva za izobraževanje Ljudske republike Kitajske. Njegovi začetki segajo v leto 1954, vlada Ljudske republike Kitajske ga je uradno sprejela 11. februarja 1958, Mednarodna organizacija za standardizacijo ga je kot ISO 7098:1982 Documentation – Romanization of Chinese⁶ potrdila leta

⁶ Več na: <https://www.iso.org/standard/13682.html>.

1982, dopolnila v ISO 7098:1991 Information and documentation – Romanization of Chinese⁷ ter ponovno revidirala leta 2015 kot ISO 7098:2015 Information and documentation – Romanization of Chinese.⁸ Od 1. januarja 2009 velja tudi kot uradni latinični zapis kitajske pisave tudi na Tajvanu.

Tajvanska različica se imenuje *Tongyong pinyin*. Leta 1998 jo je zasnoval Yu Boquan, pri čemer je poskušal ohraniti značilnosti sistema *Hanyu pinyin*, in istočasno izboljšati zapis zlogov, ki so tujcem povzročali največ težav pri izgovarjavi. Sem sodijo na primer zlogi z začetnicama q in x. Od *Hanyu pinyin*a se torej razlikuje le v nekaj podrobnostih, kot uradni latinični zapis kitajske pisave na Tajvanu pa je veljal v letih med 2002 in 2008, preden ga je nadomestil *Hanyu pinyin*.

Zelo znan črkovni zapis kitajske pisave je tudi *Wade-Giles*, ki ga je sredi 19. stoletja osnoval britanski sinolog Thomas Francis Wade, nekaj desetletij kasneje pa nadgradil britanski sinolog Herbert Allen Giles. *Wade-Giles* je v 20. stoletju postal najbolj razširjena latinična transkripcija kitajščine, predvsem v angleško govorečem delu sveta. Večina pomembnih publikacij s kitajsko tematiko, ki so nastale v letih med 1912 in 1979, uporablja ta sistem.

Na Univerzi Yale so med drugo svetovno vojno za potrebe ameriške vojske razvili sistem Yale, ki je zloge mandarinske kitajščine, kantonščine, korejščine in japonščine zapisoval podobno kot angleščino, da se vojakom ni bilo treba učiti dodatnih pravil, temveč so zapisane zloge prebrali kar »po domače«. Z vidika posameznih jezikov tako obstajajo štiri istoimenski sistemi Yale, ki so jih izdelali različni avtorji. Pod kitajsko različico se je leta 1943 podpisal sinolog George A. Kennedy.⁹ Ker Yale ne temelji na zakonitostih kitajske fonetike, temveč se zanaša na to, da je mogoče marsikatero nejasnost razčistiti s pomočjo konteksta, se vprašanju sozvočnic niso preveč posvečali. Za kitajščino so Yale uporabljali v zahodnih publikacijah v letih med 1943 in 1979, za transkripcijo kantonščine pa se uporablja še danes.

V letih 1925 in 1926 je vodilni kitajski jezikoslovec Y. R. Chao (*Hanyu pinyin*: Zhao Yuanren) s sodelavci razvil sistem *Gwoyue Romatzyh*. Uporabljali so ga v učbenikih, slovarjih in publikacijah, ki so bili v letih od 1942 do 2000 natisnjeni v Hongkongu. Kitajska je ta sistem kot uradni latinizirani zapis svoje pisave sprejela leta 1928, a se v širših krogih ni nikoli trdno prijel. *Gwoyue Romatzyh*

⁷ Več na: <https://www.iso.org/standard/13683.html>.

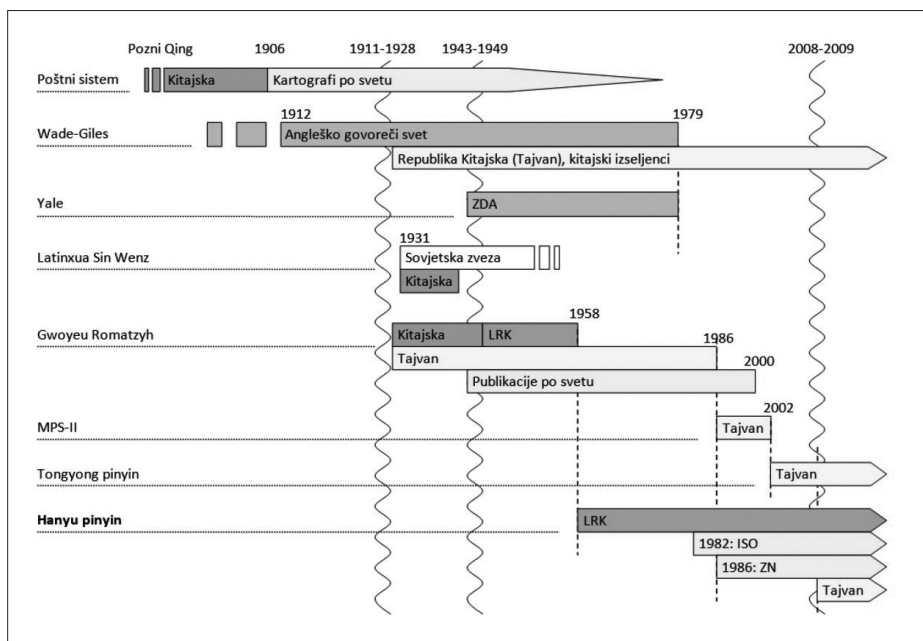
⁸ Več na: <https://www.iso.org/standard/61420.html>.

⁹ Yale za korejščino je razložen v nadaljevanju.

je v Ljudski republiki Kitajski prenehal veljati leta 1958, ko ga je nasledil *Hanyu pinyin*. Na Tajvanu je bil v veljavi še do leta 1986, ko ga je zamenjal sistem *MPS II*.

Sistem *MPS-II* (Mandarin Phonetic Symbols 2) je bil na Tajvanu uradno v veljavi od leta 1986 do 2002. Uporabljali so ga predvsem za zapisovanje zemljepisnih imen v tajvanskih publikacijah.

Nazadnje omenimo še nekoliko manj poznan sistem *Latinxua Sin Wenz*. Osnoval ga je Sovjetski znanstvenoraziskovalni inštitut za kitajske študije med leti 1929 in 1931. Kitajske železnice so ga leta 1949 sprejele za telekomunikacijske storitve, zato se lahko pojavi v publikacijah s tega področja. Slika 1 prikazuje soobstoj in rabo različnih sistemov skozi čas.



Slika 1: Grafični prikaz rabe črkovnih zapisov kitajske pisave (Petrovčič in Lin, 2018, str. 14)

Natančna primerjava vseh zlogov po omenjenih sistemih je opisana na spletnih straneh portala Pinyin.info.¹⁰ Na tem mestu si oglejmo le, kako se v sistemu Hanyu pinyin in Wade-Giles napiše ime avtorja 张憲然.¹¹ Mednarodno sprejet

¹⁰ Dostopno na: <http://www.pinyin.info/romanization/compare/hanyu.html>.

¹¹ To je pravo ime avtorja, ne izmišljeni primer. Vir: WorldCat.

zapis v pinjinu je »Zhang Queran«, isto ime bi se po sistemu Wade-Giles zapisalo kot »Chang Ch'üehjan«. Če odmislimo izgubo informacij, ki je nastala zaradi črkovnega zapisa, je tu še dodatna dilema, kateri sistem imamo pred seboj. Zapis »Chang« namreč obstaja tudi v sistemu Hanyu pinyin, pri čemer se seveda nanaša na drug zlog in s tem priimek (Chang – Zhang).

Z latinizacijo kitajske pisave izgubimo veliko informacij, a težave potekajo tudi v obratni smeri. V nekaterih primerih ne moremo vedeti, kateri latinizirani zapis je v določeni situaciji pravilen. Takih primerov v kitajščini ni veliko, vendar obstajajo. Pismenko 朝 lahko izgovarjamo kot chao ali zhao. Oseba, ki tak zapis latinizira, ne more biti povsem prepričana, kateri zlog naj izbere. Zaradi kompleksnega razmerja zapis-izgovor je zato možnost neposrednega iskanja v azijskih pisavah neprecenljivega pomena.

4.2 Latinični zapis korejščine

Za zapis korejščine v latinici obstaja več sistemov. Med temi je danes najbolj razširjen sistem, ki ga je leta 2000 predlagal Nacionalni inštitut za korejski jezik (National Institute of Korean Language oziroma 국립국어원, v nadaljevanju NIKL, ki deluje pod okriljem južnokorejskega Ministrstva za kulturo), z angleškim imenom *Revised Romanization of Korean* (국어의 로마자 표기법, v nadaljevanju RR). Sistem so dopolnili leta 2014 in je trenutno uradni sistem Koreje za zapis korejščine v latinici. Vključuje tako pravila za transliteracijo, tj. enoznačno pretvorbo znakov pisave *hangul* v latinico, kot tudi za transkripcijo, ki zapisuje tudi glasovne spremene, ki se v pisavi *hangul* ne zapišejo eksplicitno. Pri tem uporablja samo črke nabora ASCII brez diakritičnih znakov, tako da je široko uporaben v različnih računalniških sistemih.

Korejska nacionalna knjižnica (National Library of Korea ali NLK oziroma 국립중앙도서관 oziroma 國立中央圖書館) je osrednja knjižnica, ki v Koreji hrani obvezne izvode vsega, kar se tiska v Koreji, ter od leta 2015 prispeva podatke v VIAF (Virtual International Authority File) ter od leta 2016 določa ISNI za korejske avtorje,¹² pri latiničnem zapisu korejskih imen pa uporablja RR. ISNI v Koreji določa tudi Knjižnica korejskega državnega zbora (National Assembly Library of the Republic of Korea 대한민국 국회도서관), ki ravno tako uporablja standard RR.

Pred uveljavitvijo sistema RR je bil najbolj razširjen sistem *McCune-Reischauer* (MR), ki sta ga razvila ameriška jezikoslovca George M. McCune in Edwin O.

¹² Dostopno na: <http://www.nl.go.kr/authorities/main/index>.

Reischauer (1939). To je sistem transkripcije, ki je namenjen temu, da čim bolj zvesto zapiše dejansko izgovarjavo in tako bralcu olajša branje, ne pa enoznačno pretvorbo (transliteracijo) pisave *hangul*. Dopolnjeno različico tega sistema, ki bolj podrobno določa delitev besed, uporablja ameriška Kongresna knjižnica (Library of Congress) za latinični zapis korejščine.¹³

V Severni Koreji je Akademija za družbene vede (Academy of Social Sciences oziroma 사회과학원 ali Sahoe Kwahagwŏn), ki deluje pod Akademijo znanosti Demokratične ljudske republike Koreje (Academy of Sciences of the Democratic People's Republic of Korea oziroma 조선인민민주주의공화국과학원), razvila svoj sistem za latinični zapis korejščine. Od leta 1980 dalje je bilo več pogovorov in poskusov, da bi poenotili latinični zapis korejščine v Severni in Južni Koreji, in izoblikovali so standard ISO/TR 11941:1996, ki pa je bil ukinjen leta 2013, saj ni prišlo do dogovora o poenotenju med Severno in Južno Korejo (Jeon, 1996; Lee, 1993).

Obstajajo še drugi sistemi. V jezikoslovnih razpravah se pogosto uporablja sistem *Yale*, ki ga je razvil Samuel E. Martin s kolegi na Univerzi Yale med 2. svetovno vojno in smo ga omenili že v povezavi s kitajščino. V tem sistemu se podobno kot pri pisavi *hangul* vedno uporablja isti znak za isti fonem, ki pa se lahko v različnih kontekstih izgovori drugače. Uporablja piko za mejo med zlogi, kjer bi bila sicer v latiničnem zapisu brez presledkov meja zabrisana. Za pravilno izgovarjavo tako latiniziranega besedila mora torej bralec poznati glasoslovna pravila korejščine, obenem pa je tako latinizirana besedila lažje pretvoriti nazaj v *hangul*. Poleg teh obstajajo še drugi, manj razširjeni sistemi.

Poseben problem pri latinizaciji je raba presledkov. V pisavi *hangul* je namreč vsak sklop znakov, ki zapisuje zlog, vidno združen v enoto, ki je rahlo odmaknjena od ostalih zlogov (podobno kot je med vsako črko latinice v isti besedi zelo ozek prazen prostor), med besedami (oziroma sklopi, ki jih sestavlja polnopomenska beseda in slovnični členek ali pregibna končnica) pa se uporablja presledek.

Za ponazoritev si oglejmo naslov knjige 아내를 닮은 도시 – 류블라나.¹⁴ V Preglednici 4 je v prvi vrsti zapis v pisavi *hangul*, v drugi zapis po sistemu RR, ki smo mu dodali pike, ki označujejo meje med zlogi. V tretji vrstici pa je običajni latinični zapis, v katerem so samo presledki med besedami.

¹³ ALA-LC romanization tables: transliteration schemes for non-roman scripts. Korean, (rev. 2009). Dostopno na: <https://www.loc.gov/catdir/cpsa/romanization/korean.pdf>.

¹⁴ Kang, Byoung Yoong: Anaereul dalmeun dosi, Ljubljana = Ljubljana: mesto, ki spominja na mojo ženo. Gyeonggi: Nanda, 2015. COBISS.SI-ID 67213410.

Preglednica 4: Primer zapisa korejskega naslova v latinici

Zapis v pisavi <i>hangul</i>	아내를 닮은 도시 – 류블라나
Latinični zapis po sistemu RR z dodanimi oznakami za meje med zlogi	a-nae-reul-dalm-eun-do-si – ryu-beul-ly-na
Običajni latinični zapis po sistemu RR brez oznak za meje med zlogi	anaereul dalmeun dosi – ryubeullyana
Prevod naslova v slovenščino	Ljubljana: mesto, ki spominja na mojo ženo

Ko bi radi zdaj npr. prvo besedo *anaereul* pretvoriti nazaj v *hangul*, bi morali najprej določiti meje med zlogi, možnosti pa je veliko in vsaka se zapiše drugače v pisavi *hangul*, kot je razvidno iz Preglednice 5.

Preglednica 5: Zapisi v pisavi *hangul*, ki se prečkrujejo v isti latinični zapis brez presledkov

anaereul	아내를	a-nae-reul
	아나에레울	a-na-e-re-ul
	아내레울	a-nae-re-ul
	안아에레울	an-a-e-re-ul
	안아에를	an-a-e-reul
	안애레울	an-ae-re-ul
	안애를	an-ae-reul

Do podobnega problema pride tudi pri ostalih sistemih za latinični zapis korejščine, saj nobeden od njih ne omogoča nedvoumne preslikave iz pisave *hangul* v latinico in obratno iz latinice v *hangul*. Tudi pri imenih se pojavlja ta problem. Tako npr. latinični zapis imena Bomi lahko odgovarja zlogoma Bom-i (v pisavi *hangul* 봄이) ali zlogoma Bo-mi (v pisavi *hangul* 보미), kar je drugo ime. Takih imen in priimkov je še veliko.

Še posebej problematična je pretvorba osebnih imen. Reforma leta 2000 je priporočila, naj se tudi lastna imena v latinici zapisujejo po sistemu RR, toda za lastna imena standard ni obvezen. Večina tistih, ki so rojeni pred tem datumom, je pred letom 2000 že uporabljala drug zapis, ki ga še ohranjajo v potnem listu in v stikih s tujino; nekateri uporabljajo enega od zgoraj opisanih standardov, nekateri pa si izberejo zapis, ki ne ustreza nobenemu standardu. Tako se npr. za tri najbolj pogoste korejske priimke uporabljajo različice, ki so prikazane v Preglednici 6.

Preglednica 6: Različni latinični zapisi treh najbolj pogostih korejskih priimkov

Kitajske pismenke	Hangul	Sistem RR	Drugi zapisi
金	김	Gim	Kim
朴	박	Bak	Pak, Park
李	이	I	Lee, Yi, Ye, Rhee
	리	Ri	

Pri prvih dveh priimkih ponovna pretvorba iz latinice v hangul ni problematična, saj v pisavi hangul ni drugih zapisov, ki bi odgovarjali kateri od teh latiničnih oblik.

Pri tretjem priimku pa se pojavi problem, ker je uradni zapis po sistemu RR zelo kratek (samo črka I, zato se ga nosilci tega imena izogibajo) in ker se kitajska pismenka, s katero so izvorno zapisovali ta priimek, lahko izgovori tudi Ri (v pisavi hangul se ta zapiše ㄹ이). Večina Korejcev s tem priimkom ga v latinici zapiše Lee, nekateri pa tudi Yi, Ye ali Rhee. Če torej imamo samo podatek, da se v latinici priimek zapiše Lee, ne moremo vedeti, ali gre za priimek, ki se v pisavi *hangul* zapiše ㄹ이, ali za tistega, ki se zapiše ㄹ이.

4.3 Latinični zapis japonščine

Za zapis japonščine v latinici obstaja več sistemov, pri čemer se danes v svetovnem merilu najbolj pogosto uporablja sistem *hepburn*, ki se imenuje po ameriškem jezikoslovcu Jamesu C. Hepburnu, ki je sistem uporabil v japonsko-angleškem slovarju leta 1886. Sistem zapisuje fonotaktično pogojene različice istih fonemov z različnimi črkami, npr. za fonem /s/ uporablja črko s pred glasovi /a/, /u/, /e/ in /o/, toda pred glasom /i/, ko se isti fonem izgovori kot nebni pripornik [š], uporablja dvočrkje *sh*. Za zapis soglasnikov sistem uporablja črke podobno kot angleški pravopis (znak *j* za glas [dž], znaka *ch* za glas [č] ipd.), samoglasnike pa zapisuje podobno kot romanski jeziki in slovenščina (Hmeljak Sangawa idr., 2003).

Manj razširjen, a na Japonskem uradni sistem *kunrei-shiki* je ime dobil po tem, da ga je za uradnega razglasil vladni odlok leta 1937 (*kunrei* pomeni *vladni odlok*, *shiki* pa *sistem*), sloni pa na sistemu *nihon-siki*, ki ga je predlagal Aikitsu Tanakadate leta 1884. Ta sistem uporablja isti znak za različne fonotaktične variante fonemov in pri tem dosledno sledi razporeditvi znakov v tabeli japonske zlogovnice hiragana. Tako npr. fonem /s/ vedno zapisuje z znakom *s*, tudi pred glasom /i/, ko se zlog izgovori [ši]. Ta sistem je leta 1954 japonska vlada razglasila za uradni sistem latiničnega zapisa japonščine, toda v istem odloku tudi dopustila rabo sistema *hepburn*, ki je bil takrat mednarodno razširjen, zato sta se v rabi ohranila oba.

Sistem *kunrei* je bil osnova za standard ISO 3602:1989 Documentation – Romanization of Japanese (kana script) in je tudi prvi sistem, ki se ga naučijo otroci v 4. razredu japonskih osnovnih šol pri pouku japonščine po kurikulumu, ki ga določa ministrstvo, pristojno za izobraževanje. V mednarodnem poslovanju pa je bolj uveljavljen sistem *hepburn*, zato japonsko ministrstvo za zunanje zadeve

priporoča rabo sistema *hepburn* za zapis osebnih imen v potnem listu. Pri tem lahko sicer posameznik, ki si prvič pridobi potni list, izbere en zapis od dveh (ali tudi nekaj nestandardnih različic), toda po prvi izbiri menjava latiničnega zapisa za uradne dokumente ni več dovoljena.

Knjižnica japonskega državnega zbora (tj. japonska osrednja nacionalna knjižnica) uporablja sistem *kunrei* oziroma ISO 3602 za latinični zapis bibliografskih podatkov, ki jih kot referenčne podatke ponuja v formatu JAPAN/MARC MARC21,¹⁵ če v gradivu samem latinica ni prisotna. Če pa je v gradivu že tudi latinični zapis, kot referenčni podatek navede le-tega. Taki napisi v gradivu so večinoma po sistemu *hepburn*. Za vsako knjižnično enoto navaja tako osnovni zapis (v standardnem japonskem sistemu pisave, ki vključuje kitajske pismenke ter znake zlogovnic hiragane in katakane) kot tudi fonetični zapis v katakani in isto še v latinici.

Library of Congress za latinični zapis japonščine uporablja sistem, ki je podrobno opisan v tabeli za latinično prečrkovanje¹⁶ ter sloni na tabeli slovarske založbe Kenkyusha in ameriškega standarda American National Standard System, ki sta oba osnovana na sistemu *hepburn*.

Razlike med sistemom *hepburn* in *kunrei* so majhne, veliko besed in lastnih imen (npr. Haruki Murakami) se zapiše identično v obeh sistemih. Morda tudi zaradi tega nekateri ne ločujejo dobro med obema in se tudi v knjižničnih katalogih včasih pojavlja nedosledna raba in različni zapisi za isto besedo v istem kataložnem zapisu, npr. shakai in syakai (prvi v sistemu *hepburn*, drugi v sistemu *kunrei*, oba odgovarjata isti izgovorjavi [šakai]).

V Sloveniji obstaja še en sistem, ki ga predpisuje slovenski pravopis (Slovenski pravopis, 2003) za zapisovanje lastnih imen iz jezikov, ki se zapisujejo z nelatiničnimi pisavami, to je poslovenjeni zapis, ki naj bi slovenskemu bralcu omogočal pravilno izgovarjavo. Dejansko pride pri pretvorbi iz enega zapisa v drugega pogosto do napak, saj se nekatere črke uporabljajo drugače v vsakem od teh zapisov: tako se npr. moško osebno ime, ki se v sistemih *hepburn* in *kunrei* zapiše Yō ali Yō, v poslovenjeni obliki zapiše Jo (brez naglasa). Zapis Jo pa v sistemih *hepburn* in *kunrei* odgovarja drugemu imenu, ki se izgovori [džo]. Predlog novega slovenskega pravopisa, ki je trenutno še v obravnavi, ne predpisuje več

¹⁵ JAPAN/MARC MARC21 Format Manual. (2019). Dostopno na: https://www.ndl.go.jp/jp/data/JAPANMARC_MARC21manual_MS.pdf.

¹⁶ ALA-LC romanization tables: transliteration schemes for non-roman scripts. Japanese. (rev. 2012). Dostopno na: <https://www.loc.gov/catdir/cpsd/romanization/japanese.pdf>.

slovenjenja japonskih lastnih imen, zato upamo, da bo v bodoče manj nedoslednosti pri latinizaciji.

5 Zasnova raziskave in njene omejitve

Kot smo videli v razdelkih 2, 3 in 4, se prvi del informacij izgubi pri pretvorbi iz nečrkovne v črkovno pisavo, drugi del informacij pa se lahko izgubi zaradi zmede med različnimi sistemi latinizacije. Zaradi omejitve sistema COBISS na rabo okrnjene podmnožice latinice imamo v knjižnici Oddelka za azijske študije FF UL vzporedni katalog vzhodnoazijskega gradiva v originalnem zapisu, ki pa žal zaradi kadrovske stiske ni niti ažuren niti v celoti dostopen uporabnikom. Ker smo zaradi prostorskih omejitev več kot polovico knjižnične zbirke prestavili v dislocirana skladišča, je uporabnikom knjižnice, večinoma študentom vzhodnoazijskih jezikov s še nepopolnim znanjem teh jezikov, izredno težko priti ne samo fizično do gradiva, ampak sploh do informacije o gradivu v knjižnični zbirki. Ko na primer v gradivu, ki ga berejo, srečajo ime in bi ga radi poiskali v katalogu, morajo najprej ugotoviti, kako se prebere in zapiše v latinici. Zlasti pri japonskih imenih je edini zanesljiv način, kako ugotoviti izgovarjavo lastnega imena avtorja, poizvedba v normativni datoteki enega od japonskih vzajemnih kataložnih sistemov, tako da morajo dejansko vsako iskanje izvesti v dveh katalogih. Ko pa v katalogu COBISS najdejo ime, ki bi ga radi citirali, morajo ravno tako poiskati originalni zapis v tujem kataložnem sistemu. Tudi za vnašalce je zelo nerodno, da ne morejo zapisa v izvorni pisavi shraniti z ostalimi podatki takrat, ko podatke vnašajo in imajo pri roki originalno gradivo. Neprimerno bolj praktično bi torej bilo, ko bi tako pri vnosu v COBISS (ko bi vnesli oboje, izvorni zapis in latinizirano obliko) kot pri poizvedovanju po njem lahko vnašali direktno izvirne znake.

Da bi preverili praktičnost in izvedljivost rabe nelatiničnih znakov v knjižničnem katalogu, smo raziskali, kako to vprašanje rešujejo v tujih bibliografskih sistemih, ter iste primere preverili še v COBISS-u. V naslednjem razdelku predstavljamo rezultate te raziskave.

6 Rezultati

Na primeru zloga »li« v kitajščini smo predstavili razmerje med pismenkami in latiničnim zapisom. Omenili smo tudi, da se lahko vse pismenke z isto izgovarjavo pojavijo kot del lastnega imena, v praksi pa so pogosteje zastopane tiste

pismenke, ki nosijo pozitivne konotacije. V katalogu WorldCat smo preverili, koliko avtorjev s priimkom »Wang« in imenom »Li« je zabeleženih v tej bazi.

Za lažjo predstavo smo avtorje, ki so prisotni v WorldCatu, razporedili v 16 skupin. V isto skupino sodijo imena, ki se zapisujejo z istimi pismenkami, a različnimi letnicami. To lahko pomeni, da gre dejansko za različne osebe s popolnoma identičnim priimkom in imenom, ne izključujemo pa tudi možnosti napak pri vnosu letnice rojstva avtorja.

1. 王力, 1900-1986; 王力, 1921-; 王力, 1951-; 王力 1968-; 王力, 1971-;
2. 王立, 1932-; 王立, 1953-; 王立, 1964-; 王立, 1966-;
3. 王利, 1953-; 王利, 1957-; 王利, 1978-; 王利, 1981-
4. 王莉, 1971-; 王莉, 1973-; 王莉, 1975-; 王莉, 1976-; 王莉, 1977-
5. 王丽, 1979-; 王丽
6. 王骊, 1928-; 王骊
7. 王犁, 1970-; 王犁
8. 王禮, 1314-1389
9. 王礼
10. 王里, 1936-
11. 王理
12. 王黎
13. 王醴
14. 王荔 (16. stol.)
15. 汪离
16. 汪莉, 1966-.

Rezultati kažejo, da latinični zapis »Wang Li« ustreza 16 različnim zapisom v kitajščini. Po drugi strani je lahko tudi zlog Li priimek, ki se zapiše s pismenkami 力利离立里李礼黎栗励厉郪 ali 澧. Oseba »Wang Li« torej ni isto kot »Li Wang«. Tudi v slovenščini obstajajo imena, ki so lahko tudi priimki, kot npr. Anton Janko oziroma Janko Anton, le da je takih primerov v našem okolju bistveno manj kot v kitajščini, kjer so lastna imena dvo- ali trizložna.

Poglejmo si še en primer trizložnega lastnega imena. Latinični zapis 'Wang Yaping' se lahko nanaša na vsaj naslednje tri osebe, kot izvemo iz kataloga WorldCat:

1. 王亚平 (王 Wang, priimek; 亚平 Yaping, ime)
2. 王雅平 (王 Wang, priimek; 雅平 Yaping, ime)
3. 王雅萍 (王 Wang, priimek; 雅萍 Yaping, ime)

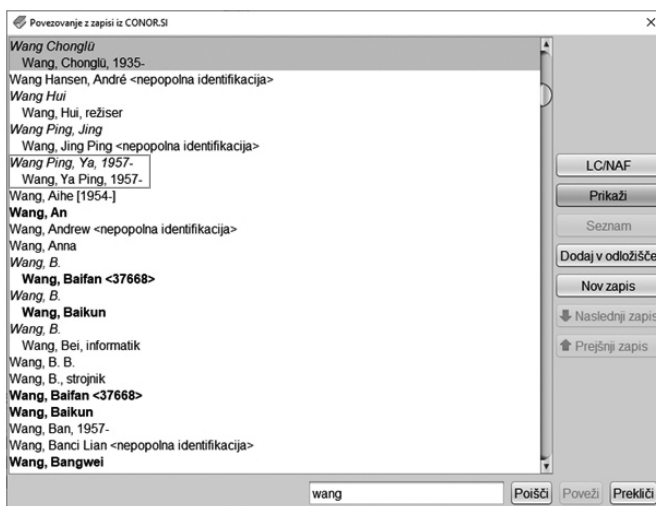
Vse tri osebe se pišejo Wang, prva in druga se razlikujeta v pismenki za zlog 'ya' (亚 ~ 雅), druga in tretja oseba pa v pismenki za zlog 'ping' (平 ~ 萍). Iz tega je

očitno, da iskanje v pismenkah precej izboljša natančnost zadetkov, pa še v tem primeru lahko naletimo na osebi s popolnoma identičnim imenom. Wang Yaping (王亚平) se lahko nanaša na profesorja na Univerzi v Glasgowu¹⁷ ali na pilotko in astronautko,¹⁸ ki je kot druga Kitajka poletela v vesolje.

Da bi bila zmeda še večja, pri latiničnem zapisu azijskih imen pogosto ni jasno, kaj je ime in kaj priimek. Profesor Wang Yaping z Univerze v Glasgowu na svoji osebni strani uporablja 'zahodni' sistem, kjer najprej navedemo ime in nato priimek, kot npr. Ivan Cankar. Pri vzhodnoazijskih jezikih je standardni besedni red priimek – ime. Poleg tega prisotnost presledkov dodatno zamegli informacije, kateri zlogi spadajo k imenu in kateri k priimku (Slika 2).



Slika 2: Posnetek osebne strani prof. Wang Yapinga
(vir: <https://www.gla.ac.uk/schools/socialpolitical/staff/yapingwang/>)



Slika 3: Pogovorno okno Povezovanje z zapisi iz CONORSI v COBISS-u

¹⁷ Wang, Yaping 1957-. Dostopno na: <http://www.centreforsustainablecities.ac.uk/about/staff-and-partners/yaping-wang/> in <http://news.pkusz.edu.cn/article-143-4167.html>.

¹⁸ Wang, Yaping 1980-. Dostopno na: https://en.wikipedia.org/wiki/Wang_Yaping.

Za konkretni primer smo proučili še zapis v COBISS-u. V normativni bazi osebnih in korporativnih imen CONOR.SI zasledimo dva različna zapisa za isto osebo, ki pa nista popolnoma pravilna.

Po zgledu predvidenih oblik vnosa »*Janez Novak*« in »*Novak, Janez*« bi sklepali, da se oseba »*Wang Ping, Ya*« piše *Wang Ping* in ji je ime *Ya*. Oseba »*Wang, Ya Ping*« pa se piše *Wang* in nosi ime *Ya Ping* (Slika 3). Po primerjavi naslovov del, avtorjeve bibliografije in zapisov o gostujočih predavanjih razberemo, da gre za isto osebo.

Z vidika uporabnika je zelo koristen tudi WorldCat Identities,¹⁹ ki nam za izbrana avtorja ponudi tudi alternativne različice. Tako na primer za iskalni niz *Chiang Kai-shek* (1887–1975) zasledimo vrsto alternativnih vnosov, na primer *Chan Kaishi*, *Chan Kayshi*, *Chang Kai-shek*, *Chang Kai-scek*, *Chang Kai-scek*, *Chiang Chieh-shih*, *Čang Kai-šek*, *Čang Kaj-šek*, *Jiang Jieshi*, *Shō Kaiseki*, *Tchang Kai-shek*, *Tschiang Kaischek*, *Tšiang Kai-šek* in še mnogo drugega. Tudi za to osebo bomo našli zmotne vnose, da je *Chiang* ime in *Kai-shek* priimek. V slovenskem okolju bi se temu zapisu pridružila še oblika *Čankajšek*.

Rezultati raziskave so nadalje pokazali, da je COBISS tudi izredno omejen in tog sistem, kar se tiče nabora znakov za črkovne pisave, če nečrkovnih sploh ne omenjamo. Za primer vzemimo iskanje avtorja, čigar ime je zapisano z znaki, ki presejajo domet osnovnih črk.

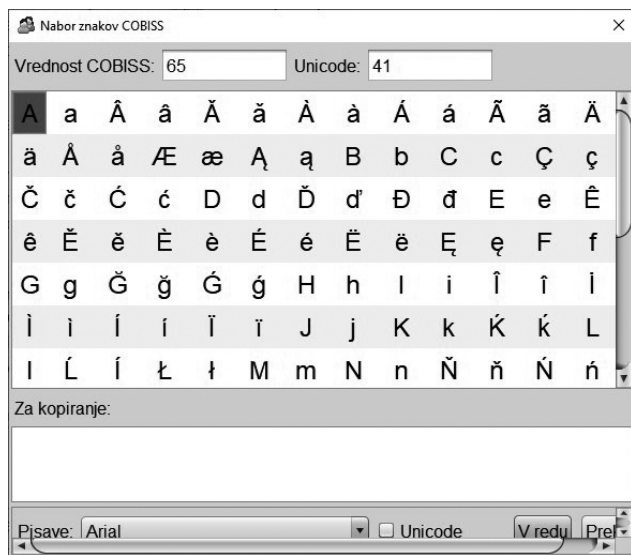
Če se v imenu avtorja nahajajo 'posebni' znaki, do katerih ne pridemo z neposrednim pritiskom na eno izmed tipk tipkovnice, ima knjižničar na voljo pogovorno okno z naborom znakov COBISS (Slika 4). Izbrani znak se pojavi v polju za kopiranje, ki ga s potrditvijo vnesemo v iskalno polje avtorja.

Ob izbiri znaka obenem izvemo, kakšna je njegova vrednost v okviru COBISS-a in katero kodno točko ima ta znak v Unicodu.²⁰ Unicode je standard za kodiranje znakov in simbolov, ki določa prikaz teksta v programskih izdelkih in standardih. Unicode za zapis znaka ali simbola uporablja edinstvena števila in črke, ne

¹⁹ Dostopno na: <https://worldcat.org/identities/>.

²⁰ Danes prevladujoči standard Unicode je prvi uspel enolično predstaviti vse znake vseh pisav sveta znotraj istega nabora in sistema kodiranja. Teoretično gledano je prostora za širitev še precej, saj trenutna verzija 12.0.0 definira »le« 137.928 od skupno možnih 1.112.064 kodnih točk. Tej nalogi bi bili kos tudi zapuščinski sistemi vzhodne Azije, če Unicode ne bi prevzel vodilne vloge. Že od samih začetkov dalje so vključevali tudi latinico, cirilico, grško abecedo in podobno. Kitajski državni standard GB 18030–2005 razpolaga z več kot 1,5 milijona kodnih točk.

glede na platformo, ne glede na programsko opremo, ne glede na jezik. Je edini sistem, ki kodira vse obstoječe znake za računalniško rabo. S Slike 4 na primer izvemo, da ima velika tiskana črka 'A' v COBISS-u vrednost 65, v Unicodu pa ima ta znak kodo <U+0041>. Povezava z Unicodom je do določene mere že prisotna, vendar so številne funkcije omejene. V polje 'Unicode' na Sliki 4 namreč ne moremo ničesar vpisati, čeprav se s kurzorjem lahko postavimo tja.



Slika 4: Nabor posebnih znakov, ki so na voljo v COBISS-u

Nabor znakov COBISS je zelo okrnjen in ne zadostuje niti za popolni latinični zapis azijskih imen. V povezavi s črko 'A' oziroma 'a' so na voljo še: Â, â, Ã, ã, Ä, ä, Á, á, ã, ã, Ä, ä, À, à, Æ, æ, À, à, manjkata pa makrona Ā in ā, s čimer označujemo na primer dolge samoglasnike v japonsščini ali prvi ton v kitajščini. Enak oziroma nekoliko bolj okrnjen nabor velja za ostale samoglasnike. Na spletnih straneh Library of Congress²¹ vidimo, da je makrone možno ohraniti (Slika 5).

Tudi pri japonsščini se znak za dolžino samoglasnikov ohranja (Slika 6). Še več, uporabnik lahko išče izbrani niz tudi v japonski pisavi (Slika 7).

²¹ Dostopno na: <https://catalog.loc.gov/>.



CATALOG

LC Catalog Quick Search

BOOK

Bhāratīya Nepālī nāṭaka sañcayana

Full Record MARC Tags

Main title
Bhāratīya Nepālī nāṭaka sañcayana / sañkalana evaṃ sampādana, Lakshmaṇa Śrīmala = Bharatiya Nepali natak sanchayan : an anthology of Nepali plays / compiled & edited by Laxman Sirmal.

Edition
Prathama saṃskaraṇa.

Published/Produced
Naī Dillī : Sāhitya Akadēmī, 2016.

Request this Item LC Find It Item Availability

Slika 5: Iskalni niz 'Bhāratīya Nepālī nāṭaka sañcayana' na straneh Library of Congress



CATALOG

LC Catalog Quick Search

BOOK

Saishin Ei-Wa rōdō yōgo jiten

Full Record MARC Tags

Main title
Saishin Ei-Wa rōdō yōgo jiten / kanshū Takahashi Sakutarō ; Rōdō Daijin Kanbō Kokusai Rōdōka hen.
最新・英和労働用語辞典 / 監修高橋欄太郎 ; 労働大臣官房国際労働課編。

Published/Created
Tōkyō : Nikkan Rōdō Tsūshinsha, Heisei 4 [1992]
東京 : 日刊労働通信社, 平成 4 [1992]

Request this Item LC Find It Item Availability

Slika 6: Izpis zadetka v latinici in japonščini na straneh Library of Congress

LIBRARY OF CONGRESS CATALOG

LC Catalog Quick Search

SEARCH OPTIONS

Browse

Titles, Authors/Creators, Subjects, Call Numbers, Standard Numbers

Browse

AUTHORS/CREATORS beginning with (enter last name first)

高橋柵太郎

Search Clear

Slika 7: Iskalni niz '高橋柵太郎' na straneh Library of Congress

Z zavihka 'MARC Tags' je razvidno, da ta zapis vsebuje tudi vse osnovne podatke v azijskih pisavah. Na tem mestu si oglejmo le kazalke o avtorju, naslovu dela, kraju in založbi (Preglednica 7).

Preglednica 7: Izsek zavihka MARC Tags pri Sliki 6

naslov	245	00	6 880-01 a Saishin Ei-Wa rōdō yōgo jiten / c kanshū Takahashi Sakutarō ; Rōdō Daijin Kanbō Kokusai Rōdōka hen.
kraj/založba	260	__	6 880-02 a Tōkyō : b Nikkan Rōdō Tsūshinsha, c Heisei 4 [1992]
avtor	700	1_	6 880-03 a Takahashi, Sakutarō, d 1935-
	710	1_	6 880-04 a Japan. b Rōdōshō. b Kokusai Rōdōka.
naslov	880	00	6 245-01/\$1 a 最新・英和労働用語辞典 / c 監修高橋柵太郎 ; 労働大臣官房国際労働課編.
kraj/založba	880	__	6 260-02/\$1 a 東京 : b 日刊労働通信社, c 平成 4 [1992]
avtor	880	1_	6 700-03/\$1 a 高橋柵太郎, d 1935-
	880	1_	6 710-04/\$1 a Japan. b 労働省. b 国際労働課.

Pogovorno okno *Povezovanje z zapisi iz CONOR.SI* v COBISS-u (Slika 3) obenem prikazuje, da se tudi COBISS lahko poveže z bazo imen Library of Congress (LC/NAF: Library of Congress, Name Authority File). Za primer vzemimo latinični vnos nepalske knjige (Slika 5). Zadetki, ki so na voljo, v tem pogovornem oknu prikazujejo tudi znake, ki jih COBISS ne podpira. Ob potrditvi vnosa se 'nedovoljeni' znaki obarvajo rumeno (Slika 8).

ID=-1 001 V1
001 an - nov zapis ba - tekstovno gradivo, tiskano cm - monografska publikacija d0 - ni hierarhičnega odnosa eocn988900683 7ba - latinica
010 a9788126050369
010 a8126050365
100 bd - publikacija, zaključena ob izidu ali v enem koledarskem letu c2016 hslv - slovenski
1010 anep - nepalski
102 aind - Indija
105 fh - govor
2001 aBhāratīya Nepālī nāṅaka sañcayana fsaṅkalana evaṅ sampādana, Lakshmaṅa Śrīmala = Bharatiya Nepali natak sanchayan : an anthology of Nepali plays / compiled & edited by Laxman Srimal
205 aPrathama saṅskaraṅa.
210 aNaj Dilli cSāhitya Akademi d2016
215 a458 str. d22 cm
300 aPlays
300 aIn Nepali.
330 aAnthology of Indian Nepali plays
5101 aBharatiya Nepali natak sanchayan ean anthology of Nepali plays
70211 aŚrīmala bLakshmaṅa 4340 - urednik
71202 aSahtya Akademi, cpublisher, ccopyright holder. 4publisher, 4260 - lastnik avtorskih pravic

Slika 8: Zapis z nedovoljenimi znaki.

Za nekatere znake COBISS nima niti ustreznih glifov – ti se prikazujejo kot kvadratki –, vendar prava kodna točka znaka na tem mestu obstaja, saj kopiranje besedila v Word prikaže manjkajoče glife:

2001 aBhāratīya Nepālī nāṅaka sañcayana fsaṅkalana evaṅ sampādana, Lakshmaṅa Śrīmala = Bharatiya Nepali natak sanchayan : an anthology of Nepali plays / compiled & edited by Laxman Srimal
205 aPrathama saṅskaraṅa.
70211 aŚrīmala bLakshmaṅa 4340 – urednik

Iz prikazanega je razvidno, da vključitev nelatiničnih pisav v bibliotekarske sisteme ni nemogoče. Glede na to, da povezave na LC/NAF zadetke pravilno prikažejo, ne bi smelo biti večjih težav z nadgradnjo oziroma posodobitvijo COBISS-a.

7 Razprava

Raziskava analizira problem latiničnega zapisa vzhodnoazijskih pisav na teoretični ravni, preveri stanje na zgolj nekaterih lastnih imenih ter na primeru

omenjenih katalogov. V bodočih raziskavah bi bilo zanimivo podrobneje analizirati konkretne primere oteženega ali onemogočenega iskanja in upravljanja knjižničnih virov v nelatiničnih pisavah ter proučiti, v katerih primerih prihaja do največ težav.

Da je za potrebe katalogizacije gradiv v vzhodnoazijskih jezikih latinična pisava premalo, so se v Library of Congress zavedeli že pred desetletji. Za osnovo so vzeli standard CCCII (Chinese Character Code for Information Interchange), ki so ga v osemdesetih letih prejšnjega stoletja razvili na Tajvanu. Odlika tega sistema je bila predvsem v tem, da je deloval v okviru 7-bitnih znakov in je bil zaradi te lastnosti primeren za bibliotekarske sisteme (Lunde, 2008; Petrovčič, 2019). Library of Congress je na podlagi tega standarda razvila EACC (East Asian Coded Character, ANSI Z39.64-1989).

Vključitev standarda EACC kot alternativni nabor znakov v formatu MARC 21 je omogočila knjižnicam z vzhodnoazijskimi katalogi, da so lahko uporabljale in delile te kataložne zapise preko kataloga OCLC (Lam, 2006). Lam (2006) nadalje omenja, da je bilo te kataložne zapise težje vključiti v integrirane sisteme za avtomatizacijo knjižnice (ILS). Težava je bila predvsem v tem, da so bili vzhodnoazijski znaki v EACC dolgi 3 bajte. Prvotna odlika standarda CCCII se je tako izkazala za oviro, EACC pa se tudi ni več razvijal in posodabljal.

S popularizacijo Unicoda je postalo obravnavanje vzhodnoazijskih pisav zopet lažje. Več kot 25-letna raba EACC ni bila ukinjena kar čez noč, temveč je prehod na Unicode trajal nekaj časa (Lam, 2006). Na straneh Library of Congress, natančneje na straneh standarda MARC 21, preberemo, da lahko MARC 21 za shemo kodiranja uporablja ali MARC-8 ali Unicode. Tudi s spletnih strani Library of Congress in kataloga WorldCat je razvidno, da so kataložni zapisi v latiničnih in nelatiničnih pisavah izvedljivi.

8 Zaključek

V prispevku smo obrazložili problematiko, zakaj latinični zapisi ne zadostujejo pri upravljanju knjižničnih zbirk v vzhodnoazijskih pisavah. Videli smo, da je latinična pisava zgolj eden od sistemov pisave, ki pa ni primeren za vse jezike. Pri pretvorbi iz nečrkovnega v črkovni sistem pisave prihaja do izgub informacij, ki jih rekonstruiramo s težavo ali pa sploh ne. Dodatne težave nastanejo tudi zaradi tega, ker je sistemov latinizacij več. Če temu dodamo še pravilo trenutno veljavnega slovenskega pravopisa, da moramo vse izraze iz nelatiničnih pisav sloveniti, je izguba informacij že kritična.

Unicode je uspešno rešil vprašanje istočasne rabe različnih pisav, da to v današnjem času ni več nepremostljiva tehnična ovira. Prehod iz sheme kodiranja MARC-8 na Unicode ni bilo zanemarljivo, a tudi ne nemogoče delo. Rezultati so namreč izrednega pomena za uporabnike katalogov. Po svetu potekajo še podrobnejše raziskave, ki bi dodatno pripomogle k učinkovitejšemu ujemanju med latinično in nelatiničnimi pisavami.²²

Če bibliografska baza podatkov ameriške Kongresne knjižnice podpira azijske pisave, gotovo obstaja tudi rešitev, kako COBISS posodobiti do te mere, da bo deloval v okviru Unicoda. Nabor znakov Unicode je mogoče kodirati na več načinov, med katerimi so najbolj znani UTF-16, UTF-32 in UTF-8. Za zapise v formatu MARC 21 je primeren UTF-8, saj obsega zapise dolge od ene do štirih 8-bitnih enot.²³ Vrednosti od 00-FF (šestnajstiški zapis) so zapisane z eno enoto in pripadajo znakom ASCII, ki so edini dovoljeni znaki v uvodniku zapisa (leader) MARC 21 in drugih pomembnejših delih, ki označujejo posamezne dele. Ker so vsa pomembnejša mesta formata MARC 21 zapisana v 8-bitnih enotah, na primer informacija o dolžini zapisa v uvodniku zapisa (pozicije 0-4), dolžine polj in začetne pozicije v direktoriju, je prav in edino UTF-8 primeren za to, saj ustreza omejitvam formata MARC 21, obenem pa kodira vse znake nabora Unicode.²⁴

Na straneh COBISS-a preberemo, da se

[z]a izmenjavo podatkov v sistemu COBISS uporablja format COMARC/B za bibliografske podatke in format COMARC/A za normativne podatke, ki temeljita na formatu UNIMARC, ter format COMARC/H za podatke o zalogi, ki ga je razvil IZUM. Za mednarodno izmenjavo bibliografskih podatkov se uporablja format MARC 21. Možne so konverzije zapisov iz formata COMARC v MARC 21 in obratno.

Kot smo videli, so konverzije zapisov iz formata COMARC v MARC 21 možne, MARC 21 pa za shemo kodiranja lahko uporablja Unicode, natančneje UTF-8. Videli smo tudi, da številni kataložni sistemi že delujejo v okviru Unicoda. Kot pravijo snovalci COBISS-a sami:

²² Glej informacije o publikacijah in dogodkih združenja ALCTS (Association for Library Collections and Technical Services, <http://www.ala.org/alcts/>) ali raziskave kot na primer Aliprand (2005), El-Sherbini (2016), Kimura (2015), Pong in Cheung (2006), Yao (2003) in podobno.

²³ Z razliko od UTF-8 so pri kodiranju UTF-16 vsi znaki dolgi eno ali dve 16-bitni enoti, pri UTF-32 pa je za vsak znak potrebno 32 bitov.

²⁴ Za več podrobnosti glej MARC 21 *Specifications for Record Structure, Character Sets, and Exchange Media; character sets and encoding options: part 3; Unicode Encoding Environment* (<https://www.loc.gov/marc/specifications/speccharucs.html>).

Knjižnice so v informacijski družbi nepogrešljiva središča znanja, njihova naloga pa je tudi evidentiranje rezultatov »intelektualne produkcije«. Svoje vloge pa ne morejo opravljati, če njihovo delovanje ni podprto s sodobno informacijsko in komunikacijsko tehnologijo. Tudi ne morejo uspešno delovati kot osamljeni otoki, ampak se morajo povezovati ter s tem zagotavljati pogoje dialoga in sodelovanja kot najboljših podlag za razvoj vsake od držav in mednarodne skupnosti v celoti.²⁵

Avtorice prispevka zato upamo, da bo gradivo v vzhodnoazijskih jezikih nekoč tudi v okviru COBISS-a katalogizirano v pisavah teh jezikov in da bo tako mogoče evidentiranje ne samo slovenske intelektualne produkcije, ampak tudi vsega gradiva, ki se hrani v slovenskih knjižnicah.

Zahvala

Avtorice se zahvaljujemo knjižničarki Oddelka za azijske študije FF UL Darinki Baraga za pomoč pri izvedbi raziskave in za potrpežljivo spopadanje s kompleksno knjižno zbirko Oddelka. Delo Mateje Petrovčič in Kristine Hmeljak Sangawa za ta prispevek je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije (ARRS) iz državnega proračuna v okviru raziskovalnega programa št. P6-0243, delo Kristine Hmeljak Sangawa je sofinanciralo tudi Japonsko združenje za spodbujanje znanosti (JSPS) v okviru raziskovalnega projekta JSPS KAKENHI Grant Number 18F18808.

Navedeni viri

Aliprand, J. M. (2005). The structure and content of MARC 21 records in the Unicode environment. *Information technology and libraries*, 24(4), 170–179. doi: 10.6017/ital.v24i4.3381

Bekeš, A. (1999). Pojemovni okvir za klasificiranje sistemov kitajske in japonske pisave. *Azijske in afriške študije*, 3(1–2), 218–238.

Coulmas, F. (1989). *The writing systems of the world*. Oxford: Cambridge.

Daniels, P. T. (2001). Writing systems. V M. Aronoff in J. Rees-Miller (ur.), *The handbook of linguistics* (str. 43–80). Oxford: Blackwell.

²⁵ Dostopno na: <https://www.cobiss.net/si/platforma-cobiss.htm>.

DeFrancis, J. (1989). *Visible speech: the diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

El-Sherbini, M. (2016). *Improving resource discoverability for non-roman language collections*. Predstavljeno na IFLA WLIC 2016. Pridobljeno 2. 6. 2019 s spletne strani: <http://library.ifla.org/1982/1/S12-2016-el-sherbi-en.pdf>

Hannas, W. C. (1997). *Asia's orthographic dilemma*. Honolulu: University of Hawai'i Press.

Hmeljak Sangawa, K. (2019). Makrostruktura predmodernih japonskih slovarjev: kitajski vzori in japonske inovacije. V A. Bekeš, J. S. Rošker in Z. Šabič (ur.), *Procesi in odnosi v Vzhodni Aziji: zbornik EARL* (str. 191–215). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Hmeljak Sangawa, K., Kobayashi, R., Kumagai, Y., Shigemori Bučar, C., Maeno, Y. in Shukuri, Y. (2003). *Uvod v japonsko pisavo: hiragana, katakana in prvih 854 pismenk*. Ljubljana: Filozofska fakulteta, Oddelek za azijske in afriške študije.

Jeon, Y. P. (1996). ISO의 한글 로마자 표기 남북통일안 고찰 = On ISO's unified romanization system of the Korean language between the South and the North. *언어학 Eoneohag: journal of the Linguistic Society of Korea*, 19, 297–317.

Kimura, M. (2015). *Name authority data and its model for non-Latin representations with special emphasis on Chinese characters*. Doktorska disertacija. Tokyo: School of Library and Information Science of Keio University. Pridobljeno 20. 2. 2020 s spletne strani: https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO10001001-20154334-0001

Lam, K. T. (2006). *EACC to unicode migration* [spletno mesto]. OCLC CJK Users Group 2006 Annual Meeting, April 8 2006, San Francisco. Pridobljeno 2. 6. 2019 s spletne strani: <https://slideplayer.com/slide/16172569/>

Lee, H. B. (1993). 남북한 로마자 표기법 비교 연구 [= Nam.Bukhan romaja pyogibeop bigyo yeongu]. *한글 HAN-GEUL*, 222, 155–174.

Lunde, K. (2008). *CJKV information processing* (2nd ed.). Sebastopol, CA: O'Reilly Media.

McCune, G. M. in Reischauer, E. O. (1939). The romanization of the Korean language based on its phonetic structure. *Transactions of the Korea branch of the Royal Asiatic Society*, 29, 1–55. Pridobljeno 2. 6. 2019 s spletne strani: http://www.raskb.com/transactions/VOL29/KORS0749D_VOL29.pdf

Petrovčič, M. (2019). Idejni razvoj obravnavanja pisave v petih vzhodnoazijskih regijah z vidika informacijskih tehnologij. V A. Bekeš, J. S. Rošker in Z. Šabič (ur.), *Procesi in odnosi v Vzhodni Aziji* (str. 235–257). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Petrovčič, M. in Lin, M. C. (2018). *Sodobna kitajščina 1*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Pong, J. in Cheung, C. (2006). Cataloging of Chinese language materials in the digital era: the cataloging standards and practices in China, Taiwan and Hong Kong. *Tushuguanxue Yu Zixun Kexue*, 32. Pridobljeno 20. 2. 2020 s spletne strani: <https://pdfs.semanticscholar.org/93e9/2cc9f2a167fc181b5f528b21020afcaf0a68.pdf>

Slovenski pravopis. (2003). Ljubljana: Znanstvenoraziskovalni center SAZU, Založba ZRC.

Taylor, I. in Taylor, M. M. (2014). *Writing and literacy in Chinese, Korean and Japanese*. Amsterdam: Benjamins.

Yao, K. T. (2003). Issues on East Asian character codes and Unicode: what happened to the character I input?. *Journal of East Asian libraries*, (131), 1–8. Pridobljeno 20. 2. 2020 s spletne strani: <http://hdl.handle.net/10125/7976>

dr. Kristina Hmeljak Sangawa

Oddelek za azijske študije, Filozofska fakulteta, Univerza v Ljubljani, Aškerčeva 2, 1000 Ljubljana
e-pošta: kristina.hmeljak@ff.uni-lj.si

dr. Hyeonsook Ryu

Oddelek za azijske študije, Filozofska fakulteta, Univerza v Ljubljani, Aškerčeva 2, 1000 Ljubljana
e-pošta: hyeonsook.ryu@ff.uni-lj.si

dr. Mateja Petrovčič

Oddelek za azijske študije, Filozofska fakulteta, Univerza v Ljubljani, Aškerčeva 2, 1000 Ljubljana
e-pošta: mateja.petrovcic@ff.uni-lj.si