Univerza v Ljubljani
Filozofska fakulteta
Oddelek za slovenistiko
Center za slovenščino kot drugi in tuji jezik

Obdobja 45

Slovenščina v dobi pametne tehnologije

Ljubljana 2026

Obdobja 45
Slovenščina v dobi pametne tehnologije

Zbirka: Obdobja (ISSN: 1408-211X, e-ISSN: 2784-7152)
Urednica: Mojca Stritar Kučuk
Recenzenti: Špela Arhar Holdt, Mateja Eniko, Apolonija Gantar, Lara Godec Soršak, Nataša Hribar, Mateja Jemec Tomazin, Iztok Kosem, Tina Lengar Verovnik, Andrej Perdih, Eva Pori, Tadeja Rozman, Mojca Smolej, Marko Stabej, Mojca Stritar Kučuk, Jana Volk, Rok Žaucer
Lektorji izvlečkov in prispevkov v tujem jeziku: Tomislav Ćužić, Tatjana Komarova, Donald F. Reindl
Tehnična urednika: Matej Klemen, Katja Krajnc
Korekture: avtorji, Matej Klemen, Katja Krajnc
Oblikovanje: Metka Žerovnik
Prelom: Miro Pečar

Založila: Založba Univerze v Ljubljani
Za založbo: Gregor Majdič, rektor Univerze v Ljubljani
Izdala: Znanstvena založba Filozofske fakultete Univerze v Ljubljani; Center za slovenščino kot drugi in tuji jezik pri Oddelku za slovenistiko
Za založbo: Mojca Schlamberger Brezar, dekanja Filozofske fakultete

Ljubljana, 2026
Prvi natis
Naklada: 350 izvodov
Tisk: Birografika Bori, d. o. o.

Publikacija je brezplačna.
Izdajo monografije je omogočilo Ministrstvo za izobraževanje, znanost in mladino RS.

Prva e-izdaja.
Publikacija je v digitalni obliki prosto dostopna na https://e-knjige.ff.uni-lj.si/.
DOI: https://doi.org/10.4312/Obdobja.45.2784-7152

To delo je ponujeno pod licenco Creative Commons Attribution-ShareAlike 4.0 International License – CC BY-SA 4.0 (priznanje avtorstva in deljenje pod istimi pogoji; izjema so fotografije).

Kataložna zapisa o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani.

Tiskana knjiga
COBISS.SI-ID 304382211
ISBN 978-961-297-933-1

E-knjiga
COBISS.SI-ID 304548611
ISBN 978-961-297-934-8 (PDF)

Vsebina

Mojca Stritar Kučuk
Prihodnost – danes! 7

Tjaša Alič, Tanja Jerman
Uporabniki spletnega tečaja za učenje slovenščine Slovene Learning Online (Slonline) v primerjalni perspektivi 9

Špela Arhar Holdt, Kaja Dobrovoljc Zor, Nives Hüll, Cyprian Laskowski, Tomaž Erjavec
Jezikovni viri na repozitoriju CLARIN.SI za slovenistične raziskave 21

Špela Arhar Holdt, Iztok Kosem
Jezikoslovne in tehnološke rešitve pri razvoju portala jezikovnih iger: obravnava občutljivega besedišča 31

Špela Arhar Holdt, Tina Munda, Nejc Robida, Živa Štebljaj, Polona Gantar
Od avtentičnih popravkov do generiranih napak: učni podatki za slovnično pregledovanje slovenščine 41

Barbara Bednjički Rošer
Uporaba generativne umetne inteligence pri načrtovanju jezikovnih dejavnosti za predšolske otroke 49

David Blažek
Prevajanje iz češčine v slovenščino s pomočjo pametnih tehnologij: pomoč ali nadloga? 57

Milena Mileva Blažić
Fran Ramovš oz. Julij Dub: digitalizacija in literarna analiza spregledanih protipravljic 67

Jaka Čibej
Polavtomatski delotok za urejanje nabora leksemov v Digitalni slovarski bazi za slovenščino na primeru glagolov 75

Erika Dimc
S skrolanjem do slovenščine: vloga Instagrama pri učenju slovenščine kot neprvega jezika 83

Kaja Dobrovoljc Zor, Luka Terčon
Nekaj jezikovnih značilnosti umetno generiranih šolskih esejev v slovenščini 93

Rok Dovjak
Lovljenje ravnotežja: jezik RTV Slovenija in odzivi javnosti nanj 107

Mateja Eniko, Matej Klemen, Damjana Kern Andoljšek, Ina Ferbežar, Tjaša Alič, Mihaela Knez
Sporazumevalne potrebe odraslih priseljenih kot izhodišče za razvoj učnega gradiva za slovenščino kot drugi jezik 117

Ina Ferbežar
Izzivi jezikovnega testiranja nove generacije 129

Polona Gantar, Nejc Robida, Simon Krek
Oblikovno nestandardna leksika v Digitalni slovarski bazi za slovenščino 139

Lara Godec Soršak, Miha Matjašič
Razvijanje pravopisne zmožnosti z vključevanjem pristopa obrnjenega učenja in poučevanja 149

Luka Horjak
Digitalna orodja pri poučevanju artikulacijske fonetike..... 159

Marina Jajić Novogradec
Kontekstualni čimbenici u ovladavanju slovenskim jezikom: sport u jeziku / jezik u sportu..... 167

Janoš Ježovnik, Dejan Gabrovšek, Nataša Gliha Komac, Artur Stepanov, Matic Pavlič, Tina Potočnik, Andrej Perdih
Sociodemografske spremenljivke kot del digitalne jezikovne infrastrukture: primer slovenskega Besedomata in katalonske raziskave razširjenosti besed..... 175

Petra Jordan
Dileme ob zapisovanju govorjenega jezika v oddaji *Kaj dogaja?*..... 185

Neža Kočnik, Petr Plecháč
Avtomatska detekcija oblikoskladenjskih lastnosti v korpusu slovenske poezije PoeTree.sl 195

Diana Košir
Primerjalna semantično-pragmatična analiza intenzitetnih označevalcev v avtorskih in umetno generiranih presojevalnih besedilih..... 201

Simona Kranjc
Veliki jezikovni modeli in izzivi poučevanja na fakulteti 211

Domen Krvina, Mija Michelizza
Od umetne trave do lažne javne podpore: o novejši izlastnoimenski leksiki na gradivu *Sprotnega slovarja slovenskega jezika* 219

Tina Lengar Verovnik, Monika Kalin Golob
Od ločevalne norme do čustvene pristnosti: jezikovna sredstva afektivnosti v informativnih novinarskih besedilih 229

Marko Ljubešić
Intercultural Linguistic Challenges in the Twenty-First Century: A Didactic Approach to Autonomous Learning of Slovenian with the Support of an AI Tutor . 239

Varvara Magomedova
Veje: A Silver-to-Gold Annotation Pipeline and Search Platform for a Slovenian Dialect Treebank 249

Irina Makarova
Generativna umetna inteligenca v pedagoškem slovaropisju: izkušnje iz prenove slovensko-ruskega kolokacijskega slovarja za začetnike 257

Lucija Mandić
Reprezentacije spola v slovenski pripovedni prozi 19. stoletja v sodobnem srednješolskem kanonu: korpusni pristop 267

Nina B. Mečkovskaja
Грамматические и лексические сходства и различия в высокочастотной лексике: самые частые словенские, русские и английские субстантивы и глаголы 277

Anita Peti-Stantić
What Can We Learn from Each Other? Slovenian and Croatian Language Technologies 289

Ivana Petric Lasnik
Učiteljeva pisna povratna informacija in orodja umetne inteligence kot potencialni pripomoček za njeno oblikovanje 297

Eva Pori, Tomaž Toporišič
Med digitalnim *feedom* in *skrolanjem*: jezik in diskurz hibridnih uprizoritev 307

Ina Poteko
Kdo to piše? Avtorsko profiliranje umetnointeligenčnih pomočnikov..... 317

Patrizia Raveggi
The Uncanny in the Algorithmic Age: Cultural Translation, Semantic Sovereignty, and the Slovenian Response to AI 327

Nejc Robida, Janez Križaj
Od korpusa do podatkov: jezikoslovcem prijazne fonetične meritve..... 335

Martina Rodela, Sonja Starc
Branje večkodnih hiperbesedil 343

Eva Sicherl
Vzpon skritih anglizmov v slovenščini 351

Mojca Stritar Kučuk
Vsak začetek ni (enako) težak: korpusni pogled na besedišče pri začetnih govorcih slovenščine kot neprvega jezika 359

Tjaša Šoltes, Marko Bajec
Problematika diskurznega označevalca *ne* pri avtomatski razpoznavi govora in strojnem prevajanju: analiza rabe v korpusu univerzitetnih predavanj 371

Veronika Štěpánová, Tanja Mirtič
Speech Recognition of Standard Czech and Slovenian Using OpenAI Whisper..... 379

Marko Trobevšek
Slovenščina gimnazijskih spisov iz časov pouka na daljavo 389

Urška Vranjek Ošlak, Hana Mžourková
Linguists' Attitudes toward the Use of AI in Language Consulting 397

Nina Zavašnik Hofer
Priložnosti in izzivi za uporabnike brajeve pisave v dobi sodobne tehnologije 407

Zarja Zver
Digitalna opremljenost za opolnomočenje uporabnikov in uporabnic slovenščine v Italiji: primer *Primorskega dnevnika* 415

Prihodnost – danes!

Prihodnost je tukaj. Morda (še) ne tako futuristična, kot so si jo zamišljali (pop)kulturni vizionarji, z letečimi avtomobili, robotskimi služabniki, oprijetimi plastičnimi uniformami in tabletkami namesto običajne hrane. Toda živimo v dobi, v kateri vsakdan zaznamujejo digitalna tehnologija in globalno vsemrežje informacij, aplikacije za vse, kar moramo opraviti, pa še za kaj, česar nam sploh ni treba, za povrh pa družba narcističnega individualizma, napajana z neusahljivo energijo všečkov in družbenih omrežij.

Kdaj točno smo padli v približek orwellovske prihodnosti, naj povedo strokovnjaki drugih ved, sama pa lahko poročam le o tem, kdaj je pametna tehnologija zares posegla v moje delo. Kot večina tistih, ki se ukvarjajo s slovenščino in jih takoj po diplomi ne odnese v šolstvo, sem se v svoji dobri dve desetletji trajajoči karieri preizkušala v različnih službah, najdlje in najbolj predano pa poučujem slovenščino za neprve govorce. Učila sem na različnih tečajih, od kratkih do najdaljših in najintenzivnejših, na različnih ravneh in po različnih učnih gradivih. Toda nič ni mojega dela spremenilo bolj kot ne čisto natančno določen trenutek okrog leta 2020, ko je strojno prevajanje za slovenščino začelo dobro delovati. Pred tem je bilo moje poučevanje utečeno, in čeprav sem delala v razmeroma velikih skupinah univerzitetnih študentov, so redne pisne domače naloge omogočale individualno ukvarjanje z vsakim posameznikom. Nenadoma pa je to postalo vprašljivo – kakšen smisel imajo spisi, oddani v popolni slovenščini, če pa njihovi tvorci v razredu ne zmorejo povedati ene same izjave, ki bi zvenela približno slovensko? Google Translate in drugi brezplačni prevajalniki na lepem niso bili več orodje namišljenih nigerijskih vdov, ki so hrepenele po tem, da bi prav nam predale nekaj svojih milijončkov, pač pa so postali najširše dostopno in učinkovito orodje, zaradi katerega sem morala resno prevetriti učne metode. Zato me pojav prvih velikih jezikovnih modelov – poleg pandemije covida za zdaj eden od najbolj revolucionarnih dogodkov v tretjem desetletju tretjega tisočletja – ni pretirano pretresel.

Še ko sem bila globoko v procesu prilagajanja same sebe in svoje miselnosti na spremembe digitalne dobe, pa sem prišla še do enega pomembnega spoznanja. Med kramljanjem s profesorico angleščine z ugledne zahodnoevropske univerze sem jo povprašala, kako se pri njih odzivajo na uporabo strojnih prevajalnikov med študenti. Gospa me je samo debelo gledala in zdi se, da ji zares ni bilo jasno, o čem govorim. Kakšni strojni prevajalniki, kakšen problem pri delu s študenti, kakšen negativen vpliv na usvajanje jezika? Takrat sem se zavedela, da je položaj slovenščine vendarle popolnoma drugačen od pozicije, ki jo zaseda globalno vseprisotna angleščina, in da se moramo s spremenjenimi okoliščinami našega jezika ukvarjati predvsem govorci slovenščine sami.

Verjetno lahko vsak pove svojo zgodbo o tem, kako mu je (pametna) tehnologija spremenila življenje, pa naj gre za stalno uporabo zaslonov, spletno nakupovanje, umetno inteligenco, pametne domove, asistenčne storitve v avtomobilu ali kaj tretjega. Zbornik tokratnega simpozija Obdobja se teh zgodb loteva znanstveno. Kot 45-letnik je dovolj zrel, da se novosti digitalne dobe ne boji in jih suvereno izkorišča. Številni

prispevki so posvečeni jezikovnim tehnologijam in umetni inteligenci, posvečajo pa se tako njihovi uporabi v različnih analizah, aplikacijah ali celo interpretacijah leposlovnih del kot tudi s tem povezanim etičnim in metodološkim vprašanjem. Hkrati pa je letošnji zbornik kot gospod srednjih let tako izkušen, da na novotarije sodobne družbe gleda z določeno mero dvoma in se o vplivu pametne tehnologije na slovensko jezikovno, pa tudi širše družbeno okolje sprašuje tudi z nekaj zadržanosti, razvidne iz prispevkov o leksikalni podobi sodobne slovenščine, specifikah novinarskega diskurza, jezikovni politiki in še čem.

V zborniku 45. simpozija Obdobja torej raziskujemo, kakšna je današnja podoba slovenščine, jezika z razmeroma majhnim številom domačih govorcev, a visoko jezikovno zavestjo in že dolgo prisotnimi prizadevanji za čim večjo jezikovnotehnološko podprtost. Seveda se ne omejujemo samo na slovenščino, pač pa se oziramo tudi po drugih, predvsem sorodnih slovanskih jezikih. Z različnih vidikov osvetljujemo tudi slovenščino tistih, ki niso njeni domači govorci.

Ob tem se zahvaljujem vsem, ki ste omogočili nastanek tega zbornika – s svojimi mnenji, predlogi, komentarji, premisleki in recenzijami, predvsem pa seveda s svojimi besedili. Brez vas te knjige ne bi bilo.

doc. dr. Mojca Stritar Kučuk,
predsednica 45. simpozija Obdobja

Uporabniki spletnega tečaja za učenje slovenščine Slovene Learning Online (Slonline) v primerjalni perspektivi

DOI: 10.4312/Obdobja.45.9-19

Tjaša Alič, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; tjasa.alic@ff.uni-lj.si
Tanja Jerman, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; tanja.jerman@ff.uni-lj.si

Prispevek predstavlja uporabnike spletnega tečaja za učenje slovenščine Slonline, njihove demografske podatke, od kod do tečaja dostopajo, zakaj ga uporabljajo in kako ocenjujejo svojo izkušnjo učenja z njim. Pri tem prispevek rezultate nove ankete primerja z ugotovitvami predhodne raziskave med uporabniki, s čimer osvetljuje spremembe in kontinuiteto v njihovih potrebah, pričakovanjih in učnih navadah. Na tej podlagi poudarja vlogo spletnega tečaja Slonline kot brezplačno dostopnega učnega okolja za samostojno učenje slovenščine ter kot razvojno prilagodljivega orodja, ki ga je mogoče nadgrajevati tudi v skladu s podatki o rabi in odzivih uporabnikov.

spletni tečaj, Slovene Learning Online/Slonline, slovenščina kot drugi in tuji jezik, učenje slovenščine, uporabniki spletnega tečaja

This article presents the users of the Slonline online course for learning Slovenian, their demographic profiles, where they access the course from, why they use it, and how they rate their learning experience with it. The article compares the results of a new survey with the findings of a prior study among users, thereby highlighting changes and continuities in their needs, expectations, and learning habits. It highlights the role of the Slonline online course as a freely accessible learning environment for independently learning Slovenian and as a developmentally adaptable tool that can be enhanced based on usage data and user feedback.

online course, Slovene Learning Online / Slonline, Slovenian as a second and foreign language, learning Slovenian, online course users

1 Uvod

Spletna učna okolja za samostojno učenje jezikov imajo v sodobni jezikovni didaktiki pomembno vlogo, saj uporabnikom širijo dostop do učnih vsebin onkraj institucionalno, prostorsko in časovno omejenih oblik pouka. Ta vidik je posebej pomemben pri slovenščini kot neprvem jeziku, saj so možnosti organiziranega učenja slovenščine, sploh izven Slovenije, omejene. Na Centru za slovenščino kot drugi in tuji jezik (v nadaljevanju CSDTJ) zato razvoj spletnih učnih okolij za otroke, mladostnike in odrasle poteka že od leta 2001 (Alič in Kern Andoljšek 2025).

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

Slovene Learning Online (v nadaljevanju Slonline)[1] je brezplačni spletni tečaj za učenje slovenščine kot neprvega jezika na začetni ravni.[2] Namenjen je odraslim uporabnikom,[3] do njega pa lahko dostopajo na računalnikih, tablicah in mobilnih telefonih. Učenje slovenščine je podprto s prevodi v osem jezikov,[4] slikovnim in zvočnim gradivom, različnimi interaktivnimi nalogami, igrifikacijo ter razlago slovnice. Takšna zasnova uporabnika vodi skozi proces samostojnega in samousmerjenega učenja, pri katerem lahko sam izbira teme, učne sklope in tempo učenja (Alič in Jerman 2022; Brown Nielson 2022; Munday 2016; Retelj 2015).

Pomembna je dvojna vloga spletnega tečaja Slonline: po eni strani omogoča dostop do učenja slovenščine različnim skupinam uporabnikov v Sloveniji in drugod po svetu, po drugi strani pa s podatki o registracijah, aktivnosti in uporabniških odzivih odpira možnost sistematičnega spremljanja potreb in učnih praks uporabnikov. Na podlagi tega razvijalci tečaja lahko odpravljajo tehnične in vsebinske pomanjkljivosti ter prilagajajo učne vsebine dejanskim potrebam ciljnih skupin. Podatki namreč lahko ponudijo edinstven vpogled v aktivnost učečih se, njihovo uspešnost in napredek pri razvijanju jezikovne zmožnosti (Brown Nielson 2022: 233).

Prispevek se najprej osredotoča na geografsko razpršenost uporabnikov in dinamiko rasti števila registracij. V nadaljevanju predstavlja načine dostopa do tečaja Slonline in pregled obiskanih vsebin, sledi primerjalna raziskava profila uporabnikov, njihovih motivov za učenje ter uporabniške izkušnje.

2 Uporabniki

2.1 Geografska razpršenost uporabnikov

Vpogled v bazo uporabnikov spletnega tečaja Slonline omogočajo letna poročila CSDTJ[5] (gl. letna poročila za leta od 2017 do 2025; Center za slovenščino kot drugi in tuji jezik (b. d.-a)), ki kažejo na stalno rast registriranih uporabnikov. V treh tednih od objave decembra 2017 je imel tečaj 2038 registriranih uporabnikov iz 97 držav, do konca leta 2025 pa je njihovo število naraslo na 62.655 iz 200 držav[6] (Graf 1). Pri interpretaciji

1 Prvi modul (SLO 1) je bil objavljen leta 2017 (Alič idr. 2017), drugi (SLO 2) leta 2019 (Alič idr. 2019), tretji (SLO 3) pa leta 2023 (Alič idr. 2023).

2 Vsebino modulov smo primerjali s seznamom jedrnega besedišča za slovenščino kot drugi in tuji jezik (Klemen idr. 2022). Skupni pregled besedišča vseh treh modulov kaže, da vsebina tečaja ustreza začetnemu učenju slovenščine, saj je 55,5 % besed uvrščenih na raven A1, 20,8 % na raven A2, 5,5 % na raven B1, 18,3 % pa ni zajetih v seznamu jedrnega besedišča (npr. osebna in zemljepisna imena oz. besedišče, ki presega raven B1).

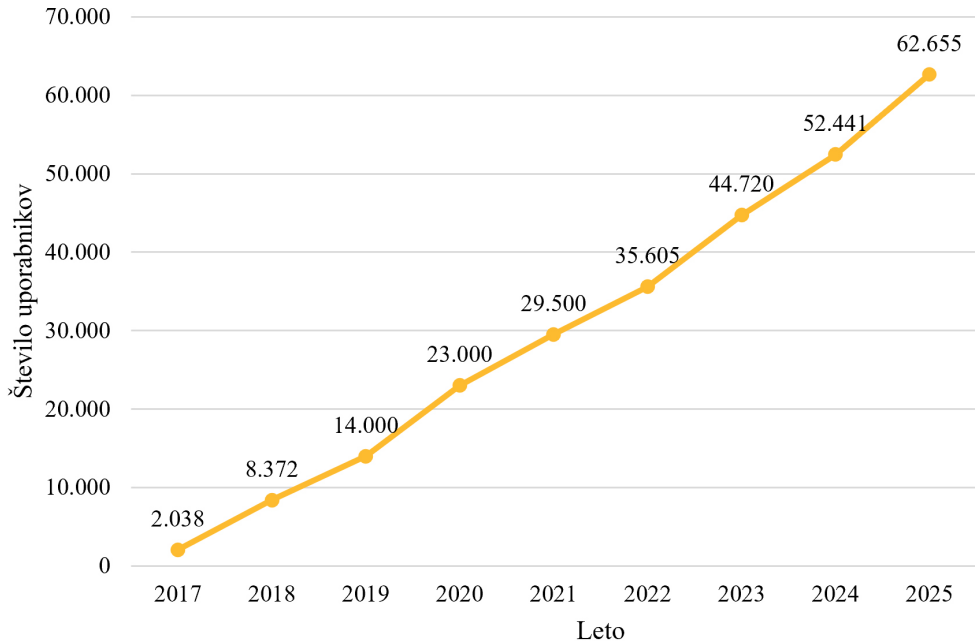
3 V nadaljevanju uporabljamo izraz uporabniki, enako tudi anketirani ipd., moška oblika se tako nanaša na katerikoli spol.

4 Uporabniki se slovenščino lahko učijo s pomočjo angleščine, francoščine, hrvaščine, italijanščine, nemščine, ruščine in španščine, od decembra 2025 pa tudi albanščine, ki je bila dodana v okviru projekta Naslon: Nadgradnja spletnega tečaja Slonline – mobilno in spletno učenje slovenščine za priseljence (https://centerslo.si/naslon/).

5 Podatki o uporabnikih so v letnih poročilih pridobljeni iz baze uporabnikov Slonline, ki jo urejamo v sodelovanju s podjetjem Digied, d. o. o.

6 Ta številka ne zajema samo mednarodno priznanih držav, temveč tudi nekatere samostojno vodene

grafa je seveda treba upoštevati, da prikazuje kumulativno število registriranih uporabnikov, ne pa nujno tudi njihove dejanske aktivnosti v tečaju.



Graf 1: Rast števila registriranih uporabnikov spletnega tečaja Slonline med letoma 2017 in 2025.

Na podlagi teh podatkov lahko trdimo, da je Slonline globalno dostopno učno okolje za učenje slovenščine. Še natančnejši vpogled v geografsko strukturo uporabnikov pa omogočajo podrobnejši podatki o državah z največ registriranimi uporabniki po posameznih letih.[7] Graf 2 prikazuje prvih deset držav z največ registriranimi uporabniki tečaja za posamezno leto v obdobju 2017–2025.[8]

oz. v analitičnih sistemih ločeno obravnavane geografske enote, npr. Kosovo.

7 Baza uporabnikov temelji na podatku o državi, ki jo uporabnik sam navede oz. izbere ob registraciji. Do leta 2023 so uporabniki lahko sami vpisali državo, iz katere prihajajo, po tem letu pa smo, predvsem zaradi boljše preglednosti, uvedli spustni seznam držav.

8 Prazna celica pomeni, da država v tistem letu ni bila uvrščena med prvih deset.



Graf 2: Deset držav z največ registriranimi uporabniki tečaja Slonline za posamezno leto v letih 2017–2025.

Največ uporabnikov prihaja iz ZDA, Slovenije, Italije, Srbije, Velike Britanije ter Bosne in Hercegovine, do leta 2025 tudi iz Severne Makedonije in od leta 2018 iz Nemčije. Obenem na Grafu 2 lahko vidimo, da se razmerja med državami, iz katerih prihajajo uporabniki, skozi leta spreminjajo. Posebej izrazita je rast števila registriranih uporabnikov iz Italije, ki se v zadnjih letih uvršča na sam vrh, ter postopna rast števila uporabnikov iz Rusije in Avstrije. Nasprotno pa se npr. po številu uporabnikov Poljska po letu 2017 ne uvršča med prvih deset držav, prav tako ne Argentina po letu 2021.

2.2 Dostop do spletnega tečaja

Podatki spletne analitike Google Analytics[9] nam pokažejo, od kod uporabniki dejansko dostopajo do tečaja. Med prvimi desetimi državami za obdobje 2017–2025 izrazito izstopa Slovenija z 29,85 % vseh aktivnih uporabnikov. Domnevamo lahko, da gre predvsem za tujce, ki se slovenščino učijo zaradi življenja, dela ali študija v Sloveniji. Pomemben delež predstavljajo tudi uporabniki iz sosednjih in bližnjih evropskih držav, kot so Italija (8,65 %), Nemčija (6,33 %), Srbija (3,69 %), Avstrija (3,63 %), in Hrvaška (1,96 %), veliko uporabnikov pa je še iz Velike Britanije (3,51 %). Med bolj oddaljenimi državami izstopata ZDA (9,54 %) in Kanada (2,10 %), med prvih deset pa se uvršča še Rusija (3,18 %). Te podatke je mogoče povezati z migracijskimi, študijskimi in delovnimi povezavami s Slovenijo ter tudi s slovenskimi skupnostmi v zamejstvu in izseljenstvu.

9 https://analytics.google.com

Uporabniki so do tečaja po podatkih Google Analytics za obdobje 2017–2025 najpogosteje dostopali prek mobilnih telefonov (54,4 %) in računalnikov (43,6 %), medtem ko je delež tistih, ki so za dostop uporabili tablico, precej manjši (2,1 %). Tečaj mora zato še naprej zagotavljati dobro uporabniško izkušnjo na obeh ključnih vrstah naprav. Podatki o virih obiska kažejo, da uporabniki do vstopne strani spletnega tečaja dostopajo neposredno (56,3 %), to pomeni, da že poznajo spletni naslov tečaja, pomemben delež pa predstavljajo tudi obiski prek spletnih iskalnikov (31,3 %). Manj uporabnikov je do tečaja prišlo prek drugih spletnih strani (8,8 %) ali družbenih omrežij (2,3 %).

Po podatkih spletne analitike Google Analytics o ogledih strani v letih 2017–2025 uporabniki največ uporabljajo začetne učne vsebine spletnega tečaja Slonline, zlasti teme in naloge modula SLO 1. Primerjava modulov SLO 2 in SLO 3 kaže vzorec, da prve teme teh dveh modulov dosežejo največ uporabnikov, nadaljnje teme pa imajo pogosto daljši povprečni čas dejavnosti. To je mogoče razumeti kot pokazatelj, da se le del uporabnikov po začetnem vstopu v modul premika naprej k zahtevnejšim vsebinam, kjer pa se pri posameznih temah zadržijo dlje časa. Pri SLO 3 je treba upoštevati tudi, da gre za pozneje dodan in vsebinsko zahtevnejši del tečaja. Testi za preverjanje znanja pri vseh treh modulih imajo praviloma manj aktivnih uporabnikov kot splošne učne strani, vendar uporabniki na teh straneh ostanejo dlje časa, kar kaže na ponavljajočo se rabo preverjanja znanja.

Obsežnejše podatke o uporabnikih tečaja in njihovi uporabniški izkušnji predstavljamo v nadaljevanju.

3 Raziskava

V letih 2022 in 2026 smo izvedli dve raziskavi, povezani z uporabniško izkušnjo.[10] Namen je bil pridobiti boljše razumevanje uporabnikov spletnega tečaja Slonline, kar bi omogočilo načrtovanje učinkovitejšega in uporabnikom bolj prilagojenega tečaja. Z drugo raziskavo pa smo želeli ugotoviti še, kako so se v štiriletnem obdobju spremenili profil uporabnikov, njihove potrebe in uporabniška izkušnja. V nadaljevanju prispevka predstavljamo ključne rezultate obeh raziskav in jih primerjamo.

3.1 Metoda

Za zbiranje podatkov smo uporabili anonimno spletno anketo, dostopno na spletni platformi 1KA. Prvo anketiranje je potekalo od 11. 4. do 4. 5. 2022, drugo, v nekoliko dopolnjeni obliki,[11] pa od 24. 2. do 25. 3. 2026. Vzorčni okvir iz leta 2022 so predstavljali uporabniki, ki so se v tečaj registrirali med 1. 2. 2021 in 31. 1. 2022 (N = 5829), iz leta 2026 pa uporabniki, registrirani med 1. 2. 2025 in 31. 1. 2026 (N = 9008). Vprašalnik je bil leta 2022 poslan 3316 uporabnikom, leta 2026 pa 4894,[12] od njih smo v prvem anketiranju prejeli 312 delno ali v celoti izpolnjenih vprašalnikov, štiri leta pozneje pa

10 Rezultati raziskave iz leta 2022 so objavljeni v Alič in Jerman (2022).

11 Na spremembo vprašalnika so vplivali predvsem leta 2023 dodani modul SLO 3 ter leta 2025 razvita nova mobilna aplikacija in dodani jezik – albanščina.

12 Izločili smo tiste, ki ob registraciji niso dali privoljenja za pošiljanje e-sporočil, nekaj pa je bilo tudi napačno vnesenih e-naslovov in nedokončanih aktivacij uporabniškega računa.

242. Zaradi vključevanja tudi delno izpolnjenih vprašalnikov se število sodelujočih (n) pri posameznih analizah razlikuje. Komunikacija z uporabniki in oba vprašalnika so bili zaradi organizacijskih in finančnih razlogov v angleškem jeziku, zavedajoč se metodološke omejitve, da s tem verjetno nismo dosegli vseh uporabnikov.

Vprašalnik iz leta 2022 je vseboval 31 vprašanj (22 zaprtega in 9 odprtega tipa), dopolnjeni vprašalnik iz leta 2026 pa 34 (od tega 8 odprtega tipa). Oba vprašalnika sta bila usmerjena v zbiranje demografskih in drugih podatkov o uporabnikih (motivacija za učenje, predhodno učenje slovenščine in drugih jezikov ipd.) ter različnih stališč do spletnega tečaja Slonline. Zaradi obsežnosti vprašalnikov in fokusa na uporabnike v prispevku predstavljamo le del rezultatov.

3.2 Rezultati raziskave in interpretacija

3.2.1 Geografska razpršenost in demografske značilnosti anketiranih

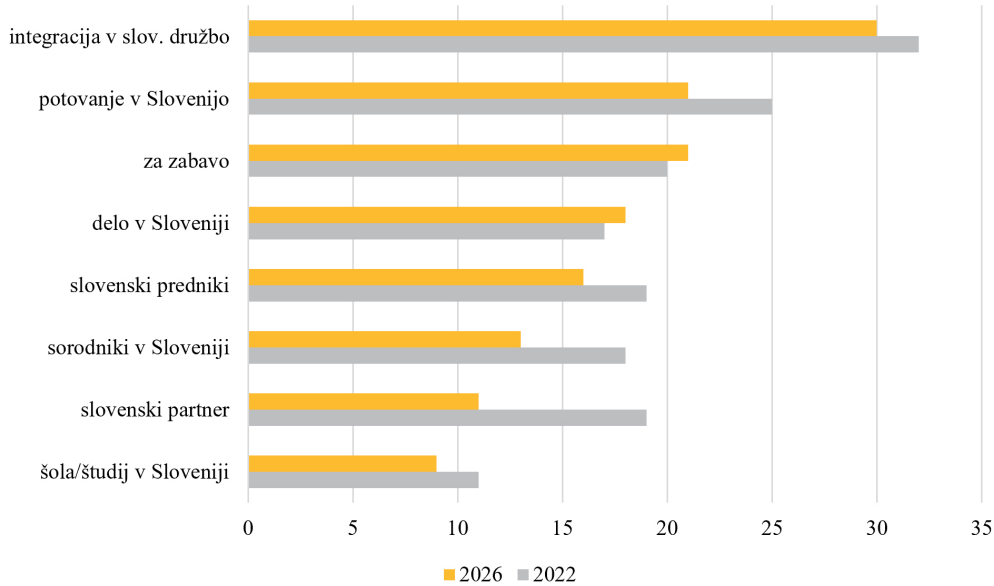
Porast števila uporabnikov in njihova geografska razpršenost se odražata tudi pri sodelujočih v obeh anketah. Anketirani so leta 2022 prihajali iz 23 držav, leta 2026 pa iz 45. Največ jih prihaja iz Italije (21 % leta 2022 oz. 14 % leta 2026), sledijo ZDA, Nemčija, Avstrija, Rusija in Velika Britanija, kar sovpada s podatki o registriranih uporabnikih. V zadnji anketi je opaznejši delež anketiranih še iz Srbije, Kanade, Severne Makedonije, Španije, Madžarske, Kitajske in Filipinov. Večina jih ni živela v Sloveniji (80 % oz. 74 %), tako verjetno mnogim tečaj omogoča edini način učenja slovenščine.

Leta 2022 je bil delež žensk in moških enak (48 %), štiri leta pozneje pa je bilo žensk nekoliko več (51 %). Opazne so spremembe v starostni strukturi. Leta 2026 je bil višji delež sodelujočih, starih od 35 do 54 let (36 % leta 2022 oz. 50 % leta 2026), konstanten je delež starejših od 65 let (11 % oz. 12 %), medtem ko je bilo mlajših od 35 let manj (37 % oz. 23 %). To verjetno lahko pripišemo tudi temu, da se mlajše generacije redkeje odzivajo na spletne vprašalnike. Dobljeni podatki se namreč razlikujejo od rezultatov raziskav aplikacij za učenje tujih jezikov, po katerih mlajši odrasli v starosti med 18. in 24. letom predstavljajo kar 48 % vseh uporabnikov.[13]

3.2.2 Motivi za učenje, učne navade in razlogi za učenje prek spletnega tečaja Slonline

V obeh anketah so sodelujoči kot najpogostejši motiv za učenje slovenščine (gl. Graf 3) navedli družinske vezi: imajo slovenske prednike, slovenskega partnerja ali sorodnike v Sloveniji (56 % leta 2022 oz. 40 % leta 2026). Sledi želja po vključitvi v slovensko družbo, še vedno pomembna, čeprav nekoliko manj, sta načrtovana pot v Slovenijo ter učenje za zabavo. Leta 2026 je nekoliko več anketiranih, katerih motiv za učenje je delo v Sloveniji,[14] večina le-teh v Sloveniji že živi. Rahlo manjši pa je delež tistih, ki se slovenščino učijo zaradi šole oz. študija, kar sovpada s prej predstavljeno starostno strukturo anketiranih.

13 Več gl. Language learning app market (2024).
14 Več o rabi slovenščine na delovnem mestu gl. Ferbežar in Huber (2024).



Graf 3: Motivi za učenje slovenščine v % (2022: n = 312; 2026: n = 169); možnih je bilo več odgovorov.

Delež anketiranih, ki se različne tuje jezike učijo prek spleta oz. aplikacij, je nekoliko višji – če se jih je leta 2022 na ta način učilo 49 %, se jih je štiri leta pozneje že 53 %, najpogosteje uporabljena aplikacija ostaja Duolingo. Slovenščine se velika večina vprašanih (85 % leta 2022 oz. 72 % leta 2026) prej ni učila ne na spletu ne z drugimi aplikacijami. S tečajem Slonline pa se je večina učila kontinuirano, največji delež, skoraj polovica, se je učila daljše časovno obdobje, tj. nekaj mesecev ali več (46 % oz. 47 %), večina od 11 do 30 minut na dan (37 % oz. 43 %). V raziskavi nas je zanimalo tudi, koliko vsebin so vprašani predelali.[15] Vse ali skoraj vse tri module je leta 2026 predelalo 20 % anketiranih, med njimi je bilo največ takih, ki se slovenščino učijo zaradi dela in bivanja v Sloveniji. Večina je predelala le posamezne teme tečaja SLO 1 (49 % oz. 38 %),[16] je pa v drugi anketi izrazito več tistih, ki so modul SLO 1 predelali v celoti (17 % oz. 35 %). Lahko predvidevamo, da gre za uporabnike, ki so se slovenščino učili kontinuirano, sistematično in poglobljeno.

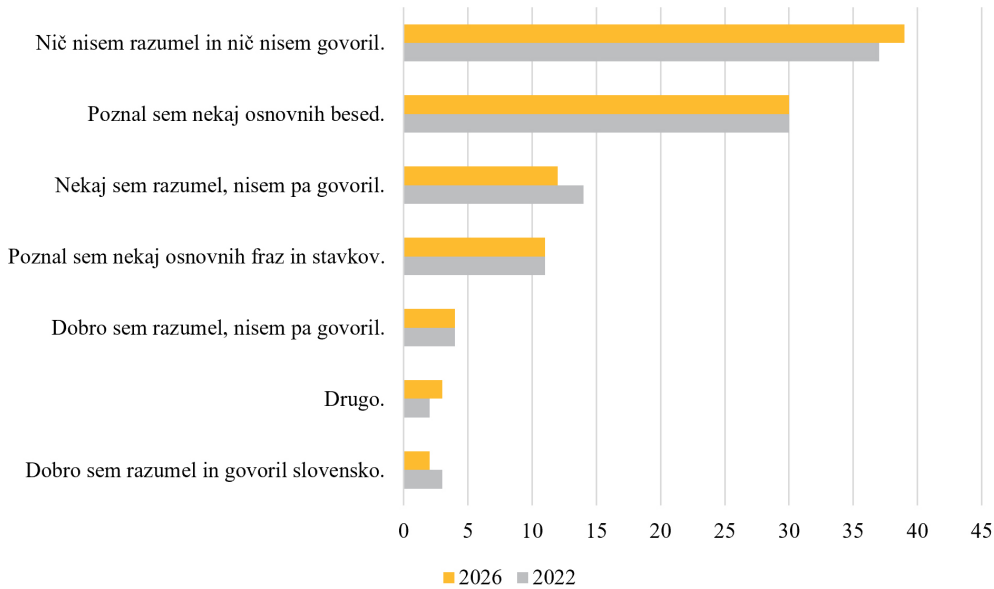
Ključni razlogi za izbiro tečaja Slonline ostajajo v obeh letih stabilni: brezplačnost (69 % leta 2022 oz. 60 % leta 2026), časovna fleksibilnost (67 % oz. 62 %), možnost učenja v svojem tempu (57 % oz. 49 %) in na katerikoli lokaciji (51 % oz. 44 %).

15 Modul SLO 3 je bil zaradi poznejšega nastanka vključen samo v drugi vprašalnik.

16 Glede na mednarodne raziskave 48 % uporabnikov preneha z učenjem tujih jezikov prek aplikacij oz. spleta, preden doseže nadaljevalno stopnjo (Floyd 2026). Tudi na klasičnih tečajih CSDTJ je največ učečih se začetnikov (Center za slovenščino kot drugi in tuji jezik (b. d.-a); gl. letna poročila za leta od 2017 do 2025).

3.2.3 Predznanje, učne potrebe in samoocena napredka

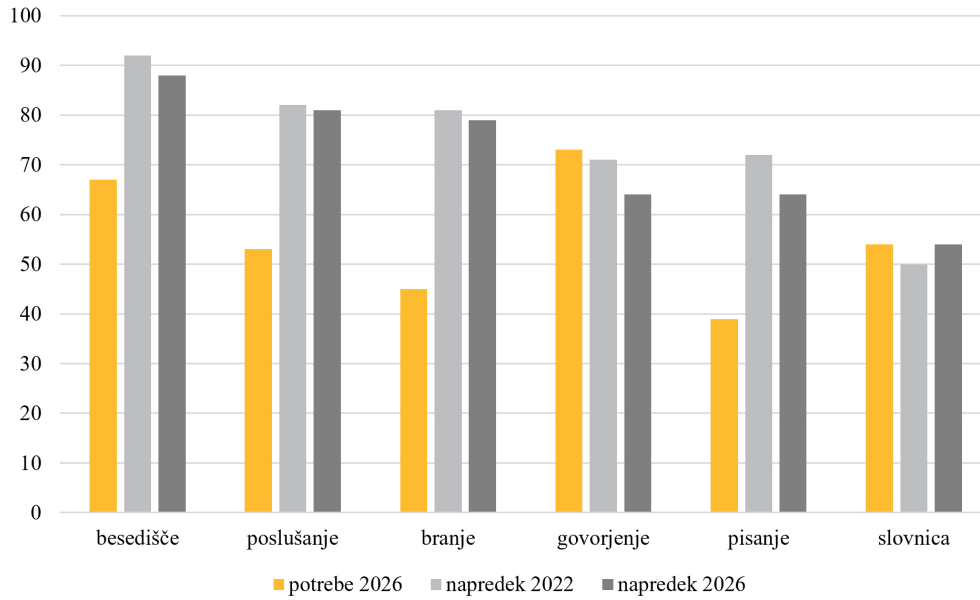
Primerjava obeh anket kaže, da je izhodiščna raven znanja slovenščine uporabnikov spletnega tečaja Slonline glede na samooceno zelo nizka, saj v obeh obdobjih velika večina (92 %) ni imela predhodnega znanja oz. je bilo to minimalno (poznali so le nekaj osnovnih besed, besednih zvez oz. stavkov ali pa so slovenščino nekaj razumeli, vendar niso nič govorili) (gl. Graf 4). Vsem tem uporabnikom, začetnikom brez znanja slovenščine oz. z minimalnim predhodnim znanjem, je tečaj tudi namenjen.



Graf 4: Predznanje slovenščine glede na samooceno anketiranih v % (2022: n = 301; 2026: n = 226).

V anketi iz leta 2026[17] se je pokazalo, da želijo vprašani izboljšati govorjenje v slovenščini in povečati obseg besedišča (to je navedlo približno 70 % vprašanih). Izrazili so visoko stopnjo strinjanja, da so po samooceni najbolj napredovali na področju besedišča, branja in poslušanja (gl. Graf 5),[18] obakrat so bile ocene od 4 navzgor,[19] najmanj pa pri slovnici (ocenjeno s 3,3 leta 2022 oz. 3,4 leta 2026), kar je primerljivo tudi z drugimi raziskavami spletnih tečajev (Jiang idr. 2024; Poveda-Balbuena 2025; Smith idr. 2024). Nizka samoocena napredka na področju slovnice v drugi anketi, ko je bil slovnični modul SLO 3 že objavljen, je sicer posledica tega, da so to ocenjevali tudi tisti, ki modula SLO 3 niso predelali, medtem ko so se tisti, ki so ga (skoraj v celoti), s trditvijo, da so napredovali v slovnici, strinjali v več kot 90 %.

17 Vprašalnik iz leta 2022 vprašanja o potrebah ni vseboval.
18 Upoštevali smo samo odgovore *strinjam se* in *popolnoma se strinjam* na petstopenjski lestvici.
19 Vse ocene so preračunane glede na petstopenjsko lestvico, pri kateri 1 pomeni *sploh se ne strinjam*, 5 pa *popolnoma se strinjam*.



Graf 5: Potrebe in napredek v znanju slovenščine po samooceni uporabnikov v % (potrebe 2026: n = 200; napredek 2022: n = 245; napredek 2026: n = 166).

3.2.4 Zadovoljstvo s tečajem in nadaljnje učenje

Uporabnike smo v obeh anketah spraševali, katere vsebine bi še potrebovali v naslednjih stopnjah tečaja. Slovnica in slovnične vaje, kar so respondenti najbolj pogrešali leta 2022 (74 %), so bile dodane kot tretji modul leta 2023, v zadnji anketi pa si respondenti že želijo nadgradnjo slovničnih vsebin. V obeh obdobjih so uporabniki želeli nadaljevanje tečaja na višjih ravneh, razširitev tem na še več vsakdanjih situacij (npr. na bencinski črpalki, hišni ljubljenčki) in tudi na specializirana področja (npr. poslovna terminologija), na slovensko zgodovino, sociokulturne informacije ipd. Predlagali so vključevanje avdio in vizualnega gradiva, še več dialogov, daljša besedila, več nalog za pisanje, več interaktivnih iger, možnosti komunikacije prek spleta z naravnim govorcem slovenščine, klepetalnico oz. forum uporabnikov, leta 2026 tudi vključitev umetne inteligence, še posebej za konverzacijo v ciljnem jeziku.

Velika večina anketiranih (87 % leta 2022 oz. 90 % leta 2026) je odgovorila, da je tečaj Slonline pripomogel k njihovemu napredku v znanju slovenščine[20] in bi z njim svoje učenje nadaljevala (98 % leta 2022 oz. 93 % leta 2026). Kot razloge za negativen odgovor so posamezni anketirani navedli npr. zanje neustrezen format tečaja, prenizko raven tečaja, pomanjkanje časa, odhod iz Slovenije, učenje z učiteljem in tehnične težave.

20 Izbrali so odgovor *strinjam se* oz. *popolnoma se strinjam*.

4 Zaključek

Analiza registracijskih podatkov, spletne analitike in obeh anket med uporabniki potrjuje, da se je Slonline uveljavil kot prepoznavno in med uporabniki dobro sprejeto spletno okolje za samostojno učenje slovenščine kot neprvega jezika. Uporabniki prihajajo iz 200 držav, do tečaja pa najpogosteje dostopajo iz Slovenije. To potrjuje, da je Slonline kljub različnim možnostim učenja slovenščine, ki jih imajo tisti, ki v Sloveniji že živijo, delajo in/ali študirajo, pomembna oblika učenja zanje, saj jim omogoča brezplačno, časovno in prostorsko fleksibilno učenje, v njihovem tempu.

Podatki o rabi vsebin kažejo, da uporabniki najpogosteje dostopajo do začetnih tem modula SLO 1, medtem ko se pri nadaljevalnih modulih oblikuje manjša, a vztrajnejša skupina uporabnikov. Upad obiska pri zahtevnejših vsebinah opozarja na težavo ohranjanja motivacije, kar je značilno za samostojno spletno učenje jezikov. Zato bi bilo smiselno natančneje spremljati, na katerih mestih uporabniki učenje opuščajo in kateri dejavniki vplivajo na to. Kvantitativne podatke spletne analitike bo nujno dopolnjevati z anketami, intervjuji in sistematičnim spremljanjem uporabniške izkušnje.

Rezultati raziskave so pokazali, da tečaj odgovarja na potrebe uporabnikov, saj so z njim zelo zadovoljni in želijo nadaljevati z učenjem. Svoj napredek v znanju slovenščine so ocenili zelo visoko, se pa zavedamo, da bodo potrebne nadaljnje raziskave, s katerimi bi merili dejansko učinkovitost tečaja. Po samooceni uporabnikov Slonline učinkovito podpira predvsem usvajanje besedišča ter bralno in slušno razumevanje, v nekoliko manjši meri pa omogoča sistematičen napredek pri govorjenju in pisanju, kar je splošna značilnost spletnega učenja tujih jezikov. Pomembna izhodišča za nadaljnji razvoj so zato tudi predlogi uporabnikov, npr. možnost komunikacije z naravnimi govorci, uporabniški forum, še več govornih in pisnih dejavnosti, avtentične vsebine ipd. Ob tem se odpira tudi vprašanje smiselnega vključevanja umetne inteligence in jezikovnih orodij, ki bi lahko omogočali simulacijo pogovorov v slovenščini, pripravo dodatnih govornih in pisnih vaj ter zagotavljanje sprotnih povratnih informacij pri pisanju. Ker več kot polovica uporabnikov do tečaja dostopa prek pametnih telefonov, bo treba zaradi hitrega razvoja mobilnih tehnologij tudi v prihodnosti vsebine in načine učenja nenehno prilagajati ter slediti sodobnim trendom mobilnega učenja.

Ugotovitve raziskave in vprašanja, ki jih odpira, bodo pomembno usmerjali nadaljnji razvoj tečaja, ki pa bo odvisen tudi oz. predvsem od pridobitve ustreznih finančnih sredstev.

Viri in literatura

Alič, T., Gradišar, B., Jerman, T. in Kete Matičič, J. (2017). *Slovene Learning Online: SLO 1*. Center za slovenščino kot drugi in tuji jezik. https://www.slonline.si/

Alič, T., Gradišar, B., Jerman, T. in Kete Matičič, J. (2019). *Slovene Learning Online: SLO 2*. Center za slovenščino kot drugi in tuji jezik. https://www.slonline.si/

Alič, T. in Jerman, T. (2022). Slovenščina kjerkoli in kadarkoli: Spletni tečaj za samostojno učenje slovenščine Slovene Learning Online. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: Slovenščina kot drugi in tuji jezik* (Obdobja 41, str. 11–22). Znanstvena založba Filozofske fakultete Univerze v Ljubljani. https://doi.org/10.4312/Obdobja.41.11-22

Alič, T., Jerman, T., Klemen, M. in Knez, M. (2023). *Slovene Learning Online: SLO 3*. Center za slovenščino kot drugi in tuji jezik. https://www.slonline.si/

Alič, T. in Kern Andoljšek, D. (2025). Interaktivna učna gradiva za (samo)učenje slovenščine kot neprvega jezika za otroke, najstnike in odrasle. V J. Jožef-Beg in D. Divjak Race (ur.), *Slavistično društvo Slovenije: Slovenistika in slavistika 90 let kasneje* (str. 327–337). Zveza društev Slavistično društvo Slovenije.

Brown Nielson, K. (2022). Language learning in online environments. V N. Ziegler in M. González-Lloret (ur.), *The Routledge handbook of second language acquisition and technology* (str. 233–245). Routledge. https://doi.org/10.4324/9781351117586

Center za slovenščino kot drugi in tuji jezik. (b. d.-a). *Letna poročila*. https://centerslo.si/knjige/letna-porocila/ (dostop 1. 5. 2026)

Center za slovenščino kot drugi in tuji jezik. (b. d.-b). *Naslon: Nadgradnja spletnega tečaja Slonline – mobilno in spletno učenje slovenščine za priseljence*. https://centerslo.si/naslon/ (dostop 15. 5. 2026)

Ferbežar, I. in Huber, D. (2024). Delati s slovenščino: Delovno mesto kot prostor učenja. *Slovenščina 2.0*, *12*(2), 269–293. https://doi.org/10.4312/slo2.0.2024.2.269-293

Floyd. (2026). *Language learning apps statistics and user data 2026*. LingoBright. https://www.lingobright.com/statistics/language-learning-apps/ (dostop 15. 5. 2026)

Google Analytics. (b. d.). https://analytics.google.com (dostop 1. 5. 2026)

Jiang, X., Hopman, E., Reuveni, B. in Kittredge, A. (2024). *Duolingo Path meets expectations for proficiency outcomes*. [Duolingo Research Report DRR2404]. Duolingo. https://duolingo-papers.s3.amazonaws.com/reports/Duolingo_whitepaper_language_read_listen_write_speak_2024.pdf

Klemen, M., Arhar Holdt, Š. in Pollak, S. (2022). *Core vocabulary for Slovenian as L2 1.0.* Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1697

Language learning app market (2024). Market.us. https://market.us/report/language-learning-app-market/ (dostop 15. 5. 2026)

Munday, P. (2016). The case for using DUOLINGO as part of the language classroom experience. *Revista Iberoamericana de Educación a Distancia*, *19*(1), 83–101. https://doi.org/10.5944/ried.19.1.14581

Poveda-Balbuena, M. L. (2025). The digital language learning platform Duolingo as an effective and motivational tool at initial stages of face-to-face instruction. V M. Banaś in G. Wlaźlak (ur.), *Language, Technology, Humanities in Society 5.0* (str. 219–232). V & R Unipress. https://doi.org/10.14220/9783737017633.219

Retelj, A. (2015). Pouk tujih jezikov v luči e-gradiv in e-storitev: od poučevanja do učenja tujega jezika z IKT. *Vzgoja in izobraževanje*, *46*(2–3), 73–78.

Smith, B., Jiang, X., in Peters, R. (2024). The effectiveness of Duolingo in developing receptive and productive language knowledge and proficiency. *Language Learning & Technology*, *28*(1), 1–26. https://hdl.handle.net/10125/73595

Jezikovni viri na repozitoriju CLARIN.SI za slovenistične raziskave

DOI: 10.4312/Obdobja.45.21-30

Špela Arhar Holdt, Fakulteta za računalništvo in informatiko in Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; arharhs@ff.uni-lj.si

Kaja Dobrovoljc Zor, Institut »Jožef Stefan«, Ljubljana; Fakulteta za računalništvo in informatiko in Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; kaja.dobrovoljc@ijs.si

Nives Hüll, Institut »Jožef Stefan«, Ljubljana; nives.hull@ijs.si

Cyprian Laskowski, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; cyp@cjvt.si

Tomaž Erjavec, Institut »Jožef Stefan«, Ljubljana; tomaz.erjavec@ijs.si

Infrastruktura CLARIN.SI vključuje repozitorij za dolgoročno hrambo jezikovnih virov, spletne konkordančnike, ki omogočajo dostop do jezikoslovno označenih korpusov, ter druga orodja in storitve. Konkordančniki so v slovenistični praksi razmeroma dobro poznani, medtem ko repozitorij ostaja slabo izkoriščen, čeprav vsebuje raznolike bogato označene jezikovne podatke in podatkovne baze. V prispevku sistematično analiziramo in popišemo repozitorijske vnose, relevantne za področje slovenističnih teoretičnih in aplikativnih raziskav. Prispevek s tem ponuja novo izhodišče za izbiro podatkovnega zaledja v jezikoslovnih študijah in prispeva k učinkovitejši rabi obstoječih jezikovnih virov za slovenščino.

CLARIN.SI, jezikovni viri, repozitorij, slovenščina, slovenistika

The CLARIN.SI infrastructure comprises a repository for the long-term preservation of language resources, online concordancers that provide access to linguistically annotated corpora, and other tools and services. The concordancers are relatively well known in Slovenian studies, whereas the repository remains underused, despite containing diverse, richly annotated language data and databases. This article systematically analyzes and catalogues repository entries relevant to theoretical and applied Slovenian studies. It offers a new starting point for selecting data for linguistic studies and contributes to the more effective use of existing Slovenian language resources.

CLARIN.SI, language resources, repository, Slovenian, Slovenian studies

1 Uvod

CLARIN.SI (Erjavec idr. 2022, 2025) je slovenski nacionalni konzorcij v mreži evropske raziskovalne infrastrukture CLARIN (Common Language Resources and Technology Infrastructure), ki zagotavlja digitalne jezikovne vire, orodja in storitve raziskovalcem na področju humanistike in družboslovja ter drugih področij, ki se ukvarjajo z jezikom (de Jong idr. 2018). Osnovna storitev CLARIN.SI je vzdrževanje

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

repozitorija jezikovnih raziskovalnih podatkov oz. jezikovnih virov, kot so besedilni korpusi, strojno berljivi leksikoni in slovarji ter računalniška orodja, vključno z modeli za orodja. Repozitorij v skladu s strategijo CLARIN ERIC implementira načela FAIR (po katerih morajo biti raziskovalni podatki najdljivi, dostopni, interoperabilni in ponovno uporabni)[1] in je, poleg Arhiva družboslovnih podatkov ADP, edini v Sloveniji akreditiran kot zaupanja vreden podatkovni repozitorij. Vnose repozitorija ustvarjajo avtorji jezikovnih virov, pri čemer morajo metapodatke in podatke oblikovati skladno s smernicami. Deponirani korpusi so tipično integrirani v spletne konkordančnike, predvsem noSketch Engine (Kilgarriff idr. 2014). Na konkordančnikih je trenutno dostopnih prek 300 korpusov, na repozitoriju pa prek 800 jezikovnih virov.

Za razliko od konkordančnikov ostaja repozitorij med jezikoslovci slabše poznan in obiskan. Na razloge za takšno stanje so opozorili sodelujoči v uporabniški anketi o izkušnjah s CLARIN.SI (Arhar Holdt idr. 2025), kjer sta bili enostavnost iskanja in jasnost metapodatkov med nižje ocenjenimi lastnostmi repozitorija. V pričujočem prispevku sistematično predstavljamo repozitorijske vnose, relevantne za področje slovenističnih teoretičnih in aplikativnih raziskav, na podlagi česar jih bo v repozitoriju CLARIN.SI lažje najti ter vključiti v bodoče raziskave in razvoj.

2 Metodologija

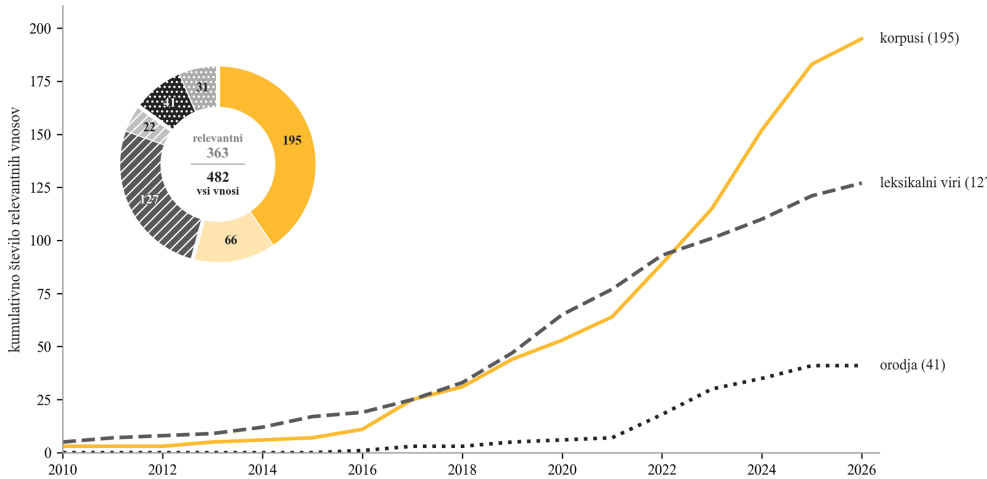
V prvem koraku metodologije smo izvozili metapodatke za vse repozitorijske vnose. Vključili smo: stalni identifikator »handle«, osnovni tip vira, jezik(e) vira, datum izida, naslov, avtorje, založnika, licenco, ključne besede in opis vira. Vsi metapodatki na repozitoriju so v angleščini.

Za analizo smo se omejili na vire, ki vključujejo (tudi) slovenščino (482 enot) in od teh le na najsodobnejše različice[2] (363 enot), pri čemer pri spremljevalnih korpusih, pri katerih je vsaka izdaja samostojen vnos, upoštevamo vse mesečne izdaje (Graf 1). V končnem naboru ima 195 virov pripisan tip *corpus* (v nadaljevanju: korpus, poglavje 3.1), 127 *lexicalConceptualResource* (v nadaljevanju: leksikalni vir, poglavje 3.2), in 41 *toolService* (v nadaljevanju: orodje, poglavje 3.3). Razmerje med tipi virov in časovno rast števila vnosov prikazuje Graf 1.

Glede na razpoložljive informacije v metapodatkih smo enotam ročno pripisali dodatne informacije: pri vseh enotah podtip vira; pri leksikalnih virih še način priprave (*ročno*, *strojno* ali *ročno in strojno*), zajete informacije (npr. *n-grami*, *pogostost*) in izhodiščni vir (npr. *Gigafida 2.0*); pri korpusih pa način pridobivanja, način označevanja (*ročno*, *strojno* ali *ročno in strojno*), nivoje označevanja (npr. *sentiment*) in tip besedila (npr. *parlamentarno*, *publicistično* itn.). Tabela z metapodatki in ročnimi oznakami je na voljo za nadaljnjo uporabo (Arhar Holdt idr. 2026).

1 Več o načelih je na strani https://odprta-knjiznica.si/storitve/raziskovalni-podatki/nacela-fair/.

2 Nekateri viri so na repozitoriju CLARIN.SI na voljo v več različicah, kar je posledica njihovega nadgrajevanja, ali pa se objavljajo po delih v rednih časovnih intervalih, kot npr. velja za spremljevalni korpus Trendi.



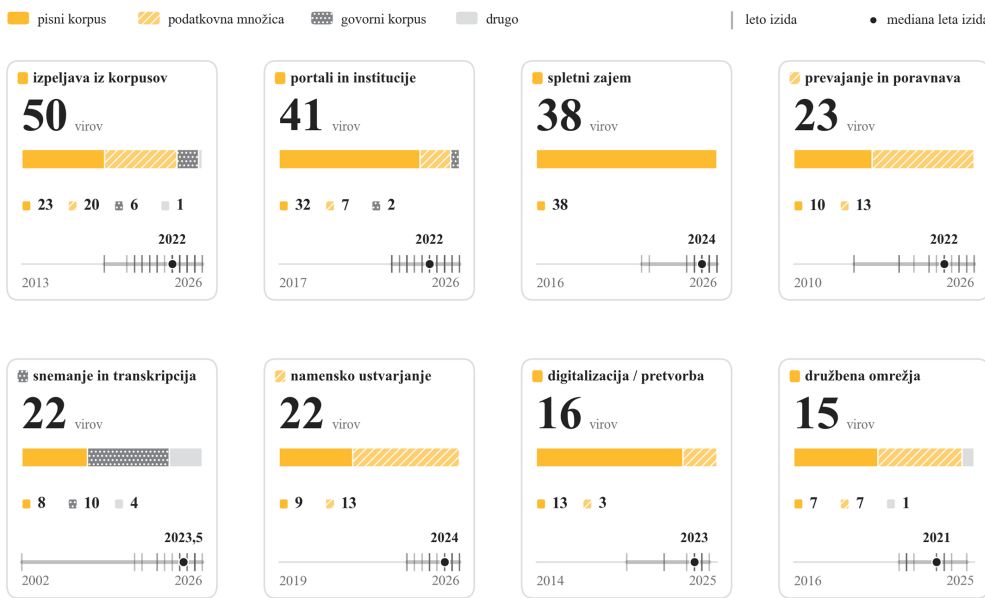
Graf 1: Tip vira in časovna rast števila vnosov za slovenščino. V kolobarnem prikazu temnejši deli predstavljajo najsodobnejše v analizo vključene različice virov, svetlejši pa starejše.

3 Rezultati

3.1 Korpusi

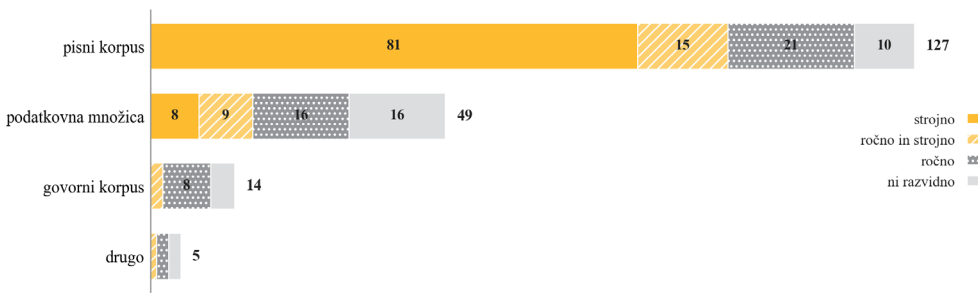
Med 195 korpusnimi vnosi prevladujejo **pisni korpusi**[3] (127 enot), sledijo **podatkovne množice**, npr. za razvoj, učenje in vrednotenje jezikovnih tehnologij (49), in **govorni korpusi** (14), **zvočne, transkripcijske, multimodalne in metapodatkovne zbirke** pa so prisotne redkeje (skupaj 5 enot). Za nekatere korpuse je na voljo tudi podatek o načinu pridobivanja gradiva, kar prikazuje Graf 2 (posamezni vir je lahko uvrščen v več skupin, zato vsota po skupinah presega skupno število virov). V posameznem okvirju je navedeno število virov v skupini, barve kažejo njihovo vrsto, časovnica pa razporeditev vnosov po letih izida. Pika označuje mediano leta izida: polovica vnosov ima zgodnejšo ali enako, polovica pa poznejšo ali enako letnico; pri vrednosti 2023,5 sta srednji letnici 2023 in 2024.

3 Oznaka *pisni korpus* v tem prispevku se nanaša na obliko gradiva, ne nujno na njegovo izvorno modalnost. Tako so npr. parlamentarni zapisi lahko uvrščeni med pisne korpuse, čeprav izvirajo iz govorjenih razprav.

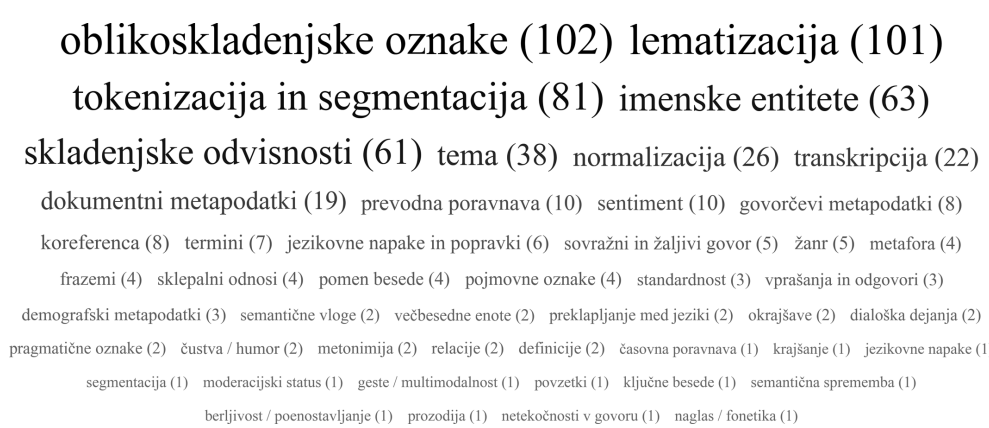


Graf 2: Skupine načinov pridobivanja korpusov glede na skupno število, vrsto podatkov in leta izida.

Označenost je tesno povezana z vrsto podatkov, kar prikazuje Graf 3. Pogostost in vrsto posameznih ravni oznak pa prikazuje Slika 1.



Graf 3: Način označevanja korpusov po vrsti podatkov.



Slika 1: Ravni oznak v korpusih.

Glede na vsebino izstopajo **publicistična besedila** (44 enot) ter **spletna besedila in spletna komunikacija** (30 enot);[4] mednje sodijo mesečne izdaje spremljevalnega korpusa Trendi, zbirke za analizo sentimenta, korpus uporabniško generiranih spletnih vsebin Janes, viri za raziskave sovražnega ali žaljivega govora (npr. IMSyPP-sl, FRENK) ter splošnejši spletni korpusi, kot so MaCoCu, CLASSLA-web in žanrsko označeni GINCO. Nekateri od teh virov vsebujejo specializirane ravni oznak, kot so normalizacija, (ne)standardnost, sentiment ter sovražni in žaljivi govor. Drugo večjo skupino predstavljajo **formalna in vsebinsko specializirana gradiva** (22 enot), predvsem zbirka KAS s podkorpusi akademskih besedil ter izpeljanimi množicami za povzemanje, strojno prevajanje in terminološko obdelavo, dopolnjujejo pa jo strokovni viri z različnih področij, npr. informatike, jezikoslovja, medicine oz. farmacije in turizma.

Soroden sklop tvorijo **parlamentarna in pravno-upravna besedila** (15 enot): parlamentarni korpusi, npr. SlovParl, siParl, ParlaMint, ParlaMeter, ter pravne in upravne zbirke, kot so DGT-UD, SI-IUS in COLESLAW, ki poleg temeljnih jezikoslovnih ravni pogosto vključujejo tudi metapodatke, uporabne za raziskave političnega in institucionalnega diskurza. **Šolska in učna gradiva** (12 enot) vključujejo korpus Šolar, učbeniška korpusa ccUčbeniki in KUUS, korpus slovenščine kot tujega jezika KOST ter pilotni korpus študentskih akademskih besedil KOŠ; nekateri vsebujejo tudi oznake jezikovnih napak in popravkov.

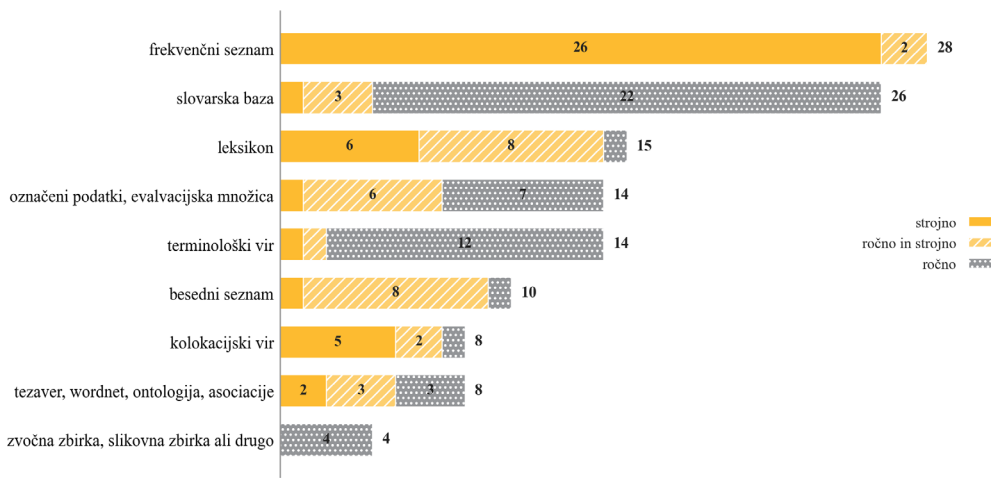
Govorni viri (17 enot), med njimi referenčni korpus Gos, zbirke za razpoznavanje govora (ARTUR, SOFES) in drugi govorni korpusi (EVA, GORDAN, ROG, Berta, G-KOMET), povezujejo raziskave govorjene slovenščine z razvojem govornih tehnologij; posamezni korpusi so pogosto razdeljeni na ločene vnose za zvočni in transkripcijski del. Značilne ravni oznak vključujejo transkripcijo, normalizacijo, dialoška dejanja in govorčeve metapodatke. Manjši sklop tvorijo **narečna in sociolingvistična gradiva** (4 enote), namenjena proučevanju regionalne in družbene jezikovne raznolikosti.

4 En vir je uvrščen v obe skupini, skupaj gre torej za 73 enot.

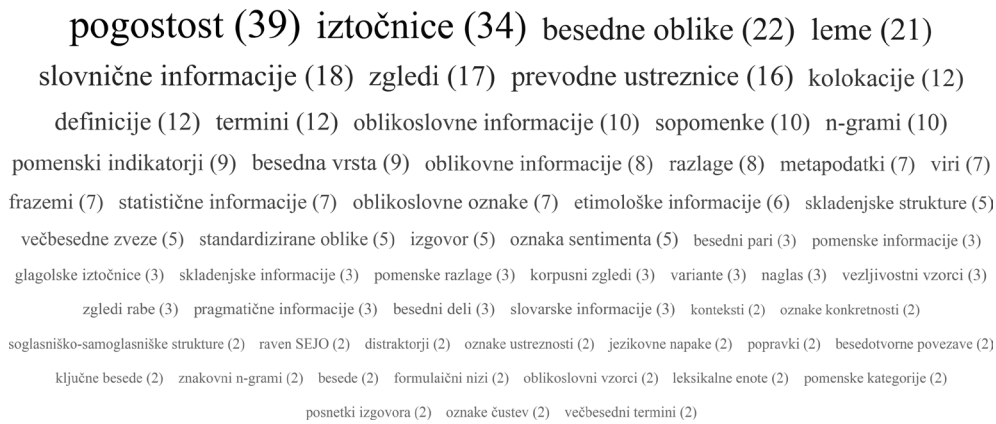
Manjše skupine predstavljajo **leposlovna besedila** (6 enot, npr. Maj68, PriLit, KDSP), **zgodovinski korpusi** (10 enot, npr. goo300k, IMP, CVET in sPeriodika; med njimi tudi verska in epigrafska gradiva) ter **folklorno-humoristična gradiva** (3 enote: Pregovori, Uganke, Šale24). Zgodovinski viri združujejo starejša besedila, prepise, normalizirane oblike oz. gradiva, ki omogočajo opazovanje jezikovne rabe v daljšem časovnem razponu; označeni so pretežno ročno, značilna pa je normalizacija med zgodovinskimi in sodobnimi oblikami. **Večžanrske zbirke** (npr. Gigafida, SUK, metaFida) zajemajo širok nabor sodobne pisne slovenščine. Poseben segment (21 enot) tvorijo **podatkovne množice vprašanj in odgovorov**, namenjene razvoju sistemov za jezikovno razumevanje, sklepanje, razdvoumljanje in prilagajanje velikih jezikovnih modelov, npr. SI-NLI, slovenski prevodi oz. razširitve zbirk SuperGLUE in SQuAD, SloATOMIC, SloWiC, GaMS-Instruct in SLO-VLM-IT-Dataset.

3.2 Leksikalni viri

Porazdelitev leksikalnih virov po vrsti in načinu priprave prikazuje Graf 4, pogostejše vrste zajetih informacij pa Slika 2.



Graf 4: Način priprave po vrsti leksikalnega vira.



Slika 2: Zajete informacije v leksikalnih virih (vsaj 2 pojavitvi).

Med leksikalnimi viri prednjačijo **frekvenčni seznami** (28 enot). Za Gigafido in GOS so na voljo izvozi znakovnih in besednih n-gramov, soglasniško-samoglasniških struktur, besednih delov in lem, za korpus Nova beseda pa seznam besednih oblik. Iz korpusa Trendi izvirajo seznami lem in n-gramov po letih, namenjeni spremljanju razvoja besedišča v medijih; iz korpusov Šolar in Učbeniki so na voljo seznami lem in skladenjskih struktur, pa tudi pogostih jezikovnih težav šolskega pisanja ter učbeniških ključnih besed in n-gramov. N-grami so na voljo tudi za korpuse IMP, Janes in Kres. Sklop dopolnjuje seznam deleksikaliziranih skladenjskih dreves iz korpusov SSJ in SST za primerjavo skladenjskih vzorcev v pisni in govorjeni slovenščini.

Sorodna kategorija so **besedni seznami** (10 enot), ki lahko vsebujejo tudi frekvenčne informacije, vendar so bolj namensko zasnovani in pogosto vključujejo ročno dodane ali pregledane podatke: referenčni seznam pogostih splošnih besed, jedrno besedišče za slovenščino kot drugi in tuji jezik, seznama iztočnic zbirke Besede slovenskega jezika in portala Franček, seznam kandidatov za nestandardno oz. govorjeno besedišče projekta Mezzanine, seznam moško-ženskih parov za poimenovanja poklicev, seznama formulaičnih nizov v pisni in govorjeni slovenščini ter iz Sloleksa pridobljena seznama besedotvorno povezanih parov in naglašenih besednih končnic.

Na repozitoriju je na voljo vrsta **slovarskih baz** (26 enot). Splošne enojezične slovarske baze so npr. Leksikalna baza za slovenščino, eSSKJ, Slovenski pravopis (slovar), Slovar novejšega besedja slovenskega jezika, Odzadnji slovar slovenskega jezika, Besedišče slovenskega jezika ter Sprotni slovar slovenskega jezika. Poseben sklop tvorijo specializirani slovarji, kot so slovarji člankov, frazemov, tviterščine in baza kolaborativnega slovarja Razvezani jezik, ter zgodovinski viri, kot so Pleteršnikov slovensko-nemški slovar, Besedje slovenskega knjižnega jezika 16. stoletja, slovar nemških izposojenk v Arcticae Horulae, Slovar jezika Janeza Svetokriškega. Šolski rabi je namenjenih več baz, izvirajočih s portala Franček, na voljo pa je tudi leksikalna baza besedišča ravni A1 po SEJO. Med večjezičnimi viri je najti Veliki slovensko-madžarski slovar, japonsko-slovenski ter slovensko-japonski učni slovar ter bazo WeSoSlaV, ki prinaša raznolike leksikalne informacije za glagole v različnih slovanskih jezikih.

Kot posebno kategorijo slovarskih baz smo ločili **terminološke vire** (14 enot), ki obsegajo večje terminološke slovarje, kot so geografski, papirniški in davčni terminološki slovar, pa tudi ožje glosarje, npr. s področij šaha, akademske integritete in konjeništva. Sklop dopolnjujejo večjezične oz. podatkovno bogate baze, kot sta TermFrame za krasoslovje in KAS-biterm iz znanstvenih besedil, ter terminološko-narečna baza oblačilnega izrazja ziljskega narečja v Kanalski dolini. Ločeni so tudi **kolokacijski viri** (8 enot): obsežni Kolokacijski slovar sodobne slovenščine, manjša baza kolokacij iz eSSKJ in izvozi kolokacij iz Gigafide 2.1, Šolarja 3.0 in Učbenikov 1.0. Sklop dopolnjujejo slMWELex, strojno zgrajena zbirka večbesednega besedišča iz korpusa Kres, Aleks, kolokacijska baza iz slovenskih znanstvenih besedil, ter zbirka ročno anotiranih kolokacijskih kandidatov iz korpusa Gigafida 2.0.

Med **leksikoni** (15 enot) prevladujejo viri z oblikoslovnimi informacijami: referenčni Sloleks 3.1, leksikona z lemami iz korpusa Beseda in s portala Franček ter večjezični leksikoni MULTEXT-East. Tem se pridružujejo leksikon večbesednih izrazov iz Gigafide 2.1, vezljivostni leksikon iz Gigafide 2.1 in leksikon večbesednih izrazov iz različnih terminoloških virov. Poseben sklop tvorijo leksikoni z informacijami o sentimentu in čustvih, kot so JOB, KSS, LiLaH in SloEmoLex. V skupini sta še leksikon z oznakami konkretnosti in predstavljivosti pojmov MEGA.HR-Crossling in leksikon zgodovinske slovenščine s historičnimi oblikami in sodobnimi ustreznicami imp25k.

Pri virih tipa **tezaver, wordnet, ontologija in asociacije** (8 enot) gre za pomensko organizirane leksikalne vire. Tu sta denimo dve različici wordneta za slovenščino, sloWNet 3.1 z dodatkom sloWNet-USAS in Open Slovene WordNet, in obsežna baza Slovarja sopomenk sodobne slovenščine. Na voljo sta tudi ontologiji SLONEST-noun in bolj namensko didaktično usmerjena ontologija ONTEM, sklop pa dopolnjuje zbirka asociacij SWOW-SL, ki jih je posredovala jezikovna skupnost.

Pri **označenih podatkih in evalvacijskih množicah** (14 enot) gre za vire, namenjene predvsem učenju, preverjanju in primerjavi jezikovnotehnoloških postopkov za različne naloge. Med njimi so množice CoSimLex, SimLex-999-sl in Multilingual Culture-Independent Word Analogy Datasets za vrednotenje leksikalne podobnosti oz. besednih vložitev, podatki za zaznavanje sopomenk in protipomenk v kontekstu, SYNDIST s trojčki iztočnica – sopomenka – distraktor, COLLDIST s trojčki iztočnica – kolokator – distraktor, KAS-term z označenimi kandidati za termine, DASSLE s primeri tipičnih jezikovnih napak in popravkov, ILS z označenimi izgovornimi variantami predsamoglasniškega *l*, ArboSloleks z besedotvornimi drevesi, podatki iz Sloleksa 2.0 z oznakami oblikoslovnih vzorcev ter KOBOS z označeno morfemsko in besedotvorno segmentacijo.

Med redkeje zastopanimi kategorijami so **zvočne, slikovne in druge zbirke**, ki vključujejo zbirko opisov starejših slovenskih slovnic in pravopisov, posnetke izgovora za Šolski slovar slovenskega jezika ter zvočno in fotografsko gradivo za slovar oblačilnega izrazja ziljskega narečja.

3.3 Orodja

Med 41 orodji so najpogostejši **modeli za strojno slovnično označevanje slovenskih besedil** (18 enot), kot so modeli orodij CLASSLA-Stanza in PyTorch, ki se uporabljajo za samodejno pripisovanje lem, besednih vrst in drugih oblikoslovnih lastnosti besedil, skladenjsko razčlenjevanje ter prepoznavanje imenskih entitet in koreferenčnih povezav v besedilu. Sledijo **splošni jezikovni modeli oz. vektorske vložitve za slovenščino** (5 enot), kot je model SloBERTa, različni **modeli za klasifikacijo besedil** (4 enote), denimo z vidika žanra in vsebovanih tem, **modeli za razpoznavanje oz. sintezo govora** (4 enote) in tudi več modelov za specializirane naloge, kot so strojno prevajanje, povzemanje, odgovarjanje na vprašanja in zdravorazumsko sklepanje v slovenščini. Poleg modelov na repozitoriju najdemo tudi različna računalniška orodja za jezikoslovno analizo besedil, spletno pajkanje besedil in evalvacijo avtomatskih razpoznavalnikov govora.

Na repozitoriju so na voljo tudi nekatere druge tehnologije, ki zaradi svoje univerzalnosti jezika v opisu ne specificirajo in zato v analizo niso bile zajete, kot so denimo orodja Q-CAT (ročno označevanje besedil), STARK (leksikalna in skladenjska analiza besedil), ColTrend in ColEmbed (analiza kolokacij v okolju Orange) ter spletna platforma za evalvacijo jezikovnih tehnologij za slovenščino SloBENCH. Tovrstna orodja se sicer tipično hranijo in diseminirajo tudi na specializiranih razvojnih platformah, za slovenske tehnologije npr. na repozitoriju CLARIN.SI na GitHub[5] ter profilih CJVT[6] in CLASSLA[7] na Hugging Face; poskus izčrpnega popisa razpršeno objavljenih modelov, orodij in drugih tehnologij za slovenščino pa je na voljo tudi na spletni strani CLARIN.SI[8] in repozitoriju GitHub.[9]

4 Sklep in nadaljnje delo

V prispevku smo predstavili popis jezikovnih virov (korpusov, leksikalnih virov in orodij), ki so dostopni v repozitoriju CLARIN.SI ter koristni za teoretične in aplikativne slovenistične raziskave. Ker se repozitorij stalno dopolnjuje, je popis začasne narave, vseeno pa nudi vpogled v tipologijo in značilnosti razpoložljivega gradiva. Naše bodoče delo se bo osredotočilo na analizo ključnih besed, ki jih posredujejo avtorji vira, in identifikacijo možnosti za boljše strukturiranje metainformacij, ki opisujejo vsebino in način priprave deponiranih virov. Končni cilj je omogočiti, da bodo uporabniki in uporabnice repozitorija raziskovalne podatke lahko iskali (in jih tudi našli) na hiter, enostaven in pregleden način.

5 https://github.com/clarinsi
6 https://huggingface.co/cjvt
7 https://huggingface.co/classla
8 https://www.clarin.si/info/k-centri/llms4ssh-sredisce-znanja-za-velike-jezikovne-modele-za-druzboslovje-in-humanistiko/#Kljucni_projekti_povezani_z_velikimi_jezikovnimi_modeli
9 https://github.com/clarinsi/Slovenian-Language-Technologies-Overview/

Delo je nastalo s finančno podporo agencije ARIS, pod okriljem raziskovalne infrastrukture CLARIN.SI (I0-E004) ter programov Jezikovni viri in tehnologije za slovenski jezik (P6-0411) in Tehnologije znanja (P2-0103).

Literatura

Arhar Holdt, Š., Dobrovoljc, K., Hüll, N., Laskowski, C. in Erjavec, T. (2026). *CLARIN.SI Repository Resource Metadata (1.0)* [Podatkovna zbirka]. Zenodo. https://doi.org/10.5281/zenodo.20338942

Arhar Holdt, Š., Meden, K. in Erjavec, T. (2025). Users' experience and development priorities for CLARIN.SI. V C. Grisot in T. Kontino (ur.), *CLARIN Annual Conference Proceedings 2025* (str. 53–57). CLARIN ERIC. https://doi.org/10.5281/zenodo.17357825

de Jong, F., Maegaard, B., De Smedt, K., Fišer, D. in Van Uytvanck, D. (2018). CLARIN: Towards FAIR and responsible data science using language resources. V N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis in T. Tokunaga (ur.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (str. 3259–3264). European Language Resources Association (ELRA). https://aclanthology.org/L18-1515/

Erjavec, T., Dobrovoljc, K., Fišer, D., Javoršek, J. J., Krek, S., Kuzman Pungeršek, T., Laskowski, C. A., Ljubešić, N. in Meden, K. (2022). Raziskovalna infrastruktura CLARIN.SI. V D. Fišer in T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika: zbornik konference* (str. 47–54). Inštitut za novejšo zgodovino. https://doi.org/10.5281/zenodo.13953420

Erjavec, T., Ljubešić, N., Meden, K., Kuzman Pungeršek, T., Laskowski, C. A., Javoršek, J. J., Krek, S., Jemec Tomazin, M., Dobrovoljc, K., Arhar Holdt, Š., Lenardič, J. in Fišer, D. (2025). CLARIN.SI, the Slovenian node of CLARIN: Ten years on. V V. Vandeghinste in T. Kontino (ur.), *Selected papers from the CLARIN Annual Conference 2024*. Linköping University Electronic Press. https://doi.org/10.3384/ecp216.07

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. in Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Jezikoslovne in tehnološke rešitve pri razvoju portala jezikovnih iger: obravnava občutljivega besedišča

DOI: 10.4312/Obdobja.45.31-39

Špela Arhar Holdt, Filozofska fakulteta in Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; arharhs@ff.uni-lj.si

Iztok Kosem, Filozofska fakulteta in Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; iztok.kosem@ff.uni-lj.si

Na interaktivnem spletnem portalu CJVT Igre je dostopnih pet slovenskih jezikovnih iger, ki spodbujajo utrjevanje in širjenje jezikovnega in metajezikovnega znanja. Portal temelji na leksikalnih podatkih Digitalne slovarske baze za slovenščino, ki so za vsako posamezno igro še dodatno strojno in jezikoslovno urejeni. Prispevek prinaša prvi celoviti opis portala in posameznih iger, pri čemer se posebej osredotočimo na identifikacijo in obravnavo občutljivega besedišča in drugega za igre neželenega gradiva.

jezikovne igre, Digitalna slovarska baza za slovenščino, občutljivo besedišče, slovenščina

The interactive web portal CJVT Igre offers five Slovenian language games designed to support consolidation and expansion of linguistic and metalinguistic knowledge. The portal draws on lexical data from the Digital Dictionary Database for Slovene; for each game, these data are further processed both automatically and through linguistic curation. This article provides the first comprehensive description of the portal and its games, with particular attention to identifying sensitive vocabulary and other material that is not suitable for the games.

language games, Digital Dictionary Database for Slovene, sensitive vocabulary, Slovenian

1 Uvod

Portal CJVT Igre[1] prinaša pet jezikovnih iger, primernih za različne starostne skupine in nivoje znanja slovenščine. Jezikovne igre podpirajo kognitivni razvoj, usvajanje besedišča in motivacijo za učenje (Bustrillo idr. 2024; Derakhshan in Khatir 2015; Hidayat 2016). Pozitiven vpliv na širjenje besedišča in usvajanje leksikalnega znanja je bil dokazan tudi za igre Hangman (sl. vislice) (Dharmayasa 2022; Munikasari idr. 2021), Wordle (Pamungkas 2021) in Semantic Boggle (Toma idr. 2017), ki so po zasnovi podobne igram na portalu.

Glavni podatkovni vir za CJVT Igre je Digitalna slovarska baza za slovenščino (DSBS; Gantar 2020), ki združuje oblikoslovne, pomenske in statistične informacije o raznolikih leksikalnih enotah, mdr. pomenske indikatorje, slovarske oznake, kolokacije, zglede rabe in semantične tipe, organizirane v ontologijo 21 kategorij, kot so *dejanje*,

1 https://igre.cjvt.si/

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

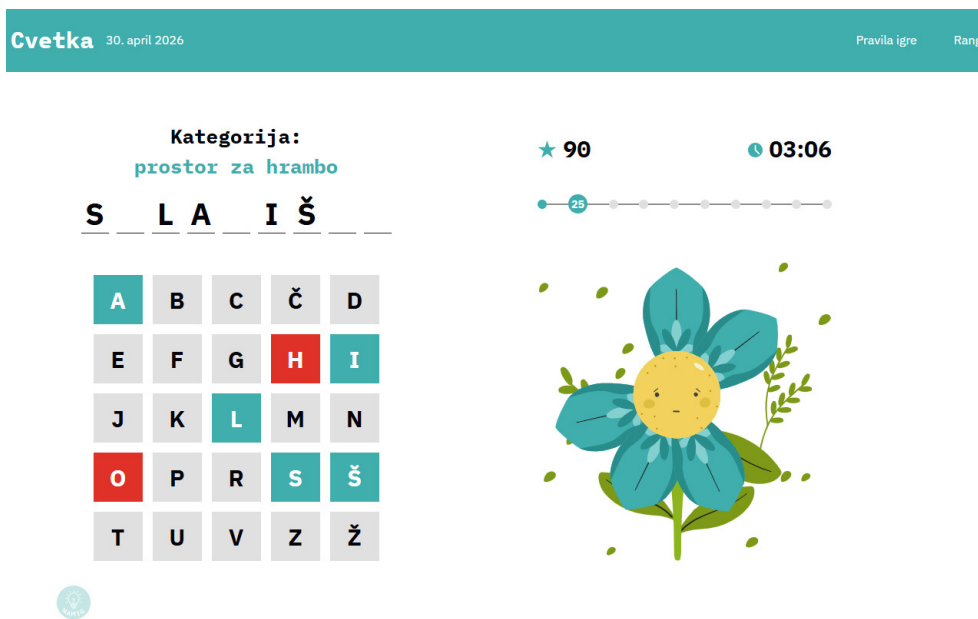
artefakt, *človek* in *lokacija* (Kosem in Pori 2021). DSBS vsebuje slovarske oznake za sovražno in grobo besedišče (Arhar Holdt idr. 2023), ki presegajo okvire slovaropisne zaznamovanosti, leksikalne enote z negativno konotacijo ali občutljivimi vsebinami pa niso izrecno označene. Pri razvoju iger je zato ključen ročni pregled, ki zagotovi družbeno občutljivost in starostno primernost gradiva, o čemer se je do sedaj pisalo zlasti v povezavi z uporabo korpusnih zgledov za izobraževalne namene (npr. Kilgarriff idr. 2014; Pilán idr. 2016).

Prispevek prinaša prvi celoviti opis iger na portalu: za vsako predstavimo osnovno mehaniko in metodologijo priprave podatkov, nato opišemo osnovna načela presojanja občutljivosti in podamo konkretne primere obravnave za različne vrste gradiva. Prispevek sklenemo z možnostmi za nadaljnje delo.

2 Predstavitev iger in njihove priprave

2.1 Cvetka

Cvetka temelji na konceptu znane igre vislice. Igralec mora v omejenem številu poskusov uganiti skrito slovensko besedo ali besedno zvezo, pri čemer mu pomaga tematska kategorija (npr. *prostor za hrambo* na Sliki 1). Mreža na levi strani zaslona kaže, katere črke je igralec že izbral. Ob napačnih izbirah z animirane cvetlice na desni strani zaslona odpadajo cvetni listi. Ko igralec besedo ugane ali ko cvetlica izgubi vseh sedem lističev, se igra premakne na naslednjo od desetih dnevnih nalog.



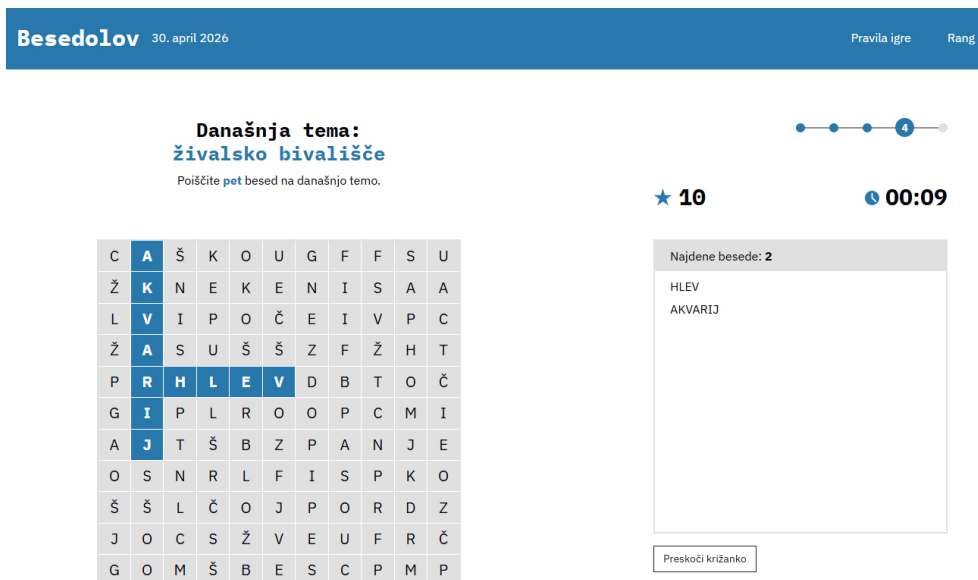
Slika 1: Igra Cvetka – kategorija: *prostor za hrambo*, odgovor: *skladišče.*

Za pripravo igre smo iz DSBS izvozili seznam samostalniških leksikalnih enot, skupaj s semantičnimi tipi, pomenskimi indikatorji in relativnimi frekvencami iz korpusa

Gigafida 2.0 (Krek idr. 2020). Podatke smo razvrstili po semantičnih tipih in indikatorjih ter jih ročno pregledali in uredili (gl. 3.1). Nato smo opredelili parametre zahtevnosti: število različnih samoglasnikov, število različnih črk in relativno frekvenco vsake od leksikalnih enot. Izračunom smo pripisali rang na petstopenjski lestvici in uporabili seštevek rangov za določitev skupne zahtevnosti besede v igri, ki posledično poteka od preprostejših do zahtevnejših primerov. Več o metodologiji je na voljo v Arhar Holdt in Kosem (2025).

2.2 Besedolov

V igri Besedolov mora igralec v črkovni mreži poiskati pomensko povezane slovenske besede. Dnevno je na voljo pet mrež z različnimi tematskimi kategorijami, vsaka vsebuje pet besed, zapisanih v katerikoli smeri. Točke se dodeljujejo glede na število najdenih besed in hitrost reševanja. Igralci lahko besede poiščejo brez pomoči ali pa uporabijo namig, ki razkrije vse ciljne besede v mreži. Ko so vse mreže rešene (možno jih je tudi preskočiti, gl. Sliko 2), se igra zaključi.

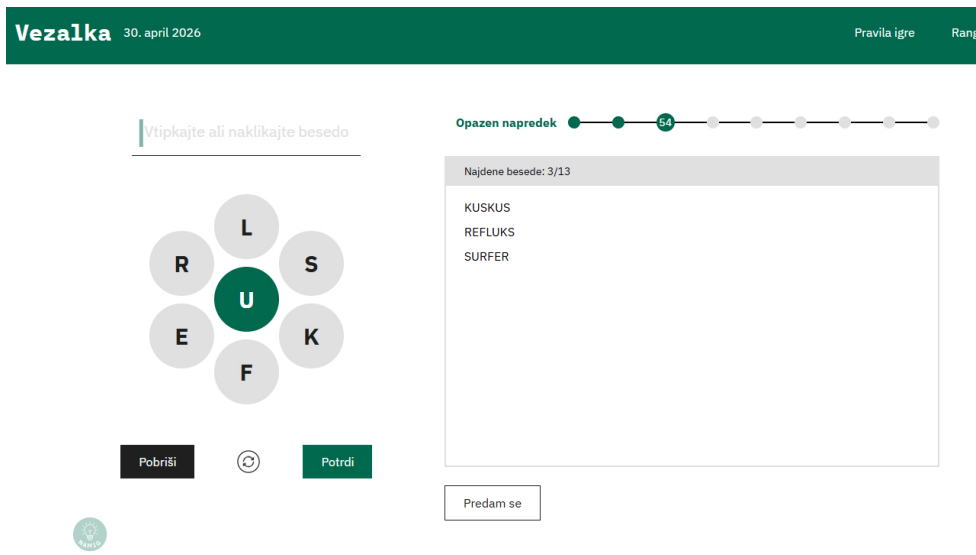


Slika 2: Igra Besedolov – kategorija: *živalsko bivališče*, odgovori: *hlev*, *brlog*, *panj*, *akvarij*, *gnezdo*.

Podatkovno izhodišče za pripravo igre je enako kot pri Cvetki (gl. 2.1), vendar smo gradivo omejili na semantične kategorije, ki vsebujejo vsaj pet visokofrekvenčnih samostalnikov. Črkovne mreže so generirane strojno, pri čemer je težavnost odvisna od tega, v katere smeri so besede skrite. Obravnavo občutljivega besedišča opisuje poglavje 3.1.

2.3 Vezalka

Pri igri Vezalka mora igralec iz danega nabora sedmih črk, med katerimi je osrednja črka obvezna, sestaviti čim več veljavnih slovenskih besed v osnovni obliki. Točke se dodeljujejo glede na dolžino besede, dodatne točke pa prinese beseda, ki uporabi vseh sedem črk. Igralec lahko igro tudi konča z gumbom *Predam se* (Slika 3). Takrat se prikažeta skupni rezultat in seznam besed, ki jih ni našel.

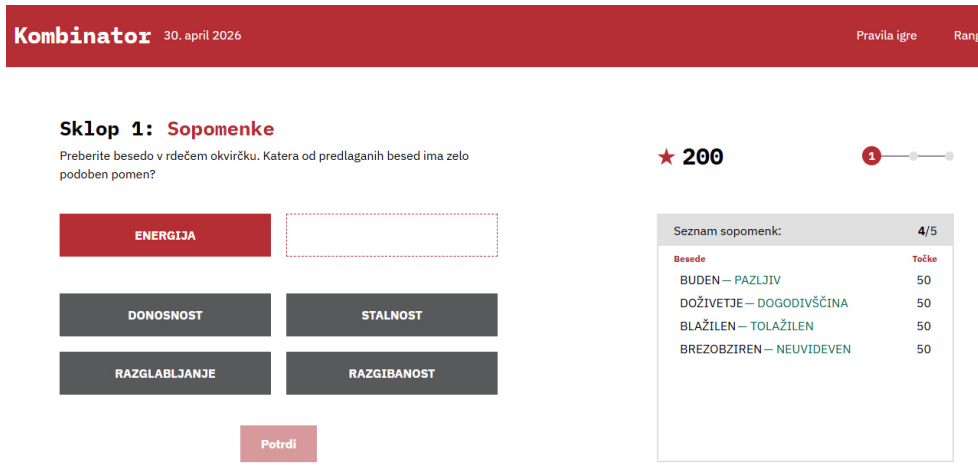


Slika 3: Igra Vezalka – odgovori: *surfer*, *refluks*, *kuskus*, *resurs*, *usek*, *surf*, *luks*, *sufle*, *ulkus*, *frufru*, *fluks*, *sufler*, *lureks*.

Kot začetni podatkovni nabor za igro Vezalka smo iz DSBS izvozili vse besedišče, ki je vključeno v Sloleks 3.0. Iz podatkov smo odstranili lastna imena in iz njih izpeljane pridevnike, enote s številkami, neslovenskimi znaki in rimskimi številkami. Frekvence iz korpusa Gigafida 2.0 smo pretvorili v pet frekvenčnih razredov, pri čemer razred 5 označuje najpogostejše, razred 1 pa najredkejše besede. Na podlagi teh podatkov smo strojno pripravili kandidatne nabore tako, da se naključno izberejo osrednja črka in šest spremljevalnih črk, nato pa najdejo vse ustrezajoče besede, pri čemer mora nabor vsebovati vsaj eno besedo, ki vsebuje vseh sedem črk. Končne nabore besed smo ročno pregledali in po potrebi prilagodili, kot predstavlja poglavje 3.2. Več o metodologiji je na voljo v Arhar Holdt in Kosem (2025).

2.4 Kombinator

Igra Kombinator je sestavljena iz dveh zaporednih korakov: igralec mora najprej za pet besed izbrati ustrezno sopomenko, nato pa za pet besed najbolj tipičen kolokator. Slika 4 prikazuje vmesnik za izbiro sopomenk, kjer je na levi nabor možnih odgovorov, na desni pa trenutni rezultat.



Slika 4: Igra Kombinator, izbiranje ustrezne sopomenke – iztočnica: *energija*, odgovor: *razgibanost*.

Igra Kombinator poleg pravilnih ponuja tudi napačne odgovore, t. i. distraktorje, ki morajo biti dovolj podobni pravilnim odgovorom, da je naloga za igralca smiselno zahtevna, hkrati pa ne smejo biti dejansko ustrezni. Distraktorje smo na osnovi podatkov iz DSBS generirali z velikimi jezikovnimi modeli. Pri sopomenkah je bil cilj najti besedo, ki je po obliki podobna ustrezni sopomenki, vendar ima drugačen pomen, pri kolokacijah pa besedo, ki se z iztočnico tipično ne pojavlja. Generirane podatke smo pregledali in odstranili neustrezne primere. Več o metodologiji je mogoče prebrati v Kosem in Arhar Holdt (2025). Obravnavo občutljivega besedišča predstavlja poglavje 3.3.

2.5 Ugibanka

V igri Ugibanka igralec išče skrito besedo, ki jo opisuje pet vsebinsko raznolikih namigov, urejenih po težavnosti od najbolj do najmanj zahtevnega. Točke se dodelijo glede na to, kako hitro je beseda uganjena. Na desni strani ekrana (Slika 5) so zabeleženi čas reševanja in nepravilni poskusi.

Kandidatke za skrite besede smo izbrali s seznama pogostih polnopomenskih besed iz DSBS, zlasti tistih, ki so v preteklosti nastopale kot besede meseca ali leta.[2] Za vsako od izbranih kandidatk smo pregledali razpoložljive leksikalne informacije v DSBS in drugih virih ter pripravili namige, ki se lahko vežejo na pomen, slovnične informacije, sopomenke, protipomenke, kolokacije, korpusne zglede, črke (npr. prva, zadnja črka ali število črk) ali zanimivost (npr. o izvoru, rabi besede). Pri pripravi namigov smo se izogibali občutljivemu besedišču in zgledom, kot opisuje poglavje 3.4.

2 Gre za seznam besed, ki se glede na statistične izračune v slovenskih medijih določenega meseca ali leta pojavljajo neobičajno pogosto. Primeri so na voljo na strani https://www.cjvt.si/besede-meseca/.



Slika 5: Igra Ugibanka – odgovor: *stavba*.

3 Obravnava občutljivega besedišča

Pri identifikaciji občutljivega gradiva za CJVT Igre smo pozorni na besedišče z enim ali več pomeni, ki se nanašajo na osebne značilnosti posameznikov, zlasti v povezavi z raso, narodnostjo, spolom in spolno usmerjenostjo, verskimi in političnimi prepričanji ter migracijskim statusom. Občutljivi so lahko tudi izrazi, ki na zaznamovan način označujejo posameznikov družbenoekonomski položaj, stopnjo izobrazbe, starost, telesne in osebnostne značilnosti ter določena vedenja in duševna stanja.

Med občutljive teme sodijo tudi vsebine, ki lahko pri igralcu sprožijo nelagodje in neprijetnost ali pa so glede na kulturne norme manj primerne za mlajše uporabnike. Med te sodijo spolnost, telesno izločanje, nasilje, bolezen, smrt, različne oblike zasvojenosti in težave v družini, na drugi strani pa teme in osebe, ki ideološko, politično ali kako drugače polarizirajo javno mnenje. Pri izbiri gradiva pazimo, da ne kažemo izkrivljene, sporne ali neobčutljive upodobitve realnosti in da ne reproduciramo družbenih stereotipov.

V mejnih primerih raje izločimo več gradiva kot premalo. Cilj namreč ni dosleden in celovit leksikografski opis jezika, temveč oblikovanje jezikovnih iger, ki morajo biti prijetne, varne in vključujoče za različne skupine uporabnikov, za šolski kontekst in samostojno rabo.

3.1 Pregled tematskih kategorij in njihove vsebine

Pri pripravi podatkov za Cvetko in Besedolov smo ročno pregledali poimenovanja tematskih kategorij in njihovo vsebino. Prvi leksikograf je opravil celovit pregled in označil potencialno problematične kategorije, nato je drugi leksikograf presodil, katere od označenih kategorij je smiselno odstraniti v celoti, pri katerih pa bi bilo smiselno odstraniti samo določene primere. Tabela 1 prikazuje dve v celoti odstranjeni kategoriji:

nemoralno dejanje in *spolna aktivnost*; kategoriji *del trupa* in *romantično čustvo* je mogoče obdržati, če se odstranijo potencialno problematični primeri.

Tematska kategorija	Besedišče v tej kategoriji
nemoralno dejanje	incest, krvoskrunstvo, maltretiranje, nečistovanje, perverzija, posilstvo, varanje
spolna aktivnost	fafanje, intimnost, kavsanje, kiberseks, ljubljenje, masturbacija, oplajanje, orgazem, parjenje, ploditev, seks
del trupa	bok, dekolte, dimlje, dojka, hrbet, oprsje, pas, pazduha, pleksus, popek, prsi, prsni koš, trebuh, zadnjica
romantično čustvo	čutnost, erotika, hlepenje, hrepenenje, koprnenje, ljubezen, naklonjenost, naslada, pohota, poželenje, strast, zaljubljenost, zvestoba

Tabela 1: Primeri občutljivega besedišča pri pripravi iger Cvetka in Besedolov.

3.2 Pregled besedišča v geslovniku za igro Vezalka

Vsi nabori besed za Vezalko so bili ročno pregledani. V procesu so bili odstranjeni podatkovni šum (npr. napačno lematizirane besede, ki se trenutno še pojavljajo v DSBS), zelo redko ali terminološko besedišče, citatno prevzeto, nestandardno, sovražno in grobo besedišče ter nesistematično vključene manjšalnice in podobne tvorjenke. Ker pri tej igri polnost nabora vključenih besed temeljno vpliva na njeno igralnost, v rabi prisotnih nezaznamovanih besed ne odstranjujemo, četudi so povezane z občutljivimi temami (npr. *rana*, *napad* v Tabeli 2).

Nabor črk	Izhodiščni nabor besed za ugibanje	Odstranjene besede
P, R, Z, A, F, E, obvezna D	pred, razred, razpad, parada, radar, redar, prepad, peder, praded, dereza, padar, pezde, zapad, prapraded, drap, prda, dred, adra, feder, farad, drape, prede, efedra, prerad, rezeda, predfaza, prapraprapraded	peder, pezde, prda, feder, prapraded, adra, rezeda, prapraprapraded, prede
P, D, R, L, U, A, obvezna N	dlan, rana, plan, luna, napad, uran, nuna, runda, nadura, nula, napa, panda, puran, anala, nada, pran, runa, urna, dnar, urarna, plundra, napadan, napudran, naan, ranar, apnar, papan, nandu, pudrn, pandan, panada, panula, drnula, ararauna	anala, dnar, papan, pandan, drnula, ararauna

Tabela 2: Primeri odstranjenih besed pri igri Vezalka.

3.3 Pregled kombinacij sopomenk in kolokatorjev z distraktorji

Sezname z generiranimi distraktorji smo ročno pregledali in odstranili potencialno občutljive primere. Selekcija je potekala tako na ravni iztočnice in ustreznega odgovora kot na ravni generiranih distraktorjev, vsi nizi s potencialnimi problemi so bili odstranjeni. Odstranili smo tudi nize, kjer vsaka beseda na sebi ni bila problematična, v sopostavitvi pa so vzbujale neprimerne asociacije (npr. *siromaštvo – sramota* v Tabeli 3).

Iztočnica	Ustrezni odgovor	Neustrezni odgovor
bujen	bohoten	pohoten
grmenje	donenje	gonjenje
črn	temnopolt	premožen
grd	nelep	slep
beda	siromaštvo	sramota

Tabela 3: Primeri odstranjenih podatkovnih nizov za sopomenske podatke v igri Kombinator.

3.4 Priprava in pregled namigov za Ugibanko

V posamezne igre Ugibanke nismo vključili skritih besed, ki so problematične v vseh pomenih (npr. *evtanazija*, *genocid*); pri besedišču, ki je občutljivo v posameznih pomenih, smo se pri pripravi namigov omejili na neproblematične pomene (npr. pri *deviški* smo se ognili spolnosti, pri *napad* dali prednost namigom, povezanim s športom). Posebej smo pazili, da pri občutljivih temah nismo sugerirali pozitivne konotacije, npr. s prvo osebo množine (*atentat – *Lahko ga izvedemo, naročimo, pripravimo ali obsodimo*) ali kako drugače. Igre je pripravila ekipa petih leksikografov, dva druga leksikografa sta gradivo pregledala in odpravila potencialno problematična mesta. Nekaj primerov zamenjanih korpusnih zgledov kaže Tabela 4.

Ugibana beseda	Prvotni zgled v namigu	Zamenjani zgled
poraz	Štela bi si za največji življenjski ___, če ne bi postala mama.	Zame bi bil osebni ___, če bi holding propadel.
prestop	Organiziran ilegalen ___ meje stane od 1000 do 4000 in več nemških mark.	Na mejnih prehodih ni daljšega čakanja pri ___ meje.
plenilec	Vedno več otrok je na begu in žrtev ___.	Že trikrat je odjezdil s svojimi vitezi, da bi ujel ___ ali napadel kakšno kljubovalno utrdbo.

Tabela 4: Primeri zamenjanih korpusih zgledov v namigih Ugibanke.

4 Sklep in nadaljnje delo

V prispevku smo predstavili portal CJVT Igre, s poudarkom na obravnavi občutljivega jezikovnega gradiva. Ročni pregled trenutno izvajamo ločeno za posamezne igre, saj se portal razvija postopoma, metodologija pa se razlikuje glede na zasnovo posamezne igre. V nadaljevanju želimo proces sistematizirati in v okviru DSBS uvesti označevanje leksikalnih enot oz. pomenov z negativno konotacijo in občutljivo vsebino. Takšne oznake presegajo označevanje za slovarski opis, vendar bi bile širše uporabne za pripravo učnih gradiv in tudi različne naloge strojne obdelave jezika, npr. strojno analizo sentimenta, zaznavanje sovražnega govora, pristranskosti in stereotipov, moderiranje spletnih vsebin, varnostno učenje velikih jezikovnih modelov ipd.

Prispevek je nastal v okviru raziskovalnih programov Jezikovni viri in tehnologije za slovenski jezik (P6-0411) in MRIC UL (I0-0022) ter raziskovalnega projekta Veliki jezikovni modeli za digitalno humanistiko (GC-0002), ki jih financira Javna agencija za znanstvenoraziskovalno in

inovacijsko dejavnost Republike Slovenije (ARIS). Delo sofinancira tudi Evropska Unija v sklopu Horizon Europe (101186647 – AI4DH).

Literatura

Arhar Holdt, Š., Kosem, I., Pori, E., Gorjanc, V., Krek, S. in Gantar, P. (2023). Negativno zaznamovano besedišče v Slovarju sopomenk sodobne slovenščine 2.0. *Slovenščina 2.0: Empirične, aplikativne in interdisciplinarne raziskave*, *11*(1), 8–32. https://doi.org/10.4312/slo2.0.2023.1.8-32

Arhar Holdt, Š. in Kosem, I. (2025). CJVT Igre: New word games based on the Digital Dictionary Database of Slovene. V I. Kosem idr. (ur.), *Proceedings of the eLex 2025 Conference, 18–20 November 2025, Bled, Slovenia* (str. 636–651). Lexical Computing CZ. https://elex.link/elex2025/wp-content/uploads/eLex2025-38-Arhar-HoldtKosem.pdf

Bustrillo, H. T. B., Sumicad, R. in Cuevas, G. C. (2024). Effectiveness of word games in teaching students vocabulary. *Journal of World Englishes and Educational Practices*, *6*(3), 7–16. https://doi.org/10.32996/jweep.2024.6.3.2

Derakhshan, A. in Khatir, E. D. (2015). The effects of using games on English vocabulary learning. *Journal of Applied Linguistics and Language Research*, *2*(3), 39–47.

Dharmayasa, I. N. W. (2022). An implementing the hangman game in teaching English vocabulary to elementary school students. *Jurnal Pendidikan Bahasa Inggris Undiksha*, *10*(3), 291–297. https://doi.org/10.23887/jpbi.v10i3.58140

Gantar, P. (2020). Dictionary of Modern Slovene: From Slovene Lexical Database to Digital Dictionary Database. *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, *46*(2), 589–602. https://doi.org/10.31724/rihjj.46.2.7

Hidayat, N. (2016). Improving students' vocabulary achievement through word game. *JEES (Journal of English Educators Society)*, *1*(2), 95–104. https://doi.org/10.21070/jees.v1i2.446

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. in Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*(1), 7–36.

Kosem, I. in Pori, E. (2021). Slovenske ontologije semantičnih tipov: samostalniki. V I. Kosem (ur.), *Kolokacije v slovenščini* (str. 159–202). Znanstvena založba Filozofske fakultete.

Kosem, I. in Arhar Holdt, Š. (2025). Using large language models to generate distractors for language games. V I. Kosem idr. (ur.), *Proceedings of the eLex 2025 Conference, 18–20 November 2025, Bled, Slovenia* (str. 629–644). Lexical Computing CZ. https://elex.link/elex2025/wp-content/uploads/eLex2025-37-KosemArhar-Holdt.pdf

Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I. in Dobrovoljc, K. (2020). Gigafida 2.0: The Reference Corpus of Written Standard Slovene. V N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk in S. Piperidis (ur.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (str. 3340–3345). European Language Resources Association. https://aclanthology.org/2020.lrec-1.409/

Munikasari, M., Sudarsono, S. in Riyanti, D. (2021). The effectiveness of using hangman game to strengthen young learners' vocabulary. *Journal of English Education Program*, *2*(1), 57–65. https://jurnal.untan.ac.id/index.php/JEEP/article/view/57-65/0

Pamungkas, N. A. R. (2021). The effects of Wordle media on students' vocabulary mastery. *JETAL: Journal of English Teaching & Applied Linguistic*, *2*(2), 56–61. https://doi.org/10.36655/jetal.v2i2.531

Pilán, I., Volodina, E. in Borin, L. (2016). Candidate sentence selection for language learning exercises: From a comprehensive framework to an empirical evaluation. *Traitement Automatique des Langues*, *57*(3), 67–91.

Toma, I., Alexandru, C. E., Dascalu, M., Dessus, P. in Trausan-Matu, S. (2017). Semantic Boggle: A game for vocabulary acquisition. V É. Lavoué idr. (ur.) *Data driven approaches in digital education. EC-TEL 2017* (Lecture Notes in Computer Science, vol. 10474, str. 603–607). Springer; Cham. https://doi.org/10.1007/978-3-319-66610-5_73

Od avtentičnih popravkov do generiranih napak: učni podatki za slovnično pregledovanje slovenščine

DOI: 10.4312/Obdobja.45.41-48

Špela Arhar Holdt, Fakulteta za računalništvo in informatiko in Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; arharhs@ff.uni-lj.si

Tina Munda, Institut »Jožef Stefan«, Ljubljana; Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; tina.munda@ijs.si

Nejc Robida, Institut »Jožef Stefan«, Ljubljana; Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; nejc.robida@ijs.si

Živa Štebljaj, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; ziva.stebljaj@fri.uni-lj.si

Polona Gantar, Fakulteta za računalništvo in informatiko in Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; apolonija.gantar@ff.uni-lj.si

Z uveljavitvijo nevronskih pristopov se je temeljito spremenil razvoj jezikovnih tehnologij, tudi strojnega slovničnega pregledovanja. Novi razvojni postopki zahtevajo velike količine enotno označenih učnih primerov, česar za slovenščino trenutno še nimamo na voljo. Prispevek predstavlja metodologijo za pripravo zgledov z jezikovnimi problemi in popravki, ki združuje korpusni pristop in generiranje z velikimi jezikovnimi modeli. Postopek temelji na jezikoslovni analizi posameznega jezikovnega problema, pripravi zbirke avtentičnih korpusnih zgledov z označenim mestom napake in pozivih, na osnovi katerih izbrani veliki jezikovni model na označeno mesto vstavi ciljno napako, pa tudi omogoči učinkovito strojno podprto kontrolo kakovosti.

strojno slovnično pregledovanje, veliki jezikovni modeli, učni podatki, korpusni pristop, slovenščina

The rise of neural approaches has fundamentally reshaped language technologies, including automated grammatical error correction. The new development approaches require large amounts of consistently annotated training data, which are not yet available for Slovenian. This article presents a methodology for creating training examples with language errors and their corrections by combining a corpus-based approach with generation using large language models. The procedure builds on the linguistic analysis of a specific error type, the compilation of a set of authentic corpus examples with a pre-marked position for the insertion of the target error, and prompts that guide a selected large language model to introduce the target error at the marked position and facilitate efficient machine-assisted quality control.

grammatical error correction, large language models, training data, corpus-based approach, Slovenian

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

1 Uvod

Slovnični pregledovalniki kot ena temeljnih jezikovnih tehnologij predstavljajo pomembno digitalno infrastrukturo za sodobne jezike, tudi slovenščino. Tradicionalni slovnični pregledovalniki so temeljili na vnaprej oblikovanih jezikovnih pravilih, v zadnjih letih pa se je razvoj (tudi) slovničnega pregledovanja temeljito spremenil z uveljavitvijo nevronskih pristopov, zlasti na predučenih jezikovnih modelih. Ti pristopi so bili za slovenščino že preizkušeni in ugotovljeno je bilo, da za izboljšanje tehnologije potrebujemo boljše učne podatke, saj obstoječi korpusni viri z jezikovnimi popravki ne prinašajo dovolj gradiva za uspešno strojno učenje, njihov dodatni izziv pa so različni kriteriji označevanja popravkov in različen odnos do prvin, ki sploh zahtevajo jezikovno popravljanje.

Prispevek predstavlja metodologijo za pripravo obsežnega nabora zgledov z jezikovnimi problemi in popravki, ki združuje korpusni pristop in generiranje z velikimi jezikovnimi modeli (VJM). Postopek temelji na jezikoslovni analizi posameznega jezikovnega problema, pripravi zbirke avtentičnih korpusnih zgledov z označenim mestom napake ter pozivih, na osnovi katerih izbrani veliki jezikovni model na označeno mesto vstavi ciljno napako, v naslednjem koraku pa sledi tudi strojno podprta kontrola kakovosti. V nadaljevanju predstavimo metodologijo, ponazorimo rezultate s popravki (ne)upoštevanja preglasa v besednih oblikah, posebej pa se posvetimo tudi razpravi o dometu in omejitvah pristopa. Prispevek sklenemo z napovedjo nadaljnjega dela, ki vključuje uporabo novopripravljenih učnih podatkov za učenje modelov za slovnično popravljanje slovenskih besedil.

2 Sorodno delo

Najbolj dodelano orodje za slovnično popravljanje slovenščine je plačljiva Amebis Besana, ki temelji na ročno pripravljenih jezikovnih pravilih in podatkovni zbirki Ases (Romih in Holozan 2002). Na voljo je tudi program Loris, primarno namenjen slovenski narodni skupnosti v Italiji (Popič idr. 2024). Poleg tega je slovenščina vključena v večjezične spletne popravljalnike, kot so Smodin, wordcount.com in odprtokodni LanguageTool, vendar z razmeroma omejenim naborom pravil.

S slovničnim popravljanjem z uporabo globokih nevronskih mrež se je ukvarjalo več avtorjev, pri čemer so rezultati po nalogah precej različni. Božič (2020) je razvil model za avtomatsko popravljanje vejic (92,5 % pravilnih napovedi), ki je v nekoliko izboljšani različici objavljen kot program CJVT Vejice. Mokotar (2023) je razvil metodologijo za zaznavanje, prepoznavanje in popravljanje različnih vrst napak po poenostavljeni tipologiji iz korpusa Šolar, pri čemer je zaznavanje doseglo uspešnost (po metriki F1) 88 %, prepoznavanje 14 %, popravljanje pa (po metriki GLEU) 50 %. Petrič idr. (2024) so pri zaznavanju določenih oblikoslovnih napak z modelom SloBERTa dosegli na sintetičnih podatkih visoko natančnost, na besedilih iz dejanske rabe pa nizko, kar je izpostavilo nujo po pripravi avtentičnih evalvacijskih podatkov za strojno popravljanje slovenščine. Klemen idr. (2024) so razvili nevronski črkovalnik SloNSpell, ki je pri evalvaciji na različnih množicah presegel rezultate vseh drugih obstoječih

orodij. Klemen idr. (2025) so nato predstavili ločene modele za različne jezikovne ravni (npr. črkovanje, velike/male začetnice, besedna oblika, besedni red) ter pri popravljanju oblikoslovja in skladnje dosegli bistveno slabši rezultat kot pri nižjih ravninah.

Slovenščina ima zgledno stanje na področju korpusov z označenimi jezikovnimi popravki. Na voljo so: razvojni korpus Šolar (Arhar Holdt in Kosem 2024); korpus besedil govorcev, ki se učijo slovenščino kot drugi ali tuji jezik KOST (Stritar Kučuk 2023); korpus besedil z lektorskimi popravki Lektor (Popič 2014) in korpus besedil, ki nastajajo v obmejnem stiku slovenščine in italijanščine StikIT (Popič in Grgič 2023). Na temeljih korpusa Šolar, ki prinaša najbolj natančno označevanje popravkov, sta bili pripravljeni referenčna evalvacijska množica Šolar-Eval 1.0 (Gantar idr. 2023) in učna množica DASSLE 1.0 (Arhar Holdt idr. 2025). Pri delu s temi viri se je pokazalo, da so za uspešno učenje modelov gradivno preveč omejeni in oblikovno pomanjkljivi; to vrzel skušamo odpraviti z metodo, ki jo opisujemo v nadaljevanju.

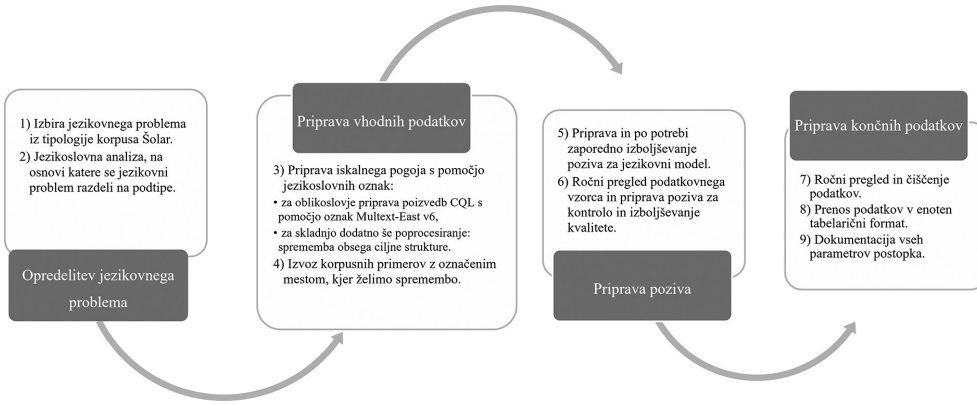
3 Metodologija

Cilj nove metode je omogočiti hitro in zanesljivo pripravo ustrezno oblikovanih učnih podatkov za strojno popravljanje posameznih jezikovnih problemov, primarno na ravneh oblikoslovja in skladnje. Metoda temelji na naslednjih premisah:

1) Izhodišče za pripravo učnih primerov mora biti avtentično gradivo, npr. iz referenčnega korpusa Gigafida (Krek idr. 2020).
2) V to gradivo želimo z izbranim velikim jezikovnim modelom vnesti avtentične jezikovne težave posameznih vrst, pri čemer trenutno sledimo tipologiji korpusa Šolar.[1]
3) Strojno podprto želimo opraviti tudi kontrolo kakovosti in odstraniti neustrezno pripravljene primere.
4) Ker je za VJM značilna manjša predvidljivost pri generiranju izhodnih podatkov, želimo strojni del naloge karseda omejiti, npr. kot vhodne podatke ponuditi izključno primere, kjer je zaželeno »kvarjenje« sploh možno; vnaprej označiti, kateri del povedi naj model spremeni; eksplicitno zahtevati, naj model opozori, kadar naloge ne more implementirati.
5) Končni podatki morajo vsebovati: informacijo o viru izvorne povedi, izvorni in »pokvarjeni« primer ter vrsto jezikovnega problema.

Postopek priprave podatkov po posameznih korakih prikazuje Slika 1.

1 https://wiki.cjvt.si/books/11-jezikovni-popravki-solar



Slika 1: Koraki priprave učnih podatkov za strojno učenje slovničnega pregledovanja.

4 Primer rezultatov: popravki (ne)upoštevanja preglasa v besednih oblikah

V nadaljevanju predstavljamo postopek pridobivanja avtentičnih korpusnih zgledov za prepoznavanje napačne rabe preglasa pri posameznih besednih oblikah. V (še neusklajenem) *Pravopisu 8.0* je v paragrafu 106 preglas opisan kot glasovna premena samoglasnika *o* v *e* za soglasniki *c*, *j*, *č*, *ž*, *š* in *dž*. Ta se izkazuje pri sklanjanju pred končnicami *-em*, *-ema* in *-ev*. Kandidatne povedi smo poiskali v korpusu Gigafida 2.1 s konkordančnikom NoSketch Engine, ki je dostopen prek spletnega servisa CLARIN.SI. Za luščenje oblik s potencialnim preglasom smo uporabili naslednjo poizvedbo CQL: [tag="S.meo|S.seo|S.mdr|S.mdd|S.sdr|S.sdd|S.mmr|S.mmd|S.smr|S.smd|S.mdo|S.sdo"& word=".*ev|.*em|.*ema"].

Na ta način smo zajeli nabor samostalniških oblik moškega in srednjega spola v relevantnih sklonih in številih, ki so v nadaljevanju predstavljali kandidate za (ustrezno) preglašene besedne oblike. Uporaba pike v oblikoskladenjskih oznakah je omogočila, da smo med zadetke vključili tako občnoimenske kot lastnoimenske samostalnike.

Pri izvozu korpusnega gradiva smo najdene oblike označili s posebnimi znaki. Model je moral v označenih besedah končnice *-ev*, *-em* oz. *-ema* zamenjati s končnicami *-ov*, *-om* oz. *-oma*, pri tem pa v povedih ni smel spreminjati ničesar drugega. Kar je spremenil, je moral v povedi označiti, v ločenem stolpcu pa izpisati konkretno zamenjavo skupaj s črko pred končnico, npr. *-cem* → *-com*. Celotni uporabljeni poziv se je glasil:

> Pripela bom nekaj povedi iz različnih besedil. Označene so besede, ki imajo preglas, npr. <coll>volivcev</coll>. Preglas se pojavlja v končnicah besed: v tistem delu besede, ki se pri sklanjanju spreminja. Oglej si označene besede in zamenjaj končnice -ev, -em ali -ema v končnice -ov, -om ali -oma. V povedih ne spreminjaj nič in čisto povsem nič drugega. Kar si spremenil, označi s simbolom ##, npr. ##volivcov##. V ločenem stolpcu izpiši, katero končnico si zamenjal, pri čemer vključi tudi črko pred končnico, npr. cem -> com. Če si pri procesiranju v dvomih, napiši NE VEM.

Tabela 1 prikazuje primere podatkov, ki smo jih generirali z modelom ChatGPT 5.2 Thinking. V vsaki vrstici so navedeni vir izvorne povedi, izvorna poved, spremenjena poved, zamenjan črkovni sklop in vrsta jezikovnega problema.

Vir	Izvorna poved	Spremenjena poved	Sprememba	Vrsta problema
GF956286 Delo, 2000	##Udeležencem## posveta ga je predstavil Igor Bevc z ministrstva za finance.	##Udeležencom## posveta ga je predstavil Igor Bevc z ministrstva za finance.	cem → com	preglas
GF7071454 internet, novice, 2017	V takšnih primerih se običajno zgodi, da država ##novoporočencema## podari 40.000 SGD (26.667 evrov) in jima tako finančno olajša nakup stanovanja.	V takšnih primerih se običajno zgodi, da država ##novoporočencoma## podari 40.000 SGD (26.667 evrov) in jima tako finančno olajša nakup stanovanja.	cema → coma	preglas
GF4257824 Dnevnik, 2002	Za to se je specializirala deželna zveza madžarskih ##rejcev## psov.	Za to se je specializirala deželna zveza madžarskih ##rejcov## psov.	cev → cov	preglas

Tabela 1: Primeri strojno pripravljenih podatkov za preglas.

S pomočjo takšne zasnove smo lahko hitro poiskali in odstranili primere, pri katerih strojna sprememba ni bila vnesena na pravo mesto, npr. pri obliki *Hočevarjeva* (*-čevarjeva → -čovarjeva* namesto *-jeva → -jova*), pri katerih so luščilni parametri zajeli oblike, ki niso povezane s preglasom, npr. *ljudem* (*-dem → -dom*), *moštev (-tev → -tov)* ipd.

Iz prečiščenega nabora smo pripravili učni nabor, ki zajema 10.000 primerov za neupoštevanje preglasa. S primerljivo metodologijo smo ločeno pripravili podatke tudi za svojilne pridevnike (npr. *ravnateljev → ravnateljov*), hiperkorekcijo (neustrezno rabo preglasa zlasti za *s* in *z*, npr. *nosom → nosem*, *Janezov → Janezev*) in primere s premeno *c – č* (npr. *prinčev → princev*).

5 Domet in omejitve pristopa

Predstavljeni primer kaže, da je metoda načeloma uspešna in koristna za pripravo učnega gradiva, tudi za jezikovne težave, ki bi jih tradicionalno reševali s pomočjo jezikovnih pravil in leksikona besednih oblik. Pri nevronskem pristopu se sistem pravila nauči posredno, iz konteksta pa se nauči identificirati tudi jezikovne napake, ki so enakopisne sicer legitimni besedni obliki, in kategorizirati vrsto jezikovnega problema.

Pri tem je treba upoštevati, da so povedi, ki smo jih uporabili za pripravo učnega gradiva, izhodiščno sicer avtentične, generirane napake pa ne odražajo nujno pogostosti in razporeditve v dejanski jezikovni rabi. Trenutno še nimamo dobre metodologije, kako bi avtentičnost »pokvarka« empirično merili in ocenjevali, možno pa bi bilo uporabiti razvojni krog, v katerem s trenutnimi primeri razvijemo pregledovalnik, z njim poiščemo avtentične napake določenega tipa, npr. v korpusu JANES (Fišer idr. 2020), izsledke uporabimo za izboljšavo učne množice itn.

Druga omejitev je, da postopek ni povsem enak za vse vrste jezikovnih problemov: vsak zahteva svojo kombinacijo jezikoslovne analize, korpusnega iskanja, filtriranja, oblikovanja ukaznega poziva in kontrole kakovosti gradiva, pripravljenega s pomočjo VJM. Pri preglasu je bilo mogoče postopek razmeroma dobro zamejiti z oblikoskladenjskimi oznakami in končnicami. Pri drugih jezikovnih vprašanjih pa smo morali v večji meri upoštevati tudi skladenjske in druge posebnosti, kar pomeni, da takšna priprava podatkov kljub strojni podpori zahteva precej časa. Na drugi strani VJM omogočijo spreminjanje gradiva, za katerega bi bilo predhodno treba razviti vrsto programskih skript.

Poseben izziv predstavljajo primeri, pri katerih se normativni opis spreminja oz. dopušča več variant. S tem smo se soočili pri pripravi učnega gradiva za predloge *s/z* in *k/h*. Prvo skupino težav predstavljajo številniki oz. številčni zapisi. Po *Slovenskem pravopisu* 2001 je bila izbira predloga *s* oz. *z* vezana na izgovor naslednje besede, kar je razvidno iz primerov tipa *deljeno z 8* in *deljeno s 3*. Takšni primeri zahtevajo posebno pripravo učnih podatkov, saj izbire predloga ni mogoče zanesljivo določiti samo na podlagi številke, temveč je treba upoštevati izgovor besede. Drugo skupino težav pa predstavljajo tujke, prevzeta imena in kratice, pri katerih se zapis in izgovor začetnega glasu oz. črke ne ujemata. V *Slovenskem pravopisu* 2001 so bili takšni primeri obravnavani z vidika zapisane oblike, npr. *z Zoisom*, *s Hugojem* ali *s SDK*. V *Pravopisu 8.0* pa je pravilo deloma drugačno, saj pri nekaterih takšnih primerih izrecneje dopušča variantnost (paragraf 132 v poglavju Glasoslovni oris). Pred črkami, ki se izgovarjajo drugače, kot so zapisane, se lahko predlog *z* zapisuje tudi kot *s* oz. obratno, npr. *s/z Zoisom*. Pri tovrstnih zapletenih primerih smo se za zdaj omejili na podatke, ki pokrijejo večje število napak, redkejše primere pa puščamo za nadaljnje delo.

6 Sklep in nadaljnje delo

V prispevku smo predstavili metodologijo za pripravo učnih podatkov za slovnično pregledovanje slovenščine, ki združuje avtentično korpusno gradivo in nadzorovano generiranje napak z velikimi jezikovnimi modeli. Na primeru (ne)upoštevanja preglasa smo pokazali, kako je mogoče izluščiti relevantne povedi, označiti mesto posega, z ustrezno oblikovanim ukaznim pozivom ustvariti ciljno napako in nato odstraniti neustrezne primere.

Tako pripravljene podatke bomo uporabili za učenje modela za slovnično popravljanje slovenskih besedil. Večino podatkov bomo objavili na CLARIN.SI kot DASSLE 2.0, manjši del pa namenili za pripravo meril uspešnosti (angl. *benchmark*), s katerimi bo mogoče oceniti uspešnost modelov za različne vrste jezikovnih težav. Ključno za kakovostno vrednotenje pa je tudi preverjanje strojnih popravkov na povsem avtentičnih jezikovnih napakah, zato bomo po vzoru množice Šolar-Eval (Gantar idr. 2023) podobne evalvacijske množice pripravili tudi za besedila odraslih piscev in piscev, ki jim je slovenščina drugi ali tuji jezik.

Prispevek je nastal v okviru raziskovalnih programov Jezikovni viri in tehnologije za slovenski jezik (P6-0411) in Slovenski jezik – bazične, kontrastivne in aplikativne raziskave (P6-0215) ter raziskovalnega projekta Veliki jezikovni modeli za digitalno humanistiko (LLM4DH; GC-0002), ki jih financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS). Delo sofinancira tudi Evropska Unija v sklopu Horizon Europe (101186647 – AI4DH).

Literatura

Arhar Holdt, Š., Antloga, Š., Gantar, P., Munda, T., Robida, N. in Zgonc, M. (2025). *Dataset of authentic and synthetic Slovene language errors (DASSLE) 1.0* [Zbirka podatkov]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/2052

Arhar Holdt, Š. in Kosem, I. (2024). Šolar, the developmental corpus of Slovene. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-024-09758-4

Božič, M. (2020). *Globoke nevronske mreže za postavljanje vejic v slovenskem jeziku* [Diplomsko delo]. Univerza v Ljubljani. https://repozitorij.uni-lj.si/IzpisGradiva.php?id=119034&lang=slv

Fišer, D., Ljubešić, N. in Erjavec, T. (2020). The Janes project: Language resources and tools for Slovene user generated content. *Language Resources and Evaluation*, *54*(1), 223–246. https://doi.org/10.1007/s10579-018-9425-z

Gantar, P., Bon, M., Gapsa, M. in Arhar Holdt, Š. (2023). Šolar-Eval: Evalvacijska množica za strojno popravljanje jezikovnih napak v slovenskih besedilih. *Jezik in slovstvo*, *68*(4), 89–108. https://doi.org/10.4312/jis.68.4.89-108

Klemen, M., Božič, M., Arhar Holdt, Š. in Robnik-Šikonja, M. (2024). Neural spell-checker: Beyond words with synthetic data generation. V E. Nöth, A. Horák in P. Sojka (ur.), *Text, Speech, and Dialogue: 27th International Conference, TSD 2024, Proceedings, Part I* (str. 85–96). Springer. https://doi.org/10.1007/978-3-031-70563-2_7

Klemen, M., Božič, M., Arhar Holdt, Š. in Robnik-Šikonja, M. (2025). Grammatical error correction of Slovenian school essays using large language models. *Sodobna pedagogika*, *76*(3), 162–176. https://doi.org/10.63384/sptB53z793a

Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I. in Dobrovoljc, K. (2020). Gigafida 2.0: The Reference Corpus of Written Standard Slovene. V N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk in S. Piperidis (ur.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (str. 3340–3345). European Language Resources Association. https://aclanthology.org/2020.lrec-1.409/

Mokotar, R. (2023). *Obvladovanje slovničnih napak v šolskih pisnih izdelkih z metodami za obdelavo naravnega jezika* [Diplomsko delo]. Univerza v Ljubljani. https://repozitorij.uni-lj.si/IzpisGradiva.php?id=144932&lang=slv

Petrič, T., Arhar Holdt, Š. in Robnik-Šikonja, M. (2024). Pomembnost realistične evalvacije: Primer popravkov sklona in števila v slovenščini z velikim jezikovnim modelom. *Slovenščina 2.0: Empirične, aplikativne in interdisciplinarne raziskave*, *12*(1), 106–130. https://doi.org/10.4312/slo2.0.2024.1.106-130

Popič, D. (2014). Revising translation revision in Slovenia. V T. Mikolič Južnič (ur.), *New horizons in translation research and education* (knj. 2, str. 72–89). University of Eastern Finland, Philosophical Faculty. https://erepo.uef.fi/handle/123456789/14340

Popič, D., Bordon, D., Franza, J. in Bezgovšek, L. (2024). Jezikovna orodja za slovenščino v Italiji: Jezikovni asistent Loris 1.1. V M. Grgič (ur.), *Slovenščina v Italiji* (str. 167–194). Založba Univerze v Ljubljani. https://ebooks.uni-lj.si/ZalozbaUL/catalog/book/676/chapter/3964

Popič, D. in Grgič, M. (2023). Korpus lektoriranih slovenskih besedil z območja italijansko-slovenskega jezikovnega stikanja – STIKit: Zasnova, gradnja in izzivi. *Jezik in slovstvo*, *68*(4), 73–87. https://doi.org/10.4312/jis.68.4.73-87

Romih, M. in Holozan, P. (2002). Infrastruktura za razvoj jezikovnih tehnologij – korpus FIDA in sistem ASES. V T. Erjavec in J. Žganec Gros (ur.), *Jezikovne tehnologije, 14.–15. oktober, Ljubljana, Slovenija* (str. 166). http://nl.ijs.si/isjt02/zbornik/sdjt02-D02amebis.pdf

Stritar Kučuk, M. (2023). Prvi korpus slovenščine kot tujega jezika KOST 1.0. V Š. Arhar Holdt in S. Krek (ur.), *Razvoj slovenščine v digitalnem okolju* (str. 93–117). Založba Univerze v Ljubljani.

Uporaba generativne umetne inteligence pri načrtovanju jezikovnih dejavnosti za predšolske otroke

DOI: 10.4312/Obdobja.45.49-55

Barbara Bednjički Rošer, Pedagoška fakulteta, Univerza v Mariboru, Maribor;
barbara.bednjicki@um.si

Generativna umetna inteligenca (GUI) vse izraziteje vpliva na spremembe v izobraževanju in spodbuja razvoj inovativnih učnih okolij in pristopov. Prispevek obravnava uporabo orodij GUI pri študentih predšolske vzgoje, ki so z njihovo pomočjo načrtovali in oblikovali didaktično gradivo za razvoj jezikovne zmožnosti predšolskih otrok. Rezultati kažejo, da so orodja, kot so ChatGPT, DeepSeek in Canva AI, uporabna predvsem za generiranje idej in vizualnih spodbud, ne nadomeščajo pa strokovne presoje, temveč ob zavedanju njihovih omejitev zahtevajo premišljeno in odgovorno rabo.

generativna umetna inteligenca, predšolska vzgoja, načrtovanje jezikovnih dejavnosti, glasovno zavedanje, rime

Generative artificial intelligence (GAI) is increasingly influencing changes in education and fostering the development of innovative learning environments and approaches. This article examines the use of GAI tools among early childhood education students, who used them to plan and design teaching materials for developing preschool children's language abilities. The results indicate that tools such as ChatGPT, DeepSeek, and Canva AI are particularly useful for generating ideas and visual prompts; however, they do not replace professional judgment and, given their limitations, require thoughtful and responsible use.

generative artificial intelligence, early childhood education, planning language activities, phonological awareness, rhymes

1 Uvod

Digitalne tehnologije so v različnih oblikah del izobraževalne krajine že vrsto let, generativna umetna inteligenca (GUI) pa z osupljivo hitrostjo vse bolj zaznamuje sodobne spremembe v izobraževanju ter pomembno vpliva na preoblikovanje poučevanja in učenja. Njena prisotnost spodbuja tudi razvoj inovativnih učnih okolij ter uvajanje sodobnih pedagoških pristopov, metod in oblik izobraževanja bodočih vzgojiteljev. Vršnik Perše (2021) poudarja, da proces poučevanja študentov v obdobju hitrih družbenih sprememb dobiva nove razsežnosti, zlasti zaradi oblikovanja sodobnih učnih okolij in uvajanja inovativnih praks v visokem šolstvu. Pomembna dejavnika pri tem sta tudi vse večja razširjenost tehnologije in njena smiselna raba v akademskem okolju, kar zahteva razvoj novih kompetenc, ki študentom omogočajo aktivno soustvarjanje in rekonstrukcijo znanja. Krajnc (2025) ter Šumak idr. (2026) ob tem poudarjajo pomen

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

celostnega razvoja digitalnih kompetenc iz okvira DigComp 2.2 (Vuorikari idr. 2023), ki v Evropski uniji in širše na področju izobraževanja nudi temelje za razvoj in merjenje digitalnih kompetenc.

Izsledki prve sistematične raziskave o uporabi GUI v visokošolskem izobraževanju za pedagoške poklice v Sloveniji (Licardo idr. 2025: 376–384) kažejo, da je poznavanje teh orodij med študenti razmeroma omejeno. Manj kot polovica jih GUI uporablja za iskanje informacij, dodatna pojasnila učne snovi ali prevajanje besedil, kar je povezano z redko uporabo GUI v študijskem procesu na fakultetah. Rezultati niso presenetljivi, saj je bila usposabljanj o uporabi GUI v izobraževanju deležna le četrtina visokošolskih učiteljev, pri čemer jo pri poučevanju uporablja le polovica. Tako študenti kot učitelji poudarjajo potrebo po dodatnih izobraževanjih, predstavitvah orodij in primerih dobre prakse za neposredno vključevanje v pedagoško delo. Med študenti je najpogosteje uporabljeno orodje ChatGPT, hkrati pa večina svoje znanje o uporabi GUI ocenjuje kot pomanjkljivo. Kot ključni etični izziv navajajo nezanesljivost informacij in potrebo po kritični presoji v akademskem okolju, obenem pa prepoznavajo prednosti, kot sta hitrejše opravljanje nalog in lažji dostop do izbranih vsebin. Polovica visokošolskih učiteljev pa je do uporabe GUI v poučevanju neopredeljena.

Na tej osnovi se prispevek osredinja na uporabo orodij GUI pri didaktičnem načrtovanju jezikovnih dejavnosti študentov predšolske vzgoje. Poudarek je na načinih njihove rabe ter na prepoznanih prednostih in omejitvah pri pripravi didaktičnih gradiv za predšolske otroke.

2 Teoretična izhodišča

V zadnjem obdobju se vse več raziskav posveča vlogi GUI pri načrtovanju vzgojno-izobraževalnega dela. ChatGPT in drugi jezikovni modeli, podprti z umetno inteligenco, odpirajo nove možnosti tudi vzgojiteljem predšolskih otrok (Bafralı in Tanju Aslışen 2025; Jambunathan in Jayaraman 2025; Kölemen in Yildirim 2025; Nikolopoulou 2026; Su in Yang 2024; Uğraş 2024; Zhang 2024), saj omogočajo pripravo raznolikih didaktičnih gradiv.

V skladu s teoretičnimi spoznanji o zgodnjem učenju otrok, ki poteka na različnih, med seboj tesno prepletenih področjih dejavnosti (jezik, družba, gibanje, matematika, narava in umetnost), je zasnovan nacionalni dokument za delo v vrtcih, *Kurikulum za vrtce* (Antič idr. 2025). Jezik kot temeljni gradnik otrokove identitete določa jezikovna zmožnost (Antič idr. 2025: 28, 38), ki zajema učenje besedišča, slovnice jezika in glasovnega zavedanja. Kranjc (2006, 1999: 53) jezikovno zmožnost opredeljuje kot sestav slovnične in sporazumevalne zmožnosti, *Skupni evropski jezikovni okvir* (Svet Evrope 2011: 132–133) pa jezikovno zmožnost definira znotraj sporazumevalne in k njej prišteva leksikalno, slovnično, semantično, fonološko, pravopisno in pravorečno zmožnost. Bešter Turk (2011: 118) piše, da je v slovensko strokovno literaturo pojem sporazumevalna zmožnost s t. i. komunikacijskim modelom jezikovne vzgoje vpeljala Olga Kunst Gnamuš, ki je razlikovala med jezikovno in sporazumevalno zmožnostjo, Bešter pa v okviru sporazumevalne zmožnosti, ki je eden temeljnih ciljev sodobnega pouka in jo

opredeljuje kot »zmožnost govorca in poslušalca, da se ob upoštevanju okoliščinskih dejavnikov različnih vrst lahko sporazumevata«, loči jezikovno in pragmatično zmožnost.

V okviru razvoja jezikovne zmožnosti je za predšolsko obdobje posebej pomembno glasovno zavedanje, ki predstavlja eno izmed ključnih predbralnih spretnosti. Zemljak Jontes in Bednjički Rošer (2020) pišeta, da se sposobnosti glasovnega zavedanja razvijajo postopoma v zaporedju od največjih glasovnih enot – zavedanje o dolžini besed; zaznavanje in tvorjenje rim ter aliteracij, členjenje povedi na besede; zlogovno zavedanje (zlogovanje in združevanje zlogov v besedo) – do najmanjših – zavedanje o začetkih in koncih besed in fonemično zavedanje, ki vključuje prepoznavanje oz. identifikacijo glasov v celotni glasovni verigi (Čok in Sila 2015). V *Kurikulumu za vrtce* (2025: 39) je glasovno zavedanje zajeto pri razvijanju predbralnih in predpisalnih spretnosti, in sicer: slušno zaznavanje; slušno razločevanje in razčlenjevanje; pri igrah z glasovi in razvijanju artikulacije glasov. V skladu s kurikularnimi izhodišči smo se v raziskavi osredotočili na področje spodbujanja glasovnega zavedanja. Namen prispevka je predstaviti dejavnosti, ki so jih študenti razvijali s pomočjo uporabe orodij GUI.

3 Metodologija

Cilj raziskave je bil preučiti možnosti uporabe orodij GUI pri oblikovanju didaktičnih gradiv za spodbujanje glasovnega zavedanja v predšolskem obdobju ter opisati zaznane prednosti in omejitve. Raziskava je bila zasnovana kot kvalitativna študija primera, v kateri je sodelovalo 58 študentov drugega letnika študijskega programa Predšolska vzgoja Pedagoške fakultete Univerze v Mariboru. Študenti so s pomočjo orodij ChatGPT, DeepSeek in Canva AI načrtovali dejavnosti in ob tem kritično presojali ustreznost generiranih predlogov glede na kurikularne cilje in razvojne zmožnosti otrok.

Podatki so bili zbrani z opazovanjem procesa načrtovanja in analizo didaktičnih gradiv. Analiza temelji na deskriptivni interpretaciji. V času pedagoškega procesa smo s študenti didaktična gradiva sproti pregledovali in dopolnjevali, kar je omogočilo vpogled v proces njihovega nastajanja, presojo ustreznosti ter prepoznavanje prednosti in omejitev. Študenti so ob tem v skladu z okvirom DigComp 2.2 razvijali predvsem digitalne kompetence s področja ustvarjanja digitalnih vsebin (Vuorikari idr. 2023: 27–30). Rezultatov ni mogoče posploševati, saj podatki temeljijo predvsem na pedagoškem opazovanju in analizi študentskih izdelkov, brez sistematičnega merjenja učinkov pri otrocih ali standardiziranih evalvacijskih instrumentov. Opažanja tako predstavljajo le strokovno interpretacijo uporabnosti z orodji GUI oblikovanih didaktičnih gradiv za izvedbo dejavnosti v vrtcih.

4 Rezultati

Rezultati predstavljajo izsledke analize procesa nastajanja didaktičnih gradiv in prikazujejo, kako so študenti ob podpori orodij GUI načrtovali dejavnosti za spodbujanje glasovnega zavedanja v skladu s cilji kurikularnega področja jezika.

Analiza kaže, da so študenti z uporabo GUI najpogosteje oblikovali dejavnosti za prepoznavanje rim, zlogovanje in prepoznavanje začetnih glasov. Pri tem se je pokazal tudi ustvarjalni potencial uporabe GUI, saj so nastala raznolika didaktična gradiva, pri čemer se je kot pomemben dejavnik izkazala tudi kakovost pozivov, ki so jih študenti oblikovali pri delu z orodji GUI. Natančnejši in strokovno premišljeni pozivi so prispevali k ustreznejšim predlogom, splošni ali premalo natančni pozivi pa so pogosteje vodili do jezikovno, didaktično ali vizualno manj ustreznih rešitev. Zato je pri uporabi GUI ključna tudi zmožnost kritičnega oblikovanja pozivov in presojanja nastalih gradiv. Ob tem so se pokazale jezikovne omejitve generiranih predlogov in velike razlike v kakovosti vizualnih gradiv, povezane z izbiro in dostopnostjo umetnointeligenčnih orodij, kar kaže na prisotnost digitalnega razkoraka (Strel idr. 2026). Študenti so oblikovali različne vizualne didaktične spodbude, namenjene zlasti spodbujanju druge, tretje in pete ravni glasovnega zavedanja (Čok in Sila 2015), kot so sličice za prepoznavanje rim, sestavljanke za spodbujanje zlogovanja ter igre in kartice za prepoznavanje začetnih glasov (Slika 1).



Slika 1: Z GUI oblikovana gradiva za spodbujanje različnih ravni glasovnega zavedanja (vir: osebni arhiv).

Pri oblikovanju gradiv za spodbujanje glasovnega zavedanja z uporabo GUI moramo biti previdni, saj ChatGPT in DeepSeek sicer ponujata številne inovativne ideje za izvedbo dejavnosti, pri preverjanju predlaganih primerov pa smo pri analizi predlogov generiranih rim opazili vrzeli. Rime so pogosto neustrezne, saj orodja ne poznajo v celoti prozodičnih lastnosti slovenskega jezika (npr. naglas, kolikost), zato je bilo treba številne primere izločiti ali prilagoditi. Slovenščina nima preprostega pravila za naglaševanje,

zato orodja predlagajo besede, v katerih se sicer ujemajo končni glasovi (črke), vendar se ujemanje ne nanaša nujno na zloge od zadnjega poudarjenega samoglasnika naprej in ne upošteva pravorečja. To potrjuje nujnost doslednega preverjanja jezikovne ustreznosti v relevantnih jezikovnih virih (npr. *Slovar slovenskega knjižnega jezika*, *Slovenski pravopis*). ChatGPT in DeepSeek pogosto ne prepoznata naglasnega mesta v besedi, niti ustreznega tipa naglasa, ne ločujeta med ozkim in širokim *e* oz. *o* oz. ustvarjata primere na podlagi črk in ne glasov. Med predlogi je bilo zato veliko število neustreznih primerov rim (npr. *rôka – móka*, *véverica – lúbenica*, *žóga – nôga*, *čebélica – déklica*, *banána – smétana*, *stòl – mól*, *mórje – pólje*, *korénček – kámenček*, *pès – lés*, *vôda – móda* ipd.).

Rezultati nadalje kažejo, da so prav jezikovne omejitve orodij GUI od študentov zahtevale dodatno strokovno presojo, kar je prispevalo k razvoju njihove kritične rabe digitalnih virov. Generirane predloge so vrednotili z vidika jezikovne in didaktične ustreznosti ter jih po potrebi prilagodili ali izločili, kar jasno kaže, da uporaba GUI ne nadomešča strokovnega znanja, temveč zahteva njegovo aktivno in kritično uporabo pri didaktičnem načrtovanju dejavnosti za spodbujanje glasovnega zavedanja.

5 Razprava

Rezultati kažejo, da so Canva AI, ChatGPT in DeepSeek pri načrtovanju vzgojno-izobraževalnega dela za kurikularno področje jezika uporabni predvsem kot pripomočki za generiranje idej in ustvarjanje vizualnih spodbud, ki podpirajo otrokovo razumevanje, krepijo motivacijo za sodelovanje in zanimanje otrok za jezikovne dejavnosti. Menimo, da so orodja GUI odlična za razširitev razmišljanja, vendar strokovna presoja pri preverjanju predlaganih primerov dejavnosti ostaja ključna naloga študentov in vzgojiteljev, ki so kot strokovni delavci neposredno odgovorni za kakovost vzgojno-izobraževalnega dela v predšolski vzgoji. Kot pomemben izziv pa izpostavljamo tudi neenako dostopnost orodij GUI, saj je trenutni dostop do njihove uporabe med študenti izrazito neenakomeren in pogosto omejen z omejitvami brezplačnih različic.

Ugotovitve študije primera kažejo, da lahko orodja GUI predstavljajo uporabno podporo pri didaktičnem načrtovanju jezikovnih dejavnosti v vrtcih. Ugotovitve so skladne tudi z raziskavami tujih avtorjev (Nikolopoulou 2026; Su in Yang 2024; Zhang 2024), ki poudarjajo potencial umetne inteligence pri pripravi gradiv in podpori načrtovanju vzgojno-izobraževalnega dela. Obenem pa rezultati kažejo, da uporaba GUI prinaša tudi določene omejitve, saj imajo veliki jezikovni modeli v slovenščini več omejitev, ki izhajajo iz manjše količine podatkov in jezikovne kompleksnosti (Kuzman idr. 2023). Analiza generiranih predlogov je pokazala, da orodja GUI ne upoštevajo zanesljivo prozodije in morfologije ter posledično pogosto generirajo primere rim, ki temeljijo na grafičnem ujemanju črk in ne na fonološki ustreznosti. Generirani predlogi zato zahtevajo dodatno strokovno presojo ter preverjanje jezikovne (jezikovni priročniki, slovarji) in didaktične ustreznosti (npr. likovni in vsebinski kriteriji vizualnih gradiv). Naše ugotovitve tako potrjujejo, da GUI ne nadomešča strokovnega znanja, ampak pri načrtovanju dejavnosti zahteva kritično, premišljeno, odgovorno, reflektirano in pedagoško

osmišljeno rabo. Kerneža (2025) poudarja, da je ravno kritično mišljenje ključni del jezikovnega pouka in je tesno povezano z razumevanjem jezika.

Rezultati raziskave so pokazali tudi, da na kakovost oblikovanih didaktičnih gradiv pomembno vpliva dostopnost orodij GUI. Razlike med prostodostopnimi in licenčnimi različicami orodij so se pokazale predvsem v obsegu in kakovosti generiranih vizualnih gradiv ter funkcionalnostih za njihovo prilagajanje, kar odpira vprašanje digitalnega razkoraka, ki lahko pomembno vpliva na možnosti uporabe orodij GUI pri študentih v visokošolskem izobraževanju in vzgojiteljih v vzgojno-izobraževalnem procesu.

6 Zaključek

Prispevek obravnava možnosti uporabe orodij GUI pri načrtovanju dejavnosti za spodbujanje glasovnega zavedanja v vrtcu. Ugotovitve študije primera kažejo, da imajo lahko orodja GUI pomemben didaktični potencial, saj lahko spodbujajo otrokovo motivacijo, razumevanje in sodelovanje v jezikovnih dejavnostih. Rezultati zaradi jezikovnih omejitev orodij GUI opozarjajo tudi na potrebo po strokovni presoji, zlasti pri načrtovanju dejavnosti in kakovosti generiranih vizualnih spodbud ter odpirajo vprašanja digitalnega razkoraka pri uporabi orodij GUI v izobraževalnem procesu. Nadaljnje raziskave bi lahko podrobneje preučile digitalni razkorak in razširile raziskovanje uporabe orodij GUI tudi na vzgojitelje v vrtcih.

Literatura

Antič, S., Bah Berglez, E., Cotič Pajntar, J., Devjak, T., Drljić, K., Hmelak, M., Hohnjec, D., Jager, J., Jaunik, M., Jerše, L., Klobasa, H., Kovaček, B., Marjanovič Umek, L., Milčinović, M., Mršnik, S., Novinec, B., Plaskan, L., Sivec, M., Vidonja, M. … Sušnik, M. (2025). *Kurikulum za vrtce*. Ministrstvo za vzgojo in izobraževanje, Zavod Republike Slovenije za šolstvo. https://www.gov.si/assets/ministrstva/MVI/Dokumenti/Sektor-za-predsolsko-vzgojo/Dokumenti-smernice/KURIKULUM-ZA-VRTCE-2025.pdf

Bafralı, G. in Tanju Aslışen, E. H. (2025). *Artificial intelligence in early childhood classrooms: Preservice teachers' perspectives on ChatGPT*. SSRN. https://doi.org/10.2139/ssrn.5364356

Bešter, M. (2011). Sporazumevalna zmožnost – eden izmed temeljnih ciljev pouka slovenščine. *Jezik in slovstvo*, *56*(3–4), 111–130. https://www.dlib.si/details/URN:NBN:SI:doc-NFS5PI7E

Čok, L. in Brčaninović, A. (2015). Z igro glasov v angleškem jeziku do razvijanja fonološkega zavedanja. *Revija za elementarno izobraževanje*, *8*(1–2), 147–166. http://rei.pef.um.si/images/Izdaje_revije/2015/1-2/REI_8_1-2_09.pdf

Jambunathan, S. in Jayaraman, J. D. (2025). Exploring the efficacy of lesson planning in early childhood education using ChatGPT. *Journal of Early Childhood Teacher Education*, 1–15. https://doi.org/10.1080/10901027.2025.2556997

Kerneža, M. (2025). Razvijanje kritičnega mišljenja pri pouku slovenščine v osnovni šoli. *Revija inovativna pedagogika*, *1*(1), 46–59. https://doi.org/10.63069/63egjb72

Kölemen, E. B. in Yildirim, B. A. (2025). New era in early childhood education (ECE): Teachers' opinions on the application of artificial intelligence. *Education and Information Technologies*, *30*, 17405–17446. https://doi.org/10.1007/s10639-025-13478-9

Krajnc, R. (2025). Izbira in načrtovanje dejavnosti za razvoj digitalnih kompetenc pri pouku. *Revija inovativna pedagogika*, *1*(2), 187–202. https://doi.org/10.63069/dh8dvd27

Kranjc, S. (1999). *Razvoj govora predšolskih otrok*. Znanstveni inštitut Filozofske fakultete.

Kranjc, S. (2006). *Poglavja iz skladnje otroškega govora*. Izolit.

Kuzman, T., Mozetič, I. in Ljubešić, N. (2023). ChatGPT: Beginning of the end of manual linguistic data annotation? Use case of automatic essay scoring for Slovene. *arXiv*. https://arxiv.org/abs/2303.03953

Licardo, M., Kranjec, E., Lipovec, A., Dolenc, K., Arcet, B., Flogie, A., Plavčak, D., Ivanuš Grmek, M., Bednjički Rošer, B., Sraka Petek, B. in Laure, M. (2025). *Generativna umetna inteligenca v izobraževanju: Analiza stanja v primarnem, sekundarnem in terciarnem izobraževanju*. Univerza v Mariboru, Univerzitetna založba. https://doi.org/10.18690/um.pef.1.2025

Nikolopoulou, K. (2026). ChatGPT for educational-academic activities: Preschool student teachers' experiences. *Intelligent Technologies in Education*, *1*(1). https://doi.org/10.53761/ITED/1.6

Slovar slovenskega knjižnega jezika: Druga, dopolnjena in deloma prenovljena izdaja. (2014). Založba ZRC, ZRC SAZU. www.fran.si

Strel, K., Lebeničnik, M. in Drljić, K. (2026). Generativna umetna inteligenca kot potencialna zmanjševalka ali povečevalka digitalnega razkoraka? V M. Radovan in A. Flogie (ur.), *Generativna umetna inteligenca v izobraževanju: Preobrazba učenja v digitalni dobi* (str. 191–208). Založba Univerze v Ljubljani. https://ebooks.uni-lj.si/ZalozbaUL/catalog/book/911/chapter/4752

Su, J. in Yang, W. (2024). Powerful or mediocre? Kindergarten teachers' perspectives on using ChatGPT in early childhood education. *Interactive Learning Environments*, *32*(10), 6496–6508. https://doi.org/10.1080/10494820.2023.2266490

Svet Evrope. (2011). *Skupni evropski jezikovni okvir: Učenje, poučevanje, ocenjevanje*. Ministrstvo RS za šolstvo in šport, Urad za razvoj šolstva. https://centerslo.si/wp-content/uploads/2015/10/SEJO-komplet-za-splet.pdf

Šumak, B., Pušnik, M., Grašič, S. in Györkös, J. (2026). *Priročnik za razvijanje digitalnih kompetenc v pedagoških študijskih programih*. Univerza v Mariboru, Univerzitetna založba. https://doi.org/10.18690/um.pef.2.2026

Toporišič, J. idr. (ur.). (2001). *Slovenski pravopis*. Založba ZRC, ZRC SAZU. https://www.fran.si

Uğraş, M. (2024). Evaluation of ChatGPT usage in preschool education: Teacher perspectives. *Eğitim ve insani bilimler dergisi: Teori ve uygulama*, *15*(30), 387–414. https://doi.org/10.58689/eibd.1537337

Vršnik Perše, T. (2021). Spoznanja in izzivi sodobnih pogledov na učenje in poučevanje v visokem šolstvu. V T. Vršnik Perše (ur.), *Učenje in poučevanje v visokem šolstvu: Spoznanja in izzivi* (str. 5–18). Univerza v Mariboru, Univerzitetna založba. https://doi.org/10.18690/978-961-286-466-8.1

Vuorikari, R., Punie, Y. in Kluzer, S. (2023). *Okvir digitalnih kompetenc za državljane: Z novimi primeri rabe znanja, spretnosti in stališč: DigComp 2.2*. Zavod Republike Slovenije za šolstvo. https://www.zrss.si/wp-content/uploads/2023/08/DigComp-2-2-Okvir-digitalnih-kompetenc.pdf

Zemljak Jontes, M. in Bednjički Rošer, B. (2020). Glasovno zavedanje: 4. gradnik. V D. Haramija (ur.), *Gradniki bralne pismenosti: Teoretična izhodišča* (str. 109–136). Univerza v Mariboru, Univerzitetna založba. https://press.um.si/index.php/ump/sl/catalog/book/515

Zhang, L. (2024). Discussion on the potential of ChatGPT in the field of early childhood education. V I. A. Khan, S. H. Halili, V. Balakrishnan in R. A. Abd. Rauf (ur.), *Proceedings of the 2024 3rd International Conference on Humanities, Wisdom Education and Service Management (HWESM 2024)* (str. 373–379). Atlantis Press. https://doi.org/10.2991/978-2-38476-253-8_45

Prevajanje iz češčine v slovenščino s pomočjo pametnih tehnologij: pomoč ali nadloga?

DOI: 10.4312/Obdobja.45.57-66

David Blažek, Slovanský ústav, Akademie věd České republiky, Praga;
david.blazek@slu.cas.cz

Prispevek obravnava prevajanje besedila neformalnega tipa iz češčine v slovenščino s pomočjo pametnih tehnologij, zlasti orodij, ki temeljijo na umetni inteligenci. V prispevku na podlagi analize izbranih prevodov primerjamo kakovost rezultatov različnih orodij ter identificiramo tipične napake na morfološki, leksikalni in pragmatični ravni. Rezultati kažejo, da so prevodi pogosto razumljivi, vendar zahtevajo kritično presojo in jezikovni pregled.

umetna inteligenca, strojno prevajanje, češčina, slovenščina, slovanski jeziki

This article examines the translation of an informal communicative text type from Czech into Slovenian using smart technologies—in particular, AI-based tools. Based on an analysis of selected translations, it compares the quality of outputs produced by different tools and identifies typical errors at the morphological, lexical, and pragmatic levels. The results show that, although the translations are often understandable, they require critical evaluation and post-editing.

artificial intelligence, machine translation, Czech, Slovenian, Slavic languages

1 Uvod

Umetna inteligenca (v nadaljevanju UI) med drugim pomembno vpliva tudi na sodobno jezikovno komunikacijo in prevajanje. Veliki jezikovni modeli omogočajo hitro generiranje in prevajanje besedil. Njihova uspešnost pa ni enakovredna pri vseh jezikih, saj je v veliki meri odvisna od količine, kakovosti in raznolikosti učnih podatkov. Posledično modeli praviloma dosegajo boljše rezultate pri jezikih z obsežnejšimi jezikovnimi viri, medtem ko pri jezikih z omejenimi jezikovnimi viri kakovost prevodov ostaja bolj neenakomerna (Iyer idr. 2024; Lu idr. 2024).

Češčina in slovenščina sicer razpolagata z dobro razvitimi jezikovnimi viri in osnovno jezikovnotehnološko infrastrukturo,[1] vendar sta v primerjavi z globalno prevladujočimi jeziki v sodobnih tehnologijah obdelave naravnega jezika precej manj zastopani.

Prevajanje med češčino in slovenščino zato predstavlja zanimiv raziskovalni primer. Gre za tipološko sorodna slovanska jezika, pri katerih je mogoče pričakovati tako prednosti zaradi njune strukturne bližine kot tudi specifične težave, povezane z medjezikovno

1 Prim. npr. Rehm idr. (2020); konkretni podatki za češčino in slovenščino npr. v Hlaváčová (2023) in Krek (2023).

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

interferenco. Namen prispevka je zato ugotoviti, v kolikšni meri sodobna prevajalska orodja, ki temeljijo na umetni inteligenci, uspešno obvladujejo te posebnosti.

2 Teoretično izhodišče

V kontekstu sodobnih prevajalskih tehnologij ključno vprašanje ni samo, kako so prevodi generirani, temveč tudi, kako njihovo kakovost ustrezno ovrednotiti. Ocenjevanje prevodov, zlasti tistih, ki nastajajo s pomočjo UI, je večplasten proces, ki presega zgolj preverjanje slovnične pravilnosti.

Po Reiss (2000) kakovosti prevoda ni mogoče presojati zgolj na podlagi formalne pravilnosti, temveč tudi glede na semantično, funkcionalno in komunikacijsko ustreznost besedila.

Sodobni pristopi k vrednotenju strojnega prevajanja združujejo različne metode ocenjevanja. Poleg standardiziranih evalvacijskih naborov (angl. *benchmarks*) se pri presoji kakovosti prevodov uporabljajo avtomatske metrike (npr. BLEU) in standardizirani okviri za človeško evalvacijo, kot je MQM (Multidimensional Quality Metrics). Ker avtomatske metrike pogosto ne zajamejo pragmatične ustreznosti, slogovne naravnosti in koherence besedila, ostaja človeška evalvacija nepogrešljiv del vrednotenja.

3 Metodologija

Obravnavana raziskava temelji na večkriterijskem vrednotenju kakovosti prevodov iz češčine v slovenščino.[2] Empirični del raziskave je bil izveden s pomočjo strukturiranega vprašalnika, v katerem so respondenti ocenjevali prevode istega izhodiščnega besedila, pripravljene z različnimi prevajalskimi orodji.

Izhodiščno besedilo predstavlja zasebno dopisno komunikacijo (e-sporočilo) med dvema osebama, ki se poznata (gl. besedilo v nadaljevanju članka). Gre za neformalni komunikacijski tip besedila, za katerega so značilni neposreden jezik, raba tikanja ter preplet praktičnih informacij in osebnega tona. Besedilo združuje referenčno (posredovanje informacij), apelativno (usklajevanje dogovorov) in fatično funkcijo (vzdrževanje socialnega stika), zato predstavlja zahtevnejši preizkus kakovosti prevajalskih orodij kot zgolj terminološka ali informativna besedila. Analiza celovitega besedila[3] omogoča vrednotenje ne le prenosa pomena, temveč tudi komunikacijske funkcije in sposobnosti prevajalskih orodij, da ohranijo bistvene značilnosti izhodiščnega besedila v ciljnem jeziku.

2 Za potrebe projekta UI v prevajalski leksikografiji – možnosti in omejitve na Slovanskem inštitutu Akademije znanosti Češke republike je bil razvit ocenjevalni okvir z enotnimi merili vrednotenja, prilagojen ciljem raziskave in primerjalni analizi vseh obravnavanih jezikov. V projekt so vključeni prevodi tudi v druge slovanske jezike (slovaščino, ruščino, poljščino, bolgarščino in lužiško srbščino); po enaki metodologiji pa je bil preveden in analiziran tudi strokovni tekst. Več o projektu: https://www.slu.cas.cz/cs/ai-v-prekladove-lexikografii-moznosti-a-hranice (zadnji dostop 19. 7. 2026).

3 Avtor se je s problematiko prevodov iz češčine v slovenščino ukvarjal že prej, vendar z drugačnega metodološkega vidika. V prejšnji raziskavi je bila analiza usmerjena predvsem na primerjavo prevodov posameznih leksemov, besednih zvez in skladenjskih enot (Blažek 2024).

V raziskavi je sodelovalo devet rojenih govorcev slovenščine s filološko izobrazbo iz različnih starostnih skupin z aktivnim znanjem češkega jezika. Respondenti so bili izbrani namensko na podlagi njihove redne uporabe češkega jezika pri študiju, prevajanju, poučevanju, raziskovalnem in strokovnem delu. Skupno je bilo zbranih devet izpolnjenih vprašalnikov, ki predstavljajo osnovo za nadaljnjo analizo.

Vprašalnik, ki je bil posredovan respondentom, je bil zasnovan kot strukturirano orodje za sistematično ocenjevanje kakovosti prevoda iz češčine v slovenščino. V prvem delu je vseboval izhodiščno besedilo v češčini in referenčni prevod v slovenščino, ki ga je pripravil rojeni govorec slovenščine z dolgoletnimi izkušnjami s prevajanjem med češčino in slovenščino. Referenčni prevod ni bil razumljen kot edina možna prevodna rešitev, temveč kot orientacijsko izhodišče za primerjalno vrednotenje posameznih prevodov.[4]

V raziskavo je bilo vključenih 21 prevodov istega izhodiščnega besedila, pripravljenih z različnimi prevajalskimi orodji. Analiza je zajela sisteme nevronskega strojnega prevajanja (DeepL, Google Translate, Microsoft Translator, Lingea), orodja, integrirana v širše platforme (Microsoft Copilot, MS Bing), in modele, temelječe na velikih jezikovnih modelih (ChatGPT, Gemini, Claude in QuillBot). Analizirane so bile brezplačne in plačljive različice posameznih orodij, pri nekaterih prevodih pa tudi uporaba eksplicitno oblikovanih vhodnih pozivov (angl. *prompt*).[5] Del prevodov je nastal aprila 2025, del aprila 2026, kar je omogočilo primerjavo kakovosti prevodov skozi čas. Zaradi nepristranskega vrednotenja identiteta uporabljenih orodij respondentom med ocenjevanjem ni bila razkrita.

Kakovost prevodov je bila ocenjena na podlagi štirih vnaprej določenih kriterijev: pravopisne in slovnične ustreznosti, vsebinske ustreznosti, slogovne in formalne ustreznosti ter splošne uporabnosti besedila. Posamezni kriteriji so bili v vprašalniku dodatno pojasnjeni s podvprašanji, ki so usmerjala presojo respondentov glede jezikovne pravilnosti, ohranjanja pomena izvirnika, ustreznosti sloga in komunikacijske funkcije ter splošne uporabnosti besedila brez dodatnih popravkov. Respondenti so posamezne prevode vrednotili na petstopenjski lestvici (1–5), pri čemer je vrednost 1 označevala najboljši, vrednost 5 pa najslabši rezultat.

Poleg numeričnega ocenjevanja je vprašalnik omogočal tudi komentarje odprtega tipa, v katerih so respondenti utemeljili svoje ocene, opozorili na konkretne napake in predlagali ustreznejše prevodne rešitve. Takšna kombinacija kvantitativnega in kvalitativnega pristopa je omogočila celovitejšo analizo rezultatov ter natančnejšo identifikacijo značilnih problemov v prevodih.

---

4 Šele med pripravo prispevka se je pokazalo, da bi bilo metodološko ustrezneje dati ocenjevalcem v ocenjevanje vse prevode, tudi referenčnega, brez informacije o tem, ali ga je izdelal človek ali UI.

5 Tu je bil uporabljen poziv (v slovenščino za potrebe tega članka prevedel D. B.): »To je običajno, manj formalno elektronsko sporočilo, napisano v češčini. Prosim, analiziraj izvirno besedilo in ga prevedi v točno, naravno, idiomatično in obče uporabljano pisno slovenščino. Predvsem pa upoštevaj izražanje slovničnega spola, frazeologijo in terminološko doslednost.«

Izvirno besedilo

Milá Mařenko,

chtěla jsem se ještě domluvit na termínu. Ideálně bych mohla v pondělí odpoledne, šla by i středa tak od dvou hodin. Hodí se ti to?

Kvůli té zprávě si zavoláme zítra, nebo si můžeme napsat. V zásadě s tím souhlasím, ještě se musíme domluvit, kdo tu zprávu sestaví. Mohla bys ji pak zkorigovat?

Uvidíme se po svátcích v Nových Hradech.

Všechny pozdravuj a zlom vaz při zkoušce!

Zdravím,

Áďa

Prevod besedila v slovenščino

Draga Mařenka,

želela sem se še zmeniti za termin. Zame bi bilo najbolje v ponedeljek popoldne, šlo bi tudi v sredo nekje od dveh naprej. Bi ti to ustrezalo?

Glede tistega poročila se slišiva jutri oziroma si lahko piševa. Načeloma se strinjam, morava se še dogovoriti, kdo bo poročilo pripravil. Ali bi ga lahko potem ti pregledala?

Se vidimo po praznikih v Novih Hradih.

Pozdravi vse in veliko sreče na izpitu!

Lep pozdrav

Áďa

4 Analiza rezultatov

V Tabeli 1 so navedena pri posameznih prevodih uporabljena orodja. V nadaljevanju prispevka so analizirani prevodi označeni s številkami.

Zaradi omejenega prostora v prispevku prevodi niso predstavljeni v celoti.[6] V nadaljevanju so obravnavani le reprezentativni primeri, ki ponazarjajo najznačilnejše jezikovne pojave, tipične napake in razlike v kakovosti prevodov.

Številka prevoda	Uporabljeno orodje
1	DeepL 2025
2	DeepL 2026
3	DeepL plačljiv 2025
4	DeepL plačljiv 2026
5	Google Translator 2025
6	Google Translator 2026
7	Microsoft Copilot 2025
8	Microsoft Copilot 2026
9	Lingea 2026
10	ChatGPT 2025
11	ChatGPT 2026
12	ChatGPT z vhodnim pozivom 2025

6 Celotna prevedena besedila so dostopna na https://doi.org/10.5281/zenodo.23413924.

Številka prevoda	Uporabljeno orodje
13	ChatGPT z vhodnim pozivom 2026
14	ChatGPT plačljiv z vhodnim pozivom 2025
15	ChatGPT plačljiv z vhodnim pozivom 2026
16	MS Bing 2025
17	MS Bing 2026
18	Gemini 2,5 Flash 2026
19	Gemini 2,5 Pro 2026
20	QuillBot 2026
21	Claude 2026

Tabela 1: Uporabljena orodja.

4.1 Pravopisna in slovnična ustreznost

Na ravni pravopisa in slovnice je večina analiziranih prevodov ocenjena kot pretežno ustrezna, vendar z opaznimi vplivi izhodiščnega jezika (češčine), ki se kažejo predvsem v dobesednem prenosu jezikovnih struktur, besednega reda in posameznih slovničnih rešitev v ciljni jezik.

Med najpogostejšimi napakami so:

- napačen besedni red (prevod 2: »še se morava dogovoriti«),
- neustrezna raba dvojine (prevod 3: »moramo se še dogovoriti« namesto dvojinske oblike) ali
- oblike zemljepisnih imen v navzkrižju s slovenskimi pravopisnimi pravili (več prevodov: npr. »v Novih Hradech«, »v Nových Hradech«).

Poleg navedenega se pojavljajo tudi izrazitejša slovnična odstopanja, kot npr. v prevodu 20: »si bova jutri poklicali«, kjer je vezava glagola neustrezna in kaže neposreden vpliv češkega izhodišča.

Kljub navedenim primerom odstopanj je mogoče ugotoviti, da nekateri prevodi (npr. 8) dosegajo visoko stopnjo slovnične pravilnosti, medtem ko drugi (npr. 17 in 21) vsebujejo večje število napak, ki pomembno vplivajo na njihovo jezikovno ustreznost.

4.2. Vsebinska ustreznost

Vsebinska ustreznost predstavlja eno ključnih šibkih točk analiziranih prevodov. Najpogostejše napake so povezane z nepravilnim prenosom pomena, kar se kaže na ravni posameznih leksemov, fraz in tudi širših besedilnih segmentov.

Med tipične probleme sodijo:

- napačen prevod ključnih pojmov (npr. »sporočilo« namesto »poročilo« – prevodi 1, 3, 7, 11, 13, 16),
- napačna interpretacija časovnih podatkov (npr. »okoli druge ure« namesto »od druge ure naprej« – več prevodov, npr. 1, 2, 3),
- neustrezna adaptacija lastnih imen (npr. Maša, Mija, Marijana, Mašenka namesto Mařenka – prevodi 10, 12, 19, 21),
- neustrezen prenos idiomatskih izrazov (npr. »slabe vezi na preizkusu«),

- dodajanje ali opuščanje informacij (npr. »še enkrat« v stavku »še enkrat sem se želela dogovoriti glede termina« v prevodu 15, ki v izvirniku nima opore).

Posebej problematični so primeri, kjer se poruši referenčna struktura besedila, npr. v prevodu 6 (nejasen spol govorca in naslovnika) ali v prevodu 5 (»novice« namesto »poročila«). Ekstremen primer predstavlja prevod 9, kjer pride do popolnega vsebinskega razpada besedila (npr. »V Novem mestu se je zgodil nov praznik« namesto »Vidimo se po praznikih v Novih Hradih« in neustrezen podpis »Črni koren«).

Kljub navedenim odstopanjem nekateri prevodi (zlasti 8, 10 in 14) ohranjajo pomen izvirnika skoraj v celoti, z minimalnimi odstopanji, kar potrjuje, da prevajalska orodja lahko zagotovijo visoko stopnjo vsebinske ustreznosti, če zmorejo prepoznati vsebinske značilnosti izhodiščnega besedila.

4.3 Slogovna in formalna ustreznost

Na ravni slogovne in formalne ustreznosti večina analiziranih prevodov ohranja osnovni komunikacijski namen izhodiščnega besedila (neformalno dopisno komunikacijo), vendar se pogosto pojavljajo odstopanja, ki vplivajo na naravnost izraza in ustreznost v konkretnem komunikacijskem kontekstu.

Med najpogostejšimi problemi so:

- mešanje formalnega in neformalnega nagovora (npr. prevod 5: »Spoštovani« v kombinaciji z neformalnimi elementi, kot je tikanje),
- okorne ali dobesedne formulacije (prevoda 1, 3: »Zaradi tega sporočila se pokličemo« ali prevod 17: »Šla sem te še dogovoriti«),
- neustrezna raba glagolskih oblik (npr. mešanje velelnika in prihodnjika v prevodu 1),
- izguba naravnosti izraza (npr. »po dveh urah« v prevodu 16).

V nekaterih primerih se slogovne in vsebinske neustreznosti prepletajo, kar dodatno zmanjšuje komunikacijsko učinkovitost besedila. To je razvidno zlasti v prevodu 5, kjer poleg neustrezne izbire nagovora prihaja tudi do neustreznega zaključka (»se zabavajte med izpitom«), ki ne ustreza pragmatični funkciji izvirnika.

Kljub navedenim odstopanjem nekateri prevodi (zlasti 8, 10, 14 in deloma 15) dosegajo visoko stopnjo slogovne ustreznosti ter delujejo naravno in prepričljivo.

4.4 Skupna ocena

Na podlagi skupnih ocen je mogoče analizirane prevode razdeliti v tri osnovne skupine glede na njihovo splošno kakovost. Približno petina prevodov dosega visoko raven ustreznosti in je skoraj neposredno uporabna brez večjih popravkov; ti prevodi združujejo slovnično pravilnost, ustrezno vsebino in naraven slog.

Največji delež (približno dve tretjini) predstavljajo srednje kakovostni prevodi, ki so sicer razumljivi, vendar vsebujejo opazne napake, zlasti na vsebinski in slogovni ravni, zato zahtevajo dodatno jezikovno ali vsebinsko obdelavo.

Preostali manjši delež (približno petina) predstavljajo nizkokakovostni oz. neuporabni prevodi, pri katerih prihaja do resnih odstopanj, bodisi zaradi napačnega prenosa pomena bodisi zaradi kombinacije slovničnih, leksikalnih in pragmatičnih napak.

4.5 Sklep analize

Analiza skladnosti ocen med posameznimi respondenti kaže razmeroma visoko stopnjo konsistentnosti pri ugotavljanju ključnih problematičnih mest. Ocenjevalci prepoznavajo iste tipe napak, zlasti vpliv izvornega jezika, težave pri prenosu časovnih izrazov, neustrezno izbiro leksikalnih ekvivalentov in napačno obravnavo lastnih imen. Razlike med ocenami se pojavljajo predvsem v stopnji strogosti pri vrednotenju slogovne ustreznosti in pri interpretaciji manjših odstopanj, kar potrjuje deloma subjektivno naravo tovrstnega vrednotenja.

Analiza rezultatov kaže, da se ključne težave analiziranih prevodov pojavljajo predvsem na ravni leksikalne izbire, pragmatične ustreznosti in slogovne naravnosti, medtem ko je na formalni (slovnični) ravni zaznati bistveno manj odstopanj (z izjemo tujih lastnih imen). To potrjuje tudi pregled tipologije napak (gl. Tabelo 2), iz katerega je razvidno, da večina problematičnih mest ne zadeva osnovne jezikovne pravilnosti, temveč višje ravni jezikovne obdelave.

Rezultati analize potrjujejo, da prevajalski sistemi še vedno težje obvladujejo višje ravni besedila, zlasti pragmatično ustreznost, slogovno naravnost in ohranjanje komunikacijske funkcije. Podobne ugotovitve potrjujejo tudi novejše raziskave strojnega prevajanja in velikih jezikovnih modelov.[7]

Na splošno rezultati kažejo, da je kakovost prevodov neenotna in da je za doseganje ustreznega končnega rezultata v večini primerov nujna dodatna jezikovna obdelava. To potrjuje tudi vlogo človeškega uporabnika kot ključnega dejavnika v procesu interpretacije in vrednotenja prevodov.

V Tabeli 2 navajamo pregled najpogostejših napak, ki temelji na agregaciji ocen in komentarjev respondentov.

Tip napake	**Opis**	**Primeri[8]**
Leksikalne napake	napačna izbira ekvivalenta	1, 3, 7, 11, 13, 16
Časovni izrazi	napačna interpretacija časa	1, 2, 3, 4, 7
Lastna imena (osebna)	napačna adaptacija imen	10, 12, 19, 21
Lastna imena (toponimi)	napačna prilagoditev slovenskemu pravopisnemu sistemu	1–8, 10–18, 20
Interferenca	dobesedni prevodi	1, 3, 16
Skladenjske napake	napačen besedni red	2, 4, 20
Dvojina	napačna raba	3, 7
Pragmatične napake	neustrezen namen	5, 9
Slogovna neustreznost	nenaravne formulacije	1, 2, 4, 16, 20
Dodajanje/opustitev	dodani ali manjkajoči deli	11, 13, 15
Globalni razpad pomena	popolna izguba pomena	9

Tabela 2: Tipologija najpogostejših napak.

7 Gl. npr. Castilho in Knowles (2025), ki poudarjata pomen širšega besedilnega konteksta, koherence in pragmatične ustreznosti pri vrednotenju sodobnih sistemov strojnega prevajanja, ter Naveen in Trojovský (2024), ki med ključnimi omejitvami prevajalskih sistemov UI izpostavljata obravnavo idiomatskih izrazov, kulturno pogojenih pomenov in diskurzivnih značilnosti besedila.

8 Številke se nanašajo na oštevilčenje prevodov, ki je predstavljeno v Tabeli 1.

4.6 Primerjava prevajalskih orodij

Na podlagi ugotovitev iz predhodnih analiz je mogoče primerjati uspešnost posameznih prevajalskih orodij in opredeliti dejavnike, ki so v analiziranem vzorcu vplivali na kakovost prevodov. Tabela 3 prikazuje povprečne ocene vseh analiziranih prevodov na podlagi prejetih vprašalnikov. Stolpec Trend označuje, ali se je kakovost posameznega prevoda v letu 2026 glede na rezultate iz leta 2025 izboljšala, poslabšala ali ostala nespremenjena.

Orodje	2025	2026	Trend
DeepL	3,25	2,42	izboljšanje
DeepL plačljiv	2,75	2,08	izboljšanje
Google Translator	4,67	4,00	izboljšanje
Microsoft Copilot	2,25	1,67	izboljšanje
Lingea	–	4,75	brez sprememb[9]
ChatGPT brez ustreznega poziva	2,00	2,00	brez sprememb
ChatGPT z ustreznim pozivom	2,67	1,83	izboljšanje
ChatGPT plačljiv z ustreznim pozivom	1,92	2,25	rahlo poslabšanje
MS Bing	3,17	3,75	poslabšanje
Gemini 2.5 Flash	–	2,08	ni primerjave
Gemini 2.5 Pro	–	2,08	ni primerjave
QuillBot	–	3,00	ni primerjave
Claude	–	3,33	ni primerjave

Tabela 3: Povprečne ocene posameznih prevajalskih orodij.

Tabela kaže, da je najboljšo povprečno oceno dosegel Microsoft Copilot 2026 (1,67), sledila pa sta ChatGPT z ustreznim pozivom 2026 (1,83) in plačljiva različica ChatGPT z ustreznim pozivom 2025 (1,92). Visoko so bili ocenjeni tudi plačljivi DeepL 2026 ter Gemini 2.5 Flash in Gemini 2.5 Pro (2,08). V sredino lestvice sta se uvrstila QuillBot (3,00) in Claude (3,33), najslabše rezultate pa sta dosegla Google Translator 2025 (4,67) in Lingea (4,75).

Primerjava različic iz let 2025 in 2026 kaže pretežno trend izboljševanja kakovosti prevodov. Napredek je opazen pri sistemih DeepL, plačljivi DeepL, Google Translator, Microsoft Copilot in ChatGPT z ustreznim pozivom, medtem ko pri osnovni različici ChatGPT bistvenih razlik ni bilo. Pri MS Bing in plačljivi različici ChatGPT z ustreznim pozivom pa sta bila novejša prevoda ocenjena nekoliko slabše. Orodja Gemini 2.5 Flash, Gemini 2.5 Pro, QuillBot in Claude so bila vključena le v različicah iz leta 2026, zato časovna primerjava ni bila mogoča. Rezultati potrjujejo, da razvoj prevajalskih sistemov ni linearen in da novejša različica posameznega orodja ne zagotavlja nujno kakovostnejšega prevoda.

Posebej opazno je, da se med najbolje ocenjenimi prevodi pojavljajo predvsem sistemi, ki temeljijo na velikih jezikovnih modelih, zlasti kadar je bil uporabljen ustrezno oblikovan poziv. To nakazuje, da lahko kakovost izhodnega besedila pomembno izboljša

9 Ker se prevoda orodja Lingea iz let 2025 in 2026 pri analiziranem besedilu v ničemer nista razlikovala, je bila ocenjena le ena različica.

ne le izbira prevajalskega sistema, temveč tudi način njegove uporabe. Hkrati rezultati kažejo, da sistemi nevronskega strojnega prevajanja praviloma dosegajo stabilne rezultate na formalni ravni, medtem ko se razlike med posameznimi orodji izraziteje pokažejo pri pomenski ustreznosti, pragmatični funkciji in slogovni naravnosti prevodov.

Kljub navedenim razlikam je treba poudariti, da predstavljeni rezultati temeljijo na analizi enega izhodiščnega besedila in omejenega vzorca prevodov. Zato jih ni mogoče neposredno posploševati na vse prevajalske sisteme ali vse vrste besedil. Ugotovitve predvsem potrjujejo, da je kakovost strojnega prevoda rezultat kompleksne interakcije več dejavnikov, med katerimi imajo pomembno vlogo vrsta prevajalskega sistema, njegova različica, način uporabe (zlasti oblikovanje poziva) in čas nastanka prevoda.

5 Zaključek

Predstavljeni rezultati se nanašajo na analizirano besedilo, vključena prevajalska orodja in izbrano skupino ocenjevalcev, zato jih ni mogoče brez pridržkov posploševati na vse sisteme strojnega prevajanja ali prevajalske situacije.

Rezultati analize potrjujejo, da sodobna orodja za strojno prevajanje in umetno inteligenco predstavljajo učinkovito podporo pri prevajanju tudi med tipološko sorodnimi slovanskimi jeziki, kot sta češčina in slovenščina. Na ravni osnovnega pomena omogočajo hitro orientacijo v besedilu in pripravo razumljivih prevodnih osnutkov, kar potrjuje njihov praktični pomen v sodobni prevajalski praksi.

Hkrati analiza jasno kaže, da njihova največja učinkovitost ostaja omejena le na nekatere ravni jezikovne obdelave. Medtem ko so formalni vidiki jezika pogosto ustrezno obvladani, ostajajo ključne omejitve na semantični, pragmatični in diskurzivni ravni.

Primerjava prevodov istega besedila, ustvarjenih z istimi orodji v razmiku enega leta, kaže izboljšanje kakovosti prevodov. Hkrati pa rezultati opozarjajo, da se lahko izhodi generativnih sistemov spreminjajo tudi pri uporabi istega orodja, zato jih je treba razumeti kot časovno pogojene.

Na podlagi predstavljene raziskave je mogoče sklepati, da so analizirana orodja (tudi) pri prevajanju iz češčine v slovenščino predvsem učinkovita podpora prevajalcu, ne pa nadomestilo za človeško prevajanje.

Prispevek je nastal v okviru programa Strategie AV 21: Umělá inteligence pro vědu a společnost.

Literatura

Blažek, D. (2024). Ali v 21. stoletju še potrebujemo dvojezične slovarje? (Ob nastanku novega slovensko-češkega slovarja na Češkem). V M. Jesenšek (ur.), *Pleteršnikova dediščina: Ob stoletnici smrti Maksa Pleteršnika* (Zora 154, str. 187–198). Univerza v Mariboru, Univerzitetna založba. https://doi.org/10.18690/um.ff.3.2024.12

Castilho, S., Knowles, R. (2025). A survey of context in neural machine translation and its evaluation. *Natural Language Processing*, *31*(4), 986–1016. https://doi.org/10.1017/nlp.2024.7

Hlaváčová, J. (2023). Language report Czech. V G. Rehm in A. Way (ur.), *European language equality: A strategic agenda for digital language equality* (str. 115–118). Springer. https://doi.org/10.1007/978-3-031-28819-7_10

Iyer, V., Malik, B., Stepachev, P., Chen, P., Haddow, B. in Birch, A. (2024). Quality or quantity? On data scale and diversity in adapting large language models for low-resource translation. V B. Haddow, T. Kocmi, P. Koehn in C. Monz (ur.), *Proceedings of the ninth conference on machine translation* (str. 1393–1409). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.wmt-1.128

Krek, S. (2023). *Language report Slovenian*. V G. Rehm in A. Way (ur.), *European language equality: A strategic agenda for digital language equality* (str. 211–214). Springer. https://doi.org/10.1007/978-3-031-28819-7_34

Lu, Y., Zhu, W., Li, L., Qiao, Y. in Yuan, F. (2024). LLaMAX: Scaling linguistic horizons of LLM by enhancing translation capabilities beyond 100 languages. V Y. Al-Onaizan, M. Bansal in Y.-N. Chen (ur.), *Findings of the Association for Computational Linguistics: EMNLP 2024* (str. 10748–10772). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.631

Naveen, N. in Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. *iScience*, *27*(10), članek 110878. https://doi.org/10.1016/j.isci.2024.110878

Rehm, G., Marheinecke, K., Hegele, S., Piperidis, S., Bontcheva, K., Hajič, J., Choukri, K., Vasiļjevs, A., Backfried, G., Prinz, C., Gómez-Pérez, J. M., Meertens, L., Lukowicz, P., van Genabith, J., Lösch, A., Slusallek, P., Irgens, M., Gatellier, P., Köhler, J., … Yvon, F. (2020). The European language technology landscape in 2020: Language-centric and human-centric AI for cross-cultural communication in multilingual Europe. V N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk in S. Piperidis (ur.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (str. 3322–3332). European Language Resources Association. https://aclanthology.org/2020.lrec-1.407/

Reiss, K. (2000). *Translation criticism: The potentials and limitations: Categories and criteria for translation quality assessment*. St. Jerome Publishing.

Fran Ramovš oz. Julij Dub: digitalizacija in literarna analiza spregledanih protipravljic

DOI: 10.4312/Obdobja.45.67-74

Milena Mileva Blažić, Pedagoška fakulteta, Univerza v Ljubljani, Ljubljana;
milena.blazic@pef.uni-lj.si

Fran Ramovš (1890–1952), osrednja osebnost slovenskega jezikoslovja 20. stoletja, je svojo javno pot začel literarno: na pobudo Ivana Cankarja je pod psevdonimom Julij Dub objavil dve prozni besedili. Prispevek analizira *Božanstveno ljubezen: Japonsko tragedijo* (1910) in *Pripovedko o Mehiti* (1912), objavljeni v *Ljubljanskem zvonu* in danes dostopni na dLib.si. Besedili prevzameta pripovedni tip ATU 899A, njegove motive pa bodisi uresničita z zamenjavo nosilca bodisi jih zavrneta. Analiza je literarno-folkloristična, onomastična in simbolno-primerjalna.

Fran Ramovš, Julij Dub, protipravljica, ATU 899A, digitalizacija

Fran Ramovš (1890–1952), the central figure of twentieth-century Slovenian linguistics, began his public career as a writer: at Ivan Cankar's invitation he published two prose texts under the pseudonym *Julij Dub*. This article analyses *Božanstvena ljubezen: Japonska tragedija* (Divine Love: A Japanese Tragedy, 1910) and *Pripovedka o Mehiti* (The Tale of Mehita, 1912), both published in *Ljubljanski zvon* and now available on dLib.si. Both adopt the narrative type ATU 899A and either realize its themes with a substituted vehicle or reject them outright. The analysis is literary-folkloristic, onomastic, and symbolic-comparative.

Fran Ramovš, Julij Dub, anti–fairy tale, ATU 899A, digitization

1 Uvod

Fran Ramovš (1890–1952) je bil med prvimi profesorji novoustanovljene Univerze v Ljubljani (1919), njen rektor (1934–1935) in predsednik SAZU (1950–1952). Manj znano je, da je pod psevdonimom Julij Dub objavil dve prozni besedili: *Božanstveno ljubezen: Japonsko tragedijo* (Dub 1910) in *Pripovedko o Mehiti* (Dub 1912), ki v pregledih njegovega opusa nastopata kvečjemu kot biografska opomba.

V prispevku si zastavljamo dve vprašanji. Prvo je žanrsko in motivno: kakšno razmerje imata besedili do pripovednega tipa ATU 899A in do motivov, ki ga sestavljajo (poglavja 3.2–3.4). Drugo je gradivno in se umešča v simpozijski sklop Jezik in kulturna dediščina v digitalni dobi: kaj k obravnavi teh besedil prispeva digitalizacija periodičnega tiska (poglavje 3.1). Pri drugem ločujemo med tem, kar digitalizacija omogoča danes, in možnostmi za prihodnje korpusno delo, ki tu niso uresničene.

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

2 Gradivo in metoda

Analiza temelji na dveh virih iz Digitalne knjižnice Slovenije (dLib.si): *Božanstvena ljubezen: Japonska tragedija* (Dub 1910) in *Pripovedka o Mehiti* (Dub 1912), ki sta bili objavljeni v *Ljubljanskem zvonu*. Besedili obravnavamo z natančnim branjem izvirnikov; analiza ni računalniška, temveč tekstna. Navedki so preverjeni ob faksimilih in navedeni po izvirni paginaciji: strani 741–744 se nanašajo na *Božanstveno ljubezen*, strani 83–86 pa na *Pripovedko o Mehiti*.

Metodološko izhodišče daje uvod v katalog mednarodnih pripovednih tipov ATU (Aarne-Thompson-Uther): Uther (2011: 8–9) opozarja, da tip ni stalna merska enota ali mrtvo gradivo, temveč je prilagodljiv in ga je mogoče vključevati v nove tematske sestave in medije. Tip 899A ne sodi med čudežne, temveč med novelistične pravljice (ATU 850–999); ker sta Proppova morfologija in Lüthijev pojem protipravljice izpeljana iz čudežne pravljice, je njuna raba tu analogijska, na kar sproti opozarjamo.

Za raven motivov uporabljamo Thompsonovo (1946) opredelitev. Motiv je najmanjša prvina pripovedi, ki ima moč obstati v izročilu, in mora imeti nekaj nenavadnega in izstopajočega; motivi se po Thompsonu (1946: 415–416) delijo v tri razrede: nastopajoče like, prvine v ozadju dogajanja (čarobni predmeti, šege, verovanja) in posamezne dogodke. Delitev je odločilna: prvine v ozadju so zamenljivi nosilci, pri posameznem dogodku pa je motiv sam dogodek in zamenjava ni mogoča. Posamezna prvina tipa lahko združuje motiva iz dveh razredov; v takem primeru je zamenljiv le nosilec.[1]

Analizo dopolnjujeta Proppova (2005) morfologija in Lüthijev (2011) pojem protipravljice, simboliko pa osvetljujeta študiji Šmitka (2008) in Ferberja (2007).

3 Analiza

3.1 Digitalna dostopnost besedil

Portal dLib.si omogoča dostop do obeh besedil, delo z njima pa spreminja v dveh pogledih. Prvič, besedili sta opremljeni z identifikatorjem URN, ki omogoča stabilno sklicevanje na posamezno enoto in ne le na letnik. Drugič, iskanje po polnem besedilu razveže vezanost na letnik in številko: besedilo, objavljeno pod psevdonimom, je mogoče najti brez poznavanja avtorjevega pravega imena. Prav to je bilo odločilno pri zbiranju gradiva za pričujočo obravnavo. Hkrati velja opozoriti: pri raziskovanju avtorjev in/ali naslovov v starejši periodiki v zbirki dLib.si je treba upoštevati, da iskanje avtorjev v celoletnih letnikih ni vedno zanesljivo, zato je za večjo zanesljivost rezultatov smiselno uporabljati tudi letna kazala.

Za obrobnost besedil dostopnost ni odločilna. Zasenčila ju je avtorjeva znanstvena vloga: Ramovš je v literarnozgodovinskem spominu jezikoslovec, ne pisatelj, psevdonim pa je to ločnico utrdil. O mestu avtorjev in del v literarnem kanonu odločajo uredniška, znanstvena in pedagoška recepcija ter medbesedilna odmevnost (prim. Juvan 2000). Digitalna dostopnost besedil, na primer na dLib.si, povečuje njihovo vidnost, ne vpliva pa neposredno na njihovo kanonizacijo.

1 Oznake v oglatih oklepajih ([T41.1], [N343], [T81.6], [T91.6]) so motivi po Thompsonu (1961); črka pomeni razred (T – ljubezen, N – usoda), številka pa motiv.

3.2 Poimenovanja književnih oseb in podnaslov

V *Božanstveni ljubezni* nastopata neimenovani mladenič in mlada Girvonda, v *Pripovedki o Mehiti* pa kralj, njegova hči Mehita in mladenič Sima, ki ga kralj vzame v vojaško službo (83). Poimenovanja obravnavamo kot pomensko motivirane avtorske izbire, ne kot etimološko dokazljive tvorbe.

Sima je hipokoristično moško ime slovanskega izvora, skromno in nezaznamovano; poudarja stanovsko asimetrijo, ki je vir prepovedi.

Mehita je videti kot avtorska tvorjenka. Sklop *meh-* (prim. *mehek*) se v besedilu dvakrat dobesedno ponovi, saj junakinja seže po harfi z »mehko roko« (84), kača pa ji strupeni zob zasadi »v mehko ustnico« (85). Gre za sozvočje, ne za tvorbeno izpeljavo.

Girvonda je tvorjenka brez slovenskega ali japonskega zaledja; ime torej podnaslova ne podpira, temveč ga postavlja pod vprašaj. Podnaslov *Japonska tragedija* namreč ostaja odprt: besedilo se ne sklicuje na japonsko, temveč na indijsko-budistično izročilo. Ljubezen mladeniča in Girvonde je primerjana z ljubeznijo, »ki jo je živel le veliki Buddha« (743), konec pa ju položi »v večni Karmi« (744). Pridevnik *japonski* tu deluje kot splošna oznaka za daljno in nezahodno, kar je značilno za recepcijo vzhodnih motivov v slovenski moderni.

3.3 Tip ATU 899A in realizacija njegovih motivov

Obe besedili se navezujeta na tip ATU 899A, *Piram in Tizba*. Po Utherju (2011: 522–523) se zaljubljenca iz sosednjih družin ljubita, očeta pa jima prepovesta srečanja; sporazumevata se skozi razpoko v zidu [T41.1]; dogovorita se za nočno srečanje pod murvo; Piram najde krvavo ruto, sklepa, da je Tizba mrtva, in se ubije [N343]; Tizba se vrne, najde mrtvega ljubimca in se ubije z njegovim mečem [T81.6]; rdeči plodovi murve ohranjajo spomin na nesrečni konec. Tip je pripoved klasičnega izvora, iz Ovidijevih *Metamorfoz* (Ovidij 2019: IV, 55–166).

Besedili prevzameta pripovedno konstelacijo tipa, njegove motive pa razločujeta glede na njihovo naravo; merilo daje Thompsonova (1946: 415–416) delitev. Spominsko funkcijo, ki jo v tipu nosi murva [drevo], prevzame v *Pripovedki o Mehiti* stari hrast, v *Božanstveni ljubezni* pa morje.

Motiv sporazumevanja skozi razpoko v zidu [T41.1] se uresniči z zamenjavo nosilca. Razpoka sodi med prvine v ozadju in je zamenljiva; ohrani se dogodek, torej edini prehod skozi neprepustno pregrado. V *Božanstveni ljubezni* to vlogo prevzameta molk in dotik – »Molk mu je govoril, da ne sme, ne sme.« (742), »Roka je legla na njegovo ramo.« (742) –, v *Pripovedki o Mehiti* pa »debela vrv, ki je peljala v okno njegove hčere Mehite« (83).

Motiv zmote in samomora [N343] je zavrnjen. Pri njem motiv sovpada s samim dogodkom, torej z usodno zmoto; če zmoto odstranimo, ne dobimo različice motiva, temveč zgolj smrt. Zmote v besedilih ni. Vzrok smrti ni napačen sklep, temveč izrečena sodba: »‘Grešila!’ je potrdil kralj in sodil.« (84) Zamenjava naključja s sodbo ni motivna različica, temveč sprememba vzročne zgradbe: kjer tip smrt pripisuje nesporazumu, jo Ramovš pripiše odločitvi osebe, ki ima moč prepovedovati in kaznovati.

Motiv dekletove smrti za ljubimcem [T81.6] je uresničen le v enem besedilu. V *Božanstveni ljubezni* je smrt skupna in hotena (744). V *Pripovedki o Mehiti* junakinja ne umre po lastni odločitvi: usmrti jo kačja kraljica – »Zlobno jo je poljubila in zasadila v mehko ustnico strupeni zob.« (85) Vzročna zveza med obema smrtma se ohrani, dejanje pa preide na drugega nosilca.

Besedili tipu dodata motiv, ki ga v njem ni: ljubezen med osebama visokega in nizkega stanu [T91.6]. Tip predpostavlja zaljubljenca iz sosednjih družin, torej stanovsko izenačena; v *Pripovedki o Mehiti* je prepoved utemeljena prav v razliki med kraljevo hčerjo in vojakom. Gre za Ramovšev dodatek, ne za prvino tipa.

Prvine so povzete v Tabeli 1.

Prvina tipa (Uther 2011: 522–523)	**Razred (Thompson 1946: 415–416)**	***Božanstvena ljubezen***	***Pripovedka o Mehiti***	**Status**
prepoved očetov	dogodek	»Ne smeva, so rekli starši.« (742)	»[N]jim je bilo prepovedano iti po stopnicah v gornje prostore.« (83)	uresničen
sporazumevanje skozi razpoko v zidu [T41.1]	ozadje + dogodek	»Molk mu je govoril, da ne sme, ne sme.« (742)	»[D]ebelo vrv, ki je peljala v okno njegove hčere Mehite« (83)	uresničen, nosilec zamenjan
zmota in samomor [N343]	dogodek	»Njene ustnice so se strnile v večen poljub.« (744)	»‘Grešila!’ je potrdil kralj in sodil.« (84)	zavrnjen
dekletova smrt za ljubimcem [T81.6]	dogodek	»Izginjala sta in izginila v večnem poljubu« (744)	»[Z]asadila v mehko ustnico strupeni zob.« (85)	uresničen/ dejanje preneseno
murva [drevo] ohrani spomin	ozadje	»[V] večnem morju, sredi večnega miru« (744)	»Prišli so do starega hrasta sredi širnega polja.« (84)	nosilec zamenjan
ljubezen med osebama visokega in nizkega stanu [T91.6]	dogodek	brez ustreznice	»[K]ralj ga je vzel seboj, da bi mu služil v vojski.« (83)	dodano zunaj tipa

Tabela 1: Prvine tipa ATU 899A in njihova realizacija v Ramovševih besedilih.

Isti postopek pokaže tudi Proppova analiza funkcij (2005). Značilne so tiste, na katerih se pravljični obrazec prelomi: prepoved in kršitev prepovedi sta izvedeni v celoti, zapolnitev manka in poroka izostaneta, preobrazba pa se uresniči šele po smrti. Diagnostične funkcije prinaša Tabela 2.

Oznaka in funkcija (Propp 2005)	***Božanstvena ljubezen***	***Pripovedka o Mehiti***
γ 2 prepoved	»Ne smem priti več semkaj.« (742)	»Celo kadar so bili na dvorišču, se niso smeli ozreti navzgor« (83)
δ 3 kršitev prepovedi	»Roka je legla na njegovo ramo.« (742)	»Sima je pozabil na vse, pogledal navzgor in zagledal ob oknu Mehito« (83)
A 8 škodovanje	izostane: prepoved je brezosebna (742)	»'Grešila!' je potrdil kralj in sodil.« (84)
J 19 zapolnitev manka	izostane	izostane: »Mehita je postala njih kraljica.« (84) – pomoč, ki ne reši
Š 29 preobrazba	»[V] večni Karmi« (744)	»[T]ri svetle zvezde: Mehita, Sima in kralj« (86)
T 30 kazen (škodljivca)	izostane: brez škodljivca	»[O]b poljubu mu je počilo skesano srce.« (86)
U 31 poroka	izostane; nadomesti jo smrt (744)	izostane; nadomesti jo smrt (84–85)

Tabela 2: Diagnostične Proppove funkcije.

Diagnostične funkcije je mogoče strniti v tri nize; Ø označuje odsotnost funkcije.[2] Pravljični niz: γ (prepoved) → δ (kršitev) → A (škodovanje) → J (zapolnitev manka) → Š (preobrazba) → T (kaznovanje škodljivca) → U (poroka). *Pripovedka o Mehiti*: γ → δ → A → JØ → Š → T → UØ. *Božanstvena ljubezen*: γ → δ → AØ → JØ → Š → TØ → UØ. *Pripovedka o Mehiti* torej izpelje vse diagnostične funkcije razen zapolnitve manka in poroke; škodljivec je kaznovan (86), poroke pa ni. Kralj hkrati pripada krogu delovanja škodljivca ter krogu carične in njenega očeta, kar Propp (2005: 94) dopušča: »Ena oseba obsega več krogov delovanja.« V *Božanstveni ljubezni* je prepoved brezosebna, zato besedilo nima škodljivca in z njim niti funkcije A niti T. Prenos je tudi tu analogijski, saj so funkcije prevzete iz čudežnih pravljic.

Prepoved je v *Pripovedki o Mehiti* podvojena: poleg prepovedi hlapcem velja prepoved pogleda navzgor (83). Kršitev je dobesedno vizualna, a moralno neznatna: prestopek ni tatvina ali uboj, temveč pogled. Nesorazmerje med prestopkom in kaznijo prenese etično vprašanje z ljubimcev na tistega, ki prepoveduje (prim. Zipes 2012: 1–14).

Pojem, ki ta postopek zajame, je *protipravljica* (Lüthijev *Gegenmärchen*; v slovenskem prevodu tudi *antipravljica*). Lüthi (2011: 104) je ne opredeli kot pravljice z nesrečnim koncem, temveč kot možnost, ki je pravljici notranja: »Pravljica znotraj sebe vsebuje protipravljico.« Ob junakih stojijo protijunaki, tudi junak lahko ravna neprimerno, obstajajo pa zgodnje in razdrobljene forme, ki kažejo prav neuspeh ali propad glavnega lika (Lüthi 2011: 104–105). Ramovševi besedili torej ne stopata iz pravljice, temveč uresničita možnost, ki je v njej navzoča. Prenos je analogijski: Lüthi govori o formah ljudskega izročila, Ramovš pa piše avtorsko besedilo.

2 Grške in latinične oznake funkcij (γ, δ, A, J, Š, T, U) povzemamo po slovenski izdaji Proppove *Morfologije pravljice* (2005). Proppov T (kaznovanje) ni Thompsonov razred T.

3.4 Navpična os: drevo, kača, ptice

Tip ATU 899A postavlja prehod med ljubimcema vodoravno: sosednji hiši, zid med njima, razpoka v višini oči. Ramovš to os zasuka v navpičnico. Vrv se spušča po grajskem zidu (83), prepovedan je pogled navzgor (83), Sima je obešen nad zemljo (84), Mehitino truplo leži na stolpu (85), gavrani letajo »od hrasta na stolp, od stolpa na hrast« (85), kače so spodaj, zvezde zgoraj (86). Nasprotje zgoraj : spodaj je pri Proppu (2005) ena od stalnih parnosti pravljice; pri Ramovšu postane nosilna os.

Navpičnica ustreza slovanski predstavi o svetovnem drevesu (Šmitek 2008: 134): »Križišče je simbolično predstavljalo središče sveta, drevo ob njem pa svetovno drevo (*arbor mundi*). V koreninah tega drevesa je po slovanskih predstavah prebivalo htonsko božanstvo v podobi kače (Veles), na njegovem vrhu pa je domoval bog gromovnik.« Struktura se uresniči dobesedno: kače pridejo »do osamljenega hrasta« (85), gavrani sedijo na vejah (85).

Šmitek (2008: 138) navaja pravljice o visokem drevesu, po katerem se mladenič povzpne na nebo, v sfere zvezd, Lune in Sonca. Ramovš vzorec obrne: mladenič se po drevesu ne povzpne, temveč na njem obvisi, saj »se je Sima gugal na veji« (84); v zvezdno sfero ga povzdigne šele smrt s preobrazbo v zvezde na koncu besedila: »tri svetle zvezde: Mehita, Sima in kralj« (86). Vzpon je zamenjan z obešenjem, cilj ostane isti.

Kača sodi v Thompsonov prvi razred, med nastopajoče like, in opravi pot, kakršne pravljična nasprotnica nima. Šmitek (2008: 139) ugotavlja, da slovenske pripovedi kačam pripisujejo kralja ali kraljico in da je bil indijski drevesni kult povezan s čaščenjem kač, ki naj bi živele v hierarhiji, podobni človeški. Besedilo to hierarhijo uresniči dvojno: najprej postane kraljica kač človeška junakinja, saj »Mehita je postala njih kraljica« (84), nato nastopi izvirna kačja kraljica, ki jo usmrti (85) in nad njenim truplom izreče kesanje: »Ker nisem hotela tvoje solze ljubezni, Mehita, točim zate solzo kesanja.« (86) Ferber (2007: 186–187) ugotavlja, da kača kot notranja podoba od romantike naprej pomeni predvsem kesanje ali krivdo; Ramovš to podobo premesti iz notranjosti med nastopajoče like.

Najdosledneje je zavrnitev izpeljana v simboliki ptic. Ferber (2007: 167) dokumentira gavrana kot ptico bojišč in vislic ter kot izkljuvalca oči usmrčenih, Ramovš pa takšno prizorišče postavi in ga zanika. Gavrani priletijo k obešenemu Simi, »a nobeden se ga ni lotil, le žalostno so krakali smrtno pesem« (84). Pri Mehiti je zanikanje izrecno: kralj čaka, da bodo ptice raztrgale njeno truplo, »[p]a gavrani se niso dotaknili Mehite« (85). Namesto tega »[n]osile so besede ljubezni od Sime k Mehiti, od nje k njemu« (85). Ferber (2007: 168) dokumentira tudi nasprotno izročilo, gavrane kot dobrotnike; Ramovš ptic torej ne izumlja, temveč jih premakne z enega tradicionalnega pola simbola na drugega, prav tja, kjer se tradicionalna vloga mrhovinarja ne pričakuje.

V *Božanstveni ljubezni* sta obe vlogi ptičjega simbola združeni v enem liku. Galeb je najprej glasnik – po Ferberju (2007: 25) so ptice veljale za utelešenja ali glasnike bogov – in nato znamenje smrti: »Galeb se je mrtev zgrudil s krmila in izginil v valovih.« (744) Umre v trenutku, ko ljubimca izgineta. V *Pripovedki o Mehiti* sta vlogi

razdeljeni med ptico nad dvoriščem (83) in gavrane. Pomenljiva je leksikalna izbira: Ramovš dosledno uporablja poimenovanje *gavran* in nikjer *krokar*. O slovarski zaznamovanosti izrazov *Girvonda*, *Mehita*, *Sima* itn. ob nastanku besedil je mogoče sklepati le na podlagi tedanjih slovarskih virov, zato vprašanje ostaja odprto.

Ferber (2007: 179) med pomeni morja našteva brezčasnost: »Izginjala sta in izginila v večnem poljubu, v večni Karmi, v večnem morju« (744). V *Pripovedki o Mehiti* morja ni; njegovo mesto zasedejo solze (84, 86). Pregled prinaša Tabela 3.

Simbol (Ferber 2007)	***Božanstvena ljubezen***	***Pripovedka o Mehiti***
ptica, glasnik bogov (25)	»Ptica si je upala sesti sredi ljudi« (743)	»[N]ad njim je pela nekje prekrasna ptica nebeško pesem« (83)
gavran, ptica vislic (167–168)	»Galeb se je mrtev zgrudil s krmila« (744)	»[A] noben se ga ni lotil« (84); »[P]a gavrani se niso doteknili Mehite.« (85)
morje (179)	»[V] večnem morju, sredi večnega miru« (744)	»[S]rebrne solze, ki so padale na tla« (84)
kača, podoba kesanja (186–187)	brez ustreznice	»[T]očim zate solzo kesanja.« (86)

Tabela 3: Literarni simboli (Ferber 2007) v Ramovševih besedilih.

4 Sklep

Ramovševi besedili iz let 1910 in 1912 nista biografska zanimivost, temveč premišljena obravnava mednarodnega pripovednega tipa. Prevzameta konstelacijo tipa ATU 899A, njegove motive pa razločujeta glede na njihovo naravo: kjer je motiv vezan na prvino v ozadju, zamenjata nosilca [T41.1]; kjer motiv sovpada s samim dogodkom, ga zavrneta [N343]; motiv dekletove smrti za ljubimcem [T81.6] uresničita le v enem besedilu, motiv ljubezni med osebama visokega in nizkega stanu [T91.6] pa tipu dodata. Isti postopek deluje na ravni simbolov: vodoravna os tipa se zavrti v navpičnico svetovnega drevesa, vzpon na nebo se spremeni v obešenje, gavrani, ki bi trupli morali raztrgati, pa prenašajo besede ljubezni.

Razlika med besediloma je v tem, komu je pripisana odgovornost za razplet. V *Božanstveni ljubezni* prepoved ostaja brezosebna (»so rekli starši«, 742), zato besedilo nima škodljivca. V *Pripovedki o Mehiti* je prepoved poosebljena v kralju, ki izreče tudi sodbo (84), in prav njega na koncu doleti kazen: »[P]adel z vso težo na tla in poljubil hčerine noge. Ob poljubu mu je počilo skesano srce.« (86) Protipravljični postopek se torej izvede z različno razporeditvijo pripovednih vlog.

Digitalizacija je pogoj za dostopnost, ne metoda analize. Kot nadaljnje delo se ponujata kritična izdaja obeh besedil in korpusna primerjava Ramovševega literarnega jezika z jezikom njegovih znanstvenih del. Ta bi pokazala razlike med njegovim zgodnjim literarnim jezikom in poznejšim normativnim jezikom, vendar je izvedljiva šele po digitalizaciji besedil in njihovi pretvorbi v strojno berljivo obliko. Smiselna bi bila tudi priprava ilustrirane izdaje obeh pravljic oz. protipravljic v okviru študentskega projekta na Univerzi v Ljubljani.

Avtorica se zahvaljuje Digitalni knjižnici Slovenije (dLib.si) za dostop do digitaliziranega gradiva. Pri jezikovnem pregledu in oblikovni ureditvi besedila sta bili uporabljeni orodji InstaText in Claude.ai (Anthropic). Vse vsebinske odločitve, teoretske interpretacije, analitične ugotovitve in akademska odgovornost za objavljeno besedilo ostajajo izključno avtoričine.

Literatura

Dub, J. [Fran Ramovš]. (1910). Božanstvena ljubezen: Japonska tragedija. *Ljubljanski zvon*, *30*(12), 741–744. https://www.dlib.si/details/URN:NBN:SI:DOC-C6LPE5HS

Dub, J. [Fran Ramovš]. (1912). Pripovedka o Mehiti. *Ljubljanski zvon*, *32*(2), 83–86. http://www.dlib.si/details/URN:NBN:SI:DOC-OLB92YQS

Ferber, M. (2007). *A dictionary of literary symbols* (2. izd.). Cambridge University Press.

Juvan, M. (2000). *Intertekstualnost*. DZS.

Lüthi, M. (2011). *Evropska pravljica: Forma in narava*. Sophia.

Ovidij (2019). *Metamorfoze* (K. Gantar, prev.). Modrijan.

Propp, V. J. (2005). *Morfologija pravljice*. Studia humanitatis.

Šmitek, Z. (2008). Paralele med indijsko in slovensko mitologijo: Sledovi skupne indoevropske dediščine. *Studia Mythologica Slavica*, *11*, 127–145. https://doi.org/10.3986/sms.v11i0.1692

Thompson, S. (1946). *The folktale*. The Dryden Press.

Thompson, S. (1961). *Motif-index of folk-literature* (zv. 1–6). Indiana University Press.

Uther, H.-J. (2011). *The types of international folktales* (1. del; FF Communications 284). Academia Scientiarum Fennica.

Zipes, J. (2012). *The irresistible fairy tale*. Princeton University Press.

Polavtomatski delotok za urejanje nabora leksemov v Digitalni slovarski bazi za slovenščino na primeru glagolov

DOI: 10.4312/Obdobja.45.75-81

Jaka Čibej, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana;
jaka.cibej@ff.uni-lj.si

V prispevku predstavljamo polavtomatski postopek za odpravljanje napak pri strojno dodanih leksemih v Digitalni slovarski bazi za slovenščino na primeru glagolov. Delotok je sestavljen iz več korakov (primerjava z obstoječimi viri, preverjanje z modeli strojnega učenja in velikimi jezikovnimi modeli) in pomaga zmanjšati količino dela, ki ga morajo opraviti človeški pregledovalci, ter s tem olajša strojno širitev in sprotno posodabljanje slovarske baze s kandidati iz korpusov. Osredotočamo se na dva poglavitna problema: določanje legitimnosti leme (oz. odpravljanje lematizacijskih napak) in določanje osnovnih oblikoskladenjskih značilnosti leksema (pri glagolih je to vid).

leksemi, oblikoslovni leksikon, veliki jezikovni modeli, slovenščina, Digitalna slovarska baza za slovenščino

This article presents a semi-automatic approach to correcting errors in automatically extracted lexemes in the Digital Dictionary Database for Slovene, using verbs as an example. The workflow consists of several steps (comparison with existing resources, verification with machine learning models, and large language models) and helps reduce the amount of human work, thereby facilitating the automatic expansion and real-time updating of the dictionary database with corpus candidates. It focuses on two main problems: determining the legitimacy of lemmas (i.e., correcting lemmatization errors) and determining the basic morphosyntactic features of lexemes (aspect in the case of verbs).

lexemes, morphological lexicon, large language models, Slovenian, Digital Dictionary Database of Slovene

1 Uvod

Digitalna slovarska baza za slovenščino (DSBS; Kosem idr. 2021) je zasnovana kot osrednja odprto dostopna podatkovna baza s strojno berljivimi jezikoslovnimi informacijami o slovenskih besedah. V času pisanja tega prispevka (april 2026) vsebuje približno 380.000 enobesednih leksikalnih enot in z njimi povezanih leksemov,[1] na katere se osredotočamo v tem prispevku. Oblikoslovna podlaga DSBS je Slovenski oblikoslovni

1 Leksikalne enote v DSBS vsebujejo jezikovne informacije na nivoju pomena (npr. definicijo in kvalifikatorje na ravni pomena, pomenske povezave z drugimi leksikalnimi enotami – npr. sopomenske in protipomenske), leksemi pa oblikoslovne in izgovorne informacije (npr. pregibne in naglašene oblike ter fonetične transkripcije).

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

leksikon Sloleks (Čibej idr. 2022). Med zadnjo sistematično širitvijo nabora v okviru projekta Razvoj slovenščine v digitalnem okolju[2] (RSDO; različica Sloleks 3.0) je bilo v DSBS dodanih približno 265.000 novih leksemov, izluščenih iz frekvenčnih seznamov korpusa Gigafida 2.0 (Krek idr. 2020). Kandidati so bili pred vključitvijo le delno ročno pregledani in zato vsebujejo tudi lematizacijske (**izparivati* namesto *izpareti*) in oblikoskladenjske napake (npr. dva leksema za *zanagramirati*; enkrat kot *nedovršni glagol in drugič kot dovršni glagol) ter zatipke iz korpusa (**izkorstiti* namesto *izkoristiti*). Ker bi bilo ročno pregledovanje in urejanje zelo zamudno in tako z vidika človeških kot gmotnih virov drago, je podatke najprej smiselno čim bolj učinkovito strojno prečistiti ter na ta način čim bolj zmanjšati količino ročnih popravkov za človeške pregledovalce. Za ta namen smo zasnovali polavtomatski delotok,[3] ki sestoji iz različnih stopenj in za popravljanje leksemskih kandidatov oz. odkrivanje napačno generiranih vsebin uporablja kombinacijo hevurističnih metod (pravila, primerjava z drugimi viri), preprostejše modele strojnega učenja (npr. metoda podpornih vektorjev, logistična regresija) in velike jezikovne modele (predvsem na tistih mestih, kjer se manj zahtevne metode strojnega učenja zaradi pomanjkanja konteksta odrežejo slabše, npr. pri določanju nekaterih pregibnih vzorcev; *stanovati* – **stanovam*/*stanujem*; oz. pri določanju, ali je posamezni leksem posledica lematizacijske napake ali ne).

Delotok pokriva vse besedne vrste in upošteva njihove specifike (npr. pri pridevnikih je treba določiti, ali jih je mogoče obrazilno stopnjevati; pri samostalnikih moškega spola je pomembno, ali so slovnično označeni kot živi ali neživi), v tem kratkem prispevku pa prve stopnje v delotoku ponazarjamo na primeru urejanja glagolskih leksemov (ki jih je trenutno v DSBS približno 18.000). Med drugim s postopkom preverimo, ali so vključeni leksemi legitimni in ali imajo pripisane ustrezne osnovne oblikoskladenjske značilnosti glede na sistem MULTEXT-East v6[4] (pri glagolih je npr. glavna razlikovalna lastnost glagolski vid). Za vključitev v DSBS je namreč pomembno, da ima posamezni leksem ustrezno lemo in osnovne oblikoskladenjske značilnosti; če so podatki že na tej stopnji napačni, se v nadaljnjih korakih, ko se npr. s Pregibalnikom[5] (tj. orodjem za strojno širjenje DSBS z leksikonskimi podatki) generirajo še posamezne pregibne oblike, napake propagirajo. Hkrati so pregled lem in osnovnih oblikoskladenjskih značilnosti do zdaj opravljali človeški označevalci, kar terja precej časa in virov. Preveriti želimo torej, do kolikšne mere lahko tovrstno rutinsko pregledovanje opravimo s pomočjo jezikovnih modelov.

Najprej na kratko predstavimo podatke in metodologijo (poglavje 2), nato pa opravimo analizo (poglavje 3) in prispevek sklenemo z ugotovitvami in predlogi za nadaljnje delo (poglavje 4).

2 https://rsdo.slovenscina.eu/

3 Delotok na določenih točkah še vedno predvideva človeški pregled, zato ni povsem avtomatski, vendar znatno zmanjša količino podatkov, ki jih je treba pregledati.

4 Shema oblikoskladenjskih oznak MULTEXT-East v6: https://nl.ijs.si/ME/V6/msd/html/msd-sl.html.

5 Pregibalnik je na voljo kot odprto dostopna koda na repozitoriju Github (https://github.com/clarinsi/SloInflector) in kot API-storitev (https://orodja.cjvt.si/pregibalnik/docs).

2 Priprava podatkov in metodologija

Iz trenutne različice DSBS smo izvozili vse glagolske lekseme, ki jih je skupno 18.120. Od tega jih je bilo 10.242 (57 %) prisotnih že v različici pred strojno širitvijo v okviru RSDO, 7878 (43 %) pa jih je bilo dodanih pozneje. Osredotočili smo se na dva poglavitna podatka: na *lemo* in na *glagolski vid*, ki sta prvi korak pri odločitvi, ali določeno enoto obravnavamo kot legitimen leksem v DSBS.

Vse leme glagolskih leksemov so bile na neki točki že ročno preverjene v predhodnih projektih in raziskavah.[6] Problematični leksemi so bili ročno identificirani in zato predstavljajo dobro izhodišče za preverjanje zanesljivosti strojnih metod. Pridobljene rezultate lahko primerjamo z ročnimi oznakami in na ta način ocenimo, kako uspešni so npr. veliki jezikovni modeli pri razlikovanju med ustreznimi leksemi na eni strani (npr. *sokreirati*, *odmeditirati*) in oblikovno nestandardnimi leksemi (tj. leksemi, ki izkazujejo določene sistematične oblikovne odklone od standarda, npr. izpust *-j-* v sklopih *-lj-* in *-nj-*: **izbolšati*, **ponotraniti*) ter zatipki (**komenirati – komentirati*, **odmonitrati – odmontirati*) na drugi strani. Za preverjanje legitimnosti lem smo v prvem koraku najprej preverili, kateri glagolski leksemi v DSBS se ne končajo na *-ti*/*-či* (na ta način smo izločili lekseme, ki so bili napačno označeni kot glagolski, npr. *pekan*, *prispevalec*); preostali nabor smo primerjali z naborom glagolov v *Slovarju slovenskega knjižnega jezika* (SSKJ²)[7] in na ta način potrdili ustreznost leme pri 12.775 leksemih; za preverbo lem pri preostalih 5345 glagolskih leksemih smo zasnovali robustna navodila[8] za štiri velike jezikovne modele (Gemini 3 Flash, GPT 5.1, GaMS-9B-Instruct-Lex, GaMS3-12B-Instruct-Lex) in jih preizkusili na manjšem vzorcu (10 enot) ter ročno evalvirali rezultate, dopolnili navodila in s končno različico označili celoten preostanek nepotrjenih lem. Rezultate predstavljamo v poglavju 3.1, primer navodila (za GaMS3-12B-Instruct-Lex) pa dodajamo spodaj:

> Tu je seznam slovenskih glagolov v osnovni obliki. Si strokovnjak za slovensko leksikografijo in tvoja naloga je, da pomagaš določiti, ali gre za legitimne glagolske leme ali zatipke/nestandardne oblike. Rezultat vrni v formatu JSON na naslednji način: [{"lema":"", "popravljena_lema":"", "status":""}].

Pri določanju glagolskega vida (pri katerem po sistemu MULTEXT-East v6 razlikujemo med dovršnimi, nedovršnimi in dvovidskimi glagoli) smo nabor glagolskih leksemov prav tako najprej primerjali s SSKJ². Razvili smo tudi več modelov strojnega učenja[9] (multinomski naivni Bayesov klasifikator, linearni klasifikator s podpornimi

6 Glagolske leme so bile ročno preverjene npr. pri izdelavi seznama potencialno nestandardnega besedišča MEZZANINE-NstdLex 1.0 (Čibej idr. 2026) in pri strojni gradnji arhetipskih besedotvornih dreves (Čibej 2026).

7 *Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja*, www.fran.si (dostop 29. 4. 2026).

8 Vsa navodila za velike jezikovne modele so objavljena na platformi Github: https://github.com/jakacibej/obdobja2026_verb_lemma_and_aspect. Za Gemini in GPT smo uporabili angleška navodila, za različice modela GaMS pa slovenska.

9 Modeli strojnega učenja so bili naučeni s paketom scikit-learn (Pedregosa idr. 2011).

vektorji, logistična regresija) na podlagi statistično pomembnih *n*-gramskih značilk (tj. tipičnih končnih in začetnih besednih delov, ki statistično pomembno[10] razlikujejo med dovršnim, nedovršnim in dvovidskim vidom; npr. *-irati*, *-evati*, *-niti*), nazadnje pa določanje glagolskega vida preverili še z velikimi jezikovnimi modeli. Rezultate predstavljamo v poglavju 3.2, primer navodila (za GaMS3-12B-Instruct-Lex) pa dodajamo spodaj:

> Prevzemi vlogo strokovnjaka za slovensko leksikografijo. Podal ti bom glagol, ti pa moraš napisati, ali gre za dovršni glagol, nedovršni glagol ali dvovidski glagol. Rezultat vrni v formatu JSON na naslednji način: [{"glagol":"", "vid":""}] Glagol je »asociirati«.

3 Analiza rezultatov

3.1 Identifikacija nelegitimnih lem z velikimi jezikovnimi modeli

Med evalvacijo uspešnosti modelov pri določanju legitimnosti leme smo upoštevali, ali je model vsaj na en način signaliziral, da je lema na nek način problematična: npr. tako, da je popravil izvirno lemo (**prepljavljati – preplavljati*), ji dodal po svoji oceni nevtralnejši sinonim (*zresetirati – ponastaviti*/*zresetirati*, *pucati – čistiti*) oz. dodal komentar pri statusu (npr. *odglumiti*, ki mu je npr. Gemini 3 Flash pripisal dodatno oznako, da gre za srbohrvaški vpliv). Med 5345 glagolskimi leksemi, ki smo jih dali v ocenjevanje modelom, je 360 takšnih, ki so bili ročno označeni kot problematični z vidika oblike (zatipki ali oblikovno nestandardni leksemi). Rezultate klasifikacije prikazuje Tabela 1.

Model	TP	FP	FN	TN	A	P	R	F1
Gemini 3 Flash	282	1834	78	2865	62,20	13,33	78,33	22,78
GPT 5.1	203	500	157	4198	87,01	28,88	56,39	38,19
GaMS-9B-Instruct-Lex	8	7	352	4691	92,90	53,33	2,22	4,27
GaMS3-12B-Instruct-Lex	1	4	356	4697	92,88	20,00	0,28	0,55

Tabela 1: Rezultati klasifikacije nelegitimnih lem z velikimi jezikovnimi modeli (TP – pravi pozitivni primeri, FP – lažno pozitivni, FN – lažno negativni, TN – pravi negativni, A – točnost, P – natančnost, R – priklic, F1 – mera F1).

Visoko število lažno pozitivnih primerov pri rezultatih modela Gemini 3 Flash odraža dejstvo, da je model (morda tudi zaradi preprostih navodil) preveč ozko interpretiral koncept legitimnosti in z ustreznimi statusi (npr. sleng, pogovorni jezik) opozoril tudi npr. na veliko število primerov, ki so le registrsko nestandardni (npr. *kengurujčkati*, *rajcati*, *kiksati*). GPT 5.1 ni poleg osnovnega statusa (*legitimno* vs. *nelegitimno*) podal nobenih drugih informacij, zanj pa je značilnih več lažno negativnih primerov in manj pravih pozitivnih. Različici modela GaMS se pri nalogi nista dobro odrezali, saj izkazujeta nizek priklic in sta prepoznali le peščico primerov, morebiti tudi zato, ker sta bili doučeni s podatki iz DSBS, ki so v trenutni različici še neprečiščeni – njuno delovanje

10 Pri najboljšem modelu strojnega učenja je bilo uporabljenih 4252 *n*-gramskih značilk (do dolžine vključno 5 znakov), ki so se izkazale kot statistično pomembne glede na Kruskal-Wallisov test (Kruskal in Wallis 1952).

bomo preverili z nadaljnjimi preizkusi na drugih besednih vrstah in z drugačnimi navodili. Glede na kvalitativno analizo rezultatov je videti, da so najboljša podlaga za nadaljnjo leksikografsko analizo izhodni podatki iz modela Gemini 3 Flash, ki je sicer spregledal dobro petino problematičnih primerov (npr. **vmestiti*, **vgrizniti*, **ukleniti*, **asocirati*), a je podrobno razdelal druge nestandardne primere. Pri iskanju nelegitimnih enot na frekvenčnem seznamu iz besedilnega korpusa je torej ocena velikih jezikovnih modelov leksikografu lahko v pomoč – v zgornjem primeru bi npr. ocena Gemini 3 Flash ročni pregled omejila na približno dobro tretjino primerov.

3.2 Strojno določanje glagolskega vida

Strojno določanje glagolskega vida z modeli strojnega učenja in velikimi jezikovnimi modeli smo evalvirali na podmnožici 10.927 glagolskih leksemov, ki so prisotni v SSKJ[2] in imajo enak vid pripisan tudi v DSBS (lekseme, pri katerih so prisotne razlike v pripisanem vidu med SSKJ[2] in DSBS, smo izločili, saj gre pri nekaterih za dvoumne lekseme, pri nekaterih (predvsem na strani DSBS) pa tudi za napake; tako smo upoštevali le najbolj stabilne in zanesljive podatke). Rezultate prikazuje Tabela 2.

Model	A	P	R	F1
Gemini 3 Flash	97,22	96,08	94,51	95,26
GPT 5.1	95,56	91,87	94,50	93,09
Linearni klasifikator s podpornimi vektorji	93,02	87,66	92,09	89,50
Logistična regresija	92,82	87,27	92,43	89,38
Multinomni naivni Bayesov klasifikator	82,46	77,17	84,73	79,75
GaMS3-12B-Instruct-Lex	72,87	41,00	39,56	38,22
GaMS-9B-Instruct-Lex	66,78	35,31	38,49	35,01
Večinski klasifikator	54,45	–	–	–

Tabela 2: Rezultati klasifikacije glagolskega vida z velikimi jezikovnimi modeli in modeli strojnega učenja (razvrščeno po padajoči meri makro-F1; A – točnost, P – natančnost (makro), R – priklic (makro), F1 – mera F1 (makro)).

Tako kot pri identifikaciji nelegitimnih lem je najboljše rezultate pri napovedovanju glagolskega vida dosegel model Gemini 3 Flash, sledi mu GPT 5.1. Poudariti pa je treba, da jima z ne preveč izrazito razliko v meri F1 (le približno 3 odstotne točke) sledita precej preprostejša in računsko manj zahtevna modela strojnega učenja na podlagi *n*-gramov (linearni klasifikator s podpornimi vektorji in logistična regresija). Nekoliko slabši je multinomni naivni Bayesov klasifikator, pri dnu pa sta različici modela GaMS, pri katerih se med drugim kaže zelo očitna pomanjkljivost, da imata težave s prepoznavo dvovidskih glagolov: čeprav jih je bilo v naboru skupno 908, GaMS-9B-Instruct-Lex kot dvovidskega ni klasificiral nobenega glagola, GaMS3-12B-Instruct-Lex pa le 29. Vidna pa je še ena razlika med obema različicama: GaMS-9B-Instruct-Lex je zelo nagnjen k prekomerni klasifikaciji glagolov kot nedovršnih (62 % vseh oznak oz. 6774; v resnici jih je le 37 % oz. 4042). Pri večji različici GaMS-12B-Instruct-Lex je klasifikacija nekoliko bolj uravnotežena (63 % oznak za dovršne, približno 37 % za nedovršne,

0,2 % za dvovidske; v resnici je razporeditev naslednja: 54 % dovršnih, 37 % nedovršnih, 9 % dvovidskih).

4 Zaključek

V prispevku smo predstavili prve rezultate pri implementaciji polavtomatskega delotoka za urejanje leksemov v Digitalni slovarski bazi za slovenščino, ki je že odprto dostopno na voljo preko javnega API-ja.[11] Rezultati kažejo, da lahko veliki jezikovni modeli in modeli strojnega učenja pomagajo pri leksikografskem delu – model Gemini 3 Flash se je pri iskanju nelegitimnih lem izkazal za sicer ne povsem točnega, a trenutno ponuja najboljše izhodišče za ročni pregled – enako velja tudi za strojno določanje vida. V naslednjem koraku bomo ročno pregledali razhajanja in popravke implementirali v bazi, postopek pa bomo ponovili tudi na drugih besednih vrstah (in drugih vrstah podatkov, kot so npr. pregibne in naglašene oblike ter izgovor) in ga postopno izboljševali. To bo omogočalo še zanesljivejše strojno posodabljanje DSBS iz drugih (korpusnih) virov v prihodnje in zmanjšalo količino ročnega dela pri rutinskih opravilih. Urejena baza bo tako namenjena ne le človeškim uporabnikom, temveč bo uporabna tudi za učenje velikih jezikovnih modelov, ki za boljše delovanje zlasti pri morfološko bogatih jezikih potrebujejo jezikovnospecifično znanje (glej npr. Arnett in Bergen 2025; Xia idr. 2025). Tudi različici modela GaMS sta pri obeh nalogah (identifikacija legitimnih lem in določanje glagolskega vida) dali precej nezanesljive rezultate, kar še dodatno potrjuje potrebo po strojno berljivih podatkih o slovenskih leksemih, s katerimi lahko bodisi doučimo model ali pa jih uporabimo kot podlago za s poizvedovanjem obogateno generiranje besedil. Modele iz družine GaMs bomo na nalogi popravljanja lem in osnovnih oblikoskladenjskih značilnosti preizkusili tudi z manj robustnimi in nekoliko natančnejšimi navodili (npr. z dodatnimi primeri rabe) ter določili, na kakšen način lahko izboljšamo njihove rezultate pri leksikografskem delu na podatkih, ki jih iz korpusov uvažamo v DSBS.

Prispevek je nastal v okviru raziskovalnega programa Jezikovni viri in tehnologije za slovenski jezik (P6-0411) in raziskovalnega projekta Veliki jezikovni modeli za digitalno humanistiko (LLM4DH; GC-0002), ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS), ter projekta Nadgradnja portalov jezikovnih virov CJVT 2 (NAPOJ 2; 2026–2027), ki ga financira Ministrstvo za kulturo Republike Slovenije. Iskrena hvala tudi anonimnima recenzentoma oz. recenzentkama za konstruktivne pripombe.

11 Več o dostopu DSBS preko API-ja: https://wiki.cjvt.si/books/digital-dictionary-database-of-slovene/page/api-routes.

Literatura

Arnett, C.in Bergen, B. K. (2025). Why do language models perform worse for morphologically complex languages?. V O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. Di Eugenio in S. Schockaert (ur.), *Proceedings of the 31st international conference on computational linguistics* (str. 6607–6623). Association for computational linguistics. https://openreview.net/pdf?id=eMh0y3eciw

Čibej, J., Gantar, K., Dobrovoljc, K., Krek, S., Holozan, P., Erjavec, T., Romih, M., Arhar Holdt, Š., Krsnik, L. in Robnik-Šikonja, M. (2022). *Morphological lexicon Sloleks 3.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1745

Čibej, J., Gantar, K., Gantar, P., Šešet, J., Krek, S. in Robida, N. (2026). *List of potentially non-standard vocabulary candidates MEZZANINE-NstdLex 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/2084

Čibej, J. (2026). Metoda strojne gradnje arhetipskih besedotvornih dreves na podlagi Slovenskega oblikoslovnega leksikona Sloleks. V B. Kern (ur.), *Stopenjsko besedotvorje v slovanskih jezikih* (str. 517–531). Založba ZRC, ZRC SAZU.

Kosem, I., Krek, S. in Gantar, P. (2021). Semantic data should no longer exist in isolation: The Digital Dictionary Database of Slovenian. V Z. Gavriilidou, L. Mitits in S. Kiosses (ur.), *The XIX EURALEX international congress: Lexicography for inclusion* (str. 81–83). SynMorPhoSe Lab, Democritus University of Thrace. https://elex.is/wp-content/uploads/2021/09/Semantic-Data-should-no-longer-exist-in-isolation-the-Digital-Dictionary-Database-of-Slovenian_Kosem-Krek-Gantar_EURALEX2020.pdf

Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I. in Dobrovoljc, K. (2020). Gigafida 2.0: The Reference Corpus of Written Standard Slovene. V N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk in S. Piperidis (ur.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (str. 3340–3345). European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2020/LREC-2020.pdf

Kruskal, W. H. in Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://www.tandfonline.com/doi/pdf/10.1080/01621459.1952.10483441

Pedregosa, F., idr. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Xia, C., Wu, Q., Guan, H., Tian, S., Hao, Y. in Wu, X. (2025). Evaluating modern large language models on low-resource and morphologically rich languages: A cross-lingual benchmark across Cantonese, Japanese, and Turkish. *arXiv*. https://arxiv.org/abs/2511.10664

S skrolanjem do slovenščine: vloga Instagrama pri učenju slovenščine kot neprvega jezika

DOI: 10.4312/Obdobja.45.83-91

Erika Dimc, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; erika.dimc@ff.uni-lj.si

Prispevek obravnava vlogo Instagrama kot podpornega okolja pri učenju slovenščine kot neprvega jezika. Na podlagi anketne raziskave analizira uporabniške navade, zaznano uporabnost vsebin in motivacijski vpliv platforme. Rezultati kažejo, da Instagram pomembno prispeva k razvoju besedišča in pogostejšemu stiku z jezikom, vendar ima omejen vpliv na razvoj produktivnih jezikovnih spretnosti. Uporaba je večinoma pasivna. Ugotovitve potrjujejo, da Instagram deluje kot dopolnilno, ne pa samostojno učno okolje.

Instagram, slovenščina kot neprvi jezik, učenje jezikov, družbena omrežja, motivacija

This article examines the role of Instagram as a supportive environment for learning Slovenian as a non-first language. Based on a survey, it analyzes user habits, the perceived usefulness of content, and the platform's motivational impact. The results show that Instagram contributes significantly to vocabulary development and more frequent exposure to the language, but it has a limited impact on the development of productive language skills. Usage is mostly passive. The findings confirm that Instagram functions as a supplementary, rather than standalone, learning environment.

Instagram, Slovenian as a non-first language, language learning, social media, motivation

1 Uvod

V zadnjih letih digitalna učna okolja vse izraziteje preoblikujejo načine učenja in poučevanja jezikov. Razmah družbenih omrežij je privedel do nastanka novih, neformalnih učnih prostorov, v katerih se prepletajo vsakodnevna raba jezika, vizualna komunikacija in interaktivne oblike sodelovanja. Med temi omrežji izstopa Instagram, ki s svojo kombinacijo slikovnega gradiva, kratkih videoposnetkov in besedilnih dodatkov omogoča hitro, fragmentirano, a hkrati pogosto in ponavljajoče se izpostavljanje jeziku.

V kontekstu učenja slovenščine kot neprvega jezika družbena omrežja odpirajo nova vprašanja o vlogi neformalnega učenja, motivacije in dostopnosti jezikovnih vsebin. Uporabniki niso samo pasivni prejemniki znanja, temveč aktivno izbirajo vsebine, ki ustrezajo njihovim interesom, potrebam in ravni znanja, hkrati pa lahko z ustvarjalci vsebin sodelujejo preko sporočil, komentarjev in interaktivnih nalog. Kljub naraščajoči prisotnosti izobraževalnih vsebin na družbenih omrežjih pa ostaja vprašanje njihove dejanske didaktične vrednosti še vedno odprto.

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

2 Uporaba družbenih omrežij za učenje neprvega jezika

V sodobni družbi digitalne tehnologije bistveno preoblikujejo načine dostopa do znanja, pri čemer ima učenje jezikov posebno mesto zaradi svoje izrazite povezanosti z rabo v vsakdanjem življenju. Razvoj družbenih omrežij je omogočil premik od formalnih, institucionalno strukturiranih oblik učenja k bolj odprtim, fleksibilnim in neformalnim učnim okoljem, v katerih imajo učenci večjo avtonomijo pri izbiri vsebin in načinov učenja. Takšna okolja pogosto označujemo kot neformalno ali priložnostno učenje, ki poteka izven tradicionalnih izobraževalnih institucij.

Raziskave kažejo, da družbena omrežja ustvarjajo dinamična učna okolja, ki omogočajo stalno izpostavljenost jeziku in interakcijo z drugimi uporabniki (Ernestivita idr. 2024: 292). Uporabniki lahko v takih okoljih razvijajo različne vidike jezikovne zmožnosti, zlasti besedišče, pragmatično rabo jezika in kulturno kompetenco, saj so vsebine pogosto umeščene v avtentične komunikacijske kontekste (Zhang in Romero-Forteza 2024: 584). Pomembna značilnost teh okolij je tudi njihova dostopnost in vpetost v vsakodnevne rutine (Lee 2025: 45).

Platforme, npr. Instagram, združujejo večmodalne elemente, kot so slika, video, besedilo in zvok. Večmodalnost predstavlja pomemben dejavnik pri učenju jezikov, saj omogoča povezovanje različnih načinov predstavitve pomena in s tem olajša razumevanje ter pomnjenje (Lee 2022). Poleg tega takšne platforme omogočajo tudi t. i. mikroučenje (angl. *microlearning*), pri katerem so vsebine podane v kratkih in preglednih enotah, kar ustreza sodobnim vzorcem pozornosti in rabe digitalnih medijev (Lee 2025).

Pomemben vidik učenja je tudi motivacija, ki jo lahko družbena omrežja dodatno krepijo. Raziskave kažejo, da je stopnja vključenosti v aktivnosti na Instagramu povezana z višjo motivacijo za učenje tujega jezika, pri čemer imajo pomembno vlogo tako kognitivni kot tudi afektivni in vedenjski vidiki interakcije (npr. spremljanje vsebin, interakcija, ustvarjanje objav) (Ebadi idr. 2024).

Kljub številnim prednostim pa uporaba družbenih omrežij v jezikovnem učenju odpira tudi določene izzive. Med njimi so vprašanja kakovosti in zanesljivosti vsebin, fragmentiranosti znanja ter pomanjkanja sistematičnosti, ki je značilna za formalno izobraževanje. Kot poudarjajo sodobne raziskave, je ključno razumeti družbena omrežja kot dopolnilni, ne pa nujno nadomestni učni prostor (Lee 2025).

2.1 Slovenščina na Instagramu

Odločila sem se, da se osredotočim na Instagram, saj je najbolj priljubljen med družbenimi omrežji za sledenje in objavljanje take vsebine, hkrati pa združuje različne generacije (Kum 2026). S ključno besedo *slovenščina* v različnih jezikih sem poiskala profile, ki objavljajo vsebino, namenjeno učečim se slovenščine kot neprvega jezika. Posledično je največje število tovrstnih profilov prisotnih na platformi Instagram, saj ta omogoča doseganje širokega občinstva in visoko stopnjo uporabniške angažiranosti. V Tabelo 1 sem vključila samo profile, ki so v zadnjih treh mesecih ustvarili vsaj eno objavo, saj je za učeče se pomembno, da avtor objavlja nove vsebine ter jim tako zagotavlja konstantne dražljaje in opomnike. Profile sem razvrstila po številu sledilcev,

dopisala sem tudi metajezik, ki ga ustvarjalec uporablja za podajanje informacij, prevode in dodatne razlage.

Ime profila	Št. sledilcev[1]	Metajezik	Opombe
Slovenian with Laura	10.941	angleščina	
Be Slovenian	6124	angleščina	
Beseda na dan	4929	angleščina	besede, ki jih objavlja, ponuja tudi kot natisnjeno gradivo (kartice)
Ure slovenščine	2856	angleščina	
Pro Slovenskij	2085	ruščina	
Slovenščina s Tanjo	2022	angleščina	
Eslovenol	1731	španščina	
Slovenščina za ruskogovoreče	1248	ruščina	
That Slovenian Teacher	976	angleščina	
Slovenščina korak za korakom	741	ruščina	
Sloventura s Sašo	466	angleščina	
Šos Slovenščina	389	italijanščina	

Tabela 1: Profili na Instagramu z vsebino, namenjeno učečim se slovenščine kot neprvega jezika.

Našla sem 12 profilov, ki redno objavljajo nove jezikovne vsebine. Med vsebinami, ki jih ustvarjalci objavljajo, so najpogostejše objave z novimi besedami in objave, ki vsebujejo slovnico ali razlago pogostih težav pri določenem slovničnem poglavju. Nekateri ustvarjalci objavljajo tudi kulturno vsebino, kjer razlagajo slovenske običaje, značilnosti in tradicije. Na določenih profilih se pojavi tudi humorna vsebina, ki služi predvsem povezovanju učečih se, humor pa je običajno vezan na težave pri učenju in pri govorjenju z govorci slovenščine kot prvega jezika.

Vsebine, objavljene na profilih, niso in ne morejo biti strukturirane na način, kot so npr. v učbeniku. Profilom lahko kdorkoli kadarkoli začne slediti in sodelovati z njimi, zato ni stopnjevanja zahtevnosti, prav tako se vsebine objavljajo naključno. Določeni ustvarjalci pripravljajo objave, ki niso povezane, medtem ko drugi svoje objave grupirajo v rubrike ali pa jim dodajo določene oznake, da lahko uporabniki takoj vidijo, o čem bo objava govorila.

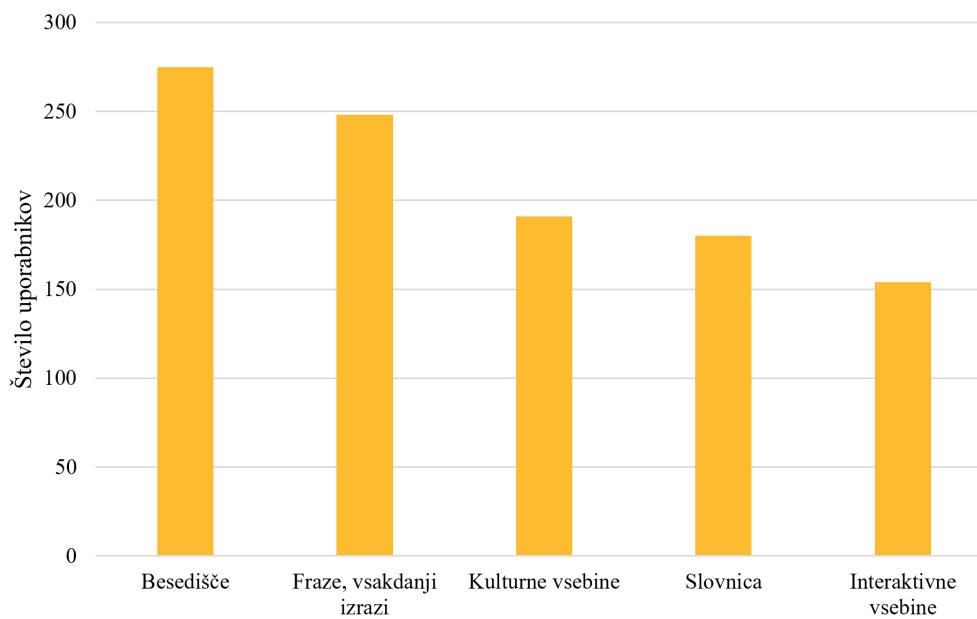
Instagram ima na voljo tri tipe objav, preko katerih lahko ustvarjalci posredujejo svojo vsebino. Najosnovnejši je tip klasične objave, kjer ustvarjalec objavi eno ali več slik, zraven pa v polje za besedilo napiše spremno besedilo. Ostali uporabniki lahko objavo komentirajo in všečkajo. Drugi tip so t. i. kratki posnetki *Instagram reel*. Kratke posnetke lahko ostali uporabniki všečkajo in komentirajo. Tretji tip objav pa so zgodbe oz. *story*, ki so uporabnikom na voljo le 24 ur. Zgodbe omogočajo številne interaktivne možnosti, ki jih lahko ustvarjalec doda objavi, kot je npr. postavljanje vprašanj, odgovarjanje na vprašanja, izbiranje med možnostmi ipd.

1 Na dan 23. 4. 2026.

3 Raziskava med učečimi se

V raziskavi je sodelovalo 312 anketirancev,[2] med katerimi je bilo 198 žensk (63,5 %), 110 moških (35,3 %) in 4 (1,2 %), ki se niso želeli opredeliti. Glede na samoocenjeno raven znanja slovenščine se jih je največ (44,6 %; 139) opredelilo kot A2–B1, 32,1 % (100) kot raven B2, 15,4 % (48) kot raven A1 in 7,9 % (25) kot raven C1 ali višje. Starost anketirancev je bila od 18 do 72 let, povprečna starost pa je znašala 36 let. Za potrebe nadaljnje analize so bili anketiranci razdeljeni v dve starostni skupini: mlajšo skupino (18–30 let) in starejšo skupino (31–72 let). Skupini sta vključevali primerljivo število anketirancev (mlajša 149 udeležencev in starejša skupina 163 udeležencev). Raziskava je bila izvedena februarja 2026 prek spletnega anketnega orodja 1KA.[3]

Rezultati kažejo, da Instagram predstavlja pomemben vir stika s slovenščino. 81,4 % anketirancev (254) poroča, da vsebine, povezane z učenjem jezika, spremlja vsaj nekajkrat tedensko, pri čemer jih 42,9 % (134) to počne vsak dan. 64,7 % udeležencev (202) do teh vsebin dostopa spontano med splošno rabo Instagrama, medtem ko jih 35,3 % (110) vsebine išče načrtno. Najpogosteje spremljane vsebine so povezane z učenjem besedišča, kar navaja 88,1 % anketirancev (275), sledijo fraze in vsakdanji izrazi (79,5 %; 248), kulturne vsebine (61,2 %; 191) ter razlage slovničnih struktur (57,7 %; 180). Skoraj polovica (49,4 %; 154) spremlja tudi interaktivne vsebine, kot so kvizi ali naloge. Pri oceni uporabnosti kot najkoristnejše omenjajo vsebine za širjenje besedišča in fraze.



Graf 1: Vsebina, ki jo na Instagramu spremljajo učeči se.

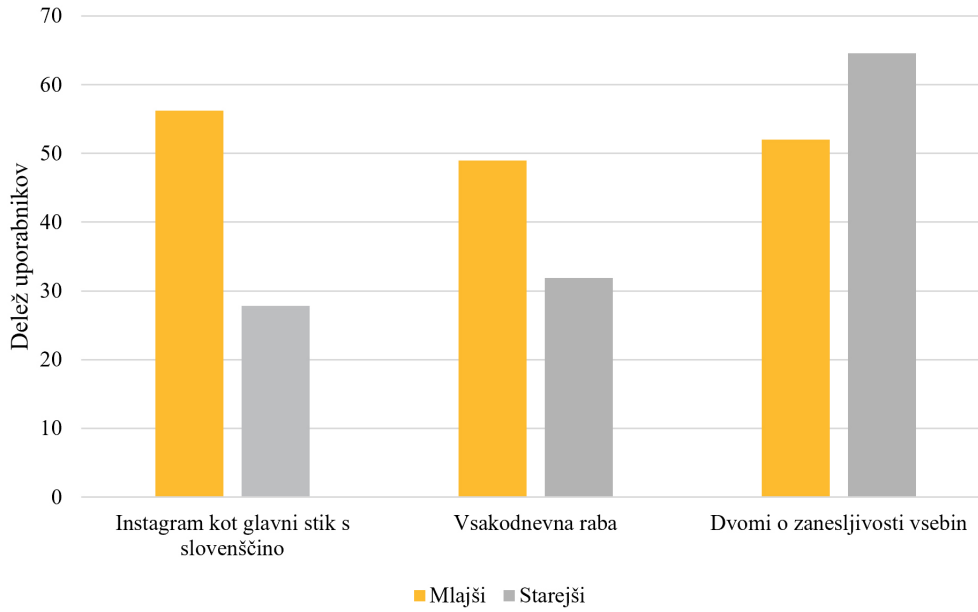
2 Anketiranci so bili k sodelovanju povabljeni preko Instagrama, tako da je bila med sledilci različnih profilov s temo učenja slovenščine kot neprvega jezika deljena povezava do spletne ankete.

3 Anketni vprašalnik v slovenščini se nahaja v Prilogi, na voljo je bil tudi v angleščini.

Anketiranci večinoma zaznavajo pozitiven vpliv Instagrama na svoje učenje. Da so z njegovo pomočjo razširili besedišče, meni 71,8 % sodelujočih (224), 63,5 % (198) pa poroča o boljšem razumevanju vsakodnevne komunikacije. Nekoliko nižji delež (48,2 %; 150) zaznava napredek pri razumevanju slovnice, medtem ko jih 36,9 % (115) opaža izboljšanje pisnega izražanja, 29,7 % (93) pa ustnega izražanja. Posebej izrazit je motivacijski vidik uporabe Instagrama. Skoraj štiri petine anketirancev (79,2 %; 247) se strinja ali popolnoma strinja s trditvijo, da jih Instagram motivira za učenje slovenščine, pri čemer povprečna ocena znaša 4,19 (od 5). Poleg tega 68,4 % (213) navaja, da zaradi Instagrama pogosteje prihaja v stik z jezikom, 54,6 % (170) pa meni, da jim takšne vsebine zmanjšujejo občutek zahtevnosti učenja.

Prevladuje pasivna raba Instagrama. 72,1 % anketirancev (225) vsebine predvsem spremlja, medtem ko 34,6 % (108) občasno rešuje naloge ali sodeluje v interaktivnih elementih. Aktivnejše oblike sodelovanja, kot je komentiranje, navaja 21,8 % (68), ustvarjanje lastnih vsebin v slovenščini pa 9,9 % (31). Pomemben vidik predstavlja tudi vprašanje zanesljivosti vsebin. Več kot polovica anketirancev (58,7 %; 183) izraža določene dvome glede njihove verodostojnosti, pri čemer jih 41,3 % (129) meni, da so razlage pogosto preveč poenostavljene. Skoraj polovica (46,8 %; 146) navaja, da težko presodi pravilnost podanih informacij, kar kaže na potrebo po kritičnem pristopu k uporabi tovrstnih virov.

Pri primerjavi po starosti se pokažejo zanimive razlike. Med anketiranci, starimi do vključno 30 let, jih 56,2 % navaja, da je Instagram njihov glavni stik s slovenščino zunaj formalnega okolja, medtem ko je med starejšimi od 30 let odstotek bistveno nižji (27,8 %). Mlajši uporabniki prav tako pogosteje poročajo o vsakodnevni rabi (49,0 %) v primerjavi s starejšimi od 30 let (31,9 %). Po drugi strani starejši anketiranci pogosteje izražajo dvome glede zanesljivosti vsebin (64,6 %) kot mlajša skupina (52,0 %), kar lahko kaže na bolj kritičen odnos do neformalnih učnih virov.



Graf 2: Primerjava med mnenji mlajših in starejših uporabnikov.

4 Diskusija

Instagram predstavlja pomemben dopolnilni prostor za učenje slovenščine kot neprvega jezika, zlasti z vidika pogostosti stika z jezikom in motivacijskega učinka. Visok delež anketirancev, ki vsebine spremlja redno (81,4 %), kaže, da so družbena omrežja postala integriran del vsakodnevnih učnih praks.

V skladu s teoretičnimi izhodišči o večmodalnosti in mikroučenju (Lee 2022) rezultati kažejo, da uporabniki kot najkoristnejše zaznavajo vsebine, povezane z besediščem in frazami. Po drugi strani pa nižji delež zaznanega napredka pri produktivnih spretnostih potrjuje omejitve tovrstnega okolja. Visok delež anketirancev (79,2 %) poroča o večji motivaciji za učenje, kar potrjuje, da kombinacija vizualne privlačnosti, kratkih vsebin in enostavne dostopnosti ustvarja spodbudno učno okolje. Takšne platforme zmanjšujejo zaznano zahtevnost jezika in omogočajo bolj sproščen stik z učno snovjo, hkrati pa lahko prav ta sproščenost vodi v površinsko učenje, kar se odraža v sorazmerno nizki zaznani učinkovitosti pri kompleksnejših jezikovnih strukturah in slovnici.

V raziskavi, v kateri je sodelovalo 1100 študentov iz Češke, Poljske, Romunije, Iraka, Malezije in Mehike (Pikhart idr. 2022), so študenti opisovali učenje preko družbenih omrežij kot fleksibilno, dostopno in zabavno, vendar so opazili slabše napredovanje v jezikovni spretnosti, manjkala jim je tudi interakcija med učitelji in sovrstniki. V raziskavi med 100 mladimi učitelji angleščine kot tujega jezika na Češkem (Janík 2022) so ugotovili, da so predvsem mlajši učitelji pripravljeni integrirati družbena omrežja, kot sta Instagram in Telegram, v učni proces.

Pri interpretaciji rezultatov je treba upoštevati nekatere metodološke omejitve raziskave. Ker je bil cilj raziskave pridobiti splošno sliko o percepciji Instagrama kot učnega

okolja, v vprašalniku ni bilo preverjanja specifičnih profilov ustvarjalcev, ki jim anketiranci sledijo. Zaznave o zanesljivosti in verodostojnosti vsebin so zato subjektivne in so lahko pogojene s posamičnimi izkušnjami z določenimi profili, kar pa ne odraža nujno kakovosti preostalih jezikovnih vsebin na platformi. Poleg tega izraženi dvomi učečih se kažejo na težavo, da sami težko objektivno evalvirajo strokovno neoporečnost gradiv. To odpira potrebo po prihodnjih raziskavah, ki bi se usmerile v analizo konkretnih profilov, hkrati pa poudarja nujnost vključevanja učiteljev, ki bi morali učečim se svetovati pri izbiri kakovostnih neformalnih virov.

Pri interpretaciji rezultatov je treba upoštevati, da je bil vzorec oblikovan med že registriranimi uporabniki Instagrama. Rezultatov zato ni mogoče neposredno posploševati na vse učence slovenščine kot neprvega jezika, saj raziskava ne vključuje perspektiv oseb, ki Instagrama ne uporabljajo. Prav tako v raziskavi niso bili zbrani podatki o morebitnem prekrivanju uporabe Instagrama z drugimi digitalnimi učnimi okolji, zato ni mogoče oceniti, v kolikšni meri se navade in stališča sodelujočih povezujejo z uporabo drugih spletnih virov za učenje slovenščine.

5 Zaključek

Prispevek je obravnaval vlogo Instagrama kot podpornega okolja pri učenju slovenščine kot neprvega jezika ter skušal osvetliti, kako uporabniki doživljajo tovrstne vsebine in kakšen pomen jim pripisujejo v lastnem učnem procesu. Rezultati raziskave kažejo, da Instagram predstavlja pomemben vir stika z jezikom, ki ga uporabniki redno vključujejo v svoje vsakdanje rutine. Pri tem se kot posebej relevanten izpostavlja njegov motivacijski potencial, saj večina anketirancev poroča, da jih takšne vsebine spodbujajo k učenju in jim omogočajo pogostejši stik s slovenščino.

Analiza je pokazala, da Instagram po samoocenah udeleženih največ prispeva k razvoju besedišča in razumevanju vsakodnevne rabe jezika, medtem ko je njegov vpliv na kompleksnejše jezikovne spretnosti, kot sta pisno in ustno izražanje, bolj omejen. Pomemben vidik predstavlja tudi vprašanje kakovosti in zanesljivosti vsebin, saj precejšen delež uporabnikov izraža dvome glede njihove strokovne ustreznosti.

Prispevek odpira širše vprašanje mesta družbenih omrežij v sodobni jezikovni didaktiki. Instagram se kaže kot orodje, ki lahko učinkovito dopolnjuje obstoječe učne prakse, vendar ne more nadomestiti strukturiranega in sistematičnega učenja. Njegova vrednost je predvsem v dostopnosti, aktualnosti in zmožnosti ustvarjanja motivacijskega okolja, ki spodbuja stik z jezikom tudi izven formalnih učnih situacij. V nadaljnjih raziskavah bi bilo smiselno podrobneje preučiti vpliv različnih tipov vsebin na posamezne jezikovne spretnosti ter raziskati načine, kako spodbuditi aktivnejšo vlogo uporabnikov. Prav tako bi bilo koristno raziskati dolgoročne učinke takšnega učenja in možnosti povezovanja družbenih omrežij z načrtovanim pedagoškim delom.

Literatura

Ebadi S., Zandi M. in Ajabshir Z. F. (2024). The effect of language learning engagement on L2 motivation over Instagram. *Acta Psychology*, *250*, članek 104512. https://doi.org/10.1016/j.actpsy.2024.104512

Ernestivita, G., Mendes, C. in Rocha, T. (2024). The role of social media in enhancing language learning. *Journal International Inspire Education Technology (JIIET)*, *3*(3), 291–303. https://ejournal.staialhikmahpariangan.ac.id/Journal/index.php/JIIET/article/view/723/383

Kum, E. (2026). Social media as micro-learning environments for Slovene as a non-first language in higher education. V J. Domenech, D. Menéndez Álvarez-Hevia, A. Martínez-Varea in D. Brunetto (ur.), *HEAd'26: 12th International Conference on Higher Education Advances* (str. 597–604). Editorial Universitat Politècnica de València. https://doi.org/10.4995/HEAd26.2026.21439

Lee, Y.-J. (2022). Language learning affordances of Instagram and TikTok. *Innovation in Language Learning and Teaching*, *17*(2), 1–16. https://doi.org/10.1080/17501229.2022.2051517

Lee, Y.-J. (2025). *Social media and language learning: Using TikTok and Instagram*. Routledge.

Janík, T., Lu, J. in Han, F. (2022). Exploring social media technologies for novice EFL school teachers to collaborate and communicate: A case in the Czech Republic. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1010686

Pikhart, M., Klimova, B., Al-Obaydi, L. H., Dziuba, S. in Cierniak Emerych, A. (2022). The quantitative evaluation of subjective satisfaction with digital media in L2 acquisition in younger adults: A study from Europe, Asia, and Latin America. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.946187

Zhang, X. in Romero-Forteza, F. (2024). Social media as a tool for informal Spanish learning: A phenomenological study of Chinese students' behaviour in Spain. *Behavioral Sciences*, *14*(7), članek 584. https://doi.org/10.3390/bs14070584

Priloga

Vprašalnik o uporabi Instagrama za učenje slovenščine

Uvod

Pozdravljeni! Ta anketa je del raziskave o tem, kako ljudje uporabljajo Instagram za učenje slovenščine kot drugega/tujega jezika. Anketa je anonimna. Hvala za vaš čas!

I. Demografski podatki

1. Spol
- Ženska
- Moški
- Ne želim se opredeliti

2. Starost
- Do vključno 30 let
- Nad 30 let

3. Kako ocenjujete svojo raven znanja slovenščine?
- A1
- A2–B1
- B2
- C1 ali višje

II. Uporaba Instagrama

4. Kako pogosto na Instagramu spremljate vsebine, povezane z učenjem slovenščine?
- Vsak dan
- Vsaj nekajkrat tedensko
- Redko ali nikoli

5. Kako običajno pridete do teh vsebin?
- Spontano (med splošno rabo Instagrama)
- Načrtno (vsebine iščem namenoma)

6. Ali je Instagram vaš glavni stik s slovenščino zunaj tečajev oziroma formalnega okolja?
- Da
- Ne

7. Kakšne vsebine za učenje slovenščine spremljate na Instagramu? (Izberete lahko več odgovorov)
- Učenje besedišča
- Fraze in vsakdanji izrazi
- Kulturne vsebine
- Razlage slovničnih struktur
- Interaktivne vsebine (kvizi, naloge)

8. Kako sodelujete pri teh vsebinah? (Izberete lahko več odgovorov)
- Vsebine predvsem pasivno spremljam (gledam/berem)
- Občasno rešujem naloge ali kvize
- Komentiram objave
- Ustvarjam lastne vsebine v slovenščini

9. Kako se strinjate z naslednjimi trditvami?
(Lestvica: 1 – Sploh se ne strinjam 5 – Popolnoma se strinjam)
- Instagram me motivira za učenje slovenščine.
- Zaradi Instagrama pogosteje prihajam v stik z jezikom.
- Takšne vsebine mi zmanjšujejo občutek zahtevnosti učenja.

10. Na katerih področjih opazite napredek zaradi uporabe Instagrama? (Izberete lahko več odgovorov)
- Razširil/-a sem besedišče
- Bolje razumem vsakodnevno komunikacijo
- Bolje razumem slovnico
- Izboljšal/-a sem pisno izražanje
- Izboljšal/-a sem ustno izražanje

11. Kako se strinjate z naslednjimi trditvami glede zanesljivosti?
(Lestvica: 1 – Sploh se ne strinjam 5 – Popolnoma se strinjam)
- Imam določene dvome glede verodostojnosti/pravilnosti vsebin na Instagramu.
- Zdi se mi, da so razlage pogosto preveč poenostavljene.
- Težko presodim, ali so podane informacije pravilne.

Nekaj jezikovnih značilnosti umetno generiranih šolskih esejev v slovenščini

DOI: 10.4312/Obdobja.45.93-105

Kaja Dobrovoljc Zor, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; Institut »Jožef Stefan«, Ljubljana; kaja.dobrovoljc@ff. uni-lj.si

Luka Terčon, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; luka.tercon@ff.uni-lj.si

V prispevku predstavljamo primerjavo izbranih jezikovnih značilnosti šolskih esejev iz korpusa Šolar in primerljivih esejev, generiranih z modelom GPT-5. Rezultati kažejo, da se umetno generirana besedila razlikujejo od človeško ustvarjenih, saj ne glede na razvojno stopnjo v povprečju vsebujejo bolj raznoliko besedišče in zapletenejše skladenjske strukture, med katerimi opažamo večji delež samostalniških zvez in upad deleža struktur, povezanih z opisovanjem dejanj in okoliščin. Posebej presenetljivo je, da so besede in skladenjske strukture v umetno generiranih besedilih le deloma prekrivne s tistimi, ki jih najdemo v besedilih slovenskih učencev, pri čemer med prvinami, ki se pojavljajo zgolj v generiranih besedilih, izstopajo abstraktno besedišče in priredni skladenjski paralelizmi.

veliki jezikovni modeli, umetno generirana besedila, jezikovne značilnosti, šolski eseji, večnivojska korpusna analiza

This article presents a comparison of selected linguistic features of school essays from the Šolar corpus and comparable essays generated by the GPT-5 model. The results show that AI-generated texts differ from human-written ones in that they contain more diverse vocabulary and more complex syntactic structures, including an increase in the proportion of nominal phrases and a decrease in structures related to the description of actions and circumstances. Particularly striking is the finding that the words and syntactic structures occurring in AI-generated texts overlap only partially with those found in texts written by Slovenian students. Among the frequent features found exclusively in the generated texts, abstract vocabulary and nested coordination are especially prominent.

large language models, AI-generated texts, linguistic characteristics, school essays, multi-level corpus profiling

1 Uvod

Uporaba velikih jezikovnih modelov se hitro širi na različnih področjih jezikovne rabe (Gasparini idr. 2024). Čeprav so generirana besedila lahko slogovno zelo prepričljiva in jih ljudje pogosto težko ločijo od človeških (Casal in Kessler 2023), danes vse več raziskav kaže na sistematične razlike med obema tipoma besedil na različnih jezikovnih ravneh (Terčon in Dobrovoljc 2025b). Poleg boljšega razumevanja vpliva sodobnih tehnologij na besedilno produkcijo je njihovo prepoznavanje pomembno tudi

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

z vidika razumevanja delovanja modelov in lažjega odkrivanja z njimi generiranih besedil (Shah idr. 2023; Wu idr. 2025).

To sta za slovenščino nedavno potrdila tudi Terčon in Dobrovoljc Zor (2026b), ki ugotavljata, da so umetno generirani eseji leksikalno in skladenjsko manj raznoliki, bolj formulaični ter skladenjsko kompleksnejši kot avtentični eseji dijakov. Čeprav njuna raziskava prinaša dragocen prvi sistematični vpogled v jezikovne značilnosti umetno generiranih besedil v slovenščini, je prenosljivost rezultatov na besedila, s katerimi se slovenski pedagogi dejansko srečujejo (Bezjak 2024; Mithans idr. 2026), precej omejena. Temelji namreč na redkeje uporabljanih modelih Gemma in GaMS ter zgolj na esejih gimnazijskih dijakov 4. letnika, pri čemer je bil model pozvan zgolj k pisanju eseja, ne pa tudi k posnemanju pisne jezikovne zmožnosti dijakov na tej razvojni stopnji.

Da bi te omejitve presegli in ponudili realnejši vpogled v to, kako uspešno sodobna tehnologija posnema jezik slovenskih šolarjev in njihov postopni razvoj pisne jezikovne zmožnosti, v tem prispevku predstavljamo primerjalno korpusno analizo nekaj izbranih jezikovnih značilnosti med avtentičnimi šolskimi eseji, zbranimi v korpusu Šolar (Arhar Holdt in Kosem 2025), in primerljivimi besedili, generiranimi z velikim jezikovnim modelom GPT-5, na katerem temelji tudi priljubljeni spletni klepetalnik ChatGPT.

2 Metodologija

2.1 Korpus človeško ustvarjenih esejev

Razvojni korpus Šolar vsebuje besedila, ki so jih učenci slovenskih osnovnih in srednjih šol samostojno tvorili pri različnih predmetih. Trenutna verzija, Šolar 3.0 (Arhar Holdt idr. 2022), vsebuje 5485 besedil učencev od 6. do 9. razreda in dijakov (15–19 let), pri čemer veliko večino besedil predstavljajo šolski spisi (eseji). Te smo kot besedilnozvrstno homogeno množico vzeli tudi za izhodišče naše raziskave, pri čemer izraz *esej* v nadaljevanju sinonimno uporabljamo tako za osnovnošolske spise kot srednješolske eseje. Čeprav korpus vsebuje številne metapodatke o šoli (osnovni ali srednji), predmetu, stopnji (razred ali letnik), vrsti besedila, regiji in letu nastanka, med njimi ni ločenega podatka o naslovu oz. izhodiščnem učiteljskem navodilu pred pisanjem besedila. Ker smo v pričujoči raziskavi ta podatek nujno potrebovali za generiranje zvrstno in tematsko čim bolj podobnih besedil, smo vse šolske eseje ročno pregledali in iz njih izpisali naslov (če je bil v besedilu jasno naveden), če je bilo to v besedilu jasno razvidno, pa dodatno tudi morebitni podnaslov in/ali obravnavano literarno delo. Na ta način smo iz celotnega nabora esejev v korpusu Šolar 3.0 (3218) identificirali 2612 esejev z jasno določljivim naslovom oz. temo, torej primernih za strojno generiranje primerljivih besedil.

2.2 Korpus umetno generiranih esejev

Kot prikazuje spodnji primer, smo za vsakega izmed človeško ustvarjenih esejev v drugem koraku z modelom GPT-5 generirali primerljivi esej z vhodnim pozivom, v katerem sta bila podana naslov in okvirna dolžina eseja, v primeru razpoložljivega podatka o podnaslovu ali izhodiščnem literarnem delu pa tudi ta dva podatka.

V nasprotju z v uvodu omenjeno sorodno raziskavo (Terčon in Dobrovoljc Zor 2026b) je bil zaradi zagotavljanja primerljivosti na vseh razvojnih stopnjah v vhodni poziv vključen tudi pričakovani razred oz. letnik pisca:

> Napiši esej z naslovom {naslov} in podnaslovom {podnaslov}, ki se nanaša na literarno delo {literarno delo}, s približno {št. besed v človeško ustvarjenem besedilu} besedami. Odgovori samo z esejem brez spremnega besedila. Esej napiši, kot bi ga napisal pisec, ki hodi v {št. razreda/letnika} razred/letnik osnovne/srednje šole.

Na ta način smo ustvarili primerljivi korpus umetno generiranih (GPT-5) šolskih esejev z 2162 besedili in nekaj več kot 1,2 milijona besed. Oba korpusa sta bila nato strojno razčlenjena z označevalnikom Trankit, prilagojenim za slovenščino (Krsnik idr. 2024), ki gola besedila samodejno razčleni na povedi in besede, slednjim pa nato pripiše še podatke o besedni vrsti, oblikoslovnih lastnostih in skladenjski vlogi po shemi Universal Dependencies (de Marneffe idr. 2021; Dobrovoljc idr. 2023).

Podrobnejši opis vsebine obeh korpusov prikazuje Tabela 1, pri čemer sta oba korpusa prosto dostopna na repozitoriju CLARIN.SI (Arhar Holdt idr. 2022; Terčon in Dobrovoljc Zor 2026a). Za zagotavljanje statistične reprezentativnosti smo iz nadaljnjih analiz v tem prispevku izključili razrede z manj kot 50 besedili (6. razred, 5. letnik in maturitetni tečaj).

	Skupno		**Šolar**		**GPT-5**	
razred	**št. besedil**	**št. razl. naslovov**	**št. povedi**	**št. besed**	**št. povedi**	**št. besed**
6. razred	25	5	694	9588	699	10.093
7. razred	67	37	2057	26.256	1746	27.425
8. razred	115	53	3707	53.258	3296	55.072
9. razred	74	20	2930	43.339	2667	47.304
1. letnik	554	60	15.349	261.567	14.383	262.632
2. letnik	418	27	11.606	203.584	10.378	206.207
3. letnik	604	82	15.563	278.060	13.783	284.205
4. letnik	691	67	19.359	353.840	17.180	374.196
5. letnik	49	9	893	17.010	795	16.895
maturitetni tečaj	15	5	460	8408	481	9488
Skupaj	2612	365	72.618	1.254.910	65.408	1.293.517

Tabela 1: Povzetek obsega človeško ustvarjenih (Šolar) in umetno generiranih esejev (GPT-5), uporabljenih v raziskavi. Število različnih naslovov predstavlja skupek vseh razpoložljivih metapodatkov o temi eseja (naslov, podnaslov in/ali literarno delo).

2.3 Metode primerjave človeško ustvarjenih in umetno generiranih esejev

Za analizo gradiva (celote oz. izbranih podkorpusov) smo uporabili več uveljavljenih korpusnojezikoslovnih metrik, ki so bile izračunane s pomočjo orodij STARK in ComparaTree (Terčon in Dobrovoljc 2025a; Krsnik in Dobrovoljc 2025) in so podrobneje že predstavljene v izvorni sorodni raziskavi (Terčon in Dobrovoljc Zor 2026b). Te za oba korpusa (Šolar in GPT-5) vključujejo:

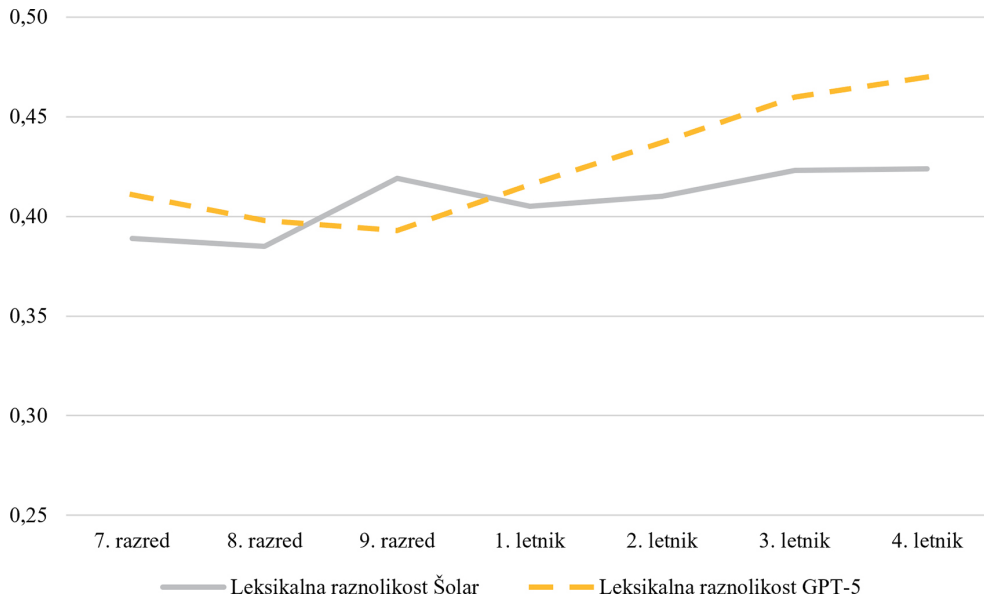
- izračun leksikalne raznolikosti (poglavje 3.1) z razmerjem med številom različnic in pojavnic (angl. *type-token ratio*), ki odraža delež različnih lem glede na vse izkazane leme;
- izračun povprečne skladenjske kompleksnosti (poglavje 3.2) povedi z mero normalizirane dolžine odvisnostnih relacij (angl. *normalized dependency distance*), ki odraža povprečno dolžino razdalj med skladenjsko odvisnimi pari pojavnic (jedri in določili);
- izračun porazdelitve besednih vrst (poglavje 3.3), ki prikazuje deleže pojavnic posameznih besednih vrst glede na število vseh pojavnic v korpusu;
- izdelavo frekvenčnega seznama vseh izkazanih lem (poglavje 3.4), na podlagi katerega primerjamo delež prekrivnosti besedišča in identificiramo najpogostejše unikatne leme v vsakem izmed obeh korpusov;
- izdelavo frekvenčnega seznama vseh izkazanih odvisnostnih pod(dreves) (poglavje 3.5), na podlagi katerega primerjamo delež prekrivnosti skladenjskih struktur in identificiramo najpogostejše unikatne skladenjske vzorce v vsakem izmed obeh korpusov.

3 Rezultati

V nadaljevanju primerjavo rezultatov izbranih metrik (poglavje 2.3) na korpusu avtentičnih šolskih esejev korpusa Šolar (poglavje 2.1) in korpusu umetno generiranih primerljivih besedil (poglavje 2.1) predstavljamo v dveh delih. V prvem se osredotočimo na splošnejšo primerjavo leksikalne raznolikosti (poglavje 3.1) in skladenjske kompleksnosti (poglavje 3.2) besedil na vseh razvojnih stopnjah, nato pa na korpusu besedil dijakov 4. letnika (srednjih šol in gimnazij) izpostavimo še nekaj konkretnejših prepoznavnih jezikovnih značilnosti umetno generiranih besedil z vidika distribucije besednih vrst (poglavje 3.3), izstopajočega besedišča (poglavje 3.4) in izstopajočih skladenjskih struktur (poglavje 3.5).

3.1 Leksikalna raznolikost

Primerjava leksikalne raznolikosti človeško ustvarjenih esejev (Šolar) in umetno generiranih ekvivalentnih besedil (GPT-5) je prikazana na Grafu 1.

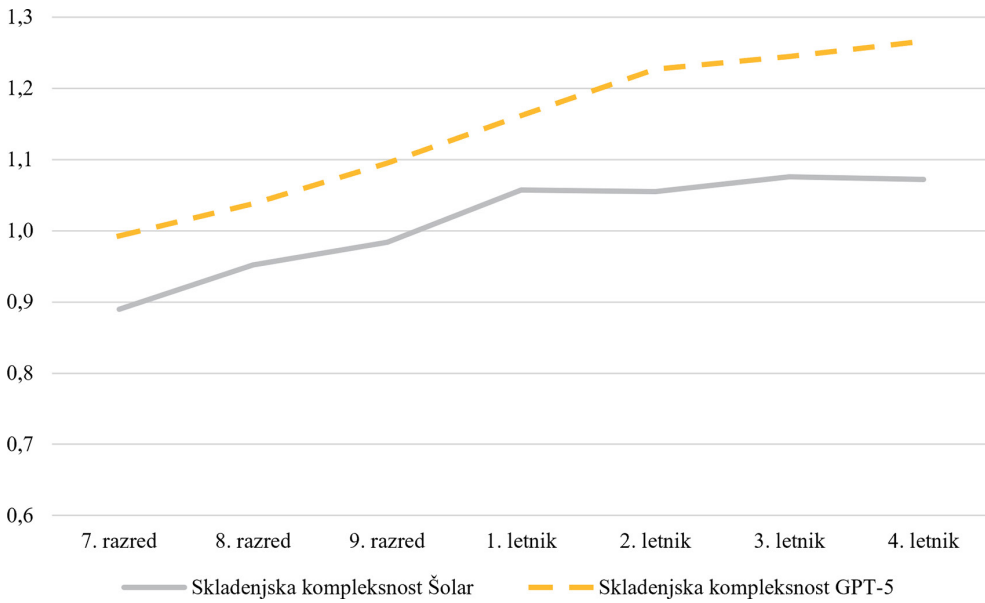


Graf 1: Primerjava raznolikosti besedišča med eseji, ki so jih napisali slovenski učenci (Šolar), in primerljivimi besedili, generiranimi z modelom GPT-5, po posameznih razredih oz. letnikih.

Graf razkriva več zanimivih ugotovitev: empirični podatki potrjujejo postopnost bogatenja besedišča učencev od 7. razreda osnovne šole do 4. letnika srednje šole, to razvojno težnjo pa s postopnim povečevanjem raznolikosti generiranih besedil delno posnema tudi uporabljeni jezikovni model, še zlasti na ravni srednješolske populacije. Manj pričakovana so nihanja na osnovnošolski stopnji (npr. navidezni upad raznolikosti na prehodu iz 7. v 8. oz. 9. razred pri obeh obravnavanih vrstah besedil), ki bi zahtevala podrobnejšo analizo, a so najverjetneje posledica dejstva, da je za te razrede na voljo bistveno manjši in manj raznolik nabor besedil (prim. Tabelo 1). Kljub temu da model precej dobro posnema postopnost razvoja leksikalne raznolikosti pri šolajoči se populaciji, pa vidimo, da generirana besedila vendarle odstopajo od besedil povprečne šolske populacije, saj generirani eseji v povprečju vsebujejo bogatejši nabor besedišča (prekinjena črta nad neprekinjeno), kot ga uporabljajo učenci oz. dijaki.

3.2 Skladenjska kompleksnost

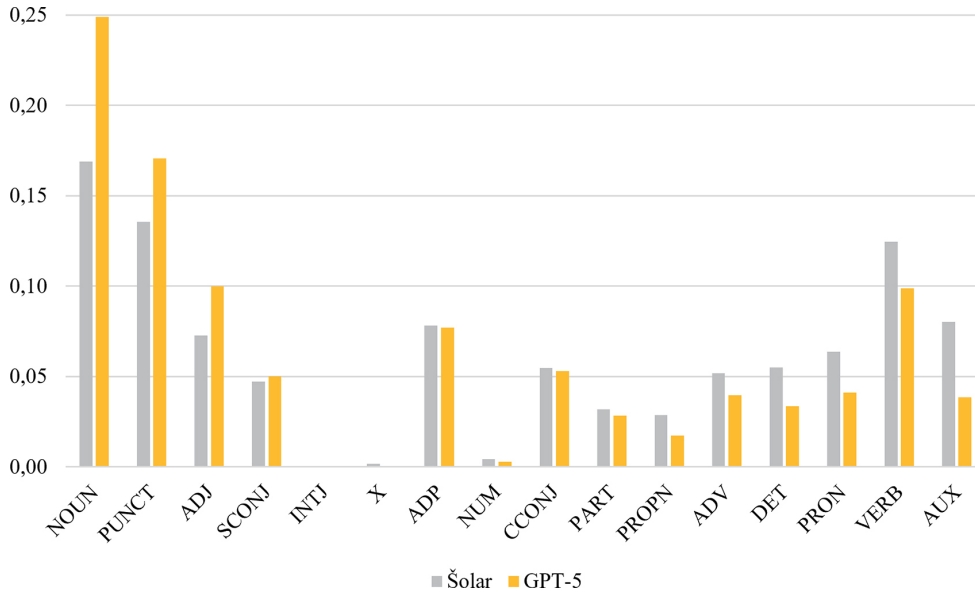
Primerjava skladenjske zapletenosti človeško ustvarjenih esejev (Šolar) in umetno generiranih ekvivalentnih besedil (GPT-5), prikazana na Grafu 2, potrjuje sklepe iz predhodnega poglavja, saj se tudi na ravni razvoja skladenjske kompleksnosti kaže, da izbrani model uspešno posnema postopnost razvoja skladenjske zmožnosti pri učencih, a to večinoma precenjuje. Sklenemo torej lahko, da umetno generirani eseji ne glede na razvojno stopnjo vsebujejo skladenjsko bolj zapletene povedi kot avtentična besedila slovenskih šolarjev.



Graf 2: Primerjava skladenjske zapletenosti povedi med eseji, ki so jih napisali slovenski učenci (Šolar), in primerljivimi besedili, generiranimi z modelom GPT-5, po posameznih razredih oz. letnikih.

3.3 Porazdelitev besednih vrst

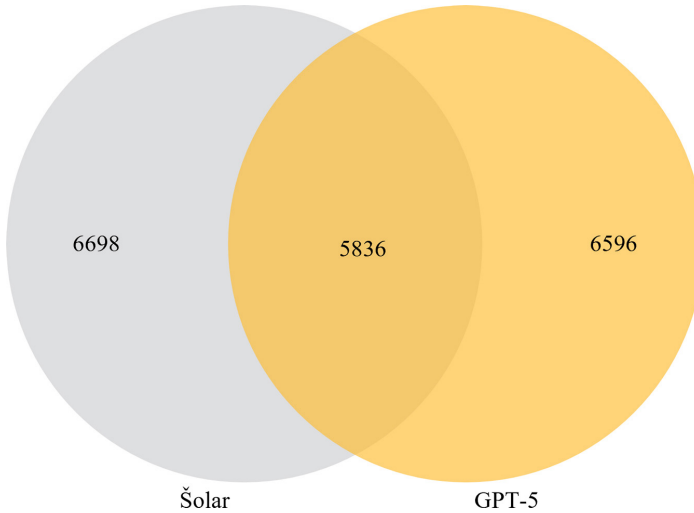
Razlike v leksikalnih oz. skladenjskih značilnosti umetno generiranih besedil v primerjavi s človeško ustvarjenimi podrobneje osvetljuje tudi primerjava distribucije besednih vrst na podkorpusu besedil 4. letnika, ki jo prikazuje Graf 3. S primerjavo razlik v deležih pojavljanja posameznih besednih vrst lahko denimo vidimo, da umetno generirani eseji v slovenščini kažejo na premik od glagolskega k bolj samostalniškemu slogu, saj vsebujejo bistveno večji delež samostalnikov (NOUN) in pridevnikov (ADJ) in bistveno manjši delež glagolov (VERB, AUX), prislovov (ADV) in zaimkov (PRON, DET). Umetno generirani eseji v slovenščini so tako bistveno bolj nagnjeni k nominaliziranemu oz. informacijsko zgoščenemu izražanju v primerjavi z avtentičnimi besedili, ki jih sicer bolj zaznamujejo strukture, povezane z opisovanjem dejanj in okoliščin ter deiktičnim navezovanjem. Čeprav ta zahteva podrobnejše empirične analize, bi to ugotovitev delno lahko pojasnili tudi s tem, da učenje tovrstnih modelov temelji zlasti na angleških besedilih.



Graf 3: Primerjava distribucije besednih vrst med eseji, ki so jih napisali slovenski dijaki 4. letnikov (Šolar), in primerljivimi besedili, generiranimi z modelom GPT-5. Besedne vrste so razporejene glede na razliko v deležu med obema korpusoma: od pogosteje zastopanih v umetno generiranih esejih (levo) do pogosteje zastopanih v esejih učencev (desno).

3.4 Izstopajoče besedišče

Za odgovor na vprašanje, kako podobno je besedišče, ki ga generira model GPT-5, v primerjavi z besediščem, ki ga slovenski učenci dejansko uporabljajo, smo iz obeh korpusov izluščili seznam lem (osnovnih oblik korpusnih pojavnic). Njuno prekrivnost prikazuje Graf 4. Presenetljivo opazimo, da se kar 6596 oz. 53 % vseh pojavnic (različnih lem), ki jih najdemo v korpusu umetno generiranih esejev, niti enkrat ne pojavi v korpusu avtentičnih šolskih besedil.



Graf 4: Prekrivnost besedišča v človeško ustvarjenih (Šolar) in umetno generiranih (GPT-5) esejih na primeru besedil dijakov 4. letnika.

Analiza desetih najpogostejših unikatnih pojavnic v Tabeli 2 nadalje kaže, da (poleg sredinskih narekovajev, ki so značilni za digitalno zapisana besedila v slovenščini) te vsebujejo predvsem abstraktne glagole in samostalnike (npr. *razpirati*, *moralizirati*, *krhkost*, *ranljivost*). Natančnejša interpretacija te ugotovitve bi seveda zahtevala podrobnejšo analizo, a ena izmed možnih hipotez je zagotovo že ugotovljeno neskladje med stopnjo jezikovne »zmožnosti« modelov in jezikovno zmožnostjo učencev oz. dijakov, prikazano na Grafih 1 in 2.

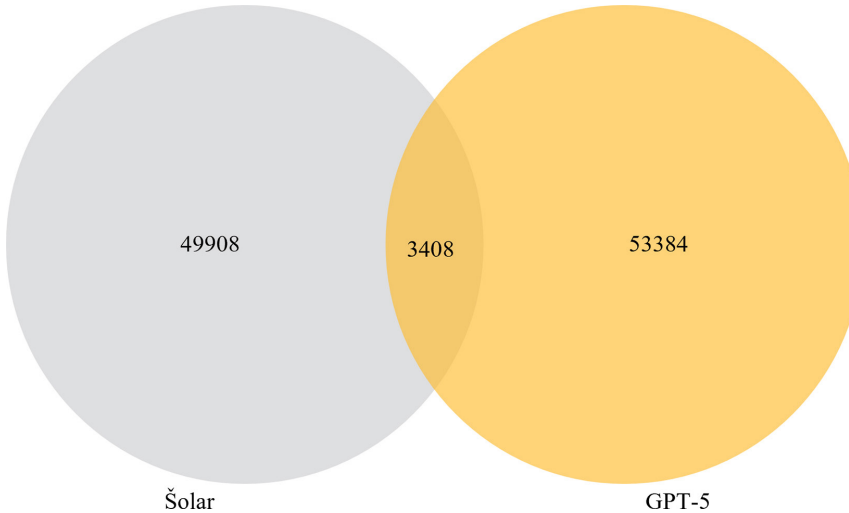
Še bolj zanimiv je seznam najpogostejših besed iz avtentičnih esejev, ki se v generiranih besedilih ne pojavijo niti enkrat (Tabela 2). Med njimi se poleg pričakovanih pojavnic, vezanih na specifike korpusa – resični narekovaji ter oznake za nerazločno zapisane (*§*) ali anonimizirane besede (npr. *XImeX*) – pojavljajo tudi določena ločila (klicaj, tropičje), presenetljivo pa tudi številna prepoznavna imena in besedišče iz slovenskih literarnih del (npr. *Tatjana*, *Fani*; *Volodje*, *zavod*), kakršna bi v generiranih esejih z naslovi, kot sta denimo *Mladostniki v romanih Prišleki in Francoski testament* ali *Drago Jančar, Veliki briljantni valček*, upravičeno pričakovali. Natančnejše pojasnilo bi seveda zahtevalo podrobnejšo analizo, a rezultati vsekakor nakazujejo na to, da je obravnavani model nagnjen k tvorjenju pomensko abstraktnejših, generičnih besedil, morda pa tudi na to, da vsebuje pomanjkljivo kulturnospecifično znanje.

Lema GPT-5	Frek. GPT-5	Frek. Šolar	Lema Šolar	Frek. Šolar	Frek. GPT-5
razpirati	197	0	*§*	553	0
»	163	0	*Tatjana*	310	0
«	163	0	*Fani*	285	0
krhek	135	0	*zavod*	249	0
krhkost	126	0	*Gregor*	236	0
jedro	124	0	*Volodje*	223	0
moralizirati	105	0	*!*	150	0
premik	99	0	*...*	149	0
ranljivost	92	0	*„*	139	0
ostrina	92	0	*XkrajX*	130	0

Tabela 2: Deset najpogostejših besed (lem), ki se pojavljajo zgolj v umetno generiranih esejih (GPT-5, levo) ali zgolj v človeško ustvarjenih esejih (Šolar, desno), na primeru besedil dijakov 4. letnika.

3.5 Izstopajoči skladenjski vzorci

Še bolj omejena je prekrivnost skladenjskih struktur, prikazana na Grafu 5, saj se v človeško ustvarjenih besedilih pojavi le 3408 oz. 6 % skladenjskih struktur, ki jih najdemo v umetno generiranih primerljivih besedilih. Ta omejena prekrivnost je glede na izbrano induktivno metodo štetja vseh možnih (pod)dreves delno pričakovana, saj se glede na kombinatoriko skladenjskih konfiguracij večina dreves v vsakem izmed korpusov pojavi le enkrat (gl. Dobrovoljc 2026), vendarle pa podrobnejši seznam unikatnih struktur v eni in drugi skupini besedil (Tabeli 3 in 4) jasno ponazarja, da med njima prihaja do razlik v skladenjski organizaciji besedil.



Graf 5: Prekrivnost skladenjskih struktur v človeško ustvarjenih (Šolar) in umetno generiranih (GPT-5) esejih na primeru besedil dijakov 4. letnika.

Vidimo denimo, da se med unikatnimi »umetno generiranimi« strukturami pojavljajo naslovi literarnih del, ki jih učenci očitno v besedilih na ta način ne ponavljajo (*Roman Mihe Mazzinija Alfred in sin*), izstopajo pa tudi različne oblike brezvezniških priredij (*kratki rezi, skopi opisi, premišljeni dialogi*) oz. prirednih paralelizmov (*med berljivostjo in kritiko, med lahkotnostjo in ostrino*) in specifične oblike odvisnikov (*da roman presega zgodovinski okvir*).

Struktura GPT-5 (s primerom)	Frek. GPT-5	Frek. Šolar
_ >flat:name _ >flat:name _ >nmod (_ >flat:name _) *Roman* ***Francea Bevka Kaplan Martin Čedermac***	36	0
_ >nmod (_ >flat:name _ >flat:name _ >nmod (_ >flat:name _)) ***Roman Francea Bevka Kaplan Martin Čedermac***	36	0
_ >nmod (_ >flat:name _) >nmod (_ >conj (_ <cc _)) ***Roman Mihe Mazzinija Alfred in sin***	35	0
_ <amod _ >conj (_ <amod _) >conj (_ <amod _) ***Kratki rezi, skopi opisi, premišljeni dialogi***	27	0
_ <case _ >conj (_ <cc _) >conj (_ <case _ >conj (_ <cc _)) most ***med berljivostjo in kritiko, med lahkotnostjo in ostrino***	26	0
_ <case _ >nmod (_ >flat:name _ >conj (_ >flat:name _) >conj (_ <cc _ >flat:name _)) ***V delih Edvarda Kocbeka, Vitomila Zupana in Alojza Rebule***	26	0
_ <cc _ <advmod (_ <advmod _) <advmod _ ***in zato še bolj zarežejo***	25	0
_ <mark _ <nsubj _ >obj (_ <amod _) Zdi se mi, ***da roman presega zgodovinski okvir***	24	0
_ <case _ >nmod (_ <case _) >conj (_ <cc _ >nmod (_ <case _)) je torej roman ***o vetru v laseh in vetru v prsih***	22	0
_ <nsubj _ <cop _ >parataxis (_ <nsubj _ <cop _) ***Pričakovanja so gorivo, resničnost je cesta***.	22	0

Tabela 3: Deset najpogostejših skladenjskih struktur (odvisnostnih dreves), ki se pojavljajo zgolj v umetno generiranih esejih (GPT-5), na primeru besedil dijakov 4. letnika. Za pojasnilo načina zapisa struktur gl. Dobrovoljc (2026).

Tabela 4 dodatno prikazuje seznam najpogostejših struktur iz besedil učencev, ki se niti enkrat ne pojavijo v umetno generiranih esejih. Poleg pričakovanih struktur, povezanih s specifičnimi oznakami korpusa (*$*) in pravopisnimi napakami učencev (*neglede na spol*), se na njem pojavljajo zlasti kratke, preprosto sestavljene povedi različnih tipov (*ki ga je napisal France Bevk*; *če bomo šli po tej poti*; *saj ni pričakoval posmeha*), kar še dodatno potrjuje večjo skladenjsko zapletenost umetno generiranih besedil, izpostavljeno v poglavju 3.2.

Struktura Šolar (s primerom)	Frek. Šolar	Frek. GPT-5
_ >flat _ >flat _ ***$ $ $***	66	0
_ <mark _ <aux _ >obl (_ <case _ <det _) ***Če bomo šli po tej poti***	20	0
_ <mark _ <obj _ <aux _ >nsubj (_ >flat:name _) Besedilo, ***ki ga je napisal France Bevk***	18	0
_ <cc _ >obj (_ <det _ <amod _) ***in ustvarja svojo lastno identiteto***	15	0
_ <expl _ >obl (_ <case _ <det _) Mislim, da bi ***se*** mati morala ***zamisliti nad njenim vedenjen***.	15	0
(_ <case _) <obl _ <expl _ <aux _ čas ***v katerem se je nahajal***	14	0
(_ >fixed _) <case _ Plače danes so ***neglede na spol*** enake	14	0
_ <advmod _ <advmod _ <case _ ***nekje globoko v gozdu***	14	0
_ <case _ <amod _ >nmod (_ >nmod _) ***v Senčnem plesu pisatelja Rebule***	14	0
_ <cc _ <aux _ >obj _ Bubi je zmeden, ***saj ni pričakoval posmeha***.	14	0

Tabela 4: Deset najpogostejših skladenjskih struktur (odvisnostnih dreves), ki se pojavljajo zgolj v človeško ustvarjenih esejih (Šolar), na primeru besedil dijakov 4. letnika. Izločene so strukture iz dobesedno navedenih odlomkov, podanih v navodilih.

4 Sklep

V prispevku smo predstavili analizo jezikovnih značilnosti šolskih esejev, generiranih z modelom GPT-5, ki kaže, da so ta besedila v primerjavi z besedili slovenskih učencev in dijakov leksikalno bolj raznolika in skladenjsko bolj zapletena, zaznamuje pa jih tudi raba specifičnega besedišča in skladenjskih struktur. Ugotovitve tako potrjujejo, da slovenska umetno generirana besedila izkazujejo specifične značilnosti glede na avtentično besedilno produkcijo, ki jih je smiselno upoštevati tako pri nadaljnjem razvoju tovrstnih tehnologij kot v razpravah o njihovi vlogi v širšem jezikoslovnem in družbenem kontekstu.

Oba korpusa sta za ponovljivost in nadaljevanje raziskav prosto dostopna na repozitoriju CLARIN.SI, med številnimi možnostmi za nadaljnje delo bi poleg novih raziskovalnih vprašanj (npr. razširitev analiz na druge jezikovne značilnosti, besedilne tipe in jezike) veljalo nadgraditi tudi samo metodologijo, npr. z uporabo vhodnih pozivov, ki zvesteje posnemajo učiteljska navodila ali demografske značilnosti piscev.

Literatura

Arhar Holdt, Š. in Kosem, I. (2025). Šolar, the developmental corpus of Slovene. *Language Resources and Evaluation*, *59*(2), 1151–1177. https://doi.org/10.1007/s10579-024-09758-4

Arhar Holdt, Š., Rozman, T., Stritar Kučuk, M., Krek, S., Krapš Vodopivec, I., Stabej, M., Pori, E., Goli, T., Lavrič, P., Laskowski, C., Kocjančič, P., Klemenc, B., Krsnik, L. in Kosem, I. (2022). *Developmental corpus Šolar 3.0* [Korpus]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1589

Bezjak, S. (2024). Prihodnost umetne inteligence v učilnicah: Pogledi srednješolskih učiteljev v Sloveniji. *Sodobna pedagogika*, *75*(4), 72–87. https://dirros.openscience.si/IzpisGradiva.php?lang=slv&id=21341

Casal, J. E. in Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, *2*(3), članek 100068. https://doi.org/10.1016/j.rmal.2023.100068

de Marneffe, M.-C., Manning, C. D., Nivre, J. in Zeman, D. (2021). Universal dependencies. *Computational Linguistics*, *47*(2), 255–308. https://doi.org/10.1162/coli_a_00402

Dobrovoljc, K. (2026). Counting trees: A treebank-driven exploration of syntactic variation in speech and writing across languages. *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/cllt-2025-0046

Dobrovoljc, K., Terčon, L. in Ljubešić, N. (2023). Universal Dependencies za slovenščino: Nove smernice, ročno označeni podatki in razčlenjevalni model. *Slovenščina 2.0: Empirične, aplikativne in interdisciplinarne raziskave*, *11*(1), 218–246. https://doi.org/10.4312/slo2.0.2023.1.218-246

Gasparini, L., Phillipson, N., Capurro, D., Rosenberg, R., Buttery, J., Howley, J., Ranganathan, S., Quinlan, C., Selvadurai, N., Wildenauer, M., South, M. in Dimaguila, G. L. (2025). A survey of large language model use in a hospital, research, and teaching campus. *Data & Policy*, *7*. https://doi.org/10.1017/dap.2025.10044

Krsnik, L. in Dobrovoljc, K. (2025). STARK: A toolkit for dependency (sub)tree extraction and analysis. V S. Jablotschkin, S. Kübler in H. Zinsmeister (ur.), *Proceedings of the 23rd International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2025)* (str. 44–51). Association for Computational Linguistics. https://aclanthology.org/2025.tlt-1.5

Krsnik, L., Dobrovoljc, K. in Terčon, L. (2024). *The Trankit model for linguistic processing of written and spoken Slovenian 1.2*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1997

Mithans, M., Ivanuš-Grmek, M. in Ograjšek, S. (2026). Uporaba umetne inteligence med osnovnošolskimi učitelji. V M. Volk, M. Ivanuš-Grmek, A. Retelj, T. Štemberger, J. Vogrinc, M. Vošnjak in T. Vršnik Perše (ur.), *Razvoj kompetenc prihodnosti v pedagoških študijskih programih* (str. 151–170). Založba Univerze na Primorskem. https://doi.org/10.26493/978-961-293-568-9.9

Shah, A., Ranka, P., Dedhia, U., Prasad, S., Muni, S. in Bhowmick, K. (2023). Detecting and unmasking AI-generated texts through explainable artificial intelligence using stylistic features. *International Journal of Advanced Computer Science and Applications*, *14*(10).

Terčon, L. in Dobrovoljc Zor, K. (2026a). *AI-generated text corpus AI-GenT 1.0* [Korpus]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/2210

Terčon, L. in Dobrovoljc Zor, K. (2026b). Evaluating the linguistic characteristics of the GaMS large language model in Slovenian. V Š. Arhar Holdt, V. Gorjanc in T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika: Zbornik konference* (str. 463–488). Inštitut za novejšo zgodovino. https://doi.org/10.5281/zenodo.22765859

Terčon, L. in Dobrovoljc, K. (2025a). ComparaTree: A multi-level comparative treebank analysis tool. V S. Jablotschkin, S. Kübler in H. Zinsmeister (ur.), *Proceedings of the 23rd International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2025)* (str. 129–139). Association for Computational Linguistics. https://aclanthology.org/2025.tlt-1.15

Terčon, L. in Dobrovoljc, K. (2025b). Linguistic characteristics of AI-generated text: A survey. *arXiv*. http://arxiv.org/abs/2510.05136

Wu, J., Yang, S., Zhan, R., Yuan, Y., Chao, L. S. in Wong, D. F. (2025). A survey on LLM-generated text detection: Necessity, methods, and future directions. *Computational Linguistics*, *51*(1), 275–338. https://doi.org/10.1162/coli_a_00549

Lovljenje ravnotežja: jezik RTV Slovenija in odzivi javnosti nanj

DOI: 10.4312/Obdobja.45.107-116

Rok Dovjak, RTV Slovenija, Ljubljana; Lektorsko društvo Slovenije, Ljubljana;
rok.dovjak@rtvslo.si

Družbeno-tehnološke spremembe v zadnjih desetletjih in z njimi povezana demokratizacija javne komunikacije jezik približujejo uporabniku. Pomemben dejavnik jezikovne politike medijev sta všečnost in upoštevanje skupnosti, o katerih se poroča. Prispevek obravnava odzive javnosti na jezik RTV Slovenija in odgovore lektorjev nanje ter daje uvid v jezikovnopolitične odločitve v zavodu, hkrati pa razkriva včasih precej različna stališča.

knjižni jezik, enotna jezikovna podoba, demokratizacija komunikacije, odzivi na jezik, politična korektnost

Social and technological changes in the last decades and the associated democratization of public communication have made public language more casual and user-friendly. Appeal and matching how audiences actually speak is of increasing importance for media language policy. This article provides an overview of public reactions to language use at RTV Slovenia and the responses of the language editors; it offers insight into the language policy of the broadcaster and highlights the differing views on how media language should sound.

standard language, unity of language image, democratization of communication, reactions to language, political correctness

1 Uvod

Po 4. členu *Zakona o Radioteleviziji Slovenija* (2005) RTV Slovenija »posebno pozornost posveča razvijanju [...] jezikovne [...] kulture«, po *Programskih standardih Radiotelevizije Slovenija* (2024: 2–3) pa je njena naloga »skrb za slovenski knjižni zborni jezik«. Tudi *Etični in poklicni kodeks RTV Slovenija* (2026) med ključna poklicna merila uvršča jezikovno dovršenost, kar opredeli kot »sledenje jezikovnim normam, predpisanim v slovnici, Slovarju slovenskega knjižnega jezika, Slovenskem pravopisu in pravorečnih priročnikih«. Čeprav zaznava razlike med govornimi položaji, je privzeta zvrst, zlasti za pisna besedila, knjižni jezik.

Glede na to, da tudi jezikoslovci lektorjem na splošno očitajo samovoljnost in nedoslednost (Majdič 2003: 340–342, 356, 371, 373, 382, 388–389; Stabej 2010: 278; Toporišič 1991: 203–204, 230–231, 402), je razumljivo, da jezikovna politika zavoda teži k enotnemu okviru knjižnega jezika in veljavni izdaji slovenskega pravopisa kot ob slovnici temeljni zunanji strokovni avtoriteti. A stroka opozarja na zastarelost normativnih priročnikov (Dobrovoljc 2010: 15; Majdič 2003: 343), kar pomeni protisloven

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

položaj: slediti zgolj pravopisu in splošnemu slovarju knjižnega jezika je lahko ozkogledno, zanašati se na svojo presojo pa subjektivno. Je že čas, da podvomimo v priročnik in smo odporni na kritike, ob tem pa mirno sprejmemo neuslovarjene zveze (*ultimat o čem*)[1] in tiste, ki pri delu govorcev vzbujajo nelagodje zaradi dvojne interpretacije (*mednarodni pravnik*, *poljub koga*)?[2] Kaj je torej porok jezikovne kakovosti? M. Pučnik (2024: 204–205) v ožjestrokovnem smislu opozarja na jezikoslovno izobrazbo, stalno izpopolnjevanje in profesionalnost, ob čemer poudarja timski pristop in medsebojno zaupanje. Tudi odziv javnosti je pri izbiri jezikovnih sredstev koristna usmeritev, celo korektiv lektorsko-uredniške interpretacije jezikovnih priročnikov ali rabe.

2 Družbeno-tehnološke spremembe od izida *Slovenskega pravopisa* (2001)

Jezikoslovci po obdobju utrjevanja norme knjižnih jezikov v Evropi opažajo tudi proces njenega rahljanja, t. i. destandardizacijo. Ne gre za jezikovno brezbrižnost, ampak za spremenjena prestižnostna razmerja med zvrstmi ter s tem svobodnejšo izbiro izraznih sredstev in dopuščanje elementov več zvrsti v formalnem govornem položaju (Androutsopoulos 2011: 156). Govorci tega ne dojamejo kot neznanje ali žaljivost, ampak kot izraz identitete (Grondelaers idr. 2016: 142). Tovrstno demokratizacijo komunikacije (in družbe) spremlja »tendenca varovanja kulturne in etnične raznovrstnosti ter zaščite marginaliziranih skupin in manjšin, pa tudi splošni porast politične samozavesti različnih socialno nedominantnih kolektivov« (Škiljan 1999: 199), ki se odražata v načelu politične korektnosti. Ta ne pomeni samo prepovedi verbalne marginalizacije posameznih družbenih skupin ali slojev v javni komunikaciji (Škiljan 1999: 130), ampak njihovo celovito opolnomočenje in priznanje samoopredelitve na podlagi prepričanja, da jezik ni le odraz realnosti, temveč jo tudi soustvarja (Erdocia 2022: 3). Ob tem je demokratizacija javne komunikacije povezana še s številom in družbenim zaledjem[3] govorcev ter povečanim deležem govorjenega jezika, ki ga dopuščajo novi mediji (Dobrovoljc 2021: 22). Škiljan (1999: 32, 48) poudarja prav pomen sodobne tehnologije, ki s spletom ponuja dostop do informacij in neposreden stik s svetom brez časovnega zamika. Hkrati pa »globalizacija kulture v pomenu skoraj neomejene mobilnosti simbolov, predmetov in ljudi« (Rizman 2003: 29) prinaša tudi nekatere normativne izenačitve (Škiljan 1999: 208), kar je najočitneje v težnji po ohranjanju tujejezične reference, npr. v napisih in grafičnih znamkah mednarodnih organizacij in prireditev (*Tokyo 2020* za poletne olimpijske igre v Tokiu),[4] pa tudi v pomensko manj

1 V SSKJ pri iztočnici *ultimat* ni primera desne vezljivosti.

2 Pri prvem izrazu ni jasno, ali gre za pravnika v mednarodnem okolju ali za strokovnjaka za mednarodno pravo. Pri drugem pa ni jasna usmerjenost dejanja (*poljub mame*), rodilnik lahko označuje vršilca dejanja (*mama poljubi*) ali pa prizadetega (*poljubiti mamo*).

3 Pri tem ne gre zanemariti višanja izobrazbene ravni svetovnega prebivalstva (Unesco 2026) in lažje dostopnosti informacij.

4 Ta zahteva temelji tudi na pravu znamk kot sistemu za rabo razlikovalnih znakov v gospodarskem prometu (Inštitut za intelektualno lastnino b. d.).

ustreznih, a na prvi pogled zvestih prevodih lastnih imen (izraelska vojaška operacija *Vzhajajoči lev*[5] po angl. *Rising Lion*, naslov blejskega strateškega foruma *A Runaway World* oz. *Pobegli svet*,[6] geslo svetovnega dneva hepatitisa *Zlomimo hepatitis*[7] po angl. *Hepatitis: Let's Break it Down*).

Glede na senzibilizacijo družbe, napredek v tehnološki obravnavi jezika ter razvoj teorije standardizacije in normativnosti se tudi ob prenovi slovenskega pravopisa omenja premik od predpisa k opisu, ki ne zahteva v imenu avtoritete, temveč pojasnjuje in razločuje pomen glede na okoliščine govornega položaja (Dobrovoljc 2021: 30–31; Dobrovoljc in Jakop 2012: 6, 18–19). Po Kozmi Ahačiču pravopisa ni mogoče vsiljevati, ampak je o pravilih treba doseči dogovor, »da bo naše življenje z jezikom čim lažje« (Bandur 2026: 11).

Dobrovoljc in Jakop (2012: 27) se strinjata, »da je v času pospešene jezikovne dinamike [...] lektorsko delo še zahtevnejše, kot je bilo nekdaj«. O tem priča npr. vprašanje zapisa *okolijski* ali *okoljski* v javni razpravi v časopisu *Delo* (Dobrovoljc 2010: 15; Lenarčič 2010: 15; Nakrst 2010: 14). Smiselno tvorjen pridevnik *okolijski* je bil predpisan s SP (2001), čeprav je dotlej v rabi prevladoval *okoljski*. Ko so poskusi lektorskega uveljavljanja oblike *okolijski* naleteli na negodovanje javnosti, so se za obliko *okoljski* zavzeli tudi na inštitutu za slovenski jezik. Po drugi strani termin *sončev mrk* v javni rabi že konkurira starejši zvezi *sončni mrk* (razmerje v Gigafidi 1033 : 1137), kljub ustaljeni zvezi z le eno obliko pridevnika (*sončni vzhod*). Prevajanje termina *sedentary behaviour* pa je pokazalo zadržek strokovnjakov, da je predlagana zveza *sedeče vedenje* preveč splošna zaradi povezovanja z izrazom *sedeči življenjski slog*, zato so predlagali zvezo *sedentarno vedenje* (Kastelic idr. 2023).

3 Varuh pravic gledalcev in poslušalcev ter odzivi na jezik

Prispevek v nadaljevanju obravnava 69 e-pisem (odzivov na jezik) uporabnikov vsebin RTV Slovenija, ki so jih lektorji prejeli od leta 2021 do konca aprila 2026. Naslovljena so bila na varuha pravic gledalcev in poslušalcev (v nadaljevanju varuh), službo za komuniciranje ali Izobraževalno središče, a so jih ti posredovali lektorjem s prošnjo, naj za pošiljatelja utemeljijo (svojo) jezikovno odločitev. Lektorji na RTV Slovenija delujejo v okviru več enot, na tovrstna pisma pa odgovarjajo skupaj. Iz pisem namreč ni vedno razvidno, na katero enoto ali oddajo se nanašajo, tudi zaradi povzemanja stanja skozi daljše obdobje.

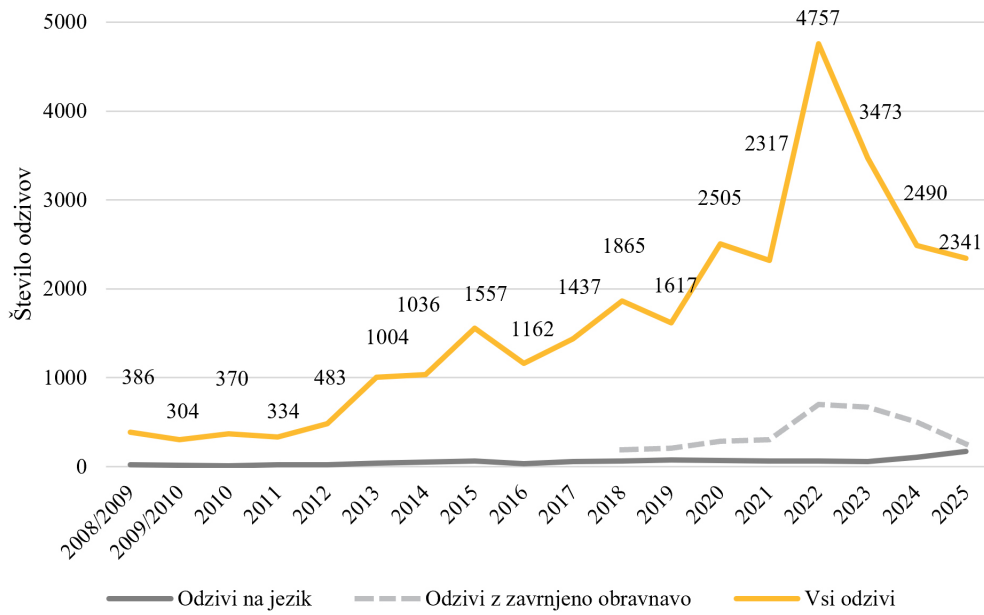
---

5 Zaradi navezave na biblijsko primero bi bilo bolj smiselno *Dvigajoči se lev*, *Vzpenjajoči se lev*, vzhaja lahko le ozvezdje Lev.

6 Teme srečanja, kot so obramba, kibernetska varnost, odziv na dezinformacije in podnebne spremembe, kažejo, da ime v angleščini izraža kaotičnost in tveganja v zaostrenih gospodarskih in geopolitičnih razmerah (torej isti pomen kot v zvezi *runaway train*, in ne v zvezi *runaway bride*), zato bi bilo pomensko bolj smiselno uporabiti *Svet, ki je ušel izpod nadzora* ali *Svet, ki je vržen s tečajev*, ker so stavčni naslovi manj učinkoviti, pa *Podivjani svet* ali *Ponoreli svet*.

7 Slovensko geslo se je pomensko oddaljilo od angleškega izvirnika, v katerem *let's break it down* označuje razčlenitev glavnih dejstev; v slovenščini bi bilo bolj smiselno *Pogovorimo se o hepatitisu*.

Varuh je poseben organ RTV Slovenija, ki s preverjanjem pritožb skrbi za upoštevanje poklicnih meril in standardov poročanja. Od leta 2008 nekatere odzive javnosti in ugotovitve pristojnih enot RTV Slovenija ob njih objavlja v spletno dostopnih mesečnih oz. od leta 2009 tudi letnih poročilih. Varuh načeloma odgovori na vse odzive (razen na žaljive in po svoji presoji nerelevantne), pri čemer se sklicuje na v poročilih že objavljene odgovore ali potrebne informacije pridobi sam. V obdobju 2008–2025 je bilo odzivov v povprečju 1635 letno (Letno poročilo 2025: 7). Odzivov na jezik je 3–8 % (izjema 1 % leta 2022), tj. od začetnih 18 do 169 leta 2025 (Graf 1).[8]



Graf 1: Odzivi javnosti, naslovljeni na varuha. Odzivi, ki jih je varuh ocenil za nerelevantne ali žaljive in zato nanje ni odgovoril, so označeni kot *odzivi z zavrnjeno obravnavo* (vir: *Letna poročila* varuha za leta 2008/2009: 8; 2009/2010: 8; 2010: 16; 2011: 28; 2012: 22; 2013: 29; 2014: 20; 2015: 8; 2016: 15; 2025: 7, 172).

4 Vsebina odzivov na jezik

Kot prikazuje Tabela 1, so prevladovali odzivi na uporabljeno besedno zvezo, nato na izgovor, od tega se štirje nanašajo na angleški izgovor kratic. Zaradi majhnega deleža odzivov na oblikoslovje je mogoče sklepati, da je na tovrstne odgovoril že varuh.

8 V starejših letnih poročilih je podatke o jezikovnih zdrsih najti tudi pri drugih poglavjih, npr. v *Letnem poročilu 2011* (10–11) pri neustreznem vodenju oddaj in poročanju navajajo rabo mašil, preglasen govor, neologizme (*jokavi plin* nam. *solzivec*), neprimerno ubeseditev (*žrtev je več, kot je poročil* nam. *kot poročajo*; *El Baradej je v zaporu v hiši* nam. *v hišnem priporu*).

Vsebina odziva	Število (N = 69)	Delež (%)
Zapis (*Čikago – Chicago*)	1	1,5
Izgovor	10	14,5
Primernost besedne zveze	38	55,0
Oblikoslovje in skladnja	6	8,7
Ženski priimki	8	11,6
Drugo	6	8,7

Tabela 1: Kategorizacija odzivov po vsebini.

17 odzivov opozarja na pretiran vpliv tujih jezikov (*naslavljati kaj* nam. *obravnavati, reševati kaj*, *pogovarjati se o čem*; *neumoren* nam. *neutruden*; *nadzor nad vozilom* nam. *oblast nad vozilom*; *redke zemlje* nam. *redke kovine*; *Mlečna cesta* nam. *Rimska cesta*; branje *NBA* [enbîêj] nam. [enbeá]; *podkast* nam. *poddaja*; angleški napisi itn.), najmanj 15 pa jih nakazuje družbeno neprimernost, celo žaljivost (*vinjen* nam. *alkoholiziran*; *starostnik* nam. *starejši*; *fantje in dekleta* [za športnike] nam. *moški in ženske*; *Harkov* nam. *Harkiv*; *ujetnik* nam. *talec*; *San Pietro del Carso* nam. *Pivka* ter raba obrazila *-ova* pri omembi žensk zgolj z njihovim priimkom).

Glede na inicialke pošiljatelja je šest uporabnikov poslalo več pisem, in sicer eden osem, eden šest in štirje po dve (trije izmed njih so svoje opozorilo ponovili). Eden je odziv poslal v imenu društva, preostali v svojem imenu. Nekateri so navedli svojo izobrazbo (zlasti jezikoslovci), drugi so se predstavili kot aktivisti (npr. *borka proti starizmu*), eden kot *konservativni občan*. Pri nekaterih odzivih je mogoče dobiti občutek, da se pritožba nanaša na kak drug medij. Nekateri so pisali tudi drugim ustanovam (in se nanje sklicevali), npr. Ministrstvu za kulturo, Uradu predsednika Republike Slovenije in Jezikovni svetovalnici.

Odzivi so predstavljeni glede na strinjanje lektorjev z njihovo vsebino (Tabela 2).

Odgovor lektorjev na odziv	Število (N = 69)	Delež (%)
Pritrditev	15	21,7
Delna pritrditev	22	31,9
Zavrnitev	13	18,9
Zadržanost	12	17,4
Drugo	7	10,1

Tabela 2: Strinjanje lektorjev z vsebino odziva: *pritrditev* pomeni popolno strinjanje z odzivom, *zavrnitev* pa popolno nestrinjanje; *zadržanost* pomeni, da se lektorji zaradi pomanjkanja informacij in nasprotujočih si argumentov niso želeli opredeliti, v kategoriji *drugo* pa so zbrani odgovori, sprejeti brez širše razprave ter dva odziva brez odgovora.

Pri osnovni presoji so se lektorji opirali na SP (2001) in SSKJ, pri manj jasnih primerih tudi na druge slovarje in svetovalnici na Franu ter na Termanio, Gigafido, Cobiss ipd. Pri šestih odzivih je bila naslovitev na lektorje manj relevantna;[9] uporabniki so RTV Slovenija namreč pozivali, naj poskrbi za:

9 Neki uporabnik je bil kritičen do rabe knjižnega jezika v nanizanki *Primeri inšpektorja Vrenka*, pri kateri pa lektorji, ki so prejeli odziv, niso sodelovali.

- pravilen izgovor imena Planica pri tujih reporterjih,
- prevod angleških napisov pri zakupljenih prenosih mednarodnih tekmovanj,
- večjo skladnost oglasnih besedil s knjižno normo,
- dosledno označbo prekmurščine kot slovenskega narečja (in ne knjižnega jezika),
- opustitev izraza *srbohrvaški*, ker da nima stvarne podlage,
- opustitev imena *San Pietro del Carso* v programu italijanske narodne skupnosti.

Če se je odziv vsebinsko ponovil in je bil odgovor vrednoten drugače, je naveden posebej.

Mnenje uporabnika uvaja kratica O (odziv), lektorjev pa P (pojasnilo).

Pritrditev

O: Neprimernost zveze *naslavljati kaj* 'obravnavati', ker gre za kalkiranje angleškega izraza *to address something*.

O: Neprimernost besede *magistra* nam. *magistrica* za žensko s 7. stopnjo izobrazbe.

O: Neprimernost besede *nadimek* nam. *vzdevek*, ker gre za srbizem.

O: Neprimerno je, da se v Dnevniku gledalce pozdravi z *Dober večer*, v Slovenski kroniki pa z *Dober dan*. P: Šlo je za napako, pri pozdravljanju se upošteva čas sončnega zahoda. Po presoji se rabi tudi *Pozdravljeni*.

O: Neprimernost zveze *nižja položnica* nam. *nižji znesek na položnici*. P: Metonimija je sicer precej razširjena tudi v besedilih vladnih služb, soroden primer *račun je visok* pa je tudi v SSKJ.

O: Neprimernost zveze *soba na vlaku*, verjetno gre za površen prevod iz angleščine. P: *Soba* je bila uporabljena, da bi se izognili ponavljanju, a je izbira neustrezna, primerneje bi bilo *kupe*.

Delna pritrditev

O (dva uporabnika v obdobju enega leta): Neprimernost zveze *enakost spolov* nam. *enakopravnost* oz. *enakovrednost spolov*. P: *Enakopravnost spolov* je res razmeroma pogosta zveza, *enakost spolov* pa eliptična poenostavitev za *enakost obravnave spolov*, a je že močno razširjena, tudi v objavah Ministrstva za delo. Prim. v SSKJ: *enakost ljudi pred zakoni* in geslo francoske revolucije *Svoboda, enakost, bratstvo*.

O: Neprimernost zveze *dvojno vodstvo*. P: Izraz je žargonski, podobno kot *dvojna zmaga*, bolje bi bilo *1. in 2. mesto* ali *dvojno zmagoslavje*, a se včasih kot sinonim vendarle dopušča.

O: Neprimernost besede *sredni* nam. *sredin*. P: *Sredin* je v SSKJ, SP (2001) pa napotuje k obliki *sredni*, verjetno po naliki *sobota – sobotni*. Zaradi jezikovne enotnosti se na RTV Slovenija priporoča obliko *sredni*.

O: Neprimernost besede *panel* 'javna razprava strokovnjakov', ker gre za nekritično prevzemanje pomena besede v angleščini. P: SP (2001) in SSKJ tak pomen res označujeta kot publicizem, a ga zaradi razširjenosti ni mogoče obiti, lahko pa se kot sinonim uporabi *posvet*.

O: Neprimernost napol podomačenega branja kratice *QR* [kúar] nam. [kúer]. P: Branje [kúêr] je doslednejše, zato bo priporočeno novinarjem. Primer odklona: PR [piár] in hibridi, kot je ChatGPT [čêt gépeté].

O (trije uporabniki): Neprimernost zveze *izraelski ujetniki* nam. *izraelski talci* za ljudi, ki jih je po 7. 10. 2023 zadrževal Hamas. P: Izraz *ujetniki* je po razlagi v SSKJ pomensko širši kot *talec*, zato pokriva tudi morebitne vojake med zajetimi, obravnava teh se namreč razlikuje od obravnave civilistov.

O: Neprimernost zveze *posmrtni ostanki* za pokojnega papeža Frančiška. P: Zveza je po SSKJ ekspresivna, med ljudmi pa je zaznati diametralno vrednotenje: pietetno – ponižujoče. Slednje je lahko povezano s predstavo razkosanosti, zato je primerno rabiti več izrazov (*telo umrlega/pokojnega*, *pokojnik*, *umrli*).

Zavrnitev

O: Neprimernost besede *vinjen* ʻalkoholiziranʼ, ker slabo vpliva na kulturo pitja vina. P: *Vinjen* je po SSKJ in SP (2001) nezaznamovan sinonim pogosteje rabljene besede *alkoholiziran* in ne namiguje na vrsto zaužite alkoholne pijače.

O: Neprimernost izgovora *tema* [təmà], *megla* [məglà], češ da so se v šoli učili drugače. P: Končniški naglas in izgovor s polglasnikom sta prvini knjižnega jezika.

O: Neprimernost tujke *podcast*. P: *Podkast* je v primerjavi s *poddajo*, ki jo ima *Islovar*, že ustaljen, ima podomačen zapis.

O: Neprimernost različice *Harkov* nam. *Harkiv*. P: Harkov je ustaljen slovenski eksonim (*Slovar slovenskih eksonimov*), ki tudi ustreza načelu prevzemanja imen po stranskosklonski osnovi.

O: Neprimernost besede *repašaž*. P: V športnih oddajah se rabi *repešaž*. Beseda je bila potrjena zaradi razširjenosti med športniki, čeprav je v SSKJ *repasaž*.

O: Neprimernost izgovora madžarskega priimka *Magyar* [mádžar] nam. [mógjar]. P: V slovenščini se madžarski zaokroženi a prevzema kot a (samo *Veliki splošni leksikon* ga prevzema z o), kar dokazujeta imeni *Bartók* in *Balassijev inštitut*, <gy> pa kot [dž].

O: Neprimernost neenotne rabe – *Paralimpijski komite Slovenije* in *paraolimpijske igre*. P: *Paralimpijski komite Slovenije* je lastno ime, zato se vanj ne posega, ohranjajo pa se ustaljena občna poimenovanja *paraolimpijske igre*, *paraolimpijka*, *paraolimpijec*.

Zadržanost

O: Neprimernost neupoštevanja rabe povratnega svojilnega zaimka (*svoj*) ob osebku v 1. osebi množine. P: Ravno v takih primerih pogosto prihaja do odklona od pričakovane oblike (*navijamo za naše*).

O: Vprašanje (leta 2021), ali stranko *Resni.ca* brati [résni píka cá] ali [resníca]. P: Izgovor imena stranke še ni ustaljen, podobna imena, ki niso spletni naslovi, npr. *Si.mobil*, pa nakazujejo, da se ime bere kot ena beseda.

O: Neprimernost rabe imena *San Pietro del Carso* nam. *Pivka* v oddaji za italijansko narodno skupnost, ker obuja spomin na fašistično raznarodovanje. P: Lektorji kot

neitalijanisti ne morejo presojati rabe v tujih jezikih, se pa izraz zdi dopusten, če gre za uveljavljen italijanski eksonim. Tudi staro slovensko ime je bilo *Sveti Peter (na Krasu)*.

O: Pet uporabnikov se je pritožilo zaradi označevanja žensk zgolj z njihovimi uradnimi priimki, dva pa zaradi rabe pridevniške oblike priimka, torej razširjenega z obrazilom *-ova*. P: Zaradi nasprotujočih si odzivov javnosti in jezikovnih utemeljitev za rabo nepridevniške oblike priimkov gre za širše sociolingvistično vprašanje, ki ni rešljivo brez pritegnitve novinarjev in urednikov oz. bi odločitev za eno od možnosti pomenila poseg v osebno opredelitev o spolni vključevalnosti. Priporočajo se nevtralne zveze imena oz. funkcije in priimka.

Različno strinjanje

O: Neprimernost zveze *žetev soli* v dnevnoinformativnem programu, ker gre za kalkiranje po tujih jezikih, domačini v primorju pa rabijo *pobiranje* ali *grabljenje soli*. P: Pritrditev, da je *žetev soli* po SSKJ žargonski izraz in zato v dnevnoinformativnem programu manj primeren. Ob vnovičnem opozorilu čez dve leti delna pritrditev, ker se je prispevek iz Sečovelj takrat navezoval na prispevek iz Pomurja o žetvi pšenice.

O: Neprimernost besede *starostnik* nam. *starejša oseba*, ker je prva žaljiva. P: Delna pritrditev, saj izrazi za starejše (*starec*, *starostnik*, *stari*, *starejši* ...) vzbujajo različna čustva, so pa ta bolj povezana s samim staranjem – po SSKJ *starostnik* ni slabšalna beseda, se pa že zdaj pogosto nadomešča s *starejši* (npr. *dom starejših*), čemur pritrjuje tudi Terminološka svetovalnica (Atelšek idr. b. d.). Ob vnovičnem opozorilu čez dve leti zavrnitev, da tudi pojavitve v korpusu Gigafida in pregled naslovov strokovnih del na Cobissu kažejo, da *starostnik* ni slabšalna beseda.

Drugo

O (dva uporabnika): Neprimernost zveze *poskus atentata*, saj atentat že pomeni poskus uboja. P: Član lektorjev je mnenju pritrdil, sklicujoč se na *Veliki slovar tujk* (»poskus umora«). Leto kasneje ob odzivu, da hude telesne poškodbe ustrezajo definiciji atentata, so lektorji podali delno pritrditev, saj iz SSKJ ni jasno razvidna opredelitev atentata glede na posledice napada.

5 Sklep

Pri jezikovnih spremembah zadnjih desetletij ne gre le za medgeneracijske razlike, temveč za družbeno-tehnološki prelom, ki se odraža tudi v teoretično-metodoloških izhodiščih nastajajočih jezikovnih normativnih priročnikov. Ustvarjalci besedil za javno objavo so z njimi opolnomočeni pri presojanju argumentov za enotno, natančno in nedvoumno, vendar ne tudi togo jezikovno podobo medija ter pri prevzemanju odgovornosti zanjo v komunikaciji z javnostjo. Glede na lektorjem posredovane odzive uporabnikov RTV Slovenija je javnost najbolj pozorna na izrazno podobo (kolokacijskost), stvarni pomen ter zaznamovanost besed in besednih zvez, med pravorečnimi napakami pa uporabniki v skoraj polovici primerov opozarjajo na angleško branje kratic. Pri tem pogosto izhajajo iz osebne izkušnje, tako izobraževanja kot stika z javnimi besedili,

manj pa se sklicujejo na normativne priročnike. Kljub temu so lektorji kot neutemeljene zavrnili le slabo petino odzivov, zavedajoč se tudi lastnih odstopov od predpisa. Pri skoraj polovici so delu argumentov pritrdili, kar kaže na zapletenost ugotavljanja upravičenosti izrazja, pa tudi odprtost, ki upošteva dojemanje izraza v posameznih skupnostih, sobesedilo in ne nazadnje avtorja besedila.

Literatura

Androutsopoulos, J. (2011). Language change and digital media: A review of conceptions and evidence. V T. Kristiansen in N. Coupland (ur.), *Standard languages and language standards in a changing Europe* (str. 145–159). Novus Press.

Atelšek, S., Fajfar, T., Jemec Tomazin, M., Sitar, J., Trojar, M. in Žagar Karer, M. (b. d.). Starejša oseba. *Terminološka svetovalnica*. https://isjfr.zrc-sazu.si/sl/terminologisce/svetovanje/starejsa-oseba

Bandur, S. (2026, 3. april). Ne uničujte Slovenski jezik. Da vas ni sram takoj izbrišite!!!! *Delo.*

Center za jezikovne vire in tehnologije. (b. d.). Gigafida 2.0: Korpus pisne standardne slovenščine [Korpus]. Viri CJVT. https://viri.cjvt.si/gigafida

Dobrovoljc, H. (2010, 2. februar). Okoljski, ne okolijski. *Delo*.

Dobrovoljc, H. (2021). Nazorska nasprotja in normativna izhodišča pri posodabljanju priročnikov knjižnega jezika. *Jezik in slovstvo*, *66*(2–3), 21–36. https://doi.org/10.4312/jis.66.2-3.21-36

Dobrovoljc, H. in Jakop, N. (2012). *Sodobni pravopisni priročnik med normo in predpisom*. Založba ZRC, ZRC SAZU. https://doi.org/10.3986/9789610504030

Erdocia, I. (2022). Participation and deliberation in language policy: The case of gender-neutral language. *Language Planning*, *23*(4), 435–455. https://doi.org/10.1080/14664208.2021.2005385

Etični in poklicni kodeks RTV Slovenija: Smernice za odgovorno ravnanje in priročnik za pripravo kakovostnih programskih vsebin. (2026). RTV Slovenija. https://www.rtvslo.si/rtv/kdo-smo/zakoni-predpisi-in-dokumenti/eticni-in-poklicni-kodeks-rtv-slovenija/782551

Grondelaers, S., van Hout, R. in van Gent, P. (2016). Destandardization is not destandardization: Revising standardness criteria in order to revisit standard language typologies in the low countries. *Taal en Tongval*, *68*(2), 119–149. https://doi.org/10.5117/TET2016.2.GRON

Inštitut za intelektualno lastnino. (b. d.). *Pravo znamk*. https://www.ipi.si/podrocja-dela/pravo-znamk/

Kastelic, K., Jurša Potocco, B., Tonin, G. in Jemec Tomazin, M. (2023). *Zidovi in mostovi – prevzemanje v slovenščini. Povzetki k okrogli mizi*. Lektorsko društvo Slovenije. https://www.lektorsko-drustvo.si/wp-content/uploads/2023/10/Povzetki-okrogle-mize_-SEDENTARNO.pdf

Lenarčič, S. (2010, 5. februar). Okoljski, ne okolijski. *Delo*.

Letno poročilo: maj 2009: Varuhinja pravic gledalcev in poslušalcev [poročilo od maja 2008 do aprila 2009]. (2009). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/varuhinja_-_letno_porocilo_maj_2009.pdf

Letno poročilo: maj 2009–maj 2010: Varuhinja pravic gledalcev in poslušalcev. (b. d.). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/letno_porocilo_maj_2009-maj_2010.pdf

Letno poročilo 2011: Varuhinja pravic gledalcev in poslušalcev. (b. d.). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/letno_porocilo_2011.pdf

Letno poročilo 2012: Varuhinja pravic gledalcev in poslušalcev. (b. d.). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/letno_porocilo_2012.pdf

Letno poročilo 2013: Varuh pravic gledalcev in poslušalcev. (b. d.). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/novo2_letno_porocilo_2013.pdf

Letno poročilo 2014: Varuh pravic gledalcev in poslušalcev. (b. d.). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/letno_porocilo_2014.pdf

Letno poročilo 2015: Varuh pravic gledalcev in poslušalcev. (b. d.). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/novo2_letno_porocilo_2015.pdf

Letno poročilo 2016: Varuh pravic gledalcev in poslušalcev. (b. d.). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/novo2_letno_porocilo_2016.pdf

Letno poročilo 2025: Varuhinja pravic gledalcev in poslušalcev RTV Slovenija (pripravila M. Uršič Zupan). (2026). RTV Slovenija. https://img.rtvslo.si/_files/2026/06/10/46_623381936491986965_letno-porocilo-2025.pdf

Majdič, V. (2003). *Pogledi na jezik: Razprave in članki o slovenskem jeziku*. Debora.

M. Pučnik, K. (2024). Lektoriranje v sodobni jezikovni kulturi: Merila kakovosti in poklicni standardi. V S. Štumberger (ur.), *Predpis in norma v jeziku* (Obdobja 43, str. 199–206). Založba Univerze. http://dx.doi.org/10.4312/Obdobja.43.199-206

Nakrst, J. (2010, 9. februar). Okoljski, ne okolijski. *Delo*.

Poročilo 2010: maj 2010–januar 2011: Varuhinja pravic gledalcev in poslušalcev. (b. d.). RTV Slovenija. https://www.rtvslo.si/files/varuhinja/novo2_poro__ilo_maj_2010_-_januar_2011.pdf

Programski standardi Radiotelevizije Slovenija. (2024). *Informator*, (99). https://img.rtvslo.si/_files/2024/05/07/9_512607432191508499_informator-24-99.pdf

Rizman, R. (2003). Globalizacija in kultura: konflikt ali sinergija? V A. Vidovič Muha (ur.), *Slovenski knjižni jezik – aktualna vprašanja in zgodovinske preizkušnje* (Obdobja 20, str. 27–42). Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete.

Toporišič, J. (1991). *Družbenost slovenskega jezika*. DZS.

Toporišič, J. idr. (ur.). (2001). *Slovenski pravopis*. Založba ZRC, ZRC SAZU. https://www.fran.si

Slovar slovenskega knjižnega jezika: Druga, dopolnjena in deloma prenovljena izdaja. (2014). Založba ZRC, ZRC SAZU. https://www.fran.si

Škiljan, D. (1999). *Javni jezik: K lingvistiki javne komunikacije*. Studia humanitatis.

Unesco. (2026, 24. februar). *What you need to know about education and why it matters*. https://www.unesco.org/en/articles/what-you-need-know-about-education-and-why-it-matters

Zakon o Radioteleviziji Slovenija (ZRTVS-1). (2005). *Uradni list RS*, št. 96/05, 109/05 – ZDavP-1B, 105/06 – odl. US, 26/09 – ZIPRS0809-B, 9/14, 163/22, 112/25, 1344/26 – odl. US. https://pisrs.si/pregledPredpisa?id=ZAKO4461

Sporazumevalne potrebe odraslih priseljenih kot izhodišče za razvoj učnega gradiva za slovenščino kot drugi jezik

DOI: 10.4312/Obdobja.45.117-128

Mateja Eniko, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; mateja.eniko@ff.uni-lj.si

Matej Klemen, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; matej.klemen@ff.uni-lj.si

Damjana Kern Andoljšek, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; damjana.kernandoljsek@ff.uni-lj.si

Ina Ferbežar, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; ina.ferbezar@ff.uni-lj.si

Tjaša Alič, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; tjasa.alic@ff.uni-lj.si

Mihaela Knez, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; mihaela.knez@ff.uni-lj.si

V prispevku predstavljamo rezultate raziskave sporazumevalnih potreb in praks odraslih priseljenih v prvih fazah življenja v Sloveniji. Na primeru interaktivnega učbenika *Naslon* pokažemo, kako je tako raziskavo mogoče uporabiti kot izhodišče za zasnovo gradiva, namenjenega za začetno učenje slovenščine odraslih priseljenih. Upoštevanje tovrstnih raziskav pri razvoju gradiv za učenje jezika je ključno, če želimo, da učeči se usvojeno znanje hitro in učinkovito prenesejo v realno življenje.

slovenščina kot drugi jezik, sporazumevalne potrebe in prakse, odrasli priseljeni, začetno učenje, scenarijski pristop

This article presents the results of an analysis of the language needs and practices of adult migrants during the initial stages of living in Slovenia. Using the interactive textbook *Naslon* as a case study, it demonstrates how such research can inform the development of Slovenian learning materials for adult migrants at the beginner's level. It is crucial to take the findings of this type of research into account when developing language-learning materials if learners are to transfer their newly acquired knowledge to real-life situations quickly and effectively.

Slovenian as a second language, language needs and practices, adult immigrants, beginner-level learning, scenarios

1 Uvod

Zaradi negotovih razmer v svetu in vse večjih potreb zahodnih družb po delovni sili se tudi v Slovenijo priseljuje vse več tujih državljank in državljanov (Razpotnik 2026). Leta 2024 se je priselilo 28.250 oseb, od tega 24.825 starejših od 15 let. 58 % jih je prišlo zaradi zaposlitve, 25 % pa zaradi združitve z družino. Podatki kažejo, da gre večinoma za državljane nekdanje skupne države (61 %), se pa v zadnjih letih povečuje

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

delež priseljenih iz oddaljenih držav (v letu 2024 približno 10 %), zlasti iz Azije (7 %) (Statistični urad Republike Slovenije b. d.).

Izzivi slovenske družbe glede vključevanja priseljenih so tako vse večji, saj mora »oblikovati pogoje, ki bodo posamezniku, ki prihaja v naše okolje, omogočali dostojno, samostojno in neodvisno delovanje v družbi, v kateri se bo počutil kot del celote« (Ministrstvo za notranje zadeve 2023: 14). Pri tem ima eno od ključnih vlog učinkovito učenje slovenščine, in to že v začetnih fazah bivanja v Sloveniji. Zapisano pomeni, da se potrebe po učenju slovenščine povečujejo; hkrati pa tudi spreminjajo, saj se v Slovenijo priseljujejo ljudje zaradi različnih razlogov, vse več pa jih prihaja tudi iz bolj oddaljenih kulturnih in jezikovnih okolij. Če namreč k zgoraj omenjeni desetini oseb, ki prihajajo iz oddaljenih držav, prištejemo še državljanke in državljane Kosova (12 % vseh priseljenih v letu 2024), lahko govorimo kar o dobri petini na novo priseljenih, ki bodo zaradi oddaljenosti svojih jezikov in kultur za učenje slovenščine predvidoma potrebovali več časa in truda, predvsem pa tudi drugačne pristope k poučevanju in strategije učenja (Slovenščina kot drugi in tuji jezik 2020: 24).

To zahteva povsem praktične rešitve, kot je ustrezna infrastruktura za učenje slovenščine z naborom jezikovnih programov, ponudbo tečajev in učnih gradiv za odrasle. Če želimo, da bo učenje učinkovito in da bodo priseljeni zanj motivirani, pa naj bi bila ta infrastruktura čim bolj usklajena z njihovimi sporazumevalnimi potrebami.

Ljudje se tujih jezikov učijo z različnimi nameni; sporazumevalne potrebe oseb, ki v novi jezikovni skupnosti že živijo in delajo ali se vanjo želijo vključiti in v njej delovati, so drugačne od potreb tistih, ki se nekega jezika morda učijo samo zaradi spoznavanja z novim jezikom ali novo kulturo. Obstoječi učbeniki in učna gradiva za slovenščino kot neprvi jezik so namenjeni obema skupinama, v njih so pogosto bolj v ospredju jezikovnosistemski vidiki (Ferbežar 2016).[1] So splošni, namenjeni kar najširšemu krogu uporabnikov za vsakovrstne namene. To dejstvo odpira vprašanje, komu so v svoji univerzalnosti (ista vsebina za vse) v resnici namenjeni (Long 2005: 1–2). Dejstvo je, da so nekatere sporazumevalne potrebe res skupne vsem, ki se učijo jezika, so splošne in so običajno povezane z jezikovnim sistemom. Na drugi strani pa imajo posamezniki tudi zelo specifične sporazumevalne potrebe, ki izhajajo iz njihovih individualnih življenjskih okoliščin. Pri poučevanju jezika je te raznolike potrebe treba uravnovesiti in učečim se omogočiti, da se lahko čim bolj identificirajo z učnimi vsebinami, jih občutijo kot smiselne oz. zase relevantne, s čimer se poveča tudi njihova motivacija za učenje. V nobenem primeru pa jih ni mogoče utemeljevati zgolj »od zgoraj navzdol«, torej s pogosto nerealnimi družbenimi pričakovanji, kaj naj bi oseba v določeni situaciji jezikovno znala. Če se osredotočimo na slovenščino kot drugi jezik oz. jezik okolja: za učinkovito jezikovno podporo je treba perspektivo tudi obrniti in z analizo ugotoviti, v katerih situacijah se ljudje dejansko znajdejo in se v njih morajo sporazumeti v novem jeziku ter kaj si želijo v njem in z njim opraviti.

1 Z njihovo pomočjo učeči se sistematično in postopno razvijajo jezikovno zmožnost, kar je osnova za dolgoročno napredovanje in doseganje tudi višjih ravni znanja slovenščine.

V prispevku predstavljamo rezultate raziskave o sporazumevalnih potrebah in praksah odraslih priseljenih v prvih fazah življenja v Sloveniji. Gre za prvo celovito tovrstno raziskavo na področju slovenščine kot drugega jezika, na podlagi katere je bil razvit interaktivni učbenik za začetno učenje slovenščine za omenjeno ciljno skupino.[2] I-učbenik *Naslon* (Alič idr. 2025) temelji na sporazumevalnih potrebah priseljenih v Slovenijo in s svojo zasnovo odgovarja na izziv, pred katerim se znajdejo – kako se z zelo omejenim znanjem slovenščine ali celo brez njega znajti v različnih formalnih okoliščinah. V nadaljevanju predstavljamo ključne izsledke raziskave in iz nje izhajajoča konceptualna izhodišča za razvoj i-učbenika *Naslon*.

2 Sporazumevalne potrebe in prakse odraslih priseljenih v Slovenijo

Podatke o sporazumevalnih potrebah in praksah odraslih priseljenih v Slovenijo smo zbirali z vprašalnikom v spletnem orodju 1KA. Za respondente je bil na voljo v slovenščini, albanščini, angleščini, bosanščini in ukrajinščini od 18. 6. do 31. 12. 2024. K sodelovanju smo povabili odrasle priseljene, vključene v različne oblike učenja slovenščine, in tiste, ki so opravljali izpit iz znanja slovenščine.[3] Na vprašalnik je odgovorilo 960 respondentov.

V nadaljevanju predstavljamo samo del rezultatov širše raziskave.[4] Osredotočamo se na vprašanja o sporazumevalnih potrebah, ki so služila kot ključna izhodišča pri razvoju i-učbenika *Naslon*. Pri tem se omejujemo na dve skupini respondentov: 1) osebe, ki v Sloveniji živijo do 11 mesecev ($N_{0 \text{ do } 11} = 88$), in 2) osebe, ki v Sloveniji živijo od 12 do 24 mesecev ($N_{12 \text{ do } 24} = 246$). Gre torej za novo priseljene, ki se v slovensko okolje šele vključujejo in zanje predvidevamo, da je njihova jezikovna zmožnost v slovenščini na nizki ravni. Prav ti priseljeni so glavni naslovniki i-učbenika *Naslon*.

Med respondenti je bilo 79 % žensk in 21 % moških. Največ je bilo starih od 35 do 44 let (37 %) in od 25 do 34 let (32 %). Kot prvi jezik jih je največ navedlo bosanščino (26 %), ukrajinščino (23 %) in srbščino (21 %). Pri vprašanju, kaj so njihovi cilji v Sloveniji, so kot odgovor največkrat izbrali dobro zaposlitev (63 %),[5] sledijo pridobitev slovenskega državljanstva (48 %), nadaljevanje šolanja (17 %), manjši delež

2 I-učbenik je nastal v okviru projekta Naslon: Nadgradnja spletnega tečaja Slonline – mobilno in spletno učenje slovenščine za priseljence (2024–2025), ki je bil usmerjen v izdelavo in nadgradnjo gradiv za mobilno, spletno in kombinirano učenje slovenščine kot drugega in tujega jezika za odrasle priseljene. I-učbenik *Naslon* je zasnovan kot interaktivni PDF, ki omogoča samostojno učenje, lahko pa ga učitelj uporabi tudi kot dodatno gradivo na tečaju ali pri kombiniranem učenju. Dostopen je brezplačno prek https://centerslo.si/knjige/ucbeniki-in-prirocniki/osnovna-stopnja/naslon/.

3 To so bili udeleženci tečajev v programu *Začetna integracija priseljencev* (ZIP), začetnih tečajev Centra za slovenščino kot drugi in tuji jezik ter kandidati na izpitih iz znanja slovenščine na vstopni in osnovni ravni.

4 Vprašalnik je obsegal različne tematske sklope (sporazumevalne prakse, sporazumevalne potrebe in učenje slovenščine).

5 Delež ustreza podatkom Statističnega urada Republike Slovenije o razlogih za priselitev v Slovenijo (gl. Uvod).

respondentov pa želi v Sloveniji ostati nekaj časa in se nato vrniti domov (9 %) oz. oditi v drugo državo (1 %).[6]

Respondenti so odgovarjali na vprašanja, ali oz. kako nujno potrebujejo slovenščino v določenih situacijah. V vprašanjih je bilo navedenih 21 situacij z javnega, poklicnega in zasebnega področja (gl. Tabelo 1), pri čemer smo izbrali predvsem situacije, povezane s formalnimi okoliščinami, saj so te za priseljene ob prihodu v novo državo neizogibne, v obstoječa učna gradiva pa niso vključene oz. so vključene le v omejenem obsegu.

Tabela 1 prikazuje, ali imajo respondenti izkušnjo s posamezno situacijo. Če je respondent pri posamezni situaciji izbral odgovor *Ne potrebujem slovenščine*, *Potrebujem slovenščino* ali *Nujno potrebujem slovenščino*, smo to interpretirali, kot da izkušnjo s to situacijo ima. Odgovor *Nikoli nisem v tej situaciji* je za nas pomenil, da izkušnje s to situacijo nima. Rezultati kažejo, da se je glavnina respondentov v večini navedenih situacij že znašla. Pričakovano jih ima največ izkušnjo z vsakodnevnimi situacijami, povezanimi s storitvami, kot so trgovina, lekarna in lokal oz. restavracija. Na drugi strani pa jih ima izrazito manj izkušnjo z organiziranimi verskimi aktivnostmi; te situacije so povezane z osebnimi odločitvami in manj s preživetjem v novem okolju. Prav tako ima manjši delež respondentov izkušnjo s situacijami v otrokovem vrtcu ali šoli, kar je morda posledica tega, da več respondentov najverjetneje nima otrok oz. njihovi otroci (še) niso vključeni v slovenski sistem vzgoje in izobraževanja.

Situacija	**Že bil v tej situaciji (%)**	**Ni bil v tej situaciji (%)**
V trgovini	100,0	0,0
V lekarni	99,4	0,6
Pogovori s sosedi	99,4	0,6
V lokalu, restavraciji	99,1	0,9
Na upravni enoti	98,5	1,5
Pri zdravniku	98,2	1,8
Na banki	97,3	2,7
Uporaba socialnih omrežij	96,7	3,3
Pogovor z uradno osebo	96,4	3,6
Na železniški ali avtobusni postaji	96,1	3,9
Iskanje stanovanja	95,5	4,5
Iskanje službe	94,6	5,4
Obisk kulturnih ustanov	94,0	6,0
Turistične rezervacije	92,1	7,9
Pri frizerju	91,2	8,8
Na servisu	88,2	11,8
Na Zavodu RS za zaposlovanje	86,1	13,9

6 Pri tem vprašanju so anketirani lahko izbrali več odgovorov, zato vsota deležev presega 100. Poleg navedenih možnosti je pri možnosti *drugo* 16 % respondentov navedlo željo po varnem in mirnem življenju, izobraževanje otrok ter družinsko življenje v Sloveniji, nekateri so tudi tukaj dodatno izpostavili zaposlitev.

Situacija	Že bil v tej situaciji (%)	Ni bil v tej situaciji (%)
V službi	86,0	14,0
Športne aktivnosti	81,2	18,8
V otrokovem vrtcu ali šoli	64,0	36,0
Organizirane religiozne aktivnosti	61,2	38,8

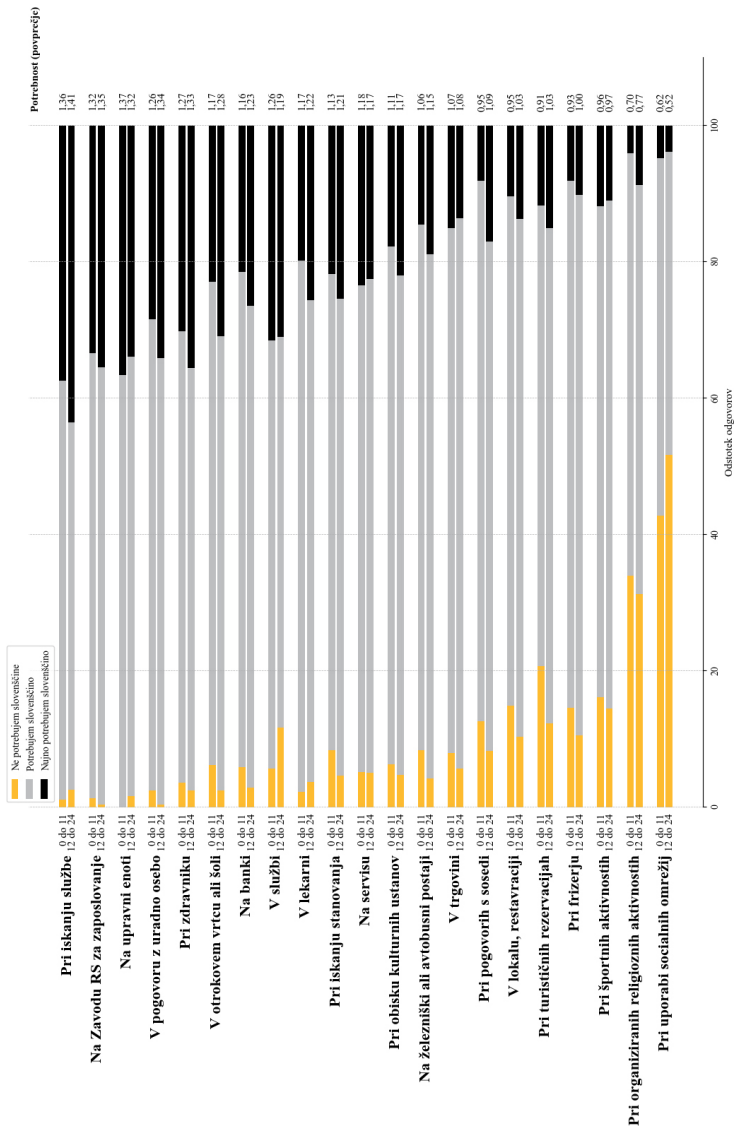
Tabela 1: Izkušnja respondentov s situacijo (N = 327–332).

Pri analizi, kako nujno priseljeni potrebujejo slovenščino, smo upoštevali samo odgovore respondentov, ki so se v določeni situaciji že znašli. Odgovore smo pretvorili v številske vrednosti: odgovor *Ne potrebujem slovenščine* v vrednost 0, *Potrebujem slovenščino* v 1 in *Nujno potrebujem slovenščino* v vrednost 2. Nato smo izračunali povprečje, ki ponazarja potrebnost znanja slovenščine (gl. Graf 1).

Primerjava potrebnosti znanja slovenščine kaže, da med skupinama – tj. respondenti, ki v Sloveniji živijo do 11 mesecev, in tistimi, ki v Sloveniji živijo od 12 do 24 mesecev – ni večjih razlik. Statistično značilna razlika se je pokazala le za situacijo *Pri pogovorih s sosedi* (p = 0,029), medtem ko se pri ostalih situacijah povprečja med skupinama niso statistično značilno razlikovala (p > 0,05). Morda iz tega lahko sklepamo, da se z daljšanjem bivanja v Sloveniji zaradi večje vključenosti v okolje veča potreba po slovenščini tudi v manj formalnih okoliščinah, kot je npr. komunikacija s sosedi.

Že v prvih fazah bivanja v Sloveniji pa se za obe skupini najvišja potreba po znanju slovenščine kaže v bolj formalnih položajih, pri urejanju osnovnih življenjskih zadev. Respondenti slovenščino najbolj potrebujejo v situacijah, povezanih z zaposlovanjem, upravnimi postopki in zdravstvenimi storitvami. Gre za situacije, v katerih imajo lahko jezikovne omejitve neposredne posledice za življenje. Med situacijami z visoko potrebo po znanju slovenščine je specifična *V otrokovem vrtcu ali šoli*. Povprečna potreba za obe skupini respondentov znaša 1,26, poudarimo pa naj, da ima izkušnjo z njo manjši delež respondentov, in sicer 64 %.

Najmanjša potreba po rabi slovenščine se kaže pri dejavnostih, povezanih z osebnimi interesi in domačim okoljem, npr. pri uporabi družbenih omrežij ter športnih in verskih aktivnostih.



Graf 1: Delež odgovorov za posamezne situacije in potrebnost znanja slovenščine – za obe skupini ($N_{0 \text{ do } 11}$ = 48–87, $N_{12 \text{ do } 24}$ = 150–245).[7]

V vprašalniku smo priseljene spraševali tudi, kako pomembno se jim zdi, da znajo v slovenščini opraviti deset izbranih jezikovnih opravil. Odgovarjali so na petstopenjski lestvici od *sploh ni pomembno* do *zelo je pomembno*. Za ista jezikovna opravila so zase

7 Na grafu so situacije razvrščene od tistih, v katerih priseljeni – ne glede na dolžino bivanja – povprečno najbolj potrebujejo slovenščino, do tistih, v katerih jo potrebujejo najmanj.

na petstopenjski lestvici od *ne znam* do *zelo dobro znam* ocenili tudi, kako dobro jih obvladajo (gl. Tabelo 2).[8]

	do 11 mesecev		od 12 do 24 mesecev	
	pomembnost	**zmožnost**	**pomembnost**	**zmožnost**
povedati osnovne informacije o sebi in drugih ljudeh	4,33	3,01	4,43	3,57
prositi za pomoč, npr. na ulici, v trgovini, pri delu	4,4	3,01	4,5	3,57
v pogovoru odgovoriti na vprašanja o sebi, npr. o svojem zdravju, težavah	4,68	2,69	4,53	3,37
razumeti informacije na odzivniku na telefonu (npr., ko kličete v zdravstveni dom)	4,65	2,78	4,63	3,48
razumeti informacije v pogovoru z otrokovo učiteljico ali vzgojiteljico	4,59	2,79	4,62	3,59
razumeti zapisane praktične informacije, npr. vozni red, delovni čas trgovine	4,46	3,4	4,39	4,0
razumeti kratka besedila v časopisu, reviji ali na internetu, npr. oglase, obvestila	4,15	3,0	4,23	3,61
napisati kratko besedilo, npr. opravičilo, prošnjo	4,42	2,62	4,53	3,37
v elektronski komunikaciji vprašati po informacijah (čas, kraj, cena)	4,33	3,06	4,49	3,63
napisati osebno sporočilo, npr. SMS	4,38	2,79	4,4	3,48

Tabela 2: Ocena pomembnosti in samoocena zmožnosti izvedbe jezikovnih opravil v slovenščini ($N_{0 \text{ do } 11}$ = 61–87, $N_{12 \text{ do } 24}$ = 202–245).

Respondenti so ne glede na dolžino bivanja v Sloveniji za vsa navedena opravila ocenili, da je zanje pomembno, da jih znajo opraviti v slovenščini (povprečne ocene se gibljejo med 4,15 in 4,68). Po drugi strani pa so svojo zmožnost izvedbe teh opravil ocenili bistveno nižje (povprečne ocene se gibljejo med 2,62 in 4,0). Za obe skupini smo s Studentovim t-testom za parne vzorce potrdili, da je razlika med oceno pomembnosti in samooceno zmožnosti izvedbe posameznega jezikovnega opravila statistično značilna (v vseh primerih $p < 0{,}001$). Razkorak je izrazito večji pri osebah, ki v Sloveniji bivajo krajši čas, in sicer znaša povprečna razlika pri respondentih, ki v Sloveniji živijo do 11 mesecev, 1,52, pri respondentih, ki so v Sloveniji med 12 in 24 meseci, pa 0,91. Analiza variance s ponovljenimi meritvami je pokazala, da je bila razlika med skupinama pri vseh navedenih opravilih statistično značilna (F = 8,49–39,60; v vseh primerih $p \leq 0{,}004$). V obeh skupinah je najmanjši razkorak pri opravilu *razumeti zapisane praktične informacije, npr. vozni red, delovni čas trgovine* (1,06 : 0,39). Posebej izrazit razkorak pa se kaže pri opravilih: *v pogovoru odgovoriti na vprašanja o sebi, npr. o svojem zdravju, težavah* (1,99 : 1,16), *razumeti informacije na odzivniku na telefonu (npr., ko kličete*

8 Tudi pri teh dveh lestvicah smo odgovore pretvorili v številske vrednosti od 1 do 5, pri čemer sta npr. odgovora *sploh ni pomembno* oz. *ne znam* dobila vrednost 1, *zelo je pomembno* oz. *zelo dobro znam* pa 5.

v zdravstveni dom) (1,87 : 1,15), *razumeti informacije v pogovoru z otrokovo učiteljico ali vzgojiteljico* (1,8 : 1,03) in *napisati kratko besedilo, npr. opravičilo, prošnjo* (1,8 : 1,16). Ti rezultati kažejo, da respondenti jezikovni primanjkljaj doživljajo predvsem pri opravilih, ki so povezana z bolj specifičnimi in/ali formalnimi sporazumevalnimi situacijami, zlasti kadar zahtevajo produktivno rabo jezika ali sprejemanje govora, kjer se pričakuje razumevanje v realnem času in morebiten odziv. Svojo zmožnost izvedbe opravil, povezanih z razumevanjem bolj predvidljivih zapisanih besedil, pa respondenti ocenjujejo kot zadovoljivejšo.

3 Razvoj gradiva za priseljene: primer i-učbenika *Naslon*

3.1 Izbor tematskih enot učbenika

Ker smo želeli z vsebino novega gradiva čim bolj nagovoriti sporazumevalne potrebe priseljenih oseb v začetnem obdobju življenja v Sloveniji, smo v i-učbenik *Naslon* vključili tiste situacije, za katere je v poglavju 2 predstavljena raziskava pokazala najvišjo potrebo po znanju slovenščine. Izbranih je bilo pet tematskih enot: *Zdravje*, *Stanovanje*, *Upravna enota*, *Delo in poklic* ter *Šola in vrtec*.[9] Gre za kontekste, pogosto formalne, ki večinoma vključujejo kompleksna jezikovna opravila, npr. najem stanovanja, urejanje statusa na upravni enoti, iskanje zaposlitve, pogovorne ure v otrokovi šoli.[10]

Pri izboru tem smo torej izhajali iz samoocene učečih se, pri natančni analizi posameznih sporazumevalnih situacij pa smo upoštevali tudi druge vire (Vandermeeren 2005: 162). Na posameznih ustanovah smo s strokovnimi službami izvedli intervjuje, s katerimi smo preverili običajni potek pogovora med njimi in njihovimi strankami. Besedila, namenjena poslušanju in branju, smo črpali iz javno dostopnih virov in uradne korespondence ustanov. Te smo v procesu didaktizacije prilagajali le deloma, saj smo želeli čim bolj ohraniti njihovo avtentičnost. Iz tako pridobljenega gradiva smo izluščili nabor besedišča in sporazumevalnih vzorcev, tipičnih za sporazumevanje v izbranih situacijah. Pri načrtovanju vsebine enot pa smo se navezovali tudi na izobraževalni program ZIP (Ferbežar idr. 2023) in priročnik *vKLJUČevanje v slovensko družbo* (Ferbežar in Pirih Svetina 2016), namenjen izvajalcem tega izobraževalnega programa.

Tako je bila razvita vsebina tematskih enot, ki so med seboj neodvisne, zato se jih v učnem procesu lahko obravnava izbirno oz. v poljubnem vrstnem redu.

3.2 Didaktični pristop: scenarij

V učnem procesu, v katerem ima učeči se aktivno vlogo in v katerem se osredotočamo na tista jezikovna sredstva, ki jih (nujno) potrebuje, da opravi določeno nalogo, je

9 Med izbranimi temami je specifična *Šola in vrtec*, pri kateri je izražena visoka potreba po znanju slovenščine, a ima izkušnjo s to situacijo pričakovano nekoliko nižji delež respondentov (64 %). V i-učbenik je bila vključena, ker je v obstoječih gradivih (verjetno prav zaradi navedenega) manj zastopana, je pa znanje slovenščine v teh kontekstih za posameznike, ki se v njih znajdejo, zelo pomembno.

10 Nekatere izbrane situacije so v učbenike za slovenščino kot drugi in tuji jezik sicer vključene, a so obravnavane parcialno oz. z drugih vidikov (jezikovna opravila v bolj formalnih situacijah so manj zastopana).

kot učinkovita oblika dela prepoznan t. i. scenarij (gl. tudi CEFR Expert Group 2023: 35). Gre za didaktični pristop, ki vključuje vrsto namišljenih aktivnosti v izbrani – za učeče se realistični – situaciji. Cilj je torej »potopitev« v situacijo, v kateri se bo posameznik najverjetneje znašel tudi v resničnem življenju (Kiefer 2018: 469). Ob tem se simulira za to situacijo tipična jezikovna opravila. Pri poučevanju odraslih priseljenih je scenarij prepoznan kot učinkovit pristop prav zato, ker omogoča, da se v čim večji meri upoštevajo njihove konkretne sporazumevalne potrebe (prim. Minuz idr. 2022: 137). Tako se priseljeni opolnomočijo za sporazumevanje v realnem življenju. Vsaka tematska enota i-učbenika je razvita v obliki scenarija, ki je sestavljen iz treh podscenarijev, ti pa iz jezikovnih opravil, tipičnih za konkretno situacijo.

»Potopitev« v situacijo podpirajo fotografije, posnete v slovenskem okolju. Te učečim se predstavljajo tudi dodatni kontekst, ki jim pomaga pri razumevanju posameznih besed, sporazumevalnih vzorcev, pa tudi pri seznanjanju z manj znanimi koncepti.

3.3 Kompleksne sporazumevalne situacije in omejena sporazumevalna zmožnost

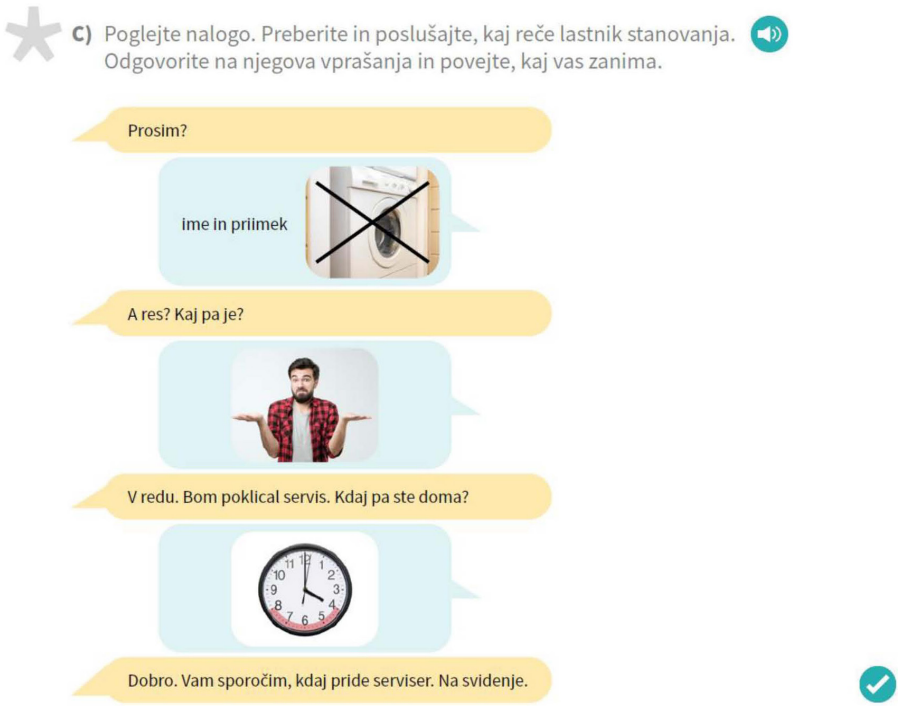
Ker smo se pri izboru tem in situacij osredotočali na formalne kontekste in se v skladu s scenarijskim pristopom trudili učečim se posredovati čim bolj avtentično jezikovno realnost, smo bili postavljeni pred izziv: kako premostiti razkorak med omejeno jezikovno zmožnostjo učečih se in potrebo po razmeroma kompleksnem izražanju, ki ga zahtevajo izbrane situacije. Nekatera jezikovna opravila v njih, pa tudi besedišče in sporazumevalni vzorci, potrebni za njihovo izvedbo, glede na SEJO namreč presegajo raven A1 (poimenovano vstopna raven). Nabor besedišča, ki se ga učenci v i-učbeniku učijo za produktivno rabo, smo primerjali s seznamom jedrnega besedišča za slovenščino kot drugi oz. tuji jezik (Klemen idr. 2022). Primerjava je pokazala, da je v i-učbeniku *Naslon* 48,6 % besed na ravni A1, 22,1 % na A2 in 11,2 % na B1. 18,1 % besed pa ni v seznamu jedrnega besedišča. Sklepamo, da to besedišče presega raven B1 (seznam jedrnega besedišča namreč obsega besede od ravni A1 do B1).[11]

Vključevanje jezikovnih opravil, ki so predvidena na višjih ravneh SEJA, je, če učenje upošteva realne sporazumevalne potrebe odraslih priseljenih, nujno (Beacco idr. 2014: 25–26). Kljub temu smo si v i-učbeniku prizadevali doseči ravnotežje med zahtevnostjo sporazumevanja v izbranih situacijah in predvideno začetno jezikovno zmožnostjo učečih se, da bi bilo gradivo zanje sploh jezikovno dostopno. Kot smiselno se je zato pokazalo dosledno ločevanje interakcijskih in tvorbnih dejavnosti od sprejemniških, ki se v gradivu navezujejo na kompleksnejše dele posameznih scenarijev oz. opravila v njih.

Interakcijskim in tvorbnim dejavnostim je v gradivu sicer namenjena posebna pozornost, saj smo želeli upoštevati rezultate raziskave, ki kažejo, da največji jezikovni primanjkljaj respondenti občutijo pri produktivni rabi jezika in slušnem razumevanju

11 Naj poudarimo, da je bil seznam jedrnega besedišča pripravljen na podlagi korpusa obstoječih učbenikov za slovenščino kot neprvi jezik. Kot že rečeno, so formalne situacije in s tem povezano besedišče tam manj zastopani.

(gl. poglavje 2, Tabela 2). Že na začetku vsake enote so zato izpostavljene najbolj ključne besede in sporazumevalni vzorci, ki jih učeči se potrebujejo za realizacijo določenega jezikovnega opravila. Izbrani so le tisti, ki jih bodo tvorili (izrekali ali zapisovali) in so značilni za družbene vloge, ki jih ti prevzemajo pri sporazumevanju v slovenskem okolju, npr. kot iskalci ali najemniki stanovanja (npr. »A je stanovanje še prosto?«, »A so stroški vključeni v najemnino?«). Ker gre pogosto za specifične izraze, smo se, kjer je bilo mogoče, odločili za tiste, ki so uporabni za konkretni kontekst, hkrati pa so dovolj splošni, da jih je možno uporabiti tudi v drugih kontekstih (npr. namesto »Kje lahko podaljšam dovoljenje za začasno prebivanje?« predlagamo »Kje uredim dovoljenje za začasno prebivanje?«). Besede in vzorci so učečim se dostopni tudi kot zvočni posnetki, tako da se lahko seznanijo z njihovo slušno podobo. Na poslušanje in ustno interakcijo pa se osredotoča tudi posebna naloga, ki je bila razvita prav za i-učbenik *Naslon*. V nalogi učeči se najprej poslušajo in urejajo dialog, nato pa izjave, ki bi jih v realnem življenju morali tvoriti sami, berejo. Sledi dejavnost za samostojno govorjenje, v kateri morajo učeči se na podlagi slikovnega gradiva samostojno govorno odreagirati na dano sogovorčevo izjavo (Slika 1). Na tak način želimo poustvariti okoliščine, v kakršnih se učeči se znajdejo tudi v resničnem življenju, ko se morajo v ustni interakciji na sogovorčevo izjavo odzvati v realnem času.



Slika 1: Primer naloge za samostojno govorjenje.

4 Sklep

SEJO v središče učenja jezika postavlja učeče se kot aktivne uporabnike jezika in družbene agense, v skladu s tem naj bi načrtovanje učnih vsebin izhajalo iz njihovih sporazumevalnih potreb, učenje pa naj bi potekalo v obliki akcijsko usmerjenega pristopa (Svet Evrope 2023: 24–25, prim. CEFR Expert Group 2023: 20, 35). To je bilo vodilo pri razvoju i-učbenika *Naslon*. Izhodišče za njegov razvoj je namreč predstavljala raziskava o sporazumevalnih potrebah in praksah odraslih priseljenih v Slovenijo. Rezultati so pokazali, da priseljeni potrebnost znanja slovenščine in pomembnost obvladovanja izbranih jezikovnih opravil v slovenščini sistematično ocenjujejo visoko. Na podlagi rezultatov so bile v i-učbenik vključene predvsem tiste teme in situacije, ki zadevajo sporazumevanje v bolj formalnih kontekstih in so v obstoječih učnih gradivih za poučevanje odraslih pomanjkljivo obravnavane ali pa vanje sploh niso vključene. V tem smislu gre za učno gradivo, ki obstoječe učbenike dopolnjuje in nadgrajuje. V i-učbeniku *Naslon* se odrasli pri začetnem učenju usposabljajo za opravljanje zanje praviloma (pre)zahtevnih jezikovnih opravil, ki pa se jim v realnosti ne morejo izogniti. V učno gradivo je vključen nabor najnujnejših tem. Upoštevaje analizo sporazumevalnih potreb in praks odraslih priseljenih kaže v prihodnje učbenik nadgraditi z dodatnimi temami in dodatnimi digitalnimi vsebinami (npr. video posnetki). Poleg tega se zdi ključno ozaveščati učitelje o pomenu prilagajanja pouka sporazumevalnim potrebam udeležencev tečajev in jih usposabljati za učinkovito rabo predstavljenega učnega gradiva v razredu. Med razvojem je bilo gradivo namreč preizkušeno in pokazalo se je, da tečaji slovenščine, ki so prilagojeni potrebam odraslih priseljenih, pozitivno vplivajo na njihovo motivacijo za učenje.

Literatura

Alič, T., Eniko, M., Kern Andoljšek, D., Klemen, M., Knez, M. in Ferbežar, I. (2025). *Naslon: Interaktivni učbenik za učenje slovenščine za odrasle priseljence*. Založba Univerze v Ljubljani. https://centerslo.si/knjige/ucbeniki-in-prirocniki/osnovna-stopnja/naslon/

Beacco, J.-C., Little, D. in Hedges, C. (2014). *Linguistic integration of adult migrants: Guide to policy development and implementation*. Council of Europe. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016802fc1cd

CEFR Expert Group. (2023). *A guide to action-oriented, plurilingual and intercultural education*. Council of Europe Publishing. https://rm.coe.int/a-guide-to-action-oriented-plurilingual-and-intercultural-education-en/1680b52354

Ferbežar, I. (2016). Kako realno je idealno? O konceptu avtentičnosti pri pouku slovenščine ob Toporišičevem Zakaj ne po slovensko. V E. Kržišnik in M. Hladnik (ur.), *Toporišičeva obdobja* (Obdobja 35, str. 489–496). Znanstvena založba Filozofske fakultete. http://centerslo.si/wp-content/uploads/2016/11/Ferbezar.pdf

Ferbežar, I. in Pirih Svetina, N. (2016). *vKLJUČevanje v slovensko družbo: Priročnik za izvajalce izobraževalnega programa Začetna integracija priseljencev*. Znanstvena založba Filozofske fakultete. https://centerslo.si/wp-content/uploads/2024/01/KLJUC_s-posnetki.pdf

Ferbežar, I., Pirih Svetina, N., Možina, E. in Pavlič, U. (2023). *Začetna integracija priseljencev (ZIP): Izobraževalni program za odrasle* (posodobljena izdaja). Center za slovenščino kot drugi in tuji jezik. https://arhiv.acs.si/programoteka/ZIP.pdf

Kiefer, K.-H. (2018). Didaktische Szenarien in der berufsbezogenen Sprachausbildung. V C. Efing in K.-H. Kiefer (ur.), *Sprache und Kommunikation in der beruflichen Aus- und Weiterbildung: Ein interdisziplinäres Handbuch* (str. 469–480). Narr Francke Attempto Verlag.

Klemen, M., Arhar Holdt, Š. in Pollak, S. (2022). *Core vocabulary for Slovenian as L2 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1697

Long, M. H. (2005). Overview: A rationale for needs analysis and needs analysis research. V M. H. Long (ur.), *Second language needs analysis* (str. 1–16). Cambridge University Press.

Ministrstvo za notranje zadeve. (2023). *Strategija vključevanja tujcev, ki niso državljani Evropske unije, v kulturno, gospodarsko in družbeno življenje Republike Slovenije*. https://www.gov.si/assets/ministrstva/MNZ/SOJ/Novice/2023/11-November/Strategija-vkljucevanja-tujcev-ki-niso-drzavljani-Evropske-unije-v-kulturno-gospodarsko-in-druzbeno-zivljenje-Republike-Slovenije_09112023.pdf

Minuz, F., Kurvers, J., Schramm, K., Rocca, L. in Naeb, R. (2022). *Literacy and second language learning for the linguistic integration of adult migrants: Reference guide*. Council of Europe Publishing. https://rm.coe.int/prems-008922-eng-2518-literacy-and-second-language-learning-couv-texte/1680a70e18

Razpotnik, B. (2026, 24. april). *Prebivalstvo, 1. januar 2026: V lanskem letu število prebivalcev naraslo*. Statistični urad Republike Slovenije. https://www.stat.si/StatWeb/News/Index/14290

Slovenščina kot drugi in tuji jezik: Izobraževalni program za odrasle. (2020). Center za slovenščino kot drugi in tuji jezik, Filozofska fakulteta, Univerza v Ljubljani. https://arhiv.acs.si/programoteka/Slovenscina_kot_drugi_in_tuji_jezik_2020.pdf

Statistični urad Republike Slovenije. (b. d.). Priseljeni prebivalci, stari 15 ali več let, po izobrazbi, državljanstvu, državi prejšnjega prebivališča in spolu, Slovenija, letno. *SiStat*. https://pxweb.stat.si/SiStatData/pxweb/sl/Data/Data/05N3108S.px/

Svet Evrope. (2023). *Skupni evropski jezikovni okvir: Učenje, poučevanje, ocenjevanje – dodatek*. Zavod RS za šolstvo. https://www.zrss.si/pdf/SEJO.pdf

Vandermeeren, S. (2005). Foreign language need of business firms. V M. H. Long (ur.), *Second language needs analysis* (str. 159–181). Cambridge University Press.

Izzivi jezikovnega testiranja nove generacije

DOI: 10.4312/Obdobja.45.129-137

Ina Ferbežar, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; ina.ferbezar@ff.uni-lj.si

Vse hitrejši razvoj pametnih tehnologij v kontekstu testiranja jezikovnega znanja zahteva premislek o vprašanjih, kaj je znanje jezika in v katerih (novih) sporazumevalnih situacijah se morajo znajti uporabniki in uporabnice jezika. V prispevku se sprašujemo o nalogah v jezikovnih izpitih, ali oz. kako jih s papirja prenesti v digitalno okolje, o varnosti in zanesljivosti uporabe pametnih tehnologij pri ocenjevanju ter o dostopnosti tehnologij za vse. V okviru javnoveljavnega testiranja za uradne namene so to vprašanja, ki se nanašajo tako na vsebinske kot tehnične vidike. Namen prispevka je predstaviti dileme, s katerimi se v hitro spreminjajočem se svetu soočamo v sistemu preverjanja in potrjevanja znanja slovenščine kot neprvega jezika, podvrženem političnim odločitvam, in rešitve, ki jih sprejemamo v tem izrazito prehodnem obdobju.

jezikovno testiranje, pametne tehnologije, slovenščina kot drugi in tuji jezik

The accelerating development of smart technologies in the context of language proficiency testing requires consideration of what language proficiency is, and the communicative situations, roles, and tasks in which language users must be able to function. This article examines language exam tasks, considering whether and how they can be transferred to a digital environment. It also considers the security and reliability of using smart technologies in assessment and their accessibility to all. In the context of large-scale standardized language testing for official purposes, these challenges relate to both content and technology. This article presents the dilemmas faced within the system of assessing and certifying Slovenian as a nonnative language, which is subject to political decisions, and the solutions being adopted during this transitional period.

language assessment, smart technologies, Slovenian as a second and foreign language

1 *In medias res*

Ena od ključnih zahtev kakovostnega preverjanja jezikovnega (in siceršnjega) znanja je, poenostavljeno, da to znanje meri na načine, ki omogočajo sklepanje o dejanski (merjeni) zmožnosti posameznikov.[1] Zelo natančno mora biti torej jasno, kaj se meri, tj. preverja in ocenjuje, in ta *konstrukt* mora biti tudi ustrezno opredeljen. Govorimo o *veljavnosti* merjenja. Hkrati pa mora biti merski instrument, tj. jezikovni test oz. izpit, zasnovan tako, da čim bolje odraža realno jezikovno rabo, saj je le na tej podlagi mogoče izpeljati ustrezne sklepe. Govorimo o *avtentičnosti* izpitnih nalog.[2] To za učitelje jezikov in razvijalce jezikovnih testov v hitro spreminjajočih se okoliščinah pomeni

1 V prispevku moško obliko uporabljam za katerikoli spol.

2 Oba koncepta sta bila v okviru področja slovenščine kot neprvega jezika slovenski strokovni javnosti že predstavljena (Ferbežar 2016; Ferbežar 2019).

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

izziv: pametne tehnologije namreč spreminjajo tudi jezik kot družbeno prakso. Sodobne sporazumevalne prakse presegajo obvladanje jezikov v njihovih normativnih okvirih in niso omejene na poznavanje jezikovnih sredstev, temveč zahtevajo dodatne spretnosti in znanja. Pri sporazumevanju se vse bolj uporabljajo pametne tehnologije, z njimi so povezane dejavnosti branja, poslušanja, govorjenja in še posebej pisanja. Jezikovna zmožnost tako vključuje tudi zmožnost smiselne in kritične uporabe pametnih tehnologij. To pri poučevanju jezikov zahteva premislek o tem, kaj so njegovi cilji ter za katere in kakšne sporazumevalne situacije, vloge in opravila je treba usposobiti uporabnike jezika, da bodo učinkovito jezikovno delovali. Posledično torej tradicionalni konstrukti jezikovne zmožnosti[3] (in znanja na sploh) niso več ustrezni in zahtevajo novo opredelitev (Unesco 2023: 2; Voss idr. 2023: 522, 523).

Zapisano sicer velja za jezikovno zmožnost ne glede na njeno »prvotnost«, torej prvi, drugi, tuji jezik (če je taka delitev ob vsej kompleksnosti pojavov še smiselna); v tem prispevku se osredotočam na slovenščino kot drugi in tuji jezik (SDTJ).

Dodatek k SEJU (Svet Evrope 2023) kot referenčni okvir za poučevanje, preverjanje in ocenjevanje znanja neprvih jezikov v tradicionalno opredelitev konstrukta sporazumevalne jezikovne zmožnosti ne posega, dodaja pa nekatere nove oblike komunikacije, kot sta spletna interakcija in uporaba telekomunikacij. Res je, da je nastajal, preden je generativna umetna inteligenca (GUI) prišla v splošno uporabo, a najrazličnejša orodja, temelječa na pametnih tehnologijah (npr. črkovalniki, prevajalniki, slovarji), so bila ob njegovi objavi že široko uporabljana. Kljub temu jih avtorji ob opredelitvi jezikovne zmožnosti niso upoštevali; je pa v lestvico Spletni pogovori in razprave vključena uporaba spletnih orodij za preverjanje pravilnosti (na spodnji B1) in spletnega prevajalnika (spodnja A2, A1 in pred A1) (Svet Evrope 2023: 84).[4] To je nekoliko nenavadno, saj je raba pametnih tehnologij pri kompleksnejših jezikovnih opravilih na višjih ravneh jezikovnega znanja morda še bolj uveljavljena, vsekakor pa je drugačna.

Zmožnost uporabe pametnih tehnologij kaže v kontekstu besednega sporazumevanja vključiti v širšo opredelitev jezikovne zmožnosti, saj je neločljivo povezana z njim, še posebej v primeru GUI; ta od uporabnika – če se omejimo le na dve ključni točki kontinuuma – zahteva po eni strani natančno ubesedovanje pozivov, po drugi pa vrednotenje strojno generiranih besedil in njihovo ločevanje od tistih, ki jih je ustvaril človek.

V sistemu javnoveljavnega preverjanja in potrjevanja jezikovnega znanja, ki ga razvija Center za slovenščino kot drugi in tuji jezik (CSDTJ), ki deluje na Filozofski fakulteti Univerze v Ljubljani, je razmislek o navedenih vprašanjih še posebej pomemben. Strokovnjaki namreč opozarjajo, da tehnologija ni zgolj nevtralni prenosnik, temveč aktivno preoblikuje temeljne koncepte v jezikovnem testiranju, kot so veljavnost, avtentičnost in pravičnost, ter zahteva njihovo teoretsko rekonceptualizacijo (Bolender

3 Pojma *jezikovna zmožnost* in *jezikovno znanje* tu uporabljam sinonimno, z zavedanjem teoretskih predpostavk za njuno ločevanje. Zvezo *jezikovno znanje* uporabljam v zelo splošnem smislu.

4 Uporaba slovarja je sicer omenjena tudi v lestvicah Pojasnjevanje podatkov, Obdelava besedila in Prevajanje pisnega besedila (Svet Evrope 2023: 95, 99, 101), a spletni medij v njih ni izrecno izpostavljen.

idr. 2023; Sadeghi in Douglas 2023: 2; Yu in Xu 2024). To pa ima nujno tudi praktične posledice.

2 *In principio*

Načela dobre prakse v sodobnem jezikovnem testiranju zahtevajo vključevanje avtentičnih nalog oz. jezikovnih opravil, ki odražajo dejanske sporazumevalne prakse ciljne populacije. Jezikovna raba vključuje tudi pomoč pametnih tehnologij in nekaterih jezikovnih opravil, povezanih predvsem s pisanjem, ki si ga ni več mogoče predstavljati brez njihove podpore. Zožimo pogled na testiranje SDTJ in izpostavimo dvoje.

Razvoj vsakega izpita iz znanja SDTJ vsebuje vprašanje, ali naj bo ta kot doslej na papirju ali naj se prenese v digitalno okolje. Govorimo o formatu izpita v čisto tehničnem smislu. Toda pri prenosu s papirja v digitalni medij ne gre le za tehnično prilagoditev, ampak ta zahteva tehten razmislek o konstruktu merjenja (gl. več prispevkov v Sadeghi in Douglas 2023 ter v Yu in Xu 2024). Kot kažejo raziskave, digitalna besedila sprožajo drugačne kognitivne procese kot linearna tiskana besedila (prav tam). Na proces in rezultat razumevanja vplivajo krmarjenje po besedilih, fragmentiranost informacij v njih in sočasna raba več virov. Če torej test, avtomatično prenesen s papirja v digitalno okolje, tega ne upošteva, obstaja tveganje, da meri nekaj drugega, kot naj bi meril, da torej konstrukt merjenja ni ustrezno opredeljen.

Drugo vprašanje zadeva pisanje kot eno od sporazumevalnih dejavnosti, ki je del vsakega učenja in preverjanja jezikovnega znanja. Tako tudi izpiti iz znanja SDTJ glede na program, na katerem temeljijo (Slovenščina kot drugi in tuji jezik 2020), vključujejo samostojno pisanje. Toda »človeška komunikacija je postala digitalna in odvisna od tehnologije« (Yu in Xu 2024: 2). Kickbusch idr. (2025) zato predlagajo novo konceptualizacijo avtentičnosti. Avtentična naloga po njihovem ne omejuje dostopa do pametnih tehnologij, temveč preverja tudi sposobnost presoje, kdaj, kako in zakaj jih je posameznik uporabil ter ali zna rezultat tudi kritično ovrednotiti (Kickbusch idr. 2025: 6; Voss idr. 2023); to pa je izjemno težko ocenjevati (Kickbusch idr. 2025: 7, 13). Če se s splošnega premaknemo k specifičnemu: Ali je torej ob splošni digitalizaciji pisne komunikacije sploh smiselno vztrajati pri izpitnih nalogah, ki predpostavljajo popolnoma samostojno pisanje brez tehnološke podpore? Kakšne naloge za pisanje vključevati v izpite in kako?

Ob razvoju programa *Slovenščina kot drugi in tuji jezik* (2020) in izpitov smo za pisanje na nižjih ravneh izbrali enostavna praktična besedila, na višjih pa esej oz. argumentativno praktično besedilo in povzemanje besedil s komentarjem. Toda ljudje prav povzemanje in pisanje vseh vrst besedil vse pogosteje zaupamo UI,[5] esejev pa – z izjemo redkih poklicnih profilov – zunaj izobraževalnih kontekstov ne pišemo. Zato bomo v prihodnje pisanje na ravneh od B2 navzgor preverjali z integriranimi nalogami:[6] te

5 Gl. npr. poročilo HEPI (2026), dostopno prek https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026.

6 Take naloge že vključuje izpit na ravni odličnosti: v njem je treba pisno povzeti besedila iz več virov in komentirati problematiko, ki jo tematizirajo.

bodo zahtevale krajše pisno komentiranje slišanih in prebranih besedil, kar se zdi za ciljno populacijo smiselno in dovolj avtentično.[7] A seveda ne pisanje na papirju.

Izpit na ravni odličnosti, ki obsega najvišji od šestih ravni SEJA (C1 in C2), je bil razvit tako, da omogoča digitalno obliko zapisa (pisanje besedila na računalnik) in uporabo različnih orodij, npr. črkovalnika in interneta. A izkušnje so pokazale, da je uporaba interneta neustrezna: udeleženci izpita spleta niso uporabljali za iskanje informacij ali razlago besed in konceptov, ampak za kopiranje celotnih delov besedila – vsaj tisti spretnejši. Seveda prav pametne tehnologije uporabljamo za odkrivanje tovrstnega plagiatorstva (gl. tudi Holmes 2023: 40–41; Kickbusch idr. 2025: 5). A uporabo interneta na izpitu smo ukinili; tudi zato, ker v ocenjevalnih merilih širša uporaba »podpornih tehnologij« ni predvidena. Ostaja možnost uporabe črkovalnika, a je velika večina naših uporabnikov ne izkoristi.

In čeprav bi bil premik s papirja v digitalni format pri testiranju prihajajočih generacij nujen,[8] še posebej za pisanje na višjih ravneh, imamo tu tehnični problem: opremljenost. Izpitov iz znanja SDTJ se udeležuje vse več ljudi (gl. letna poročila CSDTJ). Ker jih iz vrste razlogov ne izvajamo na daljavo (gl. tudi v nadaljevanju), bi morali zagotoviti ustrezno opremljenost lastnih prostorov, kar pa v našem primeru ni mogoče.

Javnoveljavni program *Slovenščina kot drugi in tuji jezik* (2020) opredeljuje standarde znanja in načine njihovega preverjanja. Vsakršna sprememba v zvezi s tem, npr. digitalna oblika ali izvedba izpita na daljavo ali bolj »vsebinske« spremembe, kot je npr. kombinacija branja/poslušanja in pisanja v isti nalogi, terja tudi spremembo programa. To pa je povezano tudi z institucionalnimi zahtevami: program mora slediti smernicam (Možina idr. 2020), kar je včasih zaradi specifike področja težko, in potrditi ga morajo ustrezni organi;[9] ob tem je potrebna tudi ustrezna regulacija.[10] Takšni postopki otežujejo sprotno prilagajanje hitrim spremembam.

3 *In processu*

Tu se osredotočamo na priložnosti, ki jih pametne tehnologije ponujajo področju jezikovnega testiranja. V tem primeru tehnologije niso nujno »integralni del konstrukta, ampak je njihova vloga izboljšati učinkovitost« postopkov (Yu in Xu 2024: 356). Odkrivanje plagiatorstva smo že omenili; tu izluščimo še tri možnosti.

Prva se nanaša na pripravo izpitnega gradiva, kot je »generiranje stimulusov, vprašanj za razumevanje, praktičnih problemov in kvizov« (Kasneci idr. 2023, v Bolender idr. 2023: 512), pri čemer se še posebej GUI zdi privlačno orodje: je dostopna in učinkovita, prihrani čas in zmanjšuje stroške avtorstva. To so v sistemih standardiziranega jezikovnega testiranja pomembni dejavniki, ki pa imajo lahko brez upoštevanja temeljnih strokovnih standardov, tj. veljavnosti, zanesljivosti in pravičnosti, številne

7 Gre večinoma za populacijo bodočih študentov in poklicnih profilov, ki zahtevajo višjo izobrazbo.

8 Nekatere institucije (npr. Goethe-Institut, Cambridge English) zato ponujajo izpite tako v papirni kot digitalni obliki, kar pa odpira vprašanje primerljivosti obeh načinov.

9 V našem primeru je to Strokovni svet Republike Slovenije za izobraževanje odraslih pri ministrstvu, pristojnem za izobraževanje.

10 Npr. sprememba pravilnika o izvajanju izpitov.

posledice (Bolender idr. 2023). Strokovnjaki pogosto opozarjajo na pristranskost izpitnih gradiv, generiranih z UI (Bolender idr. 2023: 515; Kickbusch idr. 2025: 12–13; Varsik in Vosberg 2024: 38; Voss idr. 2023: 527–528). Avtorji izpitov iz znanja SDTJ si pri pripravi izpitnega gradiva sicer pomagajo z GUI, a ne za oblikovanje vprašanj ob besedilih. Na višjih ravneh jo uporabljajo npr. pri iskanju izpitnih tem, za oblikovanje argumentov glede določenega problema pri eseju in stimulusov, namenjenih komentiranju pri govorjenju. A je pri tem nujna natančna strokovna presoja (»hibridni pristop« po Bolender idr. 2023: 517) in pogosto se zaradi prevelike klišejskosti in leporečenja tako gradivo zavrže. Avtorjem je za pripravo nalog ob besedilih na nižjih ravneh lahko v pomoč tudi orodje za poenostavljanje besedil.[11] A se to uporablja le za poenostavljanje navodil k nalogam, ne pa besedil ali vprašanj ob njih. Pri izdelavi nalog namreč veljajo stroge omejitve, povezane z varovanjem izpitne tajnosti, ki smo mu zavezani v standardiziranem jezikovnem testiranju.

Druga priložnost, ki jo jezikovnemu testiranju ponujajo pametne tehnologije, so prilagodljivi testi in individualiziran pristop, pa tudi možnost povratne informacije. Prilagodljivi testi omogočajo preverjanje dejanskega jezikovnega znanja posameznikov, individualizacija pa upoštevanje njihovih sporazumevalnih praks in potreb. S tem bi lahko zagotovili največjo možno mero avtentičnosti, v primeru oseb s posebnimi potrebami pa tudi pravičnosti.[12] Lahko pa pametna tehnologija ponudi tudi uporabno povratno informacijo glede tega, katere dele jezikovne zmožnosti naj uporabnik izboljša, posebej v primeru negativnega rezultata (ta je glede na izkušnje najpogostejši pri pisanju). Taka individualizacija izpitov iz znanja SDTJ je v tem trenutku neuresničljiva, bi pa takšna praksa bržkone vplivala na motivacijo odraslih za učenje.

In tretja prednost: pametne tehnologije omogočajo avtomatizacijo ocenjevanja. Avtomatsko, s pametno tehnologijo podprto ocenjevanje povečuje hitrost in natančnost, posebej v delih, kjer so odgovori enoznačni in se točkujejo po točkovalnih šablonah (pri objektivnih nalogah za receptivne dejavnosti). Tako ocenjevanje omogoča tudi takojšnjo analizo težavnosti, občutljivosti in zanesljivosti nalog, ki jo zdaj v primeru izpitov iz znanja SDTJ prepuščamo zunanjim strokovnjakom, in učinkovito gradnjo banke nalog.[13] Avtomatizacija ocenjevanja produktivnih dejavnosti (pisanja in govorjenja) pa odpira vrsto vprašanj o zanesljivosti, preglednosti in pravičnosti. Na prvi pogled se zdi, da algoritmi lahko ocenjujejo bolj natančno kot ljudje. Toda algoritmi se oblikujejo na podlagi človeških ocen,[14] ljudje pa smo v svojih presojah nepredvidljivi in podvrženi različnim pristranostim[15] (Kahneman idr. 2021). In »[č]e človeška presoja ni zanesljiva, bo tudi računalnik težko izračunal zanesljivo oceno« (Link 2024: 266). Prav tako različni ocenjevalci v svojih ocenah niso nujno skladni med sabo in se morajo nenehno

11 Npr. Senta, dostopno prek https://senta.cjvt.si/ (dostop 10. 5. 2026).

12 »Podporne tehnologije« so pri testiranju tudi sicer lahko v pomoč osebam s posebnimi potrebami (Voss idr. 2023: 520).

13 Gre za repozitorij nalog, iz katerega se pri sestavi prihodnjih testov jemljejo naloge, t. i. sidra.

14 Kompleksni postopek razvoja enega takih ocenjevalnih algoritmov opisuje Link (2024).

15 O pristranskosti (kot verjetni posledici stereotipov) pri ocenjevanju pisne produkcije na izpitih iz znanja SDTJ gl. Ferbežar in Likar Stanovnik (2021).

»umerjati« oz. standardizirati (Ferbežar 2019: 89). Raziskovalci sicer poročajo o razmeroma visoki stopnji ujemanja med strojnim in človeškim ocenjevanjem (O'Sullivan 2023; nekoliko drugače Link 2024 in Casabianca idr. 2026). Se pa pri razvoju modelov za avtomatsko ocenjevanje omejujejo le na izbrane vidike jezikovne produkcije (Link 2024: 259), običajno na jezikovno pravilnost (Holmes 2023: 44).[16] Za slovenščino kot jezik z majhnim številom uporabnikov se, če se omejimo na pisanje, taki modeli še razvijajo, a ne za namene standardiziranega jezikovnega testiranja, temveč za formativno spremljanje (Arhar Holdt in Munda 2025; Klemen idr. 2025). Strokovnjaki opozarjajo, da je v kontekstih testiranja jezikovnega znanja, kjer se usodno odloča o ljudeh, pri uporabi pametne tehnologije nujno »določiti minimalno sodelovanje človeka«, da bo ocenjevanje veljavno (Casabianca idr. 2026: 28), pa tudi pošteno in pravično. Človek – ob zavedanju, da je njegova presoja vedno nekoliko pristranska in da v besedilu kdaj naredi tudi nepotrebne popravke (Klemen idr. 2025: 172; Pirih Svetina in Ferbežar 2005) – namreč lahko prepozna tudi morebitno tveganje in izvirnost oz., po Cotos (2024: 297), najrazličnejše »odtenke sporazumevalnega namena«.

Dejstvo je, da je modele za avtomatsko ocenjevanje vedno mogoče povezati z zunanjimi referencami (Link 2024: 260), kot je SEJO, kar je v pomoč pri določanju ravni znanja v sistemih preverjanja jezikovnega znanja, ki take ravni predvidevajo. Od lastnega jezikovnega modela smo na izpitih SDTJ še daleč, saj se korpus učečih se, umerjen na lestvice SEJA, šele vzpostavlja (Klemen idr. 2026).

Summa summarum: Čeprav je učinkovitost pomembna, »pa morda ni dovolj 'dober razlog' za vključevanje tehnologije v jezikovno testiranje« (Yu in Xu 2024: 2). Dodajmo: še posebej kadar gre za teste, ki imajo za ljudi pomembne posledice. Na to problematiko kaže posebej opozarjati, saj se v Sloveniji ob zaostrovanju jezikovnih pogojev (Ferbežar in Krajnc 2025) in posledično vse večjih pritiskih na sistem preverjanja znanja SDTJ (množična udeležba na izpitih) vse bolj izražajo tudi želje nekaterih pristojnih institucij po poenostavljanju postopkov – ne oziraje se, da gre to na škodo strokovnosti, preglednosti in pravičnosti.[17]

In ko govorimo z vidika institucije, omenimo še možnosti, ki jih pametne tehnologije ponujajo v administrativnih postopkih. Tu ne mislimo toliko na npr. urejanje podatkovnih baz, ampak na še enega od vidikov varnosti: zlorabe v sistemih, ki so za posameznika odločujoči. Na izpitih iz znanja SDTJ jih zaznavamo vse več, od nedovoljene uporabe tehničnih pripomočkov do lažne identifikacije in ponarejanja spričeval. Izvedba izpitov v digitalnem okolju in na daljavo je za take zlorabe še bolj privlačna, zato zahteva dodatne varnostne ukrepe, kar strokovnjaki izpostavljajo kot vse bolj pereč problem (Association of Language Testers in Europe 2026; Holmes 2024: 40); so namreč dragi. Kot že rečeno, testiranje znanja SDTJ razen na najvišji ravni, ki se ga udeležuje zelo

16 Za slovenščino kot prvi jezik se razvijajo modeli za korekcijo slovničnih napak (Klemen idr. 2025), ki pa niso namenjeni ocenjevanju v standardiziranih testih. Verjamem pa, da pa bo to eden od prihodnjih ciljev.

17 Holmes (2023: III) posebej opozarja na nevarnost komercializacije takih sistemov.

malo oseb, v taki obliki za zdaj ni izvedljivo. Predvsem pa se ob tem odpira povsem nov sklop vprašanj, povezanih z varstvom osebnih podatkov (Isbell idr. 2023).

4 *In fine*

Zato na koncu obrnimo perspektivo: Kako bodo tehnološko podprto jezikovno testiranje sprejeli udeleženci izpitov? Bodo zaupali algoritmom in jim bodo pripravljeni prepustili odločitev? S kakšnimi pomisleki? Kaj bo zanje pomenil izpit v digitalnem okolju? Kako se bodo počutili pod nadzorom različnih naprav? Bodo znali uporabljati tehnologijo ali bodo pri tem potrebovali pomoč? Predvsem pa: ali bodo imeli vsi enake možnosti dostopa do nje? Velik del ciljne populacije v sistemu preverjanja znanja SDTJ je namreč precej specifičen. Prvič, izpitov se večinoma – vsaj na nižjih ravneh – udeležuje zaradi pridobivanja dovoljenj za prebivanje ali naturalizacije (gl. letna poročila CSDTJ). Drugič, za to populacijo je pogosto značilna nižja raven (besedilne in drugih) pismenosti, kar je še dodaten izziv. V tem kontekstu vprašanje digitalizacije ni zgolj tehnično, temveč izrazito etično in socialno. Kot opozarjata Varsik in Vosberg (2024), lahko neustrezna uporaba pametnih tehnologij v izobraževanju – in jezikovno testiranje je njegov del – poglablja obstoječe neenakosti, zlasti pri ranljivih skupinah.

Drugačne kompetence pa bodo nove okoliščine zahtevale tudi od vseh vključenih v postopke jezikovnega testiranja od zmožnosti smiselne in »pravilne« uporabe pametnih tehnologij za pripravo testov ali za ocenjevanje do zmožnosti ločevanja strojne jezikovne produkcije od človeške (Voss idr. 2023: 525).

5 *Ad futura*

Pametne tehnologije ne morejo nadomestiti ljudi, lahko pa dopolnijo njihovo strokovno delo (Bolender idr. 2023: 517–518). In zdi se, da smo se znašli v izrazito prehodnem obdobju. Čeprav ne gre za nove pojave,[18] o njih – gre pač za izjemno hitro »rastoče tehnologije« –, predvsem pa o posledicah njihove uporabe vemo še zelo malo. Zato se v nekaterih okoljih vračajo k papirnim načinom jezikovega testiranja, hkrati pa večjo veljavo dobiva govorjenje. Bržkone pa bo digitalna doba zahtevala tudi novo konceptualizacijo jezikovnega testiranja (Yu in Xu 2024: 357).

Dejstvo je, da se bo testiranje znanja SDTJ prilagajalo novim okoliščinam, čeprav za začetek morda počasi in v zelo omejenem obsegu. Kako, je stvar temeljnega premisleka in sodelovanja strokovnjakov različnih področij. Ob tem bo treba iskati ravnovesje med tehnološkimi možnostmi, teoretskimi utemeljitvami in družbeno odgovornostjo. Le tako bo mogoče, prosto po Lanierju (2025: 82), moč pametnih tehnologij izkoristiti za optimiziranje prav(ičn)ih stvari.

18 O razvoju »tehnološko podprtega jezikovnega testiranja« gl. Bachman 2023; zgodovinski pregled ponujajo Sadeghi in Douglas (2023: 4–10) ter Yu in Xu (2024: 1–4). Sadeghi in Douglas (2023: 1) menita, da je »globalna zdravstvena kriza povzročila revolucijo v jezikovnem testiranju«, posledično pa tudi v raziskovanju tega področja (gl. npr. tematsko številko *Language Assessment Quartely* (2023), *20*(4–5); simpozij *Next generation assessment in AI Era* 2025; konferenco ALTE 2026). Strokovna literatura na to temo je izjemno bogata, tu se omejujem na nekatera najnovejša spoznanja in premisleke.

Literatura

Association of Language Testers in Europe (ALTE). (2026). *Responsible language assessment to empower learners: Conference programme*. https://www.alte.org/resources/Documents/ALTE-Munich2026_Conference_Programme_web_130426.pdf

Arhar Holdt, Š. in Munda, T. (2025). Jezikovno popravljanje v digitalnem okolju. *Sodobna pedagogika, 76*(3), 86–106. https://www.sodobna-pedagogika.net/clanki/03-2025_jezikovno-popravljanje-v-digitalnem-okolju-kvalitativna-studija-z-uciteljicami-in-ucitelji-slovenscine/

Bachman, L. F. (2023). Foreword. V K. Sadeghi in D. Douglas (ur.), *Fundamental considerations in technology mediated language assessment* (str. xix–xxv). Routhledge, Taylor & Francis Group.

Bolender, B., Foster, C. in Vispoel, S. (2023). The criticality of implementing principled design when using AI technologies in test development. *Language Assessment Quarterly*, *20*(4–5), 512–519. https://doi.org/10.1080/15434303.2023.2288266

Casabianca, J. M., McCaffrey, D. F., Johnson, M. S., Alper, N. in Zubenko, V. (2026). From feature-based models to generative AI: Validity evidence for constructed response scoring. *arXiv*. https://doi.org/10.48550/arXiv.2603.19280

Cotos, E. (2024). Towards a validity argument for genre-based automated writing evaluation. V G. Yu in J. Xu (ur.), *Language test validation in a digital age* (str. 279–308). Cambridge University Press & Assessment.

Ferbežar, I. (2016). Kako realno je idealno? O konceptu avtentičnosti pri pouku slovenščine ob Toporišičevem *Zakaj ne po slovensko*. V E. Kržišnik in M. Hladnik (ur.), *Toporišičeva obdobja* (Obdobja 35, str. 489–496). Znanstvena založba Filozofske fakultete. https://centerslo.si/wp-content/uploads/2016/11/Ferbezar.pdf

Ferbežar, I. (2019). Kakovost v jezikovnem testiranju: Misija nemogoče? *Jezik in slovstvo*, *64*(3–4), 83–94. https://www.dlib.si/details/urn:nbn:si:doc-2g1s3gdv

Ferbežar, I., in Krajnc, K. (2025). Tudi to je Slovenija: Jezikovne zahteve in njihovo uresničevanje. V T. Balažic Bulc in V. Gorjanc (ur.), *Govoriti, razumeti, delovati: vpogled v uporabno jezikoslovje* (str. 105–120). Založba Univerze v Ljubljani. https://ebooks.uni-lj.si/zalozbaul/catalog/book/820/chapter/4386

Ferbežar, I., in Likar Stanovnik, P. (2021). How to challenge prejudice in assessing the productive skills of speakers of closely related languages: The case of Slovenia. V B. Lateigne, C. Coombe in J. D. Brown (ur.), *Challenges in language testing around the world: Insights for language test users* (str. 201–219). Springer.

Holmes, W. (2023). *The unintended consequences of artificial intelligence and education*. Education International. https://www.ei-ie.org/en/item/28115-the-unintended-consequences-of-artificial-intelligence-and-education

Isbell, D. R., Kremmel, B. in Kim, J. (2023). Remote proctoring in language testing: Implications for fairness and justice. *Language Assessment Quarterly*, *20*(4–5), 469–487. https://doi.org/10.1080/15434303.2023.2288251

Kahneman, D., Sibony, O. in Sunstein, C. R. (2021). *Hrup: Zakaj tako slabo presojamo*. UMco.

Kalonde, G., Boateng, S. in Duedu, C. (2025). Artificial intelligence in classroom assessment: Opportunities, equity challenges, and best practices for formative and summative integration. *Open Access Library Journal*, *12*, članek e14121. https://doi.org/10.4236/oalib.1114121

Kickbusch, S., Ashford-Rowe, K., Kemp, A., Boreland, J. in Huijser, H. (2025). Beyond detection: Redesigning authentic assessment in an AI-mediated world. *Education Sciences*, *15*(11), članek 1537. https://doi.org/10.3390/educsci15111537

Klemen, M., Božič, M., Arhar Holdt, Š. in Robnik Šikonja, M. (2025). Grammatical error correction of Slovenian school essays using large language models. *Sodobna pedagogika*, *76*(3), 162–176. https://doi.org/10.63384/sptb53z793a

Klemen, M., Stritar Kučuk, M., Eniko, M. in Huber, D. (2026). SEJO KOST: Razvoj novega jezikovnega vira za slovenščino kot neprvi jezik. V Š. Arhar Holdt, V. Gorjanc in T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika : Zbornik konference* (str. 224-243). Inštitut za novejšo zgodovino. https://doi.org/10.5281/zenodo.22765832

Lanier, J. (2025). *Čigava je prihodnost?*. Mladinska knjiga.
Link, S. (2024). The interface between diagnostic writing assessment systems and a socio-cognitive validity framework. V G. Yu in J. Xu (ur.), *Language test validation in a digital age* (str. 251–278). Cambridge University Press & Assessment.
Možina, E., Možina, T. in Žalec, N. (2020). *Izhodišča za pripravo javnoveljavnih izobraževalnih programov za odrasle*. Andragoški center Slovenije. https://www.acs.si/digitalna-bralnica/izhodisca-za-pripravo-javnoveljavnih-izobrazevalnih-programov-za-odrasle/
O'Sullivan, B. (2023). Reflections on the application and validation of technology in language testing. *Language Assessment Quarterly*, *20*(4–5), 501–511. https://doi.org/10.1080/15434303.2023.2291486
Pirih Svetina, N. in Ferbežar, I. (2005). Slovenščine tujejezičnih govorcev. *Jezik in slovstvo*, *50*(6), 3–15. https://www.dlib.si/details/urn:nbn:si:doc-jq8y1zkq
Sadeghi, K. in Douglas, D. (2023). *Fundamental considerations in technology mediated language assessment*. Routledge.
Slovenščina kot drugi in tuji jezik: Izobraževalni program za odrasle. (2020). Center za slovenščino kot drugi in tuji jezik. https://centerslo.si/knjige/ucbeniki-in-prirocniki/prirocniki-in-ucno-gradivo/slovenscina-kot-drugi-in-tuji-jezik_listanje/
Svet Evrope. (2023). *Skupni evropski jezikovni okvir: Učenje, poučevanje, ocenjevanje – dodatek*. Zavod Republike Slovenije za šolstvo. https://www.zrss.si/pdf/sejo.pdf
University College Dublin in Oslo Metropolitan University. (2025, 16. oktober). *Online symposium: Next generation language assessment literacy in AI era.* https://www.ucd.ie/cegl/about/symposium/
Unesco. (2023). *Guidance for generative AI in education and research*. Unesco. https://doi.org/10.54675/EWZM9535
Varsik, S. in Vosberg, L. (2024). *The potential impact of artificial intelligence on equity and inclusion in education* (OECD Artificial Intelligence Papers No. 23). OECD Publishing. https://doi.org/10.1787/15df715b-en
Voss, E., Cushing, S. T., Ockey, G. J. in Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, *20*(4–5), 520–532. https://doi.org/10.1080/15434303.2023.2288256
Yu, G. in Xu, J. (ur.). (2024). *Language test validation in a digital age*. Cambridge University Press & Assessment.

Oblikovno nestandardna leksika v Digitalni slovarski bazi za slovenščino

DOI: 10.4312/Obdobja.45.139-147

Polona Gantar, Filozofska fakulteta in Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; apolonija.gantar@ff.uni-lj.si

Nejc Robida, Institut »Jožef Stefan«, Ljubljana; Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; nejc.robida@ijs.si

Simon Krek, Institut »Jožef Stefan«, Ljubljana; Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; simon.krek@fri.uni-lj.si

V prispevku analiziramo besedne oblike, ki so v trenutni različici oblikoslovnega leksikona Sloleks, od koder prihajajo podatki v Digitalno slovarsko bazo za slovenščino, označene kot nestandardne na ravni zapisa besedne oblike. Posamezne kategorije oblikovne nestandardnosti predstavimo na vzorčnih glasoslovnih, oblikoslovnih in leksikalnih jezikovnih problemih in preverjamo njihovo zastopanost v korpusu standardnega jezika Gigafida 2.0. Evidentirani vzorci jezikovnih napak nam bodo služili za ugotavljanje sorodnih jezikovnih problemov in njihove produktivnosti v jezikovni rabi.

nestandardne besedne oblike, oblikovna nestandardnost, Digitalna slovarska baza za slovenščino, oblikoslovni leksikon Sloleks

This article analyzes word forms marked as nonstandard in the current version of the morphological lexicon Sloleks, from which the data for the Digital Dictionary Database for Slovene are derived. It presents individual categories of form-based non-standardness for sample phonological, morphological, and lexical language problems and examines their representation in the corpus of standard language Gigafida 2.0. The recorded patterns of language errors will serve to identify related language problems and their productivity in language use.

nonstandard word forms, form-based non-standardness, Digital Dictionary Database for Slovene, morphological lexicon Sloleks

1 Uvod

V slovenskem prostoru, tako strokovnem kot laičnem, je pojem *nestandardnost* na podlagi slovenske socialnozvrstne teorije, ki se je uveljavila v *Slovarju slovenskega knjižnega jezika* in njegovih naslednikih, razumljen kot odstopanje od pričakovane knjižnojezikovne norme – t. i. knjižna nestandardnost. V Digitalni slovarski bazi za slovenščino (DSBS; Gantar 2020; Kosem idr. 2021), ki predstavlja celovit leksikalno-gramatični in fonetični opis leksikalnih enot s slovenskim izhodiščem v strojno procesljivi obliki, razumemo nestandardnost kot odklon od pričakovane osnovne ali pregibne

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

oblike oz. fonetične transkripcije, npr. pri napačnem črkovanju ali zapisu sklanjatvene ali spregatvene paradigme. T. i. knjižno nestandardnost na drugi strani razumemo kot določen tip sporočanjske oz. registrske zaznamovanosti, ki se tipično kaže v neformalnih govornih položajih (prim. Gantar idr. 2026).[1] Oblikovna nestandardnost, kot jo obravnavamo v prispevku, torej zadeva besedne oblike, ki so glede na izkazano pravopisno normo (SP: Pravila; ePravopis; Pravopis 8.0) neustrezne z vidika zapisa, in sicer ne glede na svojo potencialno registrsko zaznamovanost.

DSBS primarno vključuje leksiko sodobnega slovenskega jezika, kot se odraža v splošnem korpusu standardnega jezika Gigafida 2.0 (Krek idr. 2020), poleg tega pa tudi tipično govorjeno leksiko, zlasti kot zapis govora, v določenem segmentu pa tudi leksiko, ki predstavlja odklon od pričakovanega zapisa osnovne ali pregibne besedne oblike (Gantar idr. 2026), kar je predmet našega prispevka. Pri tem je treba poudariti, da merila za vključevanje oblikovno nestandardne leksike v DSBS že v izhodišču izključujejo lematizacijske in oblikoskladenjske napake, npr. napačen pripis osnovne oblike ali oblikoskladenjskih lastnosti, ter enkratne pojavitve, ki v DSBS prihajajo iz korpusa Gigafida 2.0.

Odločitev za vključevanje oblikovno nestandardne leksike v DSBS temelji na dejstvu, da so posamezne oblike besed lahko glede na obstoječi jezikovni standard sicer neustrezne, a hkrati v pisni rabi opazno zastopane. Pristop, ki nestandardnih oblik ne izključuje iz jezikovnega opisa, zagotavlja, da nestandardnost sama na sebi ne pomeni hkrati tudi odsotnosti informacije. Še več, sporoča, kaj in zakaj je v določenih okoliščinah bolj ali manj tipično in/ali bolj ali manj sprejemljivo. Vključevanje oblikovno nestandardne leksike odpira tudi široko polje jezikovne variantnosti, ki je odraz jezikovnih trendov in napovednik potencialnih normativnih sprememb. Drugi vidik jezikovne uporabnosti vključevanja nestandardnih besednih oblik v jezikovni opis je uporabnost v jezikovnotehnološke namene, zlasti za razvoj avtomatskih pregledovalnikov in črkovalnikov pa tudi za podajanje povratne informacije v šolskem okolju (prim. Gantar in Bon 2025).

V prispevku analiziramo leksikalne enote (LE), ki so v oblikoslovnem leksikonu Sloleks (Čibej idr. 2022), od koder trenutno prihajajo oblikoslovni in naglasni podatki v DSBS, označene kot nestandardne. Glede na časovno oddaljenost smo predvidevali, da oznaka *nestandardno* v obstoječi različici Sloleksa ni več informativna, zato smo želeli tako označene LE pred vključitvijo v DSBS analizirati z vidika sistematičnosti pripisane oznake in relevantnosti za vključitev v DSBS. Poleg vsebinskih in kvalitativnih meril za vključevanje oblikovno nestandardne leksike v DSBS nam bo kategorizacija pomagala identificirati ključne probleme pišočih in nasloviti problem variantnosti v primerih, kjer raba to nakazuje.

1 Podobno uvaja razmejitev med t. i. sporazumevalnimi položaji in zaznamovanostjo glede na knjižnost tudi eSSKJ (prim. Gliha Komac in Divjak Race 2022: 614) in ePravopis (Uvod).

2 Viri in metodologija

Izhodišče analize predstavljajo LE, ki so 1) v leksikonu Sloleks označene kot nestandardne (Arhar Holdt idr. 2013; Dobrovoljc idr. 2015) na podlagi podedovane podatkovne baze ASES v različici 1.0 (Romih in Holozan 2002), in 2) LE, izluščene na podlagi govornih korpusov, ki so ročno označene glede na tipe nestandardnosti[2] (Čibej idr. 2026). Prvi seznam vsebuje 3285 nestandardnih besednih oblik sicer standardnih lem,[3] ki so v Sloleksu opredeljene kot normativno zaznamovane glede na pravopisni standard (Dobrovoljc idr. 2015: 89). Z drugega seznama, oblikovanega za analizo govorjene in nestandardne leksike za vključitev v DSBS (Gantar idr. 2026), smo vključili 630 LE. Te so bile ročno označene kot take, ki odstopajo od pravopisne norme v zapisu, npr. *pomankljiv*, *iniciranje*. Končni seznam LE, vključenih v analizo, je vseboval 3915 LE.

2.1 Kategorije oblikovno nestandardne leksike

Na podlagi analize smo posamezne primere razvrstili v vrhnje kategorije problemov: primere smo nadalje opredelili glede na tip pravopisne težave in glede na zastopanost v korpusu standardne slovenščine Gigafida 2.0 (Krek idr. 2020). Zaradi obširnosti problematike smo v posamezni vrhnji kategoriji za prikaz problema izbrali le referenčne primere,[4] ki najbolje odražajo pravopisni problem.

2.1.1 Glasoslovje

Znotraj problematike, ki se nanaša na razmerje glas – črka, srečujemo že znane težave pišočih pri zapisu besednih oblik z neobstojnim samoglasnikom, zlasti pri sklanjanju tujih lastnih imen, in problem zapisa preglasa pri samostalnikih in pridevnikih.

Zapis neobstojnega polglasnika

V Sloleksu so kot nestandardne opredeljene pridevniške oblike, ki se v imenovalniku ednine v standardnem izgovoru končujejo na *-lən* in napačno zapisujejo kot *-len*, npr. *brezposelen*, *domiselen*, *nesmiselen*, *popolen*. V obratni logiki najdemo tudi primere kot *ansambl*. V odvisnih sklonih pa je problem očiten pri lastnih imenih, npr. *Dresdena, Diesela*, *Jemena*, *Powella*; *tjulenja*. Posamezne oblike so v korpusu bodisi redkeje (npr. *lahkomiselen*; *Dresdena*, *Jemena*) bodisi opazneje zastopane (npr. *popolen*, *smiselen*; *Diesela*), s čimer se odpira vprašanje sistematične analize tovrstnega vzorca s prepoznavanjem LE, ki se v rabi dejansko kažejo kot problematične. V ta sklop problemov sodi tudi nestandarden zapis polglasnika pri samostalnikih, ki se končajo na zvočniški sklop

2 In sicer kot 1) registrsko nestandardne besede; 2) kot besede, ki odstopajo od pravopisne norme v zapisu; 3) kot neznane besede in 4) kot besede, ki se na podlagi izračuna Levenshteinove razdalje minimalno razlikujejo v zapisu.

3 Normativne oznake so v Sloleksu pripisane le 35 LE (npr. *povdariti*, *življenski* ipd.), če je nestandardna že osnovna oblika besede (Arhar Holdt idr. 2013), vendar pa te oznake v trenutnem vmesniku niso prikazane.

4 V prispevku zaradi gospodarnosti navajamo le nestandardne oblike zapisov, izjema so podatki v tabelah.

-rm, *-lm*, *-vl*. Čeprav v Sloleksu navedeni primeri, kot so *filem*, *šarem*, *alarem*, *kravel*, v korpusu niso pogosti, bi bilo smiselno preveriti rabo pri celotnem vzorcu: v korpusu najdemo npr. še *favel*, v obratni logiki pa zapise govorjenih oblik v neformalnih govornih položajih, npr. *nobl*, *žnabl*, *kurbl*.

Med sorodnimi primeri omenimo še zapis polglasnika v imenovalniku in tožilniku ednine pri samostalnikih ženskega spola, ki se v standardnem izgovoru končajo na *-əv*, npr. *bukva*, *lestva*, *letva*.[5] Zastopanost tovrstnih oblik je v standardnem korpusu sicer sorazmerno skromna, a hkrati vzbuja pozornost odsotnost oblik, kot so *cerkva*, *breskva*, *žetva* ipd. v Sloleksu, ki jih je sicer mogoče najti v zapisih govorjene slovenščine in v registrsko manj formalnih govornih položajih.

Prilikovanje nezvočnikov, prilagoditve sičnikov in samoglasniški prehodi

Velik delež jezikovnih napak na glasoslovni ravni predstavljajo različne prilagoditve zapisa izgovoru. Kot prikazuje izbor glagolov v Tabeli 1, gre za produktivne vzorce, ki zahtevajo sistematično analizo korpusa, pri oblikovanju meril za vključitev v DSBS pa tudi analizo govornih virov in virov nestandardnega besedišča.

Problem	Nestandardni zapis	Standardni zapis
vs → vz	**vs**klikati, **vs**trajati, **vs**podbuditi	**vz**klikati, **vz**trajati, **vz**podbuditi
ras → raz	**ras**porediti	**raz**porediti
is → iz	**is**koristiti, pre**is**kusiti, **is**polniti	**iz**koristiti, pre**iz**kusiti, **iz**polniti
š → ž	te**š**ko	te**ž**ko
s → š	**š**čistiti	**s**čistiti
u → o	**u**postošiti[6]	**o**pustošiti
v → u	**v**novčiti, **v**skladiti, za**v**žiti	**u**novčiti, **u**skladiti, za**u**žiti
u → v	**u**plivati	**v**plivati

Tabela 1: Primeri vzorcev prilagoditev zapisa izgovoru.

Med analiziranimi primeri glede na poizvedovanje v korpusu Gigafida izstopajo posamezni zapisi, kot npr. *vnovčiti*; *preiskusiti* (tudi *preiskus*, *preiskušanje* itn.), med izkazanimi vzorci pa prilagoditve sičnikov: s → š: *ščasoma*, *ščistiti*, *ščvekati*, v obratnem smislu tudi *sčititi*, *sčitek* in š → ž: *teško*.

2.1.2 Oblikoslovje

Pri problematiki zapisa besednih oblik sicer standardnih lem izstopajo napačni zapisi posameznih glagolskih oblik in njihovih posebnosti, pri drugih besednih vrstah pa podaljšave z *-j* in *-t*, zapis soglasniških sklopov *lj*, *nj* in posamezni trši orehi znotraj samostalniških in pridevniških sklanjatev. Na tem mestu izpostavljamo nekaj tipičnih problemskih sklopov.

5 V tem primeru bi lahko govorili tudi o izbiri obrazila po analogiji s samostalniki 1. ženske sklanjatve.

6 Primer je poseben, saj ne gre toliko za prilagoditev izgovoru kot zamenjavo samoglasnika *o* z *u*.

Posebnosti pri glagolih *biti, imeti* in *dati/dajati*

Oblike, ki so v Sloleksu opredeljene kot nestandardne pri glagolih *biti*, *imeti* in *dati*, se nanašajo na zapis dvojinskih in množinskih (zlasti ženskih) oblik, vključno z nikalnimi in zapisom glagolskih sklopov z nikalnico, npr. *sve/nisve*, *sme/nisme*, *bove/nebove*, *bome/nebome*, *neboste/nebojo/nebodo*, *nebi/nebo*. Preverba v korpusu standardnega jezika pokaže, da so te oblike bodisi redko zastopane (npr. *sve*, *nisma*) bodisi sploh ne (npr. *bome*, *bove*, *sme*, *nisme*, *nebove*, *nebome*), lahko pa jih najdemo v korpusu nestandardne slovenščine Janes (Fišer idr. 2020) in v korpusu govorjene slovenščine Gos (Verdonik idr. 2023), kjer najdemo tudi *mave*, *dave*, *greve* itn.

Pri glagolih *dati* in *imeti* izstopajo napačni zapisi sedanjiških oblik, npr. *dajam*, *dajaš*, *daja* ...; *dasta*, *daste*; *imasta*/*nimasta*, *imaste*/*nimaste* ipd., njihova zastopanost v standardnem korpusu pa je sorazmerno redka, pri čemer nekatere oblike, npr. *daja*, *dajajo*, *dajamo*; *imaste*, izstopajo in v tem smislu zahtevajo pozornost glede sistematičnega vključevanja v bazo z ustrezno normativno vrednostjo.

Zapis velelnika

Med glagolskimi oblikami je treba izpostaviti napačne tvorbe velelniških oblik. V Sloleksu so zastopani zlasti naslednji vzorci: a → e: *dejmo*, *dejte*; edi → ej: *izvej*, *izvejmo*; ž → z: *narezi*, *narezimo*, *narezite*; *razrezi*; z → ž: *vrži*, *vržimo*; c → č: *pečiva*, *pečimo*; *preteči*, *pretečimo*; *priteči*, *pritečimo*; *steči*; *teči*; emi → ami: *vzami* in i/ej → aj: *dihi*, pri čemer v Sloleksu ni nestandardnih velelniških oblik *dihej*, *dihejta/dihita*, *dihejmo/dihimo*, čeprav so izpričane v korpusu. Zapisi, ki jih najdemo po tem vzorcu v korpusu standardnega jezika, so še: *nehi*, *nehite*, *čaki*, *čakite* itn.

Podaljšave z *-j*, s *-t* in zapis končnega *-j*

Med analiziranimi primeri z oznako nestandardno v Sloleksu prevladujejo zapisi kratic moškega spola s podaljšavo z *-j* brez vezaja v odvisnih sklonih, npr. *AGRFTja*. Že hiter pregled primerov pokaže, da so bile oblike po tem principu v leksikon vključene sistematično, najverjetneje kot dediščina prvotne različice Amebisove jezikovne baze ASES (Romih in Holozan 2002), saj so vključeni tudi primeri, kjer bodisi nestandardna bodisi variantna oblika v rabi nista realni, npr. *ARNESja/ARNES-ja*, *SDSja*, *SDSju*, *SDSjem/SDS-ja*, *SDS-ju*, *SDS-jem* ipd.

Sloleks vključuje tudi nestandardne oblike samostalnikov, ki se končujejo na *-r*, kjer prihaja v odvisnih sklonih do neustrezne podaljšave z *-j* ali njene odsotnosti, npr. *filterja*; *terciara*, vendar je napak po tem vzorcu v standardnem jeziku malo, zlasti v primerjavi z neustreznim zapisom lastnih imen na nemi *e* pri podaljšavi z *-j*, kjer so v Sloleks vključene samo nekatere nestandardne oblike, npr. *Čileja*, *Presleya*, *Sydneya*; *Applea*, *Pierrea*; *Shakespeareja*, *Shakespearea*; *Stallonea*, ter v obratni logiki: *Georga*; *Bruca*; *Cambridga*, realnih zadreg v standardni slovenščini pa je verjetno še več.

Med primeri, ki jih vključuje Sloleks, naj izpostavimo še zapis končnega *-j* pri samostalnikih moškega spola, npr. *lobij*, *mahagonij*, kot rezultat hiperkorekcije po analogiji s standardnimi oblikami *kalij*, *kriterij*, *imperij* ipd. Poizvedovanje v korpusu sicer

pokaže razmeroma pogost vzorec napačnega zapisa brez končnega *-j*, npr. *scenari*, *laboratori*, *študi*, *solari*, *primari*, ki jih Sloleks ne vključuje. Podobno velja tudi za neustrezne zapise podaljšave s *-t*, npr. *očita*, *Sašota*, ki so v korpusu izkazani zlasti pri lastnih imenih, npr. *Bracota*, *Mišota*, *Božota*, *Jocota*.

Zapis soglasniških sklopov *-lj, -nj* in vrivanje *j* v samoglasniški sklop

V seznamih, ki smo jih pridobili na podlagi ročno označenega nestandardnega besedišča, najdemo predvsem napačne zapise soglasniških sklopov[7] *-lj*, *-nj* in vrivanje *j* v samoglasniški sklop: *io*, *ia* in *ea*, kot prikazuje Tabela 2.

Problem	Nestandardni zapis	Standardni zapis
lj → l	upepe**lj**iti, žu**lj**iti, razvoz**lj**ati	upepe**l**iti, žu**l**iti, razvoz**l**ati
l → lj	temp**l**arski, vk**l**učenost	temp**lj**arski, vk**lj**učenost
n → nj	nadalj**n**i, življe**n**ski, včerajš**n**i, dre**n**, svi**n**ina	nadalj**nj**i, življe**nj**ski, včerajš**nj**i, dre**nj**, svi**nj**ina
nj → n	sum**nj**ičavo, državlja**nj**ski	sum**n**ičavo, državlja**n**ski
ia → ija	univerz**ia**da, vanil**ia**, utop**ia**	univerz**ija**da, vanil**ija**, utop**ija**
ija → ia	kolon**ija**lizacija	kolon**ia**lizacija
ije → ie	rel**ije**f	rel**ie**f
ijo → io	šamp**ijo**n, šamp**ijo**nka	šamp**io**n, šamp**io**nka
ea → eja	alin**ea**, azal**ea**	alin**eja**, azal**eja**
oa → oja	al**oa**	al**oja**

Tabela 2: Primeri zapisa vrivanja *j* v soglasniške in samoglasniške sklope.

Primeri so zanimivi tako zaradi aktualnosti kot tudi zaradi variantnosti, ki se napoveduje pri nekaterih oblikah, npr. *alineja-alinea*; *aloa-aloja.*[8] Primere, kot nakazujejo vzorci v Tabeli 2, pa bi bilo za obvladovanje celotne slike smiselno obravnavati tudi z vidika hiperkorekcije, npr. *vozel*, *buksel*; *takojšenj*, *žuljiti*, *coklje* ipd.

Primeri samostalniških sklanjatev

Med posameznimi primeri samostalniških sklanjatev so v Sloleksu izpostavljeni zapisi napačnega sklanjanja izpridevniških samostalnikov, npr. (*javnega*) *dobra*, *dobru*, *dobrom*,[9] ter znani pravopisni problemi, kot so sklanjanje samostalnika *gospa*: *gospej*; samostalnikov *mati* in *hči* v imenovalniku in tožilniku ednine: *hčer*, *hčera*, *hčero/hči*; *mati*; sklanjanje samostalnika *otrok* v mestniku in orodniku množine: *pri otrokih*, *z otroci/otrocmi*; sklanjanje samostalnika *pes* v orodniku množine: *s psmi*; zapis množinske/dvojinske oblike samostalnikov srednjega spola, npr. *jabolke*; *z dvemi leti/tedni*;

7 Navedene primere bi lahko glede na to, da gre predvsem za zapis pod vplivom izgovora, obravnavali tudi kot glasoslovne probleme.

8 Prim. tudi odgovor jezikovne svetovalnice ISJFR ZRC SAZU: https://svetovalnica.zrc-sazu.si/topic/4775/zakaj-kvinoja-in-ne-kvinoa-ali-kinoa.

9 Trenutno za to obliko v Sloleksu ni navedena standardna ustreznica. O dvojničnem zapisu po vzorcu *dobra/dobrega* gl.: https://svetovalnica.zrc-sazu.si/topic/4259/sklanjanje-zveze-javno--dobro-po-besani.

sklanjanje množinskih samostalnikov: *možgan*; *na plečah*; *pljučam/pljučom*, *na pljučah*; *tlem* itn.

2.1.3 Besedoslovje

Med problemi, ki smo jih zaznali na seznamu oblikovno nestandardne leksike, gre tipično za uveljavljanje različnih stopenj podomačevanja lastnih imen in zapis skupaj oz. narazen.

(Ne)podomačevanje

Vključevanje različnih stopenj podomačevanja tujih izrazov, ki se uveljavljajo v pisni slovenščini, bi v DSBS moralo slediti realnim problemom, z vidika pravopisne norme pa je treba razmisliti o variantah in sistemskih rešitvah pri poenotenju pravil. Med opaznimi primeri lahko izpostavimo: *benefikacija – bonifikacija*; *esteblišment – establišment – establishment*; *instrumental – inštrumental*; *rokerski – rockerski*; *maskarpone – mascarpone*; *dahavski – dachavski – dachauski*, *evropski – europski*, *hapaks – hapax* itn.

Zapis skupaj in narazen

Sloleks trenutno ne vključuje večbesednih enot tipa: *na tihem*, *po malem* ipd., se pa ta problem očitno kaže v seznamu potencialno oblikovno nestandardne leksike v govornih virih (Čibej idr. 2026), če izpostavimo le nekaj pogostejših variantnih oblik: *takoimenovan – tako imenovan*; *natihem – na tihem*; *vredu – v redu* (in *uredu – u redu*); *brezveze – brez veze*, *knowhow – know how* ter tipično zveze s *čim-*: *čimdlje*, *čimbolje*, *čimvečji* itn.

3 Sklep in nadaljnje delo

Analiza primerov z oznako nestandardno v trenutni različici leksikona Sloleks je na eni strani pokazala vključenost besed, ki sicer ne nastopajo kot konkretni jezikovni problemi pišočih, saj zanje v korpusu standardnega jezika Gigafida 2.0 ni mogoče najti pojavitev, vendar je takih primerov razmeroma malo. Na drugi strani smo prepoznali posamezne produktivne vzorce napačnih zapisov besednih oblik, ki so posledica različnih glasoslovnih in oblikoslovnih zadreg. Za vključevanje v DSBS je zato smiselno upoštevati čim bolj splošno napačno rabljene zapise ter izdelati metodologijo za avtomatično luščenje le-teh iz korpusa, s čimer se izognemo vključevanju hapaksov v jezikovni opis. Tretje ključno načelo, ki ga je treba upoštevati pri vključevanju oblikovno nestandardne leksike v DSBS, je upoštevanje oblikoslovnih tendenc v sodobni jezikovni rabi, o kakršni lahko sklepamo na podlagi reprezentativnih korpusov sodobne slovenščine (Dobrovoljc idr. 2015: 98). Z vidika jezikovnih tehnologij je upoštevanje jezikovne rabe pomembno zaradi zagotavljanja čim večje pokritosti pogosto rabljenega besedišča, pri čemer je z vključitvijo napačno zapisanih LE v bazo vključena tudi informacija o njihovem obstoju, z vpeljavo normativnih informacij in povezav pa tudi njihova standardna oz. standardizirana ustreznica.

Nenazadnje lahko sistematična analiza problemov v evidentiranih vzorcih pomaga pri odkrivanju sorodnih jezikovnih problemov. V zvezi s tem pri razvoju nevronskega črkovalnika za slovenščino (Klemen idr. 2025) preizkušamo metode, pri katerih s pomočjo velikih jezikovnih modelov na podlagi vhodnih podatkov, ki zahtevajo jezikovni opis problema in ročno pripravljeno učno množico, generiramo povedi z vsebovano napako v zapisu in njeno standardizirano ustreznico.

V prispevku sicer ne ponudimo konkretnih meril za vključevanje oblikovno nestandardne leksike v DSBS, razen izhodiščnega merila, ki določa aktualnost problema v pisni produkciji, kar je pomembno tako z vidika uporabnikov kot z vidika jezikovnotehnološke uporabe. Kategorizacija problemov je namenjena predvsem ciljno usmerjenim poizvedbam v splošnem pa tudi v specializiranih korpusih, ko nas zanima npr. govorjena leksika ali zapis govorjenega jezika. Omenjena raziskava je tako izhodišče za nadaljnje analize oblikovno nestandardne leksike v slovenščini pa tudi za ugotavljanje variantnosti in sprememb v normi.

Prispevek je nastal v okviru raziskovalnih programov Slovenski jezik – bazične, kontrastivne in aplikativne raziskave (P6-0215) in Jezikovni viri in tehnologije za slovenski jezik (P6-0411), ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije.

Literatura

Arhar Holdt, Š., Dobrovoljc, K. in Popič, D. (2013). Reprezentacija standardnega in nestandardnega v virih SSJ. V A. Žele (ur.), *Družbena funkcijskost jezika (vidiki, merila, opredelitve)* (Obdobja 32, str. 19–27). Znanstvena založba Filozofske fakultete. https://centerslo.si/wp-content/uploads/2015/10/32-ArharHoldt.pdf

Čibej, J., Gantar, K., Dobrovoljc, K., Krek, S., Holozan, P., Erjavec, T., Romih, M., Arhar Holdt, Š., Krsnik, L. in Robnik Šikonja, M. (2022). *Morphological lexicon Sloleks 3.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1745

Čibej, J., Gantar, P., Šešet, J., Krek, S. in Robida, N. (2026). *List of potentially non-standard vocabulary candidates MEZZANINE-NstdLex 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/2084

Dobrovoljc, K., Krek, S. in Erjavec, T. (2015). Leksikon besednih oblik Sloleks in smernice njegovega razvoja. V V. Gorjanc, P. Gantar, I. Kosem in S. Krek (ur.), *Slovar sodobne slovenščine: Problemi in rešitve* (str. 80–105). Znanstvena založba Filozofske fakultete.

Fišer, D., Ljubešić, N. in Erjavec, T. (2020). The Janes project: Language resources and tools for Slovene user generated content. *Language Resources and Evaluation*, *54*(1), 223–246. https://doi.org/10.1007/s10579-018-9425-z

Gantar, P. (2020). Dictionary of modern Slovene: From Slovene lexical database to digital dictionary database. *Rasprave Instituta za hrvatski jezik i jezikoslovlje, 46*(2), 589–602. https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=356570

Gantar, P. in Bon, M. (2025). Dati skozi ali prestati?: Napake in jezikovne variante v rabi večbesednih enot pri samostojnem tvorjenju besedil v osnovni in srednji šoli. *Sodobna pedagogika*, *76*(142), 39–58. https://www.sodobna-pedagogika.net/arhiv/nalozi-clanek/?id=2551

Gantar, P., Čibej, J., Robida, N. in Krek, S. (2026). Govorjena in nestandardna leksika v Digitalni slovarski bazi za slovenščino. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, *14*(1), 7–33.

Gliha Komac, N. in Divjak Race, D. (2022). Tipologija slovarskih oznak za tretjo izdajo Slovarja slovenskega knjižnega jezika (eSSKJ). *Slavistična revija, 70*(4), 611–625. https://srl.si/ojs/srl/article/view/4086

Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. (2014–). *ePravopis: Slovar slovenskega pravopisa*. Založba ZRC, ZRC SAZU. https://doi.org/10.3986/2385-8621

Klemen, M., Božič, M., Arhar Holdt, Š. in Robnik-Šikonja, M. (2025). Grammatical error correction of Slovenian school essays using large language models. *Sodobna pedagogika, 76*(3), 162–176. https://doi.org/10.63384/sptB53z793a

Kosem, I., Krek, S. in Gantar, P. (2021). Semantic data should no longer exist in isolation: The digital dictionary database of Slovenian. V Z. Gavriilidou, L. Mitits in S. Kiosses (ur.), *EURALEX XIX, Congress of the European association for lexicography, Lexicography for inclusion, 7-9 September, 2021, virtual, Book of Abstracts* (str. 81–83). Democritus University of Thrace.

Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I. in Dobrovoljc, K. (2020, 11.–16. maj). Gigafida 2.0: The Reference Corpus of Written Standard Slovene. V N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk in S. Piperidis (ur.), *Proceedings of the 12th Conference on language resources and evaluation, May 11-16, 2020, Marseille, France* (str. 3340–3345). European language resources association (ELRA). http://www.lrec-conf.org/proceedings/lrec2020/LREC-2020.pdf

Pravopis 8.0: Pravila novega slovenskega pravopisa za javno razpravo. (b. d.). Založba ZRC, ZRC SAZU. https://pravopis8.fran.si/ (pridobljeno 5. 5. 2026)

Romih, M. in Holozan, P. (2002). Infrastruktura za razvoj jezikovnih tehnologij – korpus FIDA in sistem ASES. V T. Erjavec in J. Gros (ur.), *Zbornik B 5. mednarodne multikonference Informacijska družba IS'2002: Jezikovne tehnologije, 14.–15. oktober, 2002, Ljubljana, Slovenija* (str. 166). http://nl.ijs.si/isjt02/zbornik/sdjt02-D02amebis.pdf

Slovenski pravopis. (2014). Založba ZRC, ZRC SAZU. www.fran.si (pridobljeno 5. 5. 2026)

Verdonik, D., Zwitter Vitez, A., Zemljarič Miklavčič, J., Krek, S., Stabej, M., Erjavec, T., Potočnik, T., Sepesy Maučec, M., Majhenič, S., Žgank, A., Bizjak, A., Gril, L., Dobrišek, S., Križaj, J., Bajec, M., Lebar Bajec, I., Jelovšek, T., Trojar, M., Bernjak, M., … Rupnik, P. (2023). *Spoken corpus Gos 2.1 (transcriptions)*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1863

Razvijanje pravopisne zmožnosti z vključevanjem pristopa obrnjenega učenja in poučevanja

DOI: 10.4312/Obdobja.45.149-158

Lara Godec Soršak, Pedagoška fakulteta, Univerza v Ljubljani, Ljubljana;
lara.sorsak@pef.uni-lj.si

Miha Matjašič, Pedagoška fakulteta, Univerza v Ljubljani, Ljubljana;
miha.matjasic@pef.uni-lj.si

V prispevku predstavljamo mnenje študentov o razvijanju pravopisne zmožnosti z vključevanjem pristopa obrnjenega učenja in poučevanja, ki je bilo predvideno kot samostojna priprava študentov na vaje v živo. Večina študentov je bila zadovoljna s pripravljenim digitalnim gradivom za usvajanje pravopisnih vsebin (tj. z videoposnetki z učiteljevo razlago, izročki v obliki drsnic in s pravopisnimi nalogami v programu H5P), samostojno pripravo na pouk po omenjenem pristopu pa so ocenili kot pozitivno. Na podlagi pridobljenih rezultatov predstavljamo prednosti in pomanjkljivosti obrnjenega učenja in poučevanja pravopisnih vsebin pri jezikovnem pouku na visokošolski ravni.

obrnjeno učenje in poučevanje, slovenski jezik, pravopis, H5P, razredni pouk

This article presents students' opinions on the development of spelling skills using the flipped learning and teaching (FLT) approach, which was intended as independent preparation for students prior to classes. Most students were satisfied with the digital materials prepared for learning ortography (i.e., videos with the teacher's explanations, handouts in the form of slides, and exercises in the H5P program), and they found their independent preparation for class using this approach to be positive. Based on the results obtained, we present the advantages and disadvantages of FLT for ortographic content in language instruction in higher education.

flipped learning and teaching, Slovenian, orthography, H5P, primary teacher education

1 Uvod

V okviru projekta Posodobitev pedagoških študijskih programov – Posodobitev PŠP NOO smo na Pedagoški fakulteti Univerze v Ljubljani (UL PEF) v študijskem letu 2024/25 posodobili učni načrt in izvajanje predmeta Slovenski jezik 1 (SJ 1) za študente 1. letnika študijskega programa Razredni pouk (RP) z vključevanjem digitalnih vsebin. Za usvajanje pravopisnih pravil in poglabljanje pravopisne zmožnosti študentov smo pripravili videoposnetke z učiteljevo razlago, drsnice in naloge v digitalni obliki H5P[1]

1 H5P (https://h5p.org/) je odprtokodno orodje za ustvarjanje interaktivnih učnih vsebin, ki temeljijo na tehnologiji HTML5, kot so kvizi, predstavitve, videoposnetki in druge dejavnosti za sprotno preverjanje znanja. Uporabili smo ga kot vtičnik, integriran v učno okolje Moodle, kar je študentom

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

ter jih vključili v pristop obrnjenega učenja in poučevanja (OUP). Ta je bil načrtovan kot didaktična inovacija za razvijanje pravopisne zmožnosti študentov, predvsem kot samostojna priprava študentov na vaje. Te so se izvajale v treh skupinah pri istem izvajalcu. Študentje naj bi ob pripravljenem digitalnem gradivu pred posameznim srečanjem v okviru vaj usvojili, ponovili in utrdili pravopisna pravila posameznega pravopisnega sklopa, in sicer o pisanju krajšav, pisanju besed skupaj/narazen/z vezajem, deljenju besed, pisanju prevzetih besed, zapisovanju glasov s črkami in o rabi ločil.[2] Nato bi na vajah z različnimi učnimi oblikami razpravljali o težjih pravopisnih primerih, dvojnicah, primerjali veljavna pravopisna pravila z novo nastajajočimi v *Pravopisu 8.0*, sestavljali (digitalne) pravopisne naloge, primerne za učence na razredni stopnji, ipd. Ker se večina študentov pred posameznimi vajami ni ustrezno pripravila nanje, vseh načrtovanih aktivnejših oblik dela nismo mogli izvesti, ampak smo na vajah razložili vsebino, ki je bila objavljena na videoposnetkih, in razpravljali o težjih primerih. Ob koncu vseh srečanj z obravnavo pravopisnih vsebin so študentje pisali pravopisni test (PT). Na podlagi izkušenj z vaj predvidevamo, da so študentje digitalno gradivo uporabili kot pripravo na PT. V prispevku predstavljamo mnenje študentov o digitalnem gradivu, samostojnem učenju in načinu obravnave pravopisnih vsebin ter preverjamo, ali je bila domača priprava povezana z uspešnostjo študentov pri PT.

2 Obrnjeno učenje in poučevanje

Pri pouku prevladujeta frontalna in individualna učna oblika (Plešec Gasparič in Valenčič Zuljan 2019; Rot Vrhovec 2015), kar velja tudi za predmet SJ 1. Z vključevanjem večjega obsega digitalnih vsebin v SJ 1 v okviru projekta PŠP NOO pa se je pokazala priložnost za uvedbo OUP. Plešec Gasparič in Valenčič Zuljan (2019: 271) definirata OUP »kot didaktično inovacijo, pri kateri učenci doma samostojno obravnavajo novo učno vsebino s pomočjo vnaprej pripravljenega interaktivnega gradiva (npr. videoposnetek), večji del pouka ‘v živo’ pa zaradi prihranjenega časa lahko poteka v individualni in skupinski učni obliki ter v paru«. OUP od učitelja zahteva skrbno pripravo, izvedbo in evalvacijo ter zagotavljanje sprotne povratne informacije učencem, hkrati pa mu omogoča vpeljavo raznih učnih oblik pri pouku na prožnejši in bolj inovativen način ter ima pomembne pozitivne učinke za učenca. Temu omogoča samostojno organizacijo učenja, ustreznejšo časovno obremenitev, večjo miselno aktivnost, doseganje višjih taksonomskih stopenj po Bloomu, višje učne dosežke, poglobljeno in trajnejše znanje, obenem pa lahko pri učečih se spodbuja višja pojmovanja znanja in njihove sodelovalne spretnosti (Bec in Podobnikar 2015; Plešec Gasparič in Valenčič Zuljan 2019; Plešec Gasparič idr. 2020).

Bistvo OUP niso le dejavnosti, ki jih učenci namesto v učilnici izvajajo izven nje, ampak se mora kakovostno spremeniti tudi pouk v učilnici (Bishop in Verleger 2013),

omogočalo neposreden dostop do interaktivnih dejavnosti in njihovo reševanje znotraj spletne učilnice. Ustvarili smo predvsem naloge tipa *vstavi manjkajoče besede*, *kartice z vprašanji*, *kartice dialogov*, *enkratni izbor*, *večkratni izbor* in *povleci besede.*

2 Za vsak pravopisni sklop smo pripravili videoposnetek z učiteljevo razlago in drsnice ter številne raznovrstne digitalne naloge H5P (za vse vsebine skupaj več kot 60 nalog).

da omogoča več diskusije, odpravljanje vrzeli v znanju, reševanje problemov in težjih vprašanj ipd. (Plešec Gasparič idr. 2020). Pogoj za to, da učitelj lahko izvede aktivnejše učne oblike pri pouku, je, da učenci samostojno vnaprej opravijo predvidene dejavnosti, ki jih je zanje pripravil učitelj. Te so lahko po mnenju nekaterih avtorjev zelo raznolike, Bishop in Verleger (2013) pa upoštevata le tiste, ki vključujejo tehnološko komponento oz. ogled videoposnetka. OUP opredeljujeta kot asinhrona videopredavanja in vadenje za domačo nalogo ter aktivno, skupinsko zasnovano reševanje problemov pri pouku. Ugotovila sta, da so mnogi učitelji po priporočilu učencev uvedli kvize, s katerimi preverijo znanje učencev pred poukom, in to se je izkazalo kot zelo učinkovito.

Raziskave OUP v visokošolskem izobraževanju mdr. kažejo, da njegove prednosti pretehtajo nad pomanjkljivostmi, da so stališča študentov do njega večinoma pozitivna, a pri nekaterih tudi izrazito negativna (Bishop in Verleger 2013; Plešec Gasparič idr. 2020). Študentom so ljubša predavanja v živo kot na videoposnetkih, vendar imajo raje interaktivne dejavnosti v živo kot predavanja v živo (Bishop in Verleger 2013; Plešec Gasparič idr. 2020). Slednje je razumljivo, saj lahko digitalna tehnologija izboljša učenje, poleg tega sodobne generacije učencev pričakujejo uporabo digitalne tehnologije (Prensky 2001, v Bec in Podobnikar 2015). Raziskave kažejo tudi to, da študentje z OUP dosegajo višje učne dosežke kot tisti, ki so deležni tradicionalnega izobraževanja, in da je uspešnost večja z uporabo videogradiva kot z bralnim gradivom (Eryilmaz in Cigdemoglu 2019; Fisher idr. 2024; Lee in Choi 2019).

Pri vpeljevanju OUP je treba učence seznaniti z načinom dela ter jim predstaviti zahteve in pričakovanja, da se bodo zavedali svoje odgovornosti in da se bo v njih prebudila vedoželjnost (Plešec Gasparič idr. 2020). Za učitelja je pomembno, da je usposobljen za uporabo IKT ter da ima na voljo dobro tehnično opremo in podporo računalničarja pri pripravi in posredovanju e-gradiv (Plešec Gasparič idr. 2020). Pristop OUP predstavlja za učitelja poglobljeno učno izkušnjo in prispeva k njegovemu profesionalnemu razvoju, pomembni pa so tudi medsebojno sodelovanje, podpora in pomoč učiteljev z izmenjavo izkušenj ter povratna informacija (Plešec Gasparič idr. 2020; Valenčič Zuljan 1997). Za pridobitev slednje smo anketirali študente, da bi uvedli spremembe in izboljšave pri izvedbi SJ 1 v naslednjih letih.

3 Analiza ankete o mnenju študentov UL PEF o OUP

3.1 Potek anketiranja

Pri študentih 1. letnika RP na UL PEF smo želeli preveriti, kaj menijo o videoposnetkih, nalogah v digitalni obliki, samostojnem učenju in načinu obravnave pravopisnih vsebin ter ali je bila domača priprava povezana z uspešnostjo pri PT.

Izvedli smo empirično kvantitativno raziskavo z deskriptivno in kavzalno neeksperimentalno metodo pedagoškega raziskovanja. V raziskavo je bilo vključenih 145 študentov, ki so bili prvič redno vpisani v 1. letnik RP in so obiskovali vaje pri predmetu SJ 1 v študijskih letih 2024/25 in 2025/26. Vzorec je bil neslučajnostni. Podatki o številu vpisanih in sodelujočih študentov so prikazani v Tabeli 1.

Študijsko leto	Število prvič redno vpisanih študentov	n	%
2024/25	81	60	74,1
2025/26	92	85	92,4
Skupaj	173	145	83,8

Tabela 1: Število študentov glede na študijsko leto.

Na spletni strani 1KA[3] smo pripravili spletni vprašalnik z osmimi vsebinskimi vprašanji zaprtega tipa in enim vprašanjem odprtega tipa. Povezava do vprašalnika in prošnja za izpolnjevanje sta bili študentom poslani na njihove študentske e-naslove in objavljeni v spletni učilnici. Anketiranje je potekalo 25.–28. 5. 2025 in 10.–14. 5. 2026. Zbrane podatke smo obdelali s programom R 4.3.1.

3.2 Rezultati

Za analizo podatkov smo uporabili opisno in inferenčno statistiko. Pri vprašanjih z nominalnimi oz. ordinalnimi merskimi lestvicami so bile izračunane frekvence in odstotki.

Razlike med študijskima letoma pri ogledovanju posnetkov in reševanju digitalnih nalog smo testirali s hi-kvadrat testom. Ker so bile pri nekaterih kategorijah frekvence majhne, s tem pa je bila kršena tudi predpostavka hi-kvadrat testa, smo uporabili Fisher-Freeman-Haltonov test. Poleg natančne stopnje značilnosti smo izračunali tudi velikost učinka, in sicer Cramerjev V. Na koncu smo preverili še povezanost med domačo pripravo in oceno PT, in sicer tako, da smo izračunali Spearmanov koeficient korelacije rangov, ker sta bili obe spremenljivki ordinalni.

Najprej v Tabelah 2, 3 in 4 predstavljamo rezultate o tem, kaj menijo študentje o gradivu in samostojni pripravi na pouk.

Trditev	N	Povsem se strinjam.	Strinjam se.	Ne morem se odločiti.	Ne strinjam se.	Sploh se ne strinjam.
		n (%)				
Posnetki in razlaga na njih so bili razumljivi.	129	76 (58,9)	49 (38,0)	4 (3,1)	0 (0,0)	0 (0,0)
Posnetki so mi pomagali pri učenju.	127	55 (43,3)	59 (46,5)	11 (8,7)	2 (1,6)	0 (0,0)
Želim si še posnetkov o drugih temah.	128	57 (44,5)	49 (38,3)	14 (10,9)	8 (6,2)	0 (0,0)
Ljubša mi je razlaga profesorice v živo na fakulteti.	128	32 (25,0)	33 (25,8)	56 (43,8)	6 (4,7)	1 (0,8)

Tabela 2: Mnenje študentov o videoposnetkih.

Kot prikazuje Tabela 2, so študentje videoposnetke z učiteljevo razlago ocenili zelo pozitivno, saj so se skoraj vsi, ki so odgovorili na ta sklop, strinjali, da so bili razumljivi (96,9 %), večina pa tudi, da so jim pomagali pri učenju (89,8 %). Pri trditvi, da jim je ljubša razlaga profesorice v živo, je več neodločenih odgovorov, kar kaže, da študentje

3 https://1ka.arnes.si

posnetkov niso nujno razumeli kot zamenjavo za delo v živo, ampak bolj kot podporo učenju. Večina študentov (82,8 %) si želi še posnetkov o drugih temah.

Trditev	N	Povsem se strinjam.	Strinjam se.	Ne morem se odločiti.	Ne strinjam se.	Sploh se ne strinjam.
		n (%)				
Digitalne naloge so zelo koristne.	135	68 (50,4)	60 (44,4)	4 (3,0)	3 (2,2)	0 (0,0)
Digitalne naloge so mi pomagale pri učenju.	134	69 (51,5)	58 (43,3)	5 (3,7)	2 (1,5)	0 (0,0)
Raje rešujem naloge v tiskani obliki.	134	24 (17,9)	32 (23,9)	41 (30,6)	35 (26,1)	2 (1,5)
Želel/-a bi si še več takih nalog.	135	48 (35,6)	70 (51,9)	14 (10,4)	3 (2,2)	0 (0,0)

Tabela 3: Mnenje študentov o nalogah v digitalni obliki.

Rezultati v Tabeli 3 kažejo, da so študentje digitalne naloge ocenili izrazito pozitivno. Skoraj vsi so se strinjali ali povsem strinjali, da so digitalne naloge koristne (94,8 %) in da so jim pomagale pri učenju (94,8 %). Tudi želja po dodatnih digitalnih nalogah je bila precej izrazita, saj bi si jih še več želelo 87,5 % študentov. Pri trditvi, da raje rešujejo naloge v tiskani obliki, so bili odgovori bolj razpršeni, saj se je 41,8 % študentov s trditvijo strinjalo ali povsem strinjalo, 30,6 % se jih ni moglo odločiti, 27,6 % pa se jih ni strinjalo.

Trditev	N	Povsem se strinjam.	Strinjam se.	Ne morem se odločiti.	Ne strinjam se.	Sploh se ne strinjam.
		n (%)				
Tak način dela mi je všeč.	141	33 (23,4)	59 (41,8)	24 (17,0)	24 (17,0)	1 (0,7)
Spodbuja samostojno delo študentov.	141	52 (36,9)	68 (48,2)	10 (7,1)	11 (7,8)	0 (0,0)
Na vajah smo se lahko posvetili zahtevnejšim primerom in vprašanjem.	141	43 (30,5)	69 (48,9)	13 (9,2)	14 (9,9)	2 (1,4)
Motilo me je, da smo na vajah obravnavali iste primere.	141	15 (10,6)	27 (19,1)	26 (18,4)	58 (41,1)	15 (10,6)
Motilo me je, da se niso vsi študentje doma pripravili na vaje.	141	7 (5,0)	25 (17,7)	36 (25,5)	58 (41,1)	15 (10,6)
Ljubše bi mi bilo, če bi vsebine obravnavali samo na vajah.	141	11 (7,8)	29 (20,6)	36 (25,5)	53 (37,6)	12 (8,5)
Ogledovanje posnetkov in reševanje nalog doma vzameta preveč časa.	141	10 (7,1)	40 (28,4)	39 (27,7)	47 (33,3)	5 (3,5)
Digitalna priprava je prispevala k mojemu boljšemu znanju.	140	41 (29,3)	62 (44,3)	31 (22,1)	6 (4,3)	0 (0,0)
Zaradi digitalne priprave sem bolje pisal/-a PT.	140	39 (27,9)	35 (25,0)	40 (28,6)	20 (14,3)	6 (4,3)

Tabela 4: Mnenje študentov o samostojni pripravi na pouk.

Iz Tabele 4 je razvidno, da so študentje samostojno pripravo na pouk večinoma ocenili pozitivno. Najvišje strinjanje je pri trditvi, da tak način dela spodbuja samostojno delo študentov, saj se je s tem strinjalo ali povsem strinjalo 85,1 % študentov. Prav tako se jih je 79,4 % strinjalo ali povsem strinjalo, da so se lahko na vajah posvetili zahtevnejšim primerom in vprašanjem. Večina študentov meni, da je digitalna priprava prispevala k njihovemu boljšemu znanju, saj se je s to trditvijo strinjalo ali povsem strinjalo 73,6 % študentov. Nekoliko manj enotni so pri oceni, da so zaradi digitalne priprave bolje pisali PT; s tem se je strinjalo ali povsem strinjalo 52,9 % študentov, 28,6 % pa se jih ni moglo odločiti.

Po drugi strani pa približno tretjina študentov meni, da ogledovanje posnetkov in reševanje nalog doma vzameta preveč časa, skoraj enak delež pa se jih s tem ne strinja. Prav tako 28,4 % študentov navaja, da bi jim bilo ljubše, če bi vsebine obravnavali samo na vajah. To kaže, da je način dela za večino sprejemljiv, vendar ne ustreza vsem enako.

Nato nas je zanimalo, ali je samostojna priprava študentov po pristopu OUP povezana z oceno PT. V nadaljevanju predstavljamo rezultate ocene PT, podatke o tem, koliko študentov si je ogledovalo posnetke in reševalo naloge, ter ali je ocena PT povezana s samostojno pripravo študentov doma.

Ocena PT	n	%
10	13	9,0
9	10	6,9
8	17	11,7
7	18	12,4
6	18	12,4
5	64	44,1
PT ni pisal/-a	5	3,4

Tabela 5: Ocena PT.

Tabela 5 prikazuje, da je pozitivno oceno doseglo 76 študentov (52,4 %), 64 študentov (44,1 %) je pisalo 5, 5 študentov (3,4 %) pa PT ni pisalo. Študentje so kot glavne razloge za pozitivno oceno najpogosteje navedli veliko vaje in učenja ter znanje, pridobljeno na vajah. Ogledovanje posnetkov in reševanje digitalnih nalog je kot glavni razlog navedel manjši delež študentov, pri odgovorih *drugo* pa so pogosto izpostavili kombinacijo več dejavnikov, npr. vaje na fakulteti, posnetke, digitalne naloge in reševanje poskusnega testa. Študentje, ki PT niso opravili, so kot najpogostejša razloga navedli premalo natančnosti ter premalo vaje in učenja.

Odgovor	2024/25 n (%)	2025/26 n (%)	Skupaj n (%)	χ^2	df	Fisher-Freeman-Haltonov p	Cramerjev V
vse posnetke	29 (48,3)	36 (42,4)	65 (44,8)				
nekaj posnetkov	28 (46,7)	37 (43,5)	65 (44,8)	3,18	2	0,220	0,15
nobenega posnetka	3 (5,0)	12 (14,1)	15 (10,3)				

Tabela 6: Ogledovanje posnetkov glede na študijsko leto.

Iz Tabele 6 je razvidno, da si je v obeh letih večina študentov ogledala vse oz. vsaj nekaj posnetkov. V letu 2025/26 je bil delež študentov, ki si niso ogledali nobenega posnetka, nekoliko višji kot v letu 2024/25, vendar razlika med letoma ni statistično značilna (p = 0,220).

Odgovor	2024/25 n (%)	2025/26 n (%)	Skupaj n (%)	χ^2	df	Fisher-Freeman-Haltonov p	Cramerjev V
vse naloge	37 (61,7)	41 (48,2)	78 (53,8)				
nekatere naloge	22 (36,7)	35 (41,2)	57 (39,3)	5,42	2	0,064	0,19
ni reševal/-a nalog	1 (1,7)	9 (10,6)	10 (6,9)				

Tabela 7: Reševanje digitalnih nalog glede na študijsko leto.

V Tabeli 7 vidimo, da je v obeh letih večina študentov rešila vse ali nekatere digitalne naloge. V letu 2024/25 je bil delež tistih, ki so rešili vse naloge, nekoliko višji, a razlika ni statistično značilna (p = 0,064).

Spremenljivka	Odgovor	n	Pozitivna ocena PT n (%)	Negativna ocena PT n (%)
Ogledovanje posnetkov	vse posnetke	65	46 (70,8)	19 (29,2)
	nekaj posnetkov	62	24 (38,7)	38 (61,3)
	nobenega posnetka	13	6 (46,2)	7 (53,8)
Reševanje digitalnih nalog	vse naloge	78	50 (64,1)	28 (35,9)
	nekatere naloge	53	21 (39,6)	32 (60,4)
	ni reševal/-a nalog	9	5 (55,6)	4 (44,4)

Tabela 8: Delež pozitivnih in negativnih ocen PT glede na ogledovanje posnetkov in reševanje digitalnih nalog.

Tabela 8 prikazuje, da je med študenti, ki so si ogledali vse videoposnetke, delež pozitivnih ocen PT najvišji. Podoben vzorec je pri digitalnih nalogah: med tistimi študenti, ki so rešili vse naloge, je delež pozitivnih ocen PT višji kot med tistimi, ki so rešili le nekatere naloge. Kategoriji *nobenega posnetka* in *ni reševal/-a nalog* sta številčno majhni, zato ju je treba interpretirati previdno.

Spremenljivka	n	χ^2	df	p (χ^2)	Fisherjev p	Cramerjev V
Ogledovanje posnetkov	140	13,52	2	0,001	0,001	0,31
Reševanje digitalnih nalog		7,63		0,022	0,020	0,23

Tabela 9: Preverjanje povezanosti opravljanja dejavnosti s pozitivno oz. negativno oceno PT.

Rezultati v Tabeli 9 kažejo, da je uspešnost pri PT statistično značilno povezana tako z ogledovanjem posnetkov kot z reševanjem digitalnih nalog. Pri ogledovanju posnetkov je bila povezanost statistično značilna ($\chi^2(2) = 13{,}52$; p = 0,001), enako potrjuje tudi Fisherjev test (p = 0,001). Cramerjev V znaša 0,31, kar kaže na močno povezanost. To pomeni, da se delež pozitivnih in negativnih ocen opazno razlikuje glede na to, ali so si študentje ogledali vse posnetke, nekaj posnetkov ali nobenega posnetka.

Tudi pri reševanju digitalnih nalog je bila povezanost statistično značilna ($\chi^2(2) = 7{,}63$; $p = 0{,}022$; Fisherjev test $p = 0{,}020$). Cramerjev V znaša 0,23, kar kaže na srednje močno povezanost.

Povezava z oceno PT	n	Spearmanov ρ	p
Ogledovanje posnetkov	140	0,267	0,001
Reševanje digitalnih nalog		0,192	0,023

Tabela 10: Povezava med domačo pripravo in oceno PT.

Ogledovanje posnetkov je šibko pozitivno povezano z oceno PT (Tabela 10). Podobno velja tudi za reševanje digitalnih nalog, vendar je povezava še nekoliko šibkejša. Rezultati kažejo, da so študenti z več domače priprave v povprečju dosegali nekoliko višje ocene, vendar iz korelacije ni mogoče sklepati na neposreden vzročni učinek.

4 Razprava

Rezultati kažejo, da je bila uvedba OUP pri predmetu SJ 1 med študenti 1. letnika RP na UL PEF v obeh proučevanih generacijah sprejeta pretežno pozitivno, kar je skladno z ugotovitvami primerljivih visokošolskih raziskav (Bishop in Verleger 2013; Plešec Gasparič idr. 2020). Študentje so bili zadovoljni s posnetki in pravopisnimi nalogami v digitalni obliki, saj so jih ocenili kot koristne za učenje in si jih želijo tudi pri drugih vsebinah. Ugotovitve potrjuje tudi mnenje Prenskyja (2001, v Bec in Podobnikar 2015), da lahko digitalna tehnologija izboljša učenje in da sodobne generacije učencev pričakujejo uporabo digitalne tehnologije.

Ugotavljamo tudi, da so študentje samostojno pripravo na pouk večinoma ocenili pozitivno. Menijo, da tak način dela spodbuja samostojno delo in omogoča, da se lahko na vajah posvečajo zahtevnejšim primerom in vprašanjem. To kaže, da prepoznavajo glavni namen OUP: osnovno pripravo opravijo doma, čas na vajah pa je namenjen obravnavi težjih primerov in poglabljanju snovi (Bec in Podobnikar 2015; Bishop in Verleger 2013; Plešec Gasparič in Valenčič Zuljan 2019; Plešec Gasparič idr. 2020). Po drugi strani del študentov meni, da ogledovanje posnetkov in reševanje nalog doma vzameta preveč časa, in bi jim bilo ljubše, če bi vsebine obravnavali samo na vajah.

Rezultati so pokazali tudi statistično značilno povezanost med uspešnostjo pri PT ter ogledovanjem posnetkov in reševanjem nalog v digitalni obliki. Tisti, ki so si ogledali vse posnetke in rešili vse naloge, so v večjem deležu dosegli pozitivno oceno PT kot drugi. Spearmanov koeficient je pokazal, da je ogledovanje posnetkov šibko pozitivno povezano z oceno PT, podobno velja tudi za reševanje digitalnih nalog, pri čemer je povezava še nekoliko šibkejša. Ugotavljamo tudi, da so študenti z več domače priprave v povprečju dosegali nekoliko višje ocene, vendar iz korelacije ni mogoče sklepati na neposreden vzročni učinek. Tudi ti rezultati so v skladu z raziskavami, ki so pokazale, da študentje z OUP dosegajo višje učne dosežke (Eryilmaz in Cigdemoglu 2019; Fisher idr. 2024; Lee in Choi 2019).

Treba je omeniti tudi omejitev raziskave, in sicer vzorec ni bil izbran slučajnostno, zato rezultatov ne moremo posplošiti na širšo populacijo študentov. Vseeno pa zaradi

visoke odzivnosti anketirancev rezultati dobro opisujejo vključeni generaciji študentov. V raziskavi tudi nismo preverili, ali so se študentje ob digitalnem gradivu sproti pripravljali na vaje ali so ga uporabili predvsem za strnjeno pripravo na pisanje PT.

Zaključimo lahko, da se je digitalno gradivo izkazalo kot koristno za usvajanje pravopisnih vsebin, saj je skoraj tri četrtine študentov menilo, da je prispevalo k njihovemu boljšemu znanju. Ob tem je treba upoštevati tudi pomanjkljivosti OUP, predvsem časovno obremenitev študentov ter dodatno delo učitelja pri pripravi gradiva in spremljanju opravljanja nalog. Pri SJ 1 se je pokazalo, da je pristop ostal na ravni asinhronih videopredavanj in vadenja za domačo nalogo po Bishopu in Verlegerju (2013), saj se večina študentov pogosto ni vnaprej pripravila na vaje. Zato bi bilo v prihodnje smiselno izvesti manjše število vaj po tem pristopu, na začetku posameznih vaj uvesti digitalni kviz za preverjanje znanja, usvojenega s samostojno pripravo študentov, dosledneje spremljati domače delo študentov, vzpostaviti medpredmetno povezovanje, zlasti s splošno didaktiko, pri kateri bi študentje natančneje spoznali pristop OUP in njegov pomen, in učitelju zagotoviti dodatno tehnično podporo pri pripravi digitalnega gradiva.

5 Sklep

Uporaba digitalnega gradiva (tj. videoposnetkov z učiteljevo razlago, drsnic in nalog v digitalni obliki) za razvijanje pravopisne zmožnosti po pristopu OUP se je pri študentih 1. letnika RP na UL PEF izkazala kot pozitivna in koristna. Čeprav se pristop ni uresničil v polni meri, je pokazal pozitivne učinke in je dobro izhodišče za izboljšave poučevanja na visokošolski ravni v naslednjih letih.

Članek je nastal v okviru projekta Posodobitev pedagoških študijskih programov – Posodobitev PŠP NOO (1. 7. 2023–30. 6. 2026), katerega vodja je bil dr. Janez Vogrinc, sofinancirala pa sta ga Ministrstvo za vzgojo in izobraževanje in Evropska komisija (NextGenerationEU). Projekt je izvajala Univerza v Ljubljani v okviru konzorcija z Univerzo v Mariboru in Univerzo na Primorskem.

Literatura

Bec, D. in Podobnikar, T. (2015). Collaborative and individual approach in the flipped learning by assessing students on the basis of spatial data quality control. *Dela*, *44*, 103–133. https://doi.org/10.4312/dela.44.6.103-133

Bishop, J. L. in Verleger, M. A. (2013). The flipped classroom: A survey of the research. V *Proceedings of the 120th American Society for Engineering Education Annual Conference & Exposition*. ASSE. https://www.researchgate.net/publication/285935974_The_flipped_classroom_A_survey_of_the_research

Eryilmaz, M. in Cigdemoglu, C. (2019). Individual flipped learning and cooperative flipped learning: Their effects on students' performance, social, and computer anxiety. *Interactive Learning Environments*, *27*(4), 432–442. https://doi.org/10.1080/10494820.2018.1522652

Fisher, R., Tran, Q. in Verezub, E. (2024). Teaching English as a foreign language in higher education using flipped learning/flipped classrooms: A literature review. *Innovation in Language Learning and Teaching*, *18*(4), 332–351. https://doi.org/10.1080/17501229.2024.2302984

Lee, J. in Choi, H. (2019). Rethinking the flipped learning pre-class: Its influence on the success of flipped learning and related factors. *British Journal of Educational Technology*, *50*(2), 934–945. https://doi.org/10.1111/bjet.12618

Plešec Gasparič, R. in Valenčič Zuljan, M. (2019). Učne oblike v osnovni šoli in obrnjeno učenje in poučevanje. *Revija za elementarno izobraževanje*, *12*(3), 267–290. https://doi.org/10.18690/rei.12.3.267-290.2019

Plešec Gasparič, R., Valenčič Zuljan, M. in Kalin, J. (2020). Obrnjeno učenje in poučevanje kot priložnost za inovativno in prožno izvajanje učnih oblik v visokošolskem izobraževanju. *Revija za elementarno izobraževanje*, *13*(posebna številka), 51–80. https://doi.org/10.18690/rei.13.Special.51-80.2020

Rot Vrhovec, A. (2015). Sodelovalne učne oblike pri jezikovnem delu pouka slovenščine v osnovni šoli. *Center for Educational Policy Studies Journal*, *5*(3), 156–181. https://doi.org/10.26529/cepsj.133

Valenčič Zuljan, M. (1997). Kaj imajo v mislih učitelji, ko razmišljajo o inoviranju lastne učne prakse. *Pedagoška obzorja*, *12*(5–6), 228–239.

Digitalna orodja pri poučevanju artikulacijske fonetike

DOI: 10.4312/Obdobja.45.159-166

Luka Horjak, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; Luka.Horjak@ff.uni-lj.si

V prispevku predstavljamo digitalna orodja za spodbujanje poglobljenega razumevanja vsebin s področja artikulacijske fonetike, in sicer s konkretizacijo abstraktnih modelov. Obravnava temelji na praktični uporabi pri poučevanju predmetov s področja fonetike na Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani. Na spletni strani Seeing Speech so na voljo posnetki izgovora posameznih glasov, vključno z magnetno resonanco in ultrazvokom, medtem ko Pink Trombone omogoča interaktivno modulacijo govorne cevi. Orodja so primerna predvsem za visokošolski pouk fonetike, uporabna pa tudi pri pouku jezikov v osnovni in srednji šoli.

artikulacijska fonetika, poučevanje fonetike, Seeing Speech, Pink Trombone, govor

This article presents digital tools for fostering an in-depth understanding of articulatory phonetics through the concretization of abstract models. The discussion is based on the tools' practical use in teaching phonetics courses at the Department of Slovenian Studies at the University of Ljubljana's Faculty of Arts. The Seeing Speech website provides recordings of the articulation of individual speech sounds, including magnetic resonance imaging and ultrasound data, and Pink Trombone allows interactive manipulation of the vocal tract. These tools are primarily suitable for use in university-level phonetics courses but can also be used in language teaching at the primary and secondary school levels.

articulatory phonetics, phonetics teaching, Seeing Speech, Pink Trombone, speech

1 Uvod

Področje fonetike je izrazito interdisciplinarno. Veda se razvija od sredine 19. stoletja in združuje védenje več različnih področij, kot so biologija, fizika, jezikoslovje, psihologija, nevrologija idr. V okviru preučevanja in poučevanja jezikov je v ospredju jezikoslovni, fonološki vidik, ki se opira predvsem na artikulacijo in akustiko, saj naj bi bil vsak fonem oz. segment opredeljen s teh dveh vidikov (npr. razlikovalne lastnosti; Gussenhoven in Jacobs 2017: 65). Poučevanje fonetike tako poteka predvsem v okviru jezikoslovnih študijev, ali kot samostojne discipline ali pa kot enega izmed (običajno) obveznih predmetov pri študiju posameznega prvega ali neprvega jezika. Z raziskovanjem fonetike se v svetovnem merilu (zlasti v primerjavi z naravoslovno-tehniškimi disciplinami) ne ukvarja veliko raziskovalcev (okrog 1000 raziskovalcev se udeležuje Mednarodnega kongresa fonetičnih znanosti, ki poteka na štiri leta),[1] zato je tudi področje

1 Leta 2023 je bilo v Pragi 1127 udeležencev, povezava: https://www.internationalphoneticassociation.org/icphs/icphs2023

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

poučevanja fonetike, kot pišeta Ashby in Ashby (2015: 198–199), dokaj slabo razvito. Fonetika je zlasti v sklopu študija jezikoslovja zamejena na spoznavanje fonemskega inventarja posameznega jezika, učenje sistemskih premen in usvajanje nadsegmentnih prvin s ciljem, da se učeči se čim bolj približa izgovarjavi naravnega govorca. S tem je povezana tudi izvedba tovrstnih predmetov, ki imajo pridružene govorne vaje. Vsebine so velikokrat predstavljene abstraktno s prikazi, ki jim umanjka ustrezna razlaga. Poznavanje vsebin tako pogosto ostane le na teoretični ravni, saj pri pouku jezika oz. jezikoslovja umanjkajo vsebine s področja fizike (akustika) in biologije (fiziologija), ki bi lahko pomagale pojasniti posamezne pojave. Ob tem velja omeniti, da sta tudi v osnovni in srednji šoli prostor in čas za pouk vsebin s področja govora zelo omejena (gl. Tivadar 2015). Na mednarodnih konferencah prispevki s področja poučevanja fonetike obravnavajo predvsem usvajanje drugega oz. tujega jezika, manjkajo pa predstavitve metod in načinov poučevanja fonetike prvega jezika. Na Slovenskem je na to temo leta 2019 potekal simpozij Poučevanje govora, prispevki pa so objavljeni v monografiji *Govor v pedagoški praksi* (Podbevšek in Žavbi 2019), obravnava poučevanja govora sega od predšolskega vzgojno-izobraževalnega obdobja do pouka na visokošolskih zavodih in številnih drugih vidikov.

V tem kratkem prispevku želimo prikazati spletne aplikacije, ki nam pomagajo, da lahko študentom oz. študentkam pri visokošolskem pouku fonetike eksplicitno pojasnimo, razložimo fonetične vsebine in s tem omogočimo tudi globlje razumevanje in morebitno bodoče samostojno raziskovanje vsebin. Namen pouka ni le usvajanje veščin kakovostne standardne izreke, ampak poglobljeno razumevanje področja, zato mora pouk vključevati uporabo ustreznih prikazov (spektrogrami, slike in video posnetki izgovarjave), simbolov (mednarodna fonetična transkripcija) in diagramov (npr. samoglasniški trapez).

Pri predstavitvi se bomo osredotočili le na spletne strani, ki jih uporabljamo na Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani[2] pri poučevanju artikulacijskofonetičnih vsebin.[3] Drugo pomembno področje je poučevanje akustične fonetike, ki zahteva ločeno obravnavo, še posebej zaradi dostopnosti orodij, kot je brezplačni računalniški program Praat (1992–) (Boersma in Weenink 2026), in večjega števila akustičnofonetičnih raziskav slovenščine v primerjavi z artikulacijskofonetičnimi. Za poučevanje je smotrna kombinacija obeh vrst fonetike.

2 Izzivi poučevanja predvsem normativnih vsebin pri predmetu Fonetika in fonologija slovenskega knjižnega jezika pri študiju slovenistike na Filozofski fakulteti Univerze v Ljubljani so obravnavani v Horjak in Tivadar (2019).

3 Z raziskovanjem artikulacije slovenskih glasov se je v primerjavi z akustičnimi raziskavami ukvarjalo malo raziskovalcev (pregled v Horjak 2023: 242–244; 2025: 42–46), posledično je tovrstnih vsebin manj v slovničnih in drugih priročnikih za slovenščino. Toporišič (2000) v poglavju Glasoslovje izhaja iz Bezlajevih skiagramov, obravnava Omerze (1970) pa temelji na beograjski fonetični šoli.

2 Poučevanje artikulacije

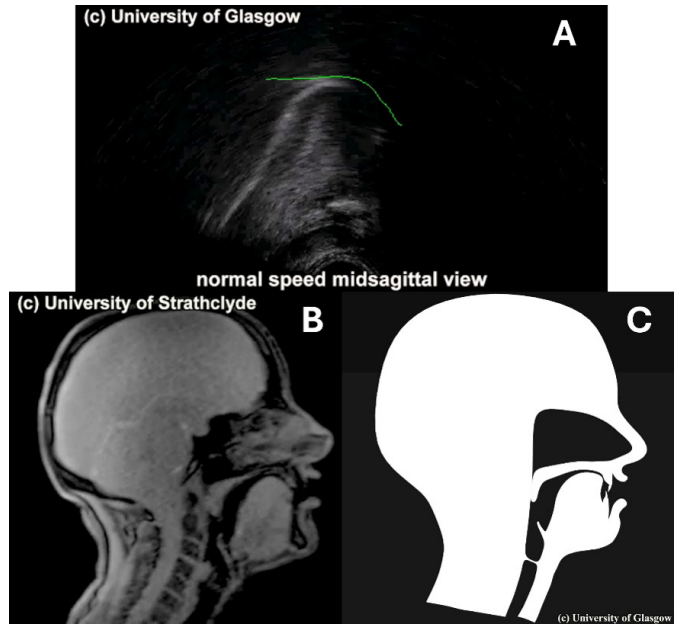
Celovit artikulacijski opis slovenskih glasov je naredil Bezlaj (1939), v novejšem času so bili raziskovani samoglasniki (Kocjančič Antolík in Tivadar 2019) in nezvočniki (Horjak 2023). Za prikaz artikulacije so uporabni predvsem Bezlajevi skiagrami, ustvarjeni na podlagi rentgenskih posnetkov, na katerih se – za razliko od ultrazvočnih – vidi več struktur v ustni votlini, na ultrazvočnih posnetkih se vidi le stik med zrakom in površino jezika in del neba. Pri poučevanju si lahko pomagamo tudi s prikazi za druge jezike, npr. za hrvaščino (Horga in Liker 2016), angleščino (Cruttenden 2014), nemščino (Krech idr. 2010) ali danščino (Grønnum 2009), kar pa lahko služi le kot demonstracija posameznih razlik in razmerij med glasovi, ne more pa nadomestiti prikazov za slovenski jezik, saj si glasovi med jeziki niso nujno zelo podobni.

Za prikaz izgovarjave posameznih glasov in razlik med njimi lahko uporabimo spletna orodja, ki niso ustvarjena za specifične jezike.

2.1 Spletna orodja s posnetki in interaktivne aplikacije

Eno od takih orodij, ki se uporablja za prikazovanje izgovarjave, je spletna stran Seeing speech. Stran vsebuje posnetke izgovora za glasove po mednarodni fonetični abecedi IPA, nastala je v sodelovanju med šestimi škotskimi univerzami. Glasovi so predstavljeni s simboli IPA, razdeljeni so na štiri skupine: pulmonalni in nepulmonalni soglasniki, samoglasniki in drugi simboli/glasovi. Za večino glasov so na voljo štirje različni videoposnetki: posnetek, posnet z ultrazvokom,[4] dva posnetka (prvi posnet leta 2014, drugi pa v letih 2022–2023, pri čemer je ločljivost drugega 70 % boljša od ločljivosti prvega), ustvarjena s slikanjem z magnetno resonanco (MRI), in animacija, ustvarjena na podlagi MRI posnetkov iz leta 2014. Vsi prikazi so sagitalni, pri čemer je sprednji del jezika na desni, zadnji pa na levi strani prikaza (gl. Sliko 1). Govorec in govorka na posnetkih sta profesorja fonetike. Spletna stran vsebuje tudi natančno predstavitev tehnik snemanja izgovora z ultrazvočno napravo in s slikanjem z magnetno resonanco (gl. Lawson idr. 2018). Na istem portalu deluje tudi spletna stran STAR Speech Therapy Animation and imaging Resource (Lawson idr. 2023), na kateri najdemo tudi video posnetke izgovora, narejene s slikanjem z magnetno resonanco, za glasove, predstavljene z razširjeno fonetično abecedo IPA (ExtIPA); posnetki so namenjeni predvsem uporabi v govorni terapiji.

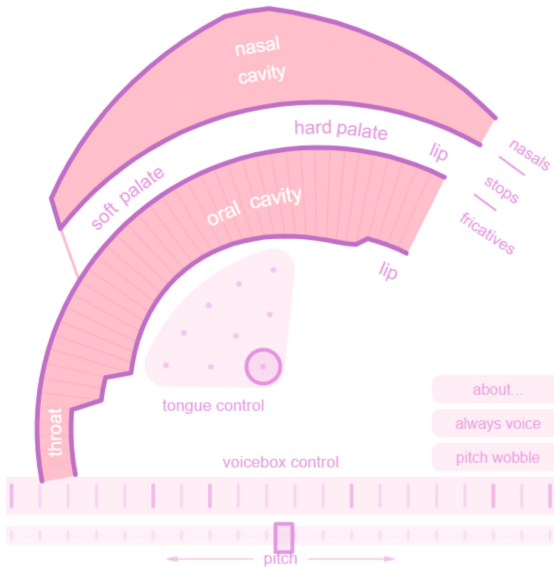
4 O uporabi ultrazvočnega slikanja za raziskovanje (iz)govora več v Horjak (2023).



Slika 1: Sagitalni prikaz pripore pri nezvenečem zadlesničnem priporniku [ʃ] s slikanjem z ultrazvočno napravo (**A**; Beck 2018a), z magnetno resonanco (**B**; Beck 2022a) in na shematizirani animaciji (**C**; Beck 2018b) (vir posnetkov: Seeing Speech; uporabljeno z dovoljenjem avtorjev).

Podoben projekt kot Seeing speech je tudi projekt The real-time MRI IPA charts (Toutios idr. b. d.), nastal v sklopu raziskovalne skupine SPAN Univerze v Južni Kaliforniji. Na spletni strani si lahko ogledamo izgovor posameznih glasov, besed in krajših stavkov, posnetih z MRI, ki jih izgovarja pet govorcev (gl. Toutios idr. 2016).

Podpora pri poučevanju artikulacijskih značilnosti glasov je tudi interaktivno orodje Pink Trombone (Thapen b. d.-a), ki sintetizira glas. Neil Thapen (Thapen b. d.-b) ga je leta 2017 razvil kot spletno igro, ki uporabniku omogoča manipulacijo govorne cevi (od grla/glasilk do ustnic). Pri tem lahko nastavimo višino osnovne frekvence (f_0) in intenziteto, nato pa spreminjamo obliko jezika za oblikovanje samoglasnikov, ožimo govorno cev, da nastanejo zaporniki ali priporniki, odpiramo oz. zapiramo prehod med ustno in nosno votlino za tvorbo nosnih glasov (gl. Sliko 2). Orodje je idealno za uporabo na zaslonih na dotik in omogoča dobro simulacijo delovanja govorne cevi, pri čemer uporabnik v realnem času lahko sliši spremembe glasu, ki nastanejo, ko spreminja konfiguracijo govorne cevi. Uporabno je zlasti za pouk o načinu izgovora soglasnikov (zapornik – pripornik) in samoglasnikov (lega jezika). Za uporabo pri pouku ga priporočata Weckwerth (2022: 63) in Jurgec (2023: 40).



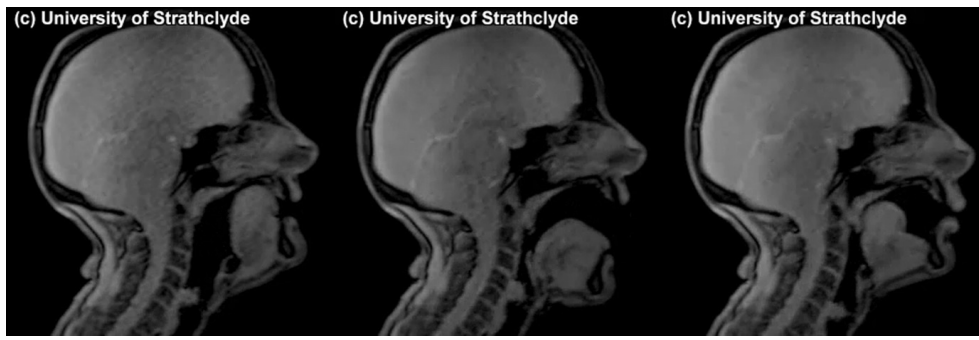
Slika 2: Spletna aplikacija Pink Trombone (vir: zajem zaslona).

Za poučevanje fonetike lahko uporabljamo tudi (prenosno) ultrazvočno napravo, s katero lahko fonetik prikaže delovanje jezika in njegovo vlogo pri tvorbi posameznih glasov v realnem času. Zlasti v razredu, v katerem so govorci različnih jezikov, se dobro pokažejo tudi razlike med glasovi, ki jih ima posamezni jezik (Proctor 2022: 88). Ultrazvok lahko uporabimo tudi za korekcijo izgovarjave ali za demonstracijo ustrezne lege jezične ploskve pri učenju drugih, tujih jezikov. S takim načinom poučevanja, predvsem usvajanja tujega jezika, se ukvarja Tanja Kocjančič Antolík, ki je preučevala učenje švedščine pri naravnih govorcih češčine (Kocjančič idr. 2025) in učenje francoščine pri naravnih govorcih japonščine (Kocjančič Antolík idr. 2019). Usvajanje artikulacije tujega jezika z uporabo ultrazvoka, pri čemer lahko učeči se vidi, v kolikšni meri lega njegovega jezika pri tvorbi posameznega glasu odstopa od pričakovane lege, se je izkazalo kot smiselno, zato bi bilo treba tovrstne raziskave narediti tudi za poučevanje slovenščine.

2.2 Uporaba pri pouku

Pri pouku fonetike posnetke s spletnih strani Seeing speech in The real-time MRI IPA charts uporabljamo za prikaz mesta artikulacije in načina artikulacije, in sicer predvsem posnetke, narejene s slikanjem z magnetno resonanco. Običajno najprej utemeljimo shematična razmerja, prikazana s samoglasniškim trikotnikom v *Slovenski slovnici* (Toporišič 2000) ali *Fonetiki 1* (Tivadar in Batista 2019), ta so namreč opredeljena artikulacijsko na podlagi lege jezične ploskve in velikosti čeljustnega kota. Abstraktna razmerja med samoglasniki razložimo s prikazom sagitalnega položaja jezika pri robnih samoglasnikih /i – a – u/, kjer se najbolje vidi, kako hidromorfen organ je jezik. Na Sliki 3 vidimo skrajne, prototipične sagitalne položaje jezične ploskve pri

tvorbi robnih samoglasnikov, s katerimi lahko artikulacijsko razložimo samoglasniški trikotnik. Ko obravnavamo višino jezične ploskve in čeljustni kot, se dobro vidi njuno povezanost, npr. pri sprednjem visokem nezaokroženem [i] je jezik približan trdemu nebu (in dlesnim), kar povzroči, da je čeljustni kot manjši (usta so manj odprta), podobno velja za zadnji visoki zaokroženi [u], le da je pri njegovi tvorbi jezik približan mehkemu nebu. Povsem nasprotno pa lahko vidimo pri nizkem centralnem [a], pri katerem je jezik nizko, čeljustni kot pa je velik (usta so bolj odprta). Pri sagitalnih prikazih se zaokroženosti ustnic ne vidi.



Slika 3: Prikaz skrajnega sagitalnega položaja pri sprednjem visokem [i] (levo; Beck 2022b), nizkem centralnem [a] (sredina; Beck 2022c) in zadnjem visokem samoglasniku [u] (desno; Beck 2022d) s slikanjem z magnetno resonanco (vir posnetkov: Seeing Speech; uporabljeno z dovoljenjem avtorjev).

Za ilustracijo izgovarjave vključimo nato še druge samoglasnike, npr. polglasnik [ə]. Iz prikazov se jasno vidi, kateri samoglasniki so trdonebni (sprednji) in kateri mehkonebni (zadnji). Prikazi osmislijo samoglasniški trikotnik, položaje jezika pri posameznih samoglasnikih lahko učeči se z introspekcijo tudi lažje uzavestijo. Posnetki nam pomagajo tudi demonstrirati različna mesta artikulacije, saj se na njih jasno vidijo različna mesta, katerim se jezik pri artikulaciji približa, npr. dlesnim pri [t d s z ʦ ʣ], alveolarnemu grebenu pri [ʃ ʒ ʧ ʤ][5] in mehkemu nebu pri [k g x]. Dobro sta vidna tudi ustnični izgovor pri [p b m] in zobnoustnični izgovor pri [f ʋ]. Demonstriramo tudi glasove, ki jih slovenski (knjižni) jezik nima, npr. različne izgovore r-jevskih glasov [ɾ r ɽ ɻ ɹ ʁ ʁ] ali tvorbo glotalnega zapornika [ʔ].[6] Prikaze uporabimo tudi za demonstracijo načina izgovarjave glasov iz bolj oddaljenih in manj znanih jezikov, npr. nepulmonalnih glasov, kot sta ustnični [ʘ] in zadlesnični klik [ǂ].

Demonstracijo izgovarjave naredimo tudi na interaktivni spletni strani Pink Trombone. S tem interaktivnim orodjem lahko prikažemo različne načine izgovarjave, kot je razlika med priporniki in zaporniki, demonstriramo lahko nosni izgovor in modeliramo položaj jezika. Orodje je primerno tudi za samostojno rabo, utrjevanje snovi. Orodje

5 Zlitnike najdemo med other symbols na straneh Seeing speech in The real-time MRI IPA charts.

6 O glotalnem zaporniku piše Jurgec (2004, 2005), ki je ugotovil, da glotalni zapornik nekateri govorci slovenščine izgovarjajo vedno pred besedami, ki se začnejo s samoglasnikom, npr. *avtobus* [ˈʔaʊ̯tɔbus] (Jurgec 2004: 139).

Pink Trombone zadostuje za osnovno predstavo o delovanju govorne cevi, ne omogoča pa prikaza vseh načinov tvorbe glasov, npr. drsnikov in r-jevskih glasov.

3 Sklep

Abstraktnost visokošolskega pouka artikulacijske fonetike lahko presežemo z uporabo posnetkov izgovarjave, pridobljenih z magnetno resonanco in ultrazvokom, saj je artikulacija fiziološki pojav. Razpoložljiva orodja študentom služijo tudi kot podpora pri samostojnem študiju. Njihova raba je smiselna tudi pri osnovnošolskem in srednješolskem pouku jezikov, zlasti za medpredmetno povezovanje slovenščine in tujih jezikov z biologijo. V prihodnosti nas čaka razvoj primernih prikazov, orodij tudi za slovenščino, do takrat pa si lahko pomagamo z v prispevku predstavljenimi orodji.

Zahvaljujem se ekipi Seeing Speech za dovoljenje za reprodukcijo sličic iz ultrazvočnih, magnetnoresonančnih in animiranih posnetkov na Slikah 1 in 3. Zahvaljujem se anonimnima recenzentoma za koristne predloge, ki so pripomogli k izboljšanju prispevka. Prispevek je nastal tudi v okviru raziskovalnega programa št. P6-0215 (Slovenski jezik – bazične, kontrastivne in aplikativne raziskave), ki ga je sofinancirala Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije iz državnega proračuna.

Viri

Beck, J. (2018a). *Voiceless postalveolar fricative: Midsagittal and coronal orientation in an /a_a/ environment* [Ultrazvočni video]. Seeing Speech. https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=2&speaker=1#location=643

Beck, J. (2018b). *Voiceless postalveolar fricative* [Animacija]. Seeing Speech. https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=643

Beck, J. (2022a). *Voiceless postalveolar fricative* [MRI video]. Seeing Speech. https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=4&speaker=1#location=643

Beck, J. (2022b). *Front close unrounded vowel* (*cardinal 1*) [MRI video]. Seeing Speech. https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=4&speaker=1#location=105

Beck, J. (2022c). *Front open unrounded vowel* (*cardinal 4*) [MRI video]. Seeing Speech. https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=4&speaker=1#location=97

Beck, J. (2022d). *Back close rounded vowel* (*cardinal 8*) [MRI video]. Seeing Speech. https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=4&speaker=1#location=117

Boersma, P. in Weenink, D. (2026). *Praat: Doing phonetics by computer* (Version 6.4.65) [Računalniški program]. https://www.fon.hum.uva.nl/praat/

Lawson, E., Cleland, J. in Stuart-Smith, J. (2023). *STAR teaching: A speech therapy animation and imaging resource*. University of Glasgow. https://seeingspeech.ac.uk/speechstar/

Lawson, E., Stuart-Smith, J., Scobbie, J. M. in Nakai, S. (2018). *Seeing Speech: An articulatory web resource for the study of phonetics*. University of Glasgow. https://www.seeingspeech.ac.uk/

Thapen, N. (b. d.-a). *Pink trombone* [Interaktivna spletna aplikacija]. https://dood.al/pinktrombone/

Thapen, N. (b. d.-b). *Games by Neil Thapen*. https://venuspatrol.nfshost.com

Toutios, A., Lingala, S. G., Vaz, C., Kim, J., Esling, J., Keating, P., Gordon, M., Byrd, D., Goldstein, L., Nayak, K. in Narayanan, S. (b. d.). *Real-time MRI IPA charts*. https://sail.usc.edu/span/rtmri_ipa/

Literatura

Ashby, P. in Ashby, M. (2015). Phonetic pedagogy. V M. J. Jones in R.-A. Knight (ur.), *The Bloomsbury companion to phonetics* (str. 198–207). Bloomsbury.

Bezlaj, F. (1939). *Oris slovenskega knjižnega izgovora*. Znanstveno društvo.

Cruttenden, A. (2014). *Gimson's pronunciation of English*. Routledge.

Grønnum, N. (2009). *Fonetik og fonologi: Almen og dansk*. Akademisk Forlag.

Gussenhoven, C. in Jacobs, H. (2017). *Understanding phonology*. Routledge.

Horga, D. in Liker, M. (2016). *Artikulacijska fonetika: Anatomija i fiziologija izgovora*. Ibis grafika.

Horjak, L. (2023). Preliminarna raziskava izgovarjave slovenskih jezičnih nezvočnikov z ultrazvočno napravo. *Jezik in slovstvo*, *68*(4), 241–258. https://doi.org/10.4312/jis.68.4.241-258

Horjak, L. (2025). *Fonetično-fonološka načela kodifikacije sodobnega slovenskega govorjenega jezika* [Doktorska disertacija]. Univerza v Ljubljani.

Horjak, L. in Tivadar, H. (2019). Odprta vprašanja kodifikacije in smiselno univerzitetno poučevanje slovenskega govorjenega jezika. V K. Podbevšek in N. Žavbi (ur.), *Govor v pedagoški praksi* (str. 151–158). Znanstvena založba Filozofske fakultete Univerze v Ljubljani. https://doi.org/10.4312/9789610602804

Jurgec, P. (2004). Antihiatski pojavi v knjižni slovenščini. *Jezikoslovni zapiski*, *10*(1), 125–144. https://doi.org/10.3986/jz.v10i1.2609

Jurgec, P. (2005). Začetne opombe k laringalizaciji v slovenščini. *Slovenski jezik/Slovene Linguistic Studies*, *5*, 135–168.

Jurgec, P. (2023). *Sounds: Their structure and patterns in language* (3. izdaja). Kendall Hunt.

Kocjančič Antolík, T. in Tivadar, H. (2019). Ultrazvočni pogled na artikulacijo slovenskih samoglasnikov. V H. Tivadar (ur.), *Slovenski javni govor in jezikovno-kulturna (samo)zavest* (str. 71–79). Znanstvena založba Filozofske fakultete Univerze v Ljubljani. https://doi.org/10.4312/Obdobja.38.71-79

Kocjančič Antolík, T., Pillot-Loiseau, C. in Kamiyama, T. (2019). The effectiveness of real-time ultrasound visual feedback on tongue movements in L2 pronunciation training: Japanese learners' progress on the French vowel contrast /y/-/u/. *Journal of Second Language Pronunciation*, *5*(1), 72–97. https://doi.org/10.1075/jslp.16022.ant

Kocjančič, T., Bořil, T. in Hofmann, S. (2025). Acoustic and articulatory visual feedback in classroom L2 vowel remediation. *Language and Speech*, *68*(4), 865–886. https://doi.org/10.1177/00238309231223736

Krech, E.-M., Stock, E., Hirschfeld, U. in Anders, L. C. (2010). *Deutsches Aussprachewörterbuch*. De Gruyter. https://doi.org/10.1515/9783110215564

Omerza, Z. (1970). *Uporabna fonetika*. DZS.

Podbevšek, K. in Žavbi, N. (ur.). (2019). *Govor v pedagoški praksi*. Znanstvena založba Filozofske fakultete. https://doi.org/10.4312/9789610602804

Proctor, M. (2022). Consonants. V R.-A. Knight in J. Setter (ur.), *The Cambridge handbook of phonetics* (str. 65–105). Cambridge University Press. https://doi.org/10.1017/9781108644198.004

Tivadar, H. (2015). Vloga pravorečja in njegovo poučevanje v slovenskem osnovno- in srednješolskem izobraževanju. *Jezik in slovstvo*, *60*(3–4), 161–172. https://doi.org/10.4312/jis.60.3-4.161-172

Tivadar, H. in Batista, U. (2019). *Fonetika 1*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Toporišič, J. (2000). *Slovenska slovnica*. Obzorja.

Toutios, A., Lingala, S. G., Vaz, C., Kim, J., Esling, J., Keating, P., Gordon, M., Byrd, D., Goldstein, L., Nayak, K. S. in Narayanan, S. S. (2016). Illustrating the production of the International Phonetic Alphabet sounds using fast real-time magnetic resonance imaging. *Interspeech 2016*, 2428–2432. https://doi.org/10.21437/Interspeech.2016-605

Weckwerth, J. (2022). Vowels. V R.-A. Knight in J. Setter (ur.), *The Cambridge handbook of phonetics* (str. 40–64). Cambridge University Press. https://doi.org/10.1017/9781108644198.003

Kontekstualni čimbenici u ovladavanju slovenskim jezikom: sport u jeziku / jezik u sportu

DOI: 10.4312/Obdobja.45.167-173

Marina Jajić Novogradec, Filozofski fakultet, Sveučilište u Zagrebu, Zagreb;
mjnovo@ffzg.unizg.hr

Prispevek obravnava vpliv kontekstualnih dejavnikov na obvladovanje slovenščine v naravnem okolju. Predstavlja študijo primera večjezične udeleženke, ki usvaja slovenščino prek obšolske dejavnosti – športa. Rezultati so pokazali pomembno interakcijo med motivacijo, izpostavljenostjo jeziku in jezikovno sorodnostjo glede razvoja znanja slovenščine in komunikacijskih veščin udeleženke v različnih fazah usvajanja jezika.

hrvaščina, slovenščina, študija primera, kontekstualni dejavniki, šport

This article discusses the influence of contextual factors on Slovenian language acquisition in a natural environment. It presents a case study of a multilingual female respondent that has been acquiring Slovenian through sports as an extra-curricular activity. The results show the significant interaction of motivation, language exposure, and typologically related languages in developing knowledge of Slovenian and the respondent's communication skills in various phases of language acquisition.

Croatian, Slovenian, case study, contextual factors, sports

1 Uvod

Proučavanje individualnih čimbenika u procesu ovladavanja inim jezikom (OVIJ) oduvijek je bilo predmet mnogobrojnih istraživanja OVIJ-a s različitih aspekata – psihološkog, psiholingvističkog, primijenjenolingvističkog, društvenog. Individualni čimbenici dijele se na dvije skupine – afektivne i kognitivne. U afektivne čimbenike ubrajaju se motivacija, stavovi, strah od stranoga jezika, dok kognitivni čimbenici obuhvaćaju inteligenciju, jezičnu sposobnost, strategije učenja, stilove učenja (Dörnyei 2005; Mihaljević Djigunović 1998). Međutim, individualni čimbenici često dolaze u interakciju s kontekstualnim čimbenicima koji također uvjetuju krajnji uspjeh u ovladavanju inim jezikom.

U našemu radu kontekstualni čimbenici zauzimaju posebno mjesto jer je naglasak na ovladavanju slovenskim jezikom u neformalnoj sredini i s nenamjernim ciljem, odnosno kroz izvanškolsku aktivnost – sport.

U nastavku rada najprije ćemo nešto reći o kontekstualnim čimbenicima u ovladavanju jezikom te opisati istraživanja bitna za naš istraživački kontekst, a potom prikazati studiju slučaja ispitanice u ovladavanju slovenskim jezikom. Radi se o longitudinalnom

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

praćenju razvoja jezične i komunikacijske kompetencije u slovenskom jeziku s obzirom na utjecaj određenih kontekstualnih čimbenika.

2 Kontekstualni čimbenici u ovladavanju drugim jezikom

U literaturi OVIJ-a kontekstualni čimbenici u uskoj su vezi s individualnim i društvenim čimbenicima, te se proučavaju s triju gledišta – gledišta jezika, učenika i procesa učenja (Walqui 2000). Što se tiče čimbenika s gledišta jezika, autorica Walqui (2000) izdvaja jezičnu udaljenost među jezicima, jezično znanje u materinskom jeziku, poznavanje drugoga jezika, dijalekt materinskog jezika, status jezika u zajednici i stavove prema jeziku. Čimbenici s gledišta učenika, prema autorici, podrazumijevaju različitu jezičnu pozadinu učenika, različite ciljeve za učenje jezika (unutarnja i vanjska motivacija), utjecaj vršnjaka i podršku roditelja. S gledišta procesa učenja oni uključuju različite stilove učenja, motivaciju i razrednu interakciju.

U našoj studiji slučaja usredotočit ćemo se na kontekstualne čimbenike s gledišta jezika, imajući u vidu srodnost među jezicima, poznavanje dijalekta materinskog jezika i drugih jezika te stavove prema jeziku, kao i čimbenike s gledišta učenika, poput specifičnog cilja za znanje jezika, odnosno vanjske motivacije i podrške roditelja. Neizostavan kontekstualni čimbenik, koji istražujemo i u našem kontekstu, jest i izloženost jezičnom unosu. Inače, izloženost jezičnom unosu u ovladavanju inim jezikom možemo odrediti kao formalnu i neformalnu. Naime, formalno, učenici su često izloženi ograničenom jezičnom unosu u inome jeziku kada se ini jezik uči s već unaprijed određenom satnicom u školama. Međutim, u današnje vrijeme, s obzirom na rast suvremenih tehnologija, učenici mogu biti izloženi i inim jezicima izvan škole. Istraživanja izloženosti jeziku uglavnom se povezuju s jezičnim znanjem, tj. ispituje se koliko izloženost jeziku izvan škole, utjecaj medija i tehnologija, pospješuje kod učenika i studenata ovladanost inim jezikom, a uglavnom se proučava engleski kao ini jezik (Bogunović i Jelčić Čolakovac 2019; Hendrih i Letica Krevelj 2019; Sundqvist 2009).

3 Pregled dosadašnjih istraživanja

Dosadašnja istraživanja o ulozi kontekstualnih čimbenika u ovladavanju inim jezikom fokusirala su se na učinak neformalnih aktivnosti i uporabu jezika izvan škole (uglavnom engleskog) na krajnji uspjeh u jeziku. U ovome poglavlju prikazujemo dvije studije slučaja hrvatske djece koja su ovladavala engleskim jezikom izvan škole. Autorice Hendrih i Letica Krevelj (2019) opisuju studiju slučaja devetogodišnje djevojčice koja je postigla visoku razinu kompetencije u engleskom jeziku putem izloženosti medijskom sadržaju izvan škole. Uz formalno učenje jezika u školi, ispitanica je, počevši u dobi od tri godine, pratila televizijske programe na engleskom jeziku tri sata dnevno, a vikendom nešto manje, jedan i pol sat dnevno. Jezična i komunikacijska kompetencija djevojčice, odnosno njezino poznavanje vokabulara provjeravalo se spontanim razgovorom između djeteta i jedne od autorica. Rezultati su pokazali da je medijski sadržaj kojemu je dijete bilo izloženo od rane dobi izvan škole

znatno utjecao na djetetovu jezičnu proizvodnju, koja je uključila više razgovorni stil te izbjegavanje metajezičnog ponašanja (objašnjavanja zašto je nešto točno ili nije u jeziku). Također, osim izloženosti jezičnom unosu, unutarnja motivacija djeteta, kao i podrška roditelja pokazali su značajan utjecaj na razvoj djetetove kompetencije u engleskom jeziku. Longitudinalno istraživanje Mihaljević Djigunović (2012), provedeno na trima studijama slučaja djece mlađe školske dobi u ovladavanju engleskim jezikom u školi i izvan, pokazalo je da se tijekom ovladavanja engleskim kod djece mlađe dobi i u različitim periodima događa stalna interakcija individualnih i kontekstualnih čimbenika, čime se potvrđuje nelinearnost jezičnog postignuća kod djece.

Istraživanja koja su se bavila utjecajem sporta i ovladavanjem inim jezikom, odnosno jezičnom interakcijom djeteta uključenog u sportsko okruženje inoga jezika izvan škole nisu pronađena. Većina istraživanja bave se ciljano značajem izvannastavnih i izvanškolskih aktivnosti za djecu (Valjan Vukić 2016), kao i povezanosti fizičkih aktivnosti, točnije sporta, s metakognitivnim sposobnostima i unutarnjom motivacijom općenito (Stam i dr. 2020).

4 Istraživanje studije slučaja: nenamjerno ovladavanje slovenskim jezikom u prirodnom jezičnom okruženju

4.1 Cilj istraživanja i istraživačka pitanja

Cilj istraživanja bio je ispitati na koji način kontekstualni čimbenici, kao što su izloženost jeziku kroz izvanškolsku aktivnost – sport, vanjska motivacija, dijalekt materinskog jezika i drugi jezici, srodnost među jezicima, stavovi prema jeziku te okruženje utječu na nenamjerno ovladavanje slovenskim jezikom u prirodnoj sredini. Radi se o longitudinalnom praćenju razvoja jezične i komunikacijske kompetencije u slovenskom jeziku koja je usko povezana s vanjskom motivacijom ispitanice i ogleda se u bavljenju sportom. Prati se stoga početak ovladavanja slovenskim jezikom (od sedme do četrnaeste godine djetetova života) kroz izloženost jeziku i okruženje u kojem se ovladava jezikom, receptivno i produktivno znanje jezika, interakciju s drugim jezicima, metajezičnu svjesnost, upotrebu jezika u različitim kontekstima i s različitim sugovornicima, unutarnju i vanjsku motivaciju te stavove ispitanice prema jeziku.

Na temelju postavljenog cilja oblikovana su sljedeća istraživačka pitanja:

1) Koja je povezanost bavljenja sportom ispitanice i ovladavanja slovenskim jezikom u prirodnom okruženju i postizanju jezičnih i komunikacijskih kompetencija u jeziku u različitim periodima?

2) Koji su kontekstualni čimbenici najviše utjecali na ovladavanje slovenskim jezikom kod ispitanice i kako su se mijenjali kroz vrijeme?

4.2 Metodologija istraživanja

4.3 Ispitanica

Ispitanica je odrastala u prirodnom jednojezičnom okruženju s hrvatskim kao prvim jezikom, međutim od rođenja je izložena engleskom kao drugom jeziku. Način na koji je ispitanica bila izložena drugom jeziku bio je ograničen. Majka ispitanice (koja nije izvorni govornik J2) bila je glavni izvor jezičnog unosa u engleskom jeziku s obzirom na svakodnevnu komunikaciju s ispitanicom, dok su otac i okolina bili izvor jezičnog unosa u hrvatskome jeziku. Riječ je tada o subjektivno motiviranoj hrvatsko-engleskoj dvojezičnosti (v. Jajić Novogradec 2015). U dobi od tri godine, ispitanica pohađa vrtić u hrvatsko-talijanskoj skupini s obzirom na to da obitelj živi na statutarno dvojezičnom hrvatsko-talijanskom području. Uz izloženost engleskom i talijanskom kao drugim jezicima, ispitanica je povremeno izložena i ruskom jeziku, kroz gledanje crtića, a kasnije i kratkim formalnim učenjem jezika iz osobnog interesa (u dobi od 11 godina), stoga u tom slučaju možemo govoriti i o razvoju slijedne višejezičnosti kod ispitanice. Također, polaskom u prvi razred osnovne škole u dobi od 6 i pol godina, ispitanica formalno uči dva obavezna druga jezika dva puta tjedno – engleski (kao obavezni J2 prema hrvatskoj obrazovnoj jezičnoj politici) i talijanski (kao J3 zbog statuta grada o dvojezičnom području na kojem se nalazi škola), a u četvrtom razredu osnovne škole počinje učiti njemački kao strani jezik i uči ga do kraja osmog razreda osnovne škole.

Što se tiče slovenskoga jezika, uz povremenu izloženost jeziku u bližoj okolini (s obzirom na život blizu hrvatsko-slovenske granice), i prema tome, receptivno usvajanje slovenskoga jezika, ispitanica započinje s nenamjernim ovladavanjem jezikom kroz individualni sport (judo) koji kreće trenirati u Republici Sloveniji sa šest i pol godina.

4.4 Istraživačka metoda

Budući da je riječ o kvalitativnom istraživanju, odnosno studiji slučaja, koristili smo se triangulacijom podataka kako bismo dobili što podrobniji uvid u razvoj jezičnih i komunikacijskih kompetencija ispitanice. To je uključilo longitudinalno praćenje ispitanice u dobi od sedme do četrnaeste godine njezina života putem promatranja interakcije ispitanice s izvornim govornicima slovenskoga jezika, razvoj receptivnog i produktivnog znanja u slovenskom jeziku, kao i unutarnje motivacije, stavova prema jeziku te uporabu jezika u okruženju izvan doticaja s izvornim govornicima, odnosno u krugu obitelji i učenjem za teorijski ispit iz sporta (judo) na slovenskom jeziku. S ispitanicom je također proveden polustrukturirani intervju kako bi se utvrdio utjecaj određenih kontekstualnih čimbenika na ovladavanje slovenskim jezikom.

5 Rezultati

U ovome poglavlju opisno se prikazuju rezultati studije slučaja na temelju postavljenih istraživačkih pitanja i dobi ispitanice u dva perioda – od 7. do 10. godine i od 11. do 14. godine života.

5.1 Prvi period: dob ispitanice od 7. do 10. godine života

Rezultati pokazuju da je u ovom periodu jezični unos i početno usvajanje slovenskog jezika u prirodnom okruženju i za specifične ciljeve obilježeno ponajprije nesvjesnim procesima usvajanja jezika – naglasak je na sportu, a potom na jeziku. Jezični unos slovenskoga jezika na početku je iznosio 2 sata tjedno, a ponekad i manje, jer se komunikacija trenera s djetetom prilagođavala s obzirom na djetetov J1, te su se upute trenera i objašnjavanja tehnika u sportu pokušale prevoditi sa slovenskog jezika na hrvatski. Treninzi su trajali od početka jedne školske godine u rujnu do kraja školske godine u lipnju. Uz treninge ispitanica je pohađala sportska natjecanja vikendom, barem jednom do dva puta mjesečno, a kasnije i više puta. Uporaba jezika kod ispitanice bila je minimalna, osobito jer je na natjecanjima prisustvovala uvijek uz pratnju roditelja. Interakcija s prijateljima postojala je, no nije bila značajna. Ispitanica bi uglavnom kratko odgovarala na pitanja na hrvatskom jeziku, no nije sama započela razgovor ni s trenerom/trenerima niti prijateljima.

Ispitanica je stoga razvila u početku receptivne vještine u jeziku, kao što je razumijevanje osnovnih uputa na treninzima i natjecanjima. Na početku, kako je izjavila, nije razmišljala o slovenskom kao nekom nepoznatom jeziku, jer je znala da je njezin J1 hrvatski sličan slovenskom, osobito što se tiče kajkavskoga dijalekta kojim se koristila u komunikaciji s majkom te rodbinom s majčine strane. Inače, metajezična i međujezična svjesnost kod ispitanice bila je i ranije razvijena s obzirom na njezino rano ovladavanje drugim jezicima, stoga je bilo i očekivano da će biti visoka i u ovladavanju slovenskim jezikom. Primjer međujezične svjesnosti između hrvatskoga kajkavskoga i slovenskoga jezika primijećena je u trenutku, kada je ispitanica samoinicijativno počela uspoređivati neke slovenske riječi s hrvatskim kajkavskim (npr. *preveč*, *zaprto*).

Motivacija ispitanice bila je uglavnom instrumentalna (vanjska), a stavovi prema jeziku, kao i govornicima toga jezika bili su neutralni. Ispitanica u tome periodu nije pokazivala dodatni interes za znanje jezika, što znači da je u prvom periodu neformalnog doticaja s jezikom, njezina motivacija bila usko povezana s kontekstom, odnosno sportom.

5.2 Drugi period: dob ispitanice od 11. do 14. godine života

U drugom periodu, ispitanica postiže stabilnu i funkcionalnu komunikacijsku kompetenciju u slovenskom jeziku. Može sudjelovati u složenijim razgovorima, razumjeti apstraktnije sadržaje (npr. učenje sportskog sadržaja za polaganje teorijskog dijela ispita za veći pojas u judu) te prilagoditi jezik različitim sugovornicima. Samoprocjena znanja pokazuje napredak ispitanice u jezičnoj proizvodnji u dobi od 14 godina, međutim ispitanica smatra da njezino znanje slovenskoga jezika još uvijek nije visoko, već osrednje. Uglavnom, sport se pokazao snažnim motivacijskim i komunikacijskim okvirom jer je jezik bio sredstvo za ostvarivanje konkretnih ciljeva (komunikacija na treninzima i sportskim natjecanjima), a ne predmet učenja. Porastom dobi ispitanice, intenzitet i redovitost treninga izravno su korelirali s napretkom u jeziku. Autentične komunikacijske situacije omogućile su razvoj pragmatičke i komunikacijske kompetencije.

Vanjska motivacija postupno je prelazila u unutarnju, tako da je u dobi od 12 godina ispitanica htjela samostalno učiti slovenski jezik u aplikaciji Duolingo. Međutim, to nije ostvareno jer aplikacija ne uključuje učenje slovenskog jezika. Nadalje, sve veća motivacija za sport utjecala je i na veću motivaciju za jezik, jer je slovenski jezik postajao sve korisnijim sredstvom komunikacije (npr. osim potrebe za razumijevanjem sportskog sadržaja, izražena je i želja za komunikacijom s poznatim sportašima). Povećana motivacija koja se ujednačila i za sport i jezik, počela se odražavati i u komunikaciji s obitelji, kada ispitanica počinje ubacivati slovenske fraze u hrvatski jezik u različitim iskazima te pritom poticati obitelj na uporabu slovenskog jezika (npr. *baš mi je všeč*; *ni res*; *kaj to pomeni*). U danom periodu ispitanica izražava pozitivan stav prema jeziku, na način da govori o tome koje su joj najdraže riječi u slovenskom (*ni res*, *lahko*). U jezičnoj interakciji s jednim od roditelja, kada je u pitanju sportski diskurs, ispitanica se ne može sjetiti nekih termina na hrvatskom jeziku, već ih objašnjava na slovenskom. Također, primjer jezične skorašnjosti, odnosno čestog utjecaja slovenskog jezika na treninzima odrazio se u formalnom kontekstu ispitanice prilikom prezentacije u školi iz Likovne umjetnosti na hrvatskom jeziku, kada je ispitanica u govoru shvatila da je rekla u jednoj rečenici *uljeta na platnu*, umjesto na hrvatskom *ulje na platnu*. Zaključno, ova studija slučaja pokazuje da je nenamjerno ovladavanje slovenskim jezikom bilo rezultat ne samo izloženosti jeziku u autentičnom kontekstu (sport), već i funkcionalne potrebe za komunikacijom, višejezične pozadine ispitanice i srodnosti među J1 i slovenskim jezikom, ali i podrške roditelja, što je time doprinijelo visokoj unutarnjoj motivaciji i pozitivnim stavovima prema jeziku te rastu jezičnih i komunikacijskih kompetencija.

6 Praktične implikacije za nastavu drugoga jezika

Rezultati provedenog istraživanja upućuju na značajnu korelaciju između izloženosti slovenskom jeziku kroz sport i postupna razvoja jezičnih i komunikacijskih kompetencija kod ispitanice te potvrđuju nalaze prethodnih istraživanja da je neformalno ovladavanje jezikom izrazito dinamičan i nelinearan proces (Mihaljević Djigunović 2012). Provedeno kvalitativno istraživanje studije slučaja doprinosi boljem razumijevanju interakcije kontekstualnih čimbenika u ovladavanju jezikom, osobito kada se ovladavanje odvija u prirodnom okruženju i to kroz izvanškolsku aktivnost, kao što je sport. Naime, istraživanje ima određena ograničenja, a najveće ograničenje predstavlja subjektivnost promatrača, koji nije samo istraživač već i roditelj. Međutim, neovisno o nemogućnosti generalizacije podataka, smatramo da kvalitativna istraživanja ovakva tipa pružaju osnovu i za daljnja slična istraživanja s većim brojem ispitanika u kojima se može primijeniti miješani pristup – i kvalitativna i kvantitativna metoda istovremeno. Nadalje, proučavanje kontekstualnih čimbenika s različitih perspektiva može pomoći nastavnicima jezika u primjeni različitih pristupa poučavanju, poput višejezičnog pristupa, i stvaranju sve raširenije individualne višejezičnosti kod učenika ili studenata. Jednako tako, interesom nastavnika za jezični repertoar učenika, kao i otvaranje mogućnosti međujezičnog poučavanja od strane nastavnika, povećava se i interes učenika za određeni jezik i u formalnom obrazovanju.

Literatura

Bogunović, I. i Jelčić Čolakovac, J. (2019). Uloga neformalnih aktivnosti u nenamjernom usvajanju jezika: Povezanost uporabe jezika i jezičnog znanja. *Fluminensia*, *31*(2), 181–199. https://doi.org/10.31820/f.31.2.15

Dörnyei, Z. (2005). *Studies in second language learning and teaching: Individual differences in second language acquisition*. Lawrence Erlbaum Associates, Inc.

Hendrih, M. i Letica Krevelj, S. (2019). Why do sparkles make a sound in English? The impact of media exposure on child L2 vocabulary acquisition. *Jezikoslovlje*, *20*(3), 447–473. https://doi.org/10.29162/jez.2019.16

Jajić Novogradec, M. (2015). Razvoj istovremene dječje dvojezičnosti u prirodnom jednojezičnom okruženju: Prikaz slučaja. U L. Cvikić, B. Filipan-Žignić, I. Gruić, B. Majhut i L. Zergollern-Miletić (ur.), *Istraživanja paradigmi djetinjstva, odgoja i obrazovanja* (str. 21–29). Učiteljski fakultet Sveučilišta u Zagrebu.

Mihaljević Djigunović, J. (1998). *Uloga afektivnih faktora u učenju stranoga jezika*. Filozofski fakultet Sveučilišta u Zagrebu.

Mihaljević Djigunović, J. (2012). Dynamics of learner affective development in early FLL. *Studies in second language learning and teaching*, *2*(2), 159–178.

Stam, F., Kouzinou, S., Visscher, C. i Elferink-Gemser, M. T. (2020). The value of metacognitive skills and intrinsic motivation for current and future sport performance level in talented youth athletes. *Psychology*, *11*, 326–339. https://doi.org/10.4236/psych.2020.112021

Sundqvist, P. (2009). *Extramural English matters. Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary* [Doktorska disertacija]. Karlstad University.

Valjan Vukić, V. (2016). Izvannastavne i izvanškolske aktivnosti učenika: Višestruke perspektive. *Školski vjesnik*, *65*(1), 33–58.

Walqui, A. (2000). *Contextual factors in second language acquisition*. ERIC Clearinghouse on languages and linguistics Washington DC.

Sociodemografske spremenljivke kot del digitalne jezikovne infrastrukture: primer slovenskega Besedomata in katalonske raziskave razširjenosti besed

DOI: 10.4312/Obdobja.45.175-183

Janoš Ježovnik, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana; janos.jezovnik@zrc-sazu.si

Dejan Gabrovšek, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana; dejan.gabrovsek@zrc-sazu.si

Nataša Gliha Komac, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana; natasa.gliha-komac@zrc-sazu.si

Artur Stepanov, Center za kognitivno modeliranje jezika, Univerza v Novi Gorici, Nova Gorica; arthur.stepanov@ung.si

Matic Pavlič, Pedagoška fakulteta, Univerza v Ljubljani, Ljubljana; matic.pavlic@pef.uni-lj.si

Tina Potočnik, Univerzitetni klinični center Ljubljana, Ljubljana; tina.potocnik@kclj.si

Andrej Perdih, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana; andrej.perdih@zrc-sazu.si

Prispevek obravnava vlogo sociodemografskih vprašalnikov pri množičnih spletnih raziskavah razširjenosti besed. Na primeru slovenskega Besedomata in katalonske raziskave razširjenosti besed pokaže, da univerzalna zasnova tovrstnih vprašalnikov ni smiselna, saj mora izbor sociodemografskih spremenljivk izhajati iz sociolingvističnih značilnosti posameznega jezikovnega okolja. Prispevek pokaže, da primerjava sociodemografskih vprašalnikov ne odpira le vprašanja izbora posameznih kategorij, temveč predvsem vprašanje njihove analitične funkcije v različnih sociolingvističnih okoljih.

sociodemografski vprašalnik, jezikovno okolje, razširjenost besed, slovenščina, katalonščina, sociolingvistika

This article discusses the role of sociodemographic questionnaires in large-scale online word prevalence studies. Using the Slovenian online experiment, or game, Besedomat and a Catalan word prevalence study as case studies, it demonstrates that the design of such questionnaires cannot be universal because the selection of sociodemographic variables must reflect the sociolinguistic characteristics of the language community under investigation. The article argues that comparing sociodemographic questionnaires involves not only the selection of individual categories, but above all an assessment of their analytical function in different sociolinguistic contexts.

sociodemographic questionnaire, language environment, word prevalence, Slovenian, Catalan, sociolinguistics

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

1 Uvod

Množične spletne raziskave razširjenosti besed omogočajo zbiranje obsežnih podatkov o poznavanju besedišča, hkrati pa odpirajo vprašanje, katere sociodemografske podatke zbirati za ustrezno interpretacijo ugotovljenih razlik. Pri pripravi slovenske raziskave, ki se metodološko naslanja na katalonsko (Guasch idr. 2023), se je kot ključno pokazalo vprašanje prilagoditve instrumenta slovenskemu jezikovnemu okolju.

Razširjenost oz. prevalenca besed označuje delež uporabnikov,[1] ki določeno besedo pozna (Keuleers idr. 2015), in se ugotavlja s psiholingvističnimi raziskavami presojanja besedišča. Tehnološki razvoj danes omogoča njihovo množično spletno izvedbo z več deset tisoč udeleženci, s pomočjo katere lahko preverimo stopnjo poznavanja več deset tisoč besed. Udeleženec posamično ocenjuje 120 besed, med katerimi so pomešane psevdobesede.[2] Podobne raziskave so bile opravljene za nizozemščino, angleščino, španščino, katalonščino in italijanščino (Aguasvivas idr. 2018, 2020; Amenta idr. 2025; Brysbaert idr. 2019; Guasch idr. 2023; Keuleers idr. 2015).

Izsledki množičnih spletnih raziskav razširjenosti besed predstavljajo nov tip digitalnih jezikovnih virov, saj omogočajo zbiranje podatkov v obsegu, ki ga s klasičnimi metodami ni mogoče doseči. Kakovost takšnih virov je odvisna tudi od zasnove sociodemografskih vprašalnikov, saj le ustrezno izbrane spremenljivke omogočajo zanesljivo interpretacijo razlik med jezikovnimi uporabniki in okolji.

Podatki o razširjenosti besed dopolnjujejo tradicionalne leksikalne vire: omogočajo vpogled ne le v pogostost besed v besedilih, temveč tudi v njihovo dejansko poznanost med uporabniki, kar je pomembno za raziskave mentalnega leksikona, preverjanje jezikovne zmožnosti, pripravo pedagoških gradiv in razvoj jezikovnih tehnologij. Medtem ko pogostost opisuje predvsem besedila v korpusih, razširjenost razkriva dejanski aktivni oz. pasivni leksikalni repertoar uporabnikov, zato je zbiranje ustreznih sociodemografskih podatkov ključen del zasnove takšnih raziskav.

Zbiranje odgovorov za slovenščino se je začelo novembra 2024 in poteka prek spletnega eksperimenta oz. igre Besedomat (www.besedomat.si). Do oktobra 2025, ko je bil opravljen prvi zajem podatkov, je 39.064 udeležencev v celoti izpolnilo 48.661 vprašalnikov, pridobljeni so bili podatki o poznavanju 42.000 slovenskih besed. Ker sta oblikovanje končnih norm in statistično modeliranje vpliva posameznih dejavnikov predmet nadaljnjih analiz, se prispevek osredotoča na predhodno metodološko vprašanje: katere sociodemografske informacije je smiselno zbirati, da bo taka analiza sploh mogoča.

Potek slovenskega eksperimenta strukturno sledi raziskavam za druge jezike, naslanja pa se predvsem na katalonsko (Guasch idr. 2023), opravljeno tik pred slovensko z enako programsko opremo, in v sodelovanju z njenimi izvajalci. Prilagojena sta besedilni opis naloge in potek eksperimenta, razlikujeta pa se struktura in vsebina

1 Moški spol se rabi kot slovnično nevtralen in označuje osebe ne glede na spol.

2 Psevdobesede so nesmiselna (tj. neobstoječa), a fonološko in fonotaktično ustrezna zaporedja črk ali glasov v danem jeziku. Uporabljajo se namesto naključnih nesmiselnih zaporedij zvokov ali črk (t. i. nebesed), saj je pomembno, da jih udeleženci ne zavrnejo iz nejezikovnih razlogov, ampak jih mentalno obravnavajo, kot da bi bile del jezika (Pavlič idr. 2025). O pripravi seznama besed gl. Perdih idr. (2025).

vprašalnika, kar izhaja iz različne sociolingvistične realnosti obeh jezikov: katalonščina deluje v izrazito večjezičnem okolju, kjer se uporabniki razlikujejo predvsem po jezikovni biografiji, starosti usvojitve in izpostavljenosti jeziku, slovenščino pa zaznamuje homogenost prvega jezika, a izrazita narečna členjenost. Ta kontrast omogoča preverjanje analitične vrednosti spremenljivk glede na strukturo jezikovne skupnosti.

Sociodemografski vprašalnik je pomemben z dveh vidikov: pove, kdo so udeleženci glede na različne dejavnike in koliko to odraža strukturo prebivalstva, obenem pa omogoča ugotavljanje, kako različni dejavniki vplivajo na poznavanje besedišča. Ni torej zgolj spremljevalni instrument za opis vzorca, temveč operacionalizacija dejavnikov s pojasnjevalno, posredovalno ali kontrolno vlogo: nekatere postavke pojasnjujejo razlike med skupinami uporabnikov, druge mehanizme teh razlik, tretje pa predvsem preverjajo strukturo vzorca. Smiselnost vključitve spremenljivke torej ni odvisna le od splošne relevantnosti, temveč predvsem od pričakovane variabilnosti in analitične funkcije v konkretnem okolju.

Iz tega izhaja splošno metodološko načelo, da morajo take raziskave vključevati tiste sociodemografske spremenljivke, za katere lahko glede na značilnosti jezikovne skupnosti pričakujemo pojasnjevalno vrednost pri razlagi razlik v leksikalnem znanju. Izbor postavk torej ni univerzalen, temveč je odvisen od sociolingvističnih značilnosti posameznega okolja.

Prispevek temelji na primerjalni analizi sociodemografskih vprašalnikov, uporabljenih v slovenski in katalonski raziskavi razširjenosti besed, in sicer glede na vrsto zajetih podatkov, predvideno analitično funkcijo ter pojasnjevalno vrednost v posameznem okolju, izhajajoč iz predpostavke, da ustreznosti spremenljivke ni mogoče presojati neodvisno od populacije, v kateri se uporablja. V nadaljevanju pokažemo, kako sociolingvistične značilnosti okolja vplivajo na izbiro spremenljivk in kako je smiselnost posameznih postavk odvisna predvsem od njihove pričakovane pojasnjevalne vrednosti znotraj konkretne skupnosti.

2 Izbor sociodemografskih spremenljivk v različnih jezikovnih okoljih

Strukturne razlike v demografskih vprašalnikih izhajajo iz dejstva, da je za katalonščino značilno heterogeno dvojezično okolje: zaradi različnega diglosičnega položaja so med njenimi uporabniki velike razlike v načinu usvajanja in učenja jezika. Za slovenščino kot večinski jezik v Republiki Sloveniji je značilen drugačen položaj, saj pri večini jezikovnih uporabnikov sovpadajo prvi jezik, dominantni jezik okolja in jezik izobraževanja.

Katalonščina je uradni jezik v Andori ter v Kataloniji, na Balearskih otokih in v Valencijski skupnosti (tam kot valencijščina), kjer si status deli s španščino. Omejeno zaščito uživa v vzhodni Aragoniji, brez uradnega statusa pa jo govorijo tudi v Severni Kataloniji v Franciji ter v Algheru/Algueru v Italiji. Večina uporabnikov je dvojezična, velik delež pa predstavljajo tisti, ki so se je kot drugega jezika priučili v šoli, prek družbenih stikov in/ali po priselitvi na tradicionalno katalonska območja (prim. Pradilla

Cardona 2022). Pri analizi razširjenosti besed (Guasch idr. 2023) so zato podatek o znanju katalonščine kombinirali s starostjo ob začetku usvajanja jezika.

Slovenščina je uradni jezik Republike Slovenije, materni in prvi jezik večine prebivalstva, prisoten v vseh domenah javne rabe že od zgodnjega otroštva. Z izjemo dvojezičnih območij na Obali in v Prekmurju (italijanščina, madžarščina) si jezikovnega okolja ne deli z nobenim drugim dominantnim jezikom, temveč v vlogi dominantnega jezika nastopa sama.[3] Posledično je dober del vprašanj o usvajanju jezika (starost ob stiku, raven znanja, izpostavljenost) manj relevanten, saj bi bili odgovori trivialno enotni (*od rojstva*, *materni jezik*, *vedno* ipd.). Slovenski vprašalnik jih zato nadomešča z vprašanjem o prvem in drugem maternem jeziku.

Demografska vprašalnika se najbolj očitno razlikujeta v načinu zajema jezikovnega ozadja udeleženca: katalonski v ospredje postavlja samoopredelitev, identiteto in rabo ter zato vključuje obsežen niz vprašanj o jezikovnem ozadju, slovenski pa bolj izpostavlja narečno členjenost ter s tem geografski izvor in kraj bivanja.

2.1 Samoopredelitev jezika staršev/maternega jezika

Ker materni jezik ni enoznačna kategorija, temveč lahko zajema izvorno, kompetenčno, funkcijsko in identifikacijsko razsežnost jezikovne pripadnosti, je njegova operacionalizacija posebej odvisna od značilnosti jezikovnega okolja.

Katalonska različica od udeleženca ne zahteva (samo)opredelitve maternega jezika, temveč sprašuje po jeziku vsakega od staršev posebej (*Llengua de la mare*, *Llengua del pare*) in pri vsakem omogoča eno izbiro. Odločitev odraža pogostost jezikovno mešanih gospodinjstev in se izogiba identitetno problematični kategoriji maternega jezika. Opredelitev jezikov staršev ponuja objektivnejši približek jezikovnega okolja v zgodnjem otroštvu, a izrecno ne upošteva možnosti, da bi bil kateri od staršev sam dvojezičen, temveč odgovornost za enoznačno opredelitev prenaša na respondenta.

Slovenska različica v nasprotju s tem neposredno sprašuje po udeležencevem lastnem maternem jeziku (*Materni jezik 1*), za dvojezične od zgodnjega otroštva pa vprašanje dopolnjuje neobvezno drugo polje (*Materni jezik 2*). Odločitev temelji na presoji, da je v okolju, kjer slovenščina nastopa kot dominantni jezik, vprašanje o maternem jeziku manj ideološko obremenjeno in da bo večina kot svoj (edini) materni jezik navedla slovenščino. Drugo polje omogoča samoopredelitev uporabnikom manjšinskih in drugih jezikov. Isti format se uporablja tudi za Slovence v sosednjih državah in po svetu, kjer je slovenščina sicer manjšinski in navadno manj prestižen jezik.

3 Razmere na območjih sosednjih držav, na katerih se slovenščina uporablja kot manjšinski jezik, so sicer drugačne, a še vedno gre za strnjen slovenski jezikovni prostor s prepoznavno slovensko narečno podobo. Za uporabnike slovenskega jezika na teh območjih je do nedavna veljalo, da se slovenščine kot maternega jezika otroci naučijo (oz. ga usvojijo) doma, kjer ga tudi prakticirajo. Razmere pa se spreminjajo. Tako npr. zadnje raziskave za Italijo (npr. Brezigar in Verša 2023) kažejo, da »je v slovenskih vzgojnoizobraževalnih ustanovah danes v povprečju že več otrok, katerih pogovorni jezik doma je pretežno italijanščina, kot otrok, katerih pogovorni jezik je pretežno slovenščina«.

2.2 Samoocena jezikovne zmožnosti

Katalonski vprašalnik vključuje postavko za samooceno jezikovne zmožnosti (*Nivell de català*) na lestvici od *kot rojeni govorec* do *nič*, saj razlike v zmožnosti med uporabniki katalonščine kot drugega jezika lahko vplivajo na prepoznavanje besed. Slovenski vprašalnik primerljivega vprašanja ne vsebuje, saj zaradi enojezičnega okolja in obveznega izobraževanja v slovenščini samoocena ne bi bistveno pripomogla k analizi, čeprav bi bilo zanimivo preveriti, ali tisti, ki zmožnost ocenjujejo visoko, izkazujejo tudi višje poznavanje besedja.

2.3 Obdobje usvojitve jezika

S samooceno je tesno povezano vprašanje o starosti, pri kateri so udeleženci prišli v stik s katalonščino (*A quina edat vas entrar en contacte amb el català?*), z izbirami od *od rojstva* do *70 ali več*. Starost je močan napovednik jezikovne zmožnosti v dvojezičnih populacijah, zato je katalonska raziskava (Guasch idr. 2023) izločila udeležence, ki so v stik s katalonščino prišli pri petih letih ali kasneje. Slovenski vprašalnik ustreznice nima, saj bi bilo vprašanje v populaciji, kjer prevladujejo rojeni govorci, neinformativno.

2.4 Izpostavljenost jeziku

Oba vprašalnika sprašujeta po izpostavljenosti jeziku, a vsak z druge perspektive. Katalonska različica sprašuje, kolikšen delež časa je udeleženec dnevno izpostavljen katalonščini (*Quina proporció del temps estàs exposat al català en el dia a dia?*), z izbirami od *vedno* do *nikoli* v 10-odstotnih korakih, kar zajema trenutno raven izpostavljenosti neodvisno od starosti usvojitve ali stopnje znanja. Slovenski vprašalnik podobno postavko zajema s samopredelitvijo najpogosteje rabljenega jezika (*Kateri jezik najpogosteje uporabljate?*); ker je pričakovani odgovor za veliko večino slovenščina, postavka deluje predvsem kot spremenljivka za prepoznavanje izjem, pri katerih prevladujoči jezik ni slovenščina in pri katerih je treba podatke potencialno obravnavati ločeno.

2.5 Večjezičnost

Katalonski vprašalnik sprašuje tudi, koliko jezikov govori udeleženec vključno s katalonščino (*Quantes llengües parles incloent el català?*), z izbiro od enega do sedem ali več. Ker mnogi uporabniki poleg katalonščine in kastiljščine uporabljajo tudi druge jezike, spremenljivka zajema širši večjezični kontekst, ki lahko vpliva na leksikalno obdelavo. Slovenski vprašalnik primerljive postavke ne vključuje, saj je večjezični profil vzorca po predpostavki strukturiran manj sistematično in zato manj verjetno pomembna napovedna spremenljivka.

2.6 Narečna členjenost

Za katalonščino je značilna nizka narečna členjenost; narečja se razlikujejo, a so v veliki meri medsebojno razumljiva. Na podlagi glasoslovja in oblikoslovja se delijo na vzhodna (severna, osrednja in balearska katalonščina ter govor Alghera/Alguera na

Sardiniji) in zahodna narečja (severozahodna katalonščina in valencijščina) (Wheeler 2005). Slovenščina je narečno izjemno razčlenjena, saj razlikuje 38 narečij v 7 narečnih skupinah (prim. Škofic idr. 2023: 11).

Razlike v vplivu narečne členjenosti se odražajo v natančnosti beleženja geografskega izvora. Katalonski vprašalnik preverja kraj odraščanja (*On has passat els primers anys de la teva vida?*) z izbiro med izpostavljenimi regijami (Katalonija, Valencija, Balearski otoki, ostala Španija) in seznamom držav sveta, kraj bivanja pa beleži s prosto vneseno poštno številko (*Codi postal actual*). Slovenski vprašalnik podatke beleži strožje, z dvema dvostopenjskima vprašanjema (*Kje ste preživeli otroštvo?*, *Trenutni kraj bivanja*): udeleženec najprej izbere državo (izpostavljena izbira je Slovenija), nato pa s seznama poštnih številk opredeli kraj; če Slovenije ne izbere, seznam nadomesti prosto vnosno polje.

Strožje preverjanje kraja odraščanja in bivanja pri slovenščini omogoča podrobnejšo prihodnjo analizo, denimo primerjavo mobilnih in manj mobilnih uporabnikov. Ker se narečna členjenost odraža tudi v leksikalni diferenciaciji, je pričakovati, da bo geografski izvor pomemben napovednik poznavanja posameznih leksemov.

Za katalonske razmere je bolj bistvena informacija, ali je uporabnik odraščal v katalonskem (dvojezičnem) okolju ali ne, saj so izpostavljene geografske enote v grobem prekrivne z jezikovnim območjem katalonščine v Španiji. Vprašalnik pri trenutnem kraju bivanja ne predvideva nedvoumne možnosti za uporabnike zunaj Španije.

3 Metodološke implikacije za oblikovanje vprašalnikov

Analiza slovenskega in katalonskega vprašalnika kaže, da razlike med njima niso zgolj posledica prilagajanja jezikovnim okoljem, temveč odražajo različno operacionalizacijo dejavnikov, ki vplivajo na leksikalno znanje. Ključno vprašanje torej ni le, katere kategorije vključiti, ampak kakšno analitično funkcijo imajo spremenljivke v raziskovalnem modelu.

Primerjava kaže tudi, da ista sociodemografska kategorija v različnih okoljih nima nujno enake analitične funkcije. V katalonskem okolju lahko podatki o starosti usvojitve, izpostavljenosti in samooceni zmožnosti delujejo kot ključne pojasnjevalne oz. posredovalne spremenljivke, saj imajo v heterogenih dvo- in večjezičnih skupnostih spremenljivke, povezane z jezikovno biografijo, večjo pojasnjevalno vrednost. V slovenskem okolju imajo iste spremenljivke zaradi manjše variabilnosti manjšo moč, to funkcijo pa prevzamejo spremenljivke, povezane z geografskim izvorom in narečno diferenciacijo, saj lahko kraj odraščanja in bivanja pomembno prispevata k razlagi poznavanja leksemov.

Na primeru katalonščine in slovenščine se tako kaže splošnejše načelo: sociodemografski vprašalnik ni univerzalni instrument, temveč mora biti prilagojen sociolingvistični strukturi raziskovane skupnosti. Kakovost digitalnih jezikovnih virov ni odvisna zgolj od velikosti vzorca, temveč tudi od ustrezne izbire spremenljivk, ki omogočajo smiselno interpretacijo razlik.

3.1 Preliminarna analiza izbranih sociodemografskih spremenljivk

Namen preliminarne analize ni ovrednotiti vpliva posameznih dejavnikov na poznavanje besed, temveč preveriti, ali struktura zbranih podatkov podpira izhodiščne predpostavke prilagoditve slovenskega vprašalnika.

V raziskavi je do 25. oktobra 2025 sodelovalo 39.064 udeležencev. Od teh je 96,14 % kot materni jezik navedlo slovenščino, od preostalih 3,86 % pa je slovenščino kot drugi materni jezik navedlo 63,51 % (tj. 2,45 % vseh udeležencev). Slovenščino kot najpogosteje rabljeni jezik je izbralo 95,51 % udeležencev. Nizka variabilnost pri teh postavkah potrjuje odločitve pri opredeljevanju jezikovne biografije udeležencev in predpostavko, da ima ta spremenljivka v vzorcu predvsem kontrolno funkcijo.

Pri analizi kraja bivanja in odraščanja smo pridobljene poštne številke zaradi anonimizacije preslikali v statistične regije po uradni klasifikaciji Statističnega urada Republike Slovenije. Regionalno sestavo vzorca smo preverili s primerjavo deležev udeležencev po regijah z uradnimi podatki o prebivalstvu Slovenije (Statistični urad Republike Slovenije b. d., stanje 1. 7. 2025), analizo pa opravili le na podatkih udeležencev, starih 18 let ali več. Znaten delež udeležencev na to vprašanje ni odgovoril (56,4 % za trenutni kraj bivanja in 58,12 % za kraj odraščanja).

Statistična regija	Delež v populaciji	Delež v vzorcu		Razmerje	
		(trenutno)	(v otroštvu)	(trenutno)	(v otroštvu)
osrednjeslovenska	26,5 %	45,4 %	33,3 %	1,71	1,25
podravska	15,8 %	10,2 %	11 %	0,65	0,70
savinjska	12,3 %	8,6 %	10,9 %	0,70	0,88
gorenjska	9,7 %	10,6 %	11,8 %	1,08	1,21
JV Slovenija	6,8 %	5 %	6,8 %	0,73	0,99
obalno-kraška	5,7 %	4,1 %	4,1 %	0,71	0,71
goriška	5,6 %	5,5 %	7,1 %	0,99	1,28
pomurska	5,4 %	2,3 %	3,5 %	0,43	0,65
posavska	3,6 %	3,6 %	2,7 %	0,57	0,75
koroška	3,3 %	1,9 %	3 %	0,58	0,90
zasavska	2,7 %	2,2 %	3,4 %	0,83	1,25
primorsko-notranjska	2,5 %	2,2 %	2,6 %	0,88	1,03

Tabela 1: Deleži udeležencev po statističnih regijah v primerjavi z deleži prebivalstva Slovenije, starega 18 let ali več (Statistični urad Republike Slovenije b. d., stanje 1. 7. 2025).

Pri vprašanju o trenutnem kraju bivanja, posplošenem na statistično regijo, se je pokazalo, da vzorec ni uravnotežen glede na dejansko regionalno sestavo prebivalstva ($\chi^2(11) = 3270{,}7$, $p < 0{,}001$). Deleži udeležencev iz posameznih regij so se v povprečju za približno 20 odstotnih točk razlikovali od pričakovanih. Najbolj izstopa osrednjeslovenska regija, zastopana skoraj dvakrat pogosteje, kot bi pričakovali (45 % proti 27 % prebivalstva), medtem ko so pomurska, posavska in koroška regija opazno podzastopane. Pri kraju bivanja v otroštvu je slika podobna, a manj izrazita: tudi ta porazdelitev se značilno razlikuje od populacijske ($\chi^2(11) = 951{,}7$, $p < 0{,}001$), a je neujemanje za dobro polovico manjše (povprečno okoli 11 odstotnih točk). Osrednjeslovenska regija je tudi

tu nekoliko prekomerno zastopana (33 % proti 27 %), podravska in pomurska regija pa sta podzastopani pri obeh spremenljivkah, kar kaže na sistematično manjšo pokritost teh dveh regij v vzorcu.

4 Zaključek

Analiza slovenskega in katalonskega vprašalnika kaže, da prilagajanje sociodemografskih vprašalnikov različnim okoljem ne pomeni zgolj izbire različnih postavk, temveč predvsem presojo njihove analitične funkcije v konkretnem kontekstu. Ista kategorija lahko glede na strukturo skupnosti deluje kot ključna pojasnjevalna spremenljivka, posredovalni dejavnik ali kontrolni podatek z omejeno pojasnjevalno močjo.

Razlike med vprašalnikoma tako ne odražajo različnih raziskovalnih ciljev, temveč prilagoditev istih metodoloških načel različnim okoljem. Preliminarna analiza omogoča prvo preverjanje predpostavk zasnove slovenskega vprašalnika, dokončno oceno pojasnjevalne vrednosti spremenljivk pa bodo omogočile nadaljnje statistične analize povezav s poznavanjem leksemov.

Rezultati imajo pomen tudi širše od zasnove vprašalnika: kažejo, da je pri oblikovanju digitalnih jezikovnih virov nujno upoštevati sociolingvistične značilnosti jezikovnega okolja. Ustrezna prilagoditev sociodemografskih instrumentov prispeva k zanesljivejši interpretaciji podatkov ter večji uporabnosti digitalnih jezikovnih virov in jezikoslovnih analiz.

Prispevek je nastal v okviru projekta Množična raziskava razširjenosti slovenskih besed (J6-50199), ki ga sofinancira ARIS.

Literatura

Aguasvivas, J. A., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E. in Duñabeitia, J. A. (2018). SPALEX: A Spanish lexical decision database from a massive online data collection. *Frontiers in Psychology*, *9*, članek 2156. https://doi.org/10.3389/fpsyg.2018.02156

Aguasvivas, J., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E. in Duñabeitia, J. A. (2020). How do Spanish speakers read words? Insights from a crowdsourced lexical decision megastudy. *Behavior Research Methods*, *52*(5), 1867–1882. https://doi.org/10.3758/s13428-020-01357-9

Amenta, S., De Varda, A. G., Mandera, P., Keuleers, E., Brysbaert, M. in Marelli, M. (2025). The Italian crowdsourcing project: Visual word recognition times for 130,495 Italian words. *Behavior Research Methods*, *57*(1), članek 26. https://doi.org/10.3758/s13428-024-02548-4

Brezigar, S. in Verša, A. (2023). Slovensko šolstvo v Italiji: včeraj, danes in jutri. V A. Istenič, M. Gačnik, B. Horvat, M. Kukanja Gabrijelčič, V. R. Kiswarday, M. Lebeničnik, M. Mezgec in M. Volk (ur.), *Vzgoja in izobraževanje med preteklostjo in prihodnostjo/Upbringing and education between the past and the future* (str. 311–327). Založba Univerze na Primorskem. https://doi.org/10.26493/978-961-293-255-8.311-327

Brysbaert, M., Mandera, P., McCormick, S. F. in Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Guasch, M., Boada, R., Duñabeitia, J. A. in Ferré, P. (2023). Prevalence norms for 40,777 Catalan words: An online megastudy of vocabulary size. *Behavior Research Methods*, *55*(6), 3198–3217. https://doi.org/10.3758/s13428-022-01959-5

Keuleers, E., Stevens, M., Mandera, P. in Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly*

Journal of Experimental Psychology, *68*(8), 1665–1692. https://doi.org/10.1080/17470218.2015.1022560

Klemen, M. (2025). Poznavanje pogostih splošnih besed v slovenščini med govorci slovenščine kot drugega in tujega jezika. *Prispevki za novejšo zgodovino*, *65*(3), 243–272. https://doi.org/10.51663/pnz.65.3.10

Meyer, D. E. in Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Pavlič, M. in Perdih, A. (2025). Lexical processing of morphologically complex Slovene words in a lexical decision task: The role of pseudowords. *Slovenski jezik / Slovene Linguistic Studies*, *17*(1), 59–91. https://doi.org/10.3986/17.1.04

Perdih, A., Gabrovšek, D. in Pavlič, M. (2025). Izdelava seznama besed za množično raziskavo razširjenosti slovenskih besed. *Slavistična revija*, *73*(1), 121–138. https://doi.org/10.57589/srl.v73i1.4231

Pradilla Cardona, M. A. (2022). *Catalan sociolinguistics: State of the art and future challenges*. John Benjamins Publishing Company. https://doi.org/10.1075/ivitra.32

Statistični urad Republike Slovenije. (b. d.). Prebivalstvo po starosti in spolu, statistične regije, Slovenija, polletno [Podatkovni vir]. *SiStat*. https://pxweb.stat.si/SiStatData/pxweb/sl/Data/-/05C2001S.px

Škofic, J., Gostenčnik, J., Hazler, V., Jakop, T., Kenda-Jež, K., Kumin Horvat, M., Nartnik, V., Pahor, N., Smole, V., Šekli, M. in Zuljan Kumar, D. (2023). *Slovenski lingvistični atlas 3: Kmetovanje. 1 Atlas.* Založba ZRC, ZRC SAZU. https://doi.org/10.3986/9789610507499

Wheeler, M. (2005). *The phonology of Catalan*. Oxford University Press. https://doi.org/10.1093/oso/9780199258147.001.0001

Dileme ob zapisovanju govorjenega jezika v oddaji *Kaj dogaja?*

DOI: 10.4312/Obdobja.45.185-193

Petra Jordan, RTV Slovenija, Ljubljana; petra.jordan@rtvslo.si

Prispevek predstavlja različne stopnje zapisa pogovornih prvin govorjenega jezika v podnapisih oddaje *Kaj dogaja?*, od ortografskega do pogovornega zapisa, na različnih ravneh jezika. Pokaže dileme, ki se v času pospešenega prodiranja pogovorne slovenščine v javna besedila porajajo ob prenašanju značilnosti govora v zapis.

govorjeni jezik, pogovorni jezik, ortografski zapis, pogovorni zapis, podnaslavljanje

This article presents different writing systems for non-standard elements of spoken language in the subtitles of the television program *Kaj dogaja?* (What's Happening?), ranging from orthographic to non-standard transcription at different levels of language. It highlights the dilemmas that arise when transferring speech characteristics into written form in the context of rapidly evolving informal Slovenian.

spoken language, informal language, orthographic transcription, non-standard transcription, subtitles

1 Uvod

Če je bil zapis spontanega govora, ki odstopa od knjižne norme, nekoč omejen predvsem na posamezne literarne eksperimente, je danes v številnih javnih diskurzih zelo razširjen in pogosto pričakovan. V digitalni dobi namreč velik del vsakodnevne zasebne in neformalne komunikacije poteka v pisni obliki, in sodobne raziskave medijskega in digitalnega diskurza kažejo, da se značilnosti govorjenega jezika sistematično prenašajo v javna pisna besedila, npr. na družbenih omrežjih, tako na ravni besedišča kot tudi zapisa (prim. Zwitter Vitez in Fišer 2015: 87–89). Ta jezik z namenom nagovarjanja in pridobivanja občinstva poustvarjajo oglaševalci, mediji in drugi tvorci javnega diskurza (Kalin Golob 2013: 201). V tem kontekstu se odpira raziskovalno vprašanje, kako pogostost pojavljanja pogovornih jezikovnih prvin,[1] odmikov od ortografskega zapisa in drugih značilnosti govorjenega jezika v javnih pisnih besedilih vpliva na oblikovanje pričakovanj naslovnikov (gledalcev) glede zapisovanja govorjenega jezika ter posledič-

1 S pogovornimi jezikovnimi prvinami mislimo prvine, ki jih *Slovenski pravopis* (2001) v členih 1048–1058 določa kot knjižno- in neknjižnopogovorne. V prispevku uporabljamo to poimenovanje zaradi usklajenosti z izrazjem, ki ga uporabljajo *Smernice za podnaslavljanje v slovenščini* Društva slovenskih filmskih in televizijskih prevajalcev (2020) in *Navodila za oblikovanje podnapisov* RTV Slovenija (2020).

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

no na spremembe standardov zapisovanja v sodobni slovenščini, kot jih opisujejo npr. *Smernice za podnaslavljanje v slovenščini* Društva slovenskih filmskih in televizijskih prevajalcev (2020).

Prispevek predstavlja dileme pri zapisovanju neknjižnih prvin v podnaslovih razvedrilne oddaje *Kaj dogaja?* na RTV Slovenija. Osredotoča se na različne stopnje pretvarjanja govora v zapis, ki jih narekujejo pravopisna norma[2] in smernice za podnaslavljanje slovenskega govora na Televiziji Slovenija. Odločitev za podnaslavljanje nekaterih delov satirične oddaje *Kaj dogaja?* temelji na smernici za podnaslavljanje slovenskih govorcev: »Govor slovenskih govorcev v televizijskih programskih vsebinah podnaslavljamo, če je nerazločen. Prednostno merilo pri odločanju, ali sogovornika podnasloviti, je razumljivost« (Radiotelevizija Slovenija 2019: 1). Pri oddaji *Kaj dogaja?* so tako podnaslovljene le nekatere rubrike, to so pevski nastopi in skeči, pri katerih je uredništvo presodilo, da bi bil govor zaradi glasbe in imitiranja (navadno) znanih oseb lahko na nekaterih mestih slabše razumljiv. V skladu s standardno televizijsko prakso so te rubrike podnaslovljene v celoti. Za to skrbi skupina podnaslavljalcev, ki besedilno predlogo iz scenarija ustrezno oblikuje v podnapise. To pomeni, da poskrbi za časovno kodiranje, ustrezno obliko in povprečno bralno hitrost 12 znakov na sekundo, kot določajo *Navodila za oblikovanje podnapisov* (Radiotelevizija Slovenija 2020), usklajenost zapisanega besedila z govorom (pregled s posnetkom) ter jezikovni pregled. Pri slednjem v prvi vrsti upošteva splošna jezikovna priporočila smernic za podnaslavljanje slovenskih govorcev, pri čemer je prvo vodilo informativnost in s tem skladnost zapisanega z dejansko izrečenim: »Če podnaslavljamo slovenske govorce, naj bo zapis kolikor se da približek povedanega. Pri tem načeloma upoštevamo pravila zbornega jezika, s prvinami knjižnega pogovornega jezika na področju leksike. Prednostnega pomena je, da podnapis prenese informacijo in vsebino povedanega, ob tem pa naj se ohranita stil in čustveni naboj povedanega« (Radiotelevizija Slovenija 2019: 1). Prav slednje navodilo je, v nasprotju z večino drugih podnaslovljenih slovenskih vsebin na RTV, pri satirični oddaji, ki vključuje elemente različnih besedilnih zvrsti, kot so skeč, stand up in verzna besedila, ključnega pomena.

Za ta prispevek smo analizirali podnaslovljene rubrike 200 oddaj *Kaj dogaja?*, ki so bile na sporedu od marca 2019 do maja 2026, pri čemer smo izpustili tiste, pri katerih podnapisi vsebinsko niso skladni z govorom ali niso v skladu z omenjenimi smernicami, torej tiste, pri katerih je zapis v celoti poknjižen. Navedeni zgledi so iz skečev Glas ljudstva (GL), Palomar (P) in Literapija (L) ter glasbenih vložkov S(H)it tedna (Sh), Uvodni song (US) in Mala terasa (MT).[3]

Pri tem nas je zanimalo, v katerih primerih in kako zapis govorjenega jezika oddaje odstopa od določil smernic za podnaslavljanje RTV ter Društva slovenskih filmskih in

2 Tj. upoštevanje v slovarskem delu *Slovenskega pravopisa* (2001) normiranega zapisa neknjižnih besed (označenih s kvalifikatorji *knjižno pogovorno*, *neknjižno pogovorno*, *slengovsko*, *narečno*), ki praviloma sledijo normi zapisovanja knjižne slovenščine, kar zadeva npr. polglasnik (prim. geslo *foter*) in dvoustnični v (prim. geslo *šravfati*).

3 Vse oddaje so dostopne v spletnem arhivu RTV Slovenija: https://365.rtvslo.si/oddaja/kaj-dogaja/173251016.

televizijskih prevajalcev, ki narekujejo, da se pogovornost nakazuje le na leksikalni ravni. Zaradi humornosti besedil gre sklepati, da sta slog in jezikovna zaznamovanost pri pretvorbi v zapis v nekem deležu ohranjena, zato smo analizirali prvine pogovornosti na treh ravneh: leksikalni, oblikoslovno-skladenjski in pravopisni. Razvrstili smo jih glede na stopnjo odmika od knjižnega jezika oz. glede na ortografski ali pogovorni zapis in glede na besedilno vrsto, tj. proza (skeči) in verzi (besedila pesmi).

2 Prvine pogovornosti v podnaslovih oddaje *Kaj dogaja?*

2.1 Leksikalna raven

Slogovno zaznamovane prvine na ravni besedišča, tj. izrazi z različnimi stopnjami pogovornosti, slengizmi ter pokrajinskopogovorni in narečni izrazi, so v podnaslovih v večini ohranjeni, npr.: *losali*, *fotkanje*, *šiht*, *štekaš*, *ziher*, *uštimat'*, *bejbe*, *sori*, *cajtengi*, *glih*, *bištek*, *špajza*, *mona*, *favl*, *fašeš*, *valda* idr.[4] Redki primeri poknjiženja na leksikalni ravni so pogovorni členki in prislovi tipa: *pol* → *potem*, *kle* in *tle* → *tu/tule*; in nekatere okrnjene glagolske oblike: *pejt*/*pejd* → *pojdi*. Ta vrsta popravka je uporabljana zelo nedosledno, srednje pogosto v prozi, v verznih besedilih pa nikoli, saj bi sprememba števila zlogov vplivala na metrum.

Posebej je pri tem treba izpostaviti tujejezične prvine. RTV Slovenija je zavezana k spoštovanju Zakona o javni rabi slovenščine, ki javnemu zavodu nalaga, da priskrbi prevod vseh tujejezičnih vsebin (Radiotelevizija Slovenija 2026: 15). To pomeni, da bo npr. skeč, v katerem nastopa imitator Aleksandra Vučića v srbščini (GL, 22. 2. 2025), v podnapisih v celoti preveden. Prevedeni so tudi deli besedil ali posamezne besede v tujih jezikih »rednih tujih gostov« oddaje, kot sta Donald in Melania Trump (oz. njuna imitatorja), ki za namene posnemanja in humornosti tipično uporabljata veliko angleških besed.

(1) Prepis govora (PG):[5] »[…] vsi, ki ste dans pršli pogledat pravga, starga in novga kinga of USA.«
Podnapis (pdn): […] vsi, ki ste danes prišli pogledat pravega, starega in novega kralja ZDA. (GL, 22. 3. 2024)

Posebnost pri tem so prepoznavne osebne fraze, ki so ohranjene, npr. pri imitiranju Borisa Pahorja:

(2) »Tanja, miška mala, poglej. Ko pride trenutek, pač moraš reči *so what* in prevzeti odgovornost.« (P, 16. 3. 2024)

Drugače je pri zapisu tujih prvin v govorici slovenskih nastopajočih. Večinoma gre za anglicizme, germanizme, hrvatizme in srbizme, ti so občasno obravnavani kot tuje besede (zgled 3) in prevedeni, pogosteje pa ohranjeni kot del pogovorne slovenščine (zgled 4). Zanimiv kazalnik razvoja pogovorne slovenščine je morda podatek, da

4 Primeri so iz različnih rubrik več epizod, našteti za ilustracijo izrazov iz neknjižnjih zvrsti, ki se v podnapisih nikoli ne poknjižijo.

5 Za prepis govora smo uporabili t. i. pogovorni zapis (Verdonik in Zwitter Vitez 2011: 57).

anglicizmi pogosteje ostanejo neprevedeni kot besedje iz drugih tujih jezikov. V besedilih pesmi tujejezične prvine nikoli niso prevedene.

(3) PG: »Đoni je dubu dojavu, da bojo pelal nekje dve milje, če ne celo tri mogoče. In če to spustiš skos … u ulomk pa to, suma sumater ti nekje pride petsto kil po glavi, tko da valda, da smo bli vsi za. Brt.«
Pdn: Đoni je izvedel, da bojo peljali nekje dva milijona, če ne celo tri mogoče. In če to spustiš v ulomek, *suma sumater* ti nekje pride 500 kil po glavi, tako da valda, da smo bili vsi za. Brat.

(4) Pdn:[6] Tako da na začetku sem vozil malo kombi, posodil kakšno *šako*, služba ko služba. Prideš na lokacijo, daš dva *šamarja*, ustreliš dva metka v steno in pol ti itak plačajo, a veš.« (L, 18. 4. 2026)

Nadalje nas je pri tujejezičnih prvinah, ki so v podnapisih ohranjene, zlasti pri anglicizmih, zanimala raba podomačenega ali citatnega zapisa. Opaziti je upoštevanje načela ustaljenosti, bodisi na podlagi slovarske normiranosti (*frend*) bodisi siceršnje razširjenosti enega ali drugega zapisa v rabi na družbenih omrežjih:[7] npr. *flet*, *lajk* ipd. so zapisani podomačeno, primeri kot *(za vse te) commente in share* pa citatno. Temu trendu sledi tudi pogostejša podomačitev prevzetih glagolov, ki jim v govoru dodajamo slovenske končnice, v primerjavi s citatno zapisanimi samostalniki v naslednjem zgledu:

(5) Pdn: Ona je takrat živela z nekimi cimri, ki so se ukvarjali s temi blockchaini pa to, pa neki so majnali, pa trejdali te coine. (L, 21. 3. 2026)

Drugi trend je uporaba ležečega tiska, ki je v prevodnih podnaslovih priporočen za zapis »manj znanih tujih besed« (Društvo slovenskih filmskih in televizijskih prevajalcev 2020: 5). Vsekakor so to popačenke (*suma sumater*) in besede, ki za lažje razumevanje potrebujejo dodaten signal, da jih je treba brati kot tujejezične, npr.: »Dobila sem *tip*,[8] zato snemam klip« (MT, 25. 10. 2025). Srbizmi in anglicizmi iz zgornjih zgledov pa kažejo, da je raba ležečega tiska močno odvisna tudi od stopnje slogovne zaznamovanosti celotnega besedila. Tako je kurziva redkejša v verznih besedilih (npr. *šamar* v verzih (S)Hita tedna (9. 3. 2024) je pisan neležeče), v katerih je zaznamovanih besed praviloma veliko, zapis pa je zaradi ohranjanja metruma pogosto bližji pogovornemu kot ortografskemu, saj bi bila dosledna raba lahko moteča pri branju.

2.2 Oblikoslovno-skladenjska raven

Smernice za podnaslovno prevajanje pri nakazovanju pogovornosti besedila izrecno odsvetujejo ohranjanje pogovornih prvin na neleksikalni ravni: nerabo rodilnika ob zanikanju, zamenjevanje glagola *morati* z *moči*[9] in kratki nedoločnik:

6 V prepisih podnapisov je ohranjen ležeči tisk iz objavljenega izvirnika. Obravnavano značilnost, ki jo zgled ilustrira, pa sem podčrtala.
7 Prim. gesli *flet* in *lajkati* v *Razvezanem jeziku*.
8 Angl. *tip* ʻnapitninaʼ.
9 Ta tip bi se najbrž moral uvrščati na leksikalno raven, vendar smo pri razvrstitvi sledili smernicam Društva slovenskih filmskih in televizijskih prevajalcev, ki pod neleksikalnimi prvinami, s katerimi

»Mater, sploh jo nisem videl« → »Mater, sploh je nisem videl.«
»Si res mogel vse skozlat?« → »Si res moral vse skozlati?« (Društvo slovenskih filmskih in televizijskih prevajalcev 2020: 10)

Analizirali smo še nekatere druge primere odstopanja od slovnične norme, ki jih kot značilne za neknjižne zvrsti jezika opredeljuje *Slovenska slovnica* (Toporišič 2004: 20–21), in še nekatere druge, ki se v govoru oddaje *Kaj dogaja?* pojavljajo najpogosteje; razvrstili smo jih glede na pogostost ohranitve.

Neknjižna oblika dvojine, npr. pri glagolih ženskega spola *sva šle*, je zelo pogosto dopuščena, ob tem so dopuščene tudi oblike, kot je *dve kante*, ki so značilne za nekatera narečja. Stoodstotno je dopuščena v besedilih pesmi, in to ne le na koncu verzov, kjer bi popravek vplival na rimo.

Opuščanje rodilnika pri zanikanju, npr.: *Z veseljem vam povemo, kar lahko in kar ne smemo* (uvodni song, 28. 2. 2026), je nekoliko redkeje ohranjeno v proznih kot v verznih besedilih, a še vedno zelo pogosto.

Pogovorno sklanjanje imen, kot je rod. *Mihata*, je obravnavano zelo različno. V verznih besedilih je zaradi ohranjanja števila zlogov vedno dopuščeno, v proznih besedilih je v prvih sezonah dosledno poknjiženo (rod. *Mihe/Miha*), v poznejših sezonah pa je večkrat ohranjena pogovorna oblika.

Raba glagola *moči* v pomenu *morati* izrazito omahuje. V začetnih sezonah je skoraj vedno poknjižena (razen v pesmih), z leti pa to močno upade.

Kratki nedoločnik se popravlja skoraj brez izjeme, najbolj vidno je to v verznih besedilih pesmi, kjer je celo označen z opuščajem; to pa sodi že na pravopisno raven.

2.3 Pravopisna raven

V osnovi so podnaslovi oddaje *Kaj dogaja?*, kjer je to mogoče, zapisani v skladu s slovenskim pravopisom, pri čemer sledijo splošnim smernicam za podnaslavljanje, ki poudarjajo razumljivost zapisa. Slednje je pri podnaslavljanju posebej pomembno zaradi le nekajsekundne vidnosti podnapisa na posameznem kadru, gledalec pa mora ob prebiranju napisanega tudi spremljati dogajanje na sliki. Nezanemarljiva pri tem ni niti jezikovnovzgojna vloga javnega medija (prim. Radiotelevizija Slovenija 2026: 14). Za stileme v leposlovju npr. Mojca Smolej ugotavlja, da v romanih *Fužinski bluz* Andreja E. Skubica in *Čefurji raus* Gorana Vojnovića »avtor pri zapisu ne zapušča varnih meja razumljivosti in berljivosti. Upošteva le tiste redukcije, ki kljub nestandardnim oblikam bralcu ne otežujejo branja in s tem razumevanja (npr. deležnik na -l za moško obliko je (nedosledno) zapisan z u → *pretegnu*, *vidu*, *prefuku*, *privošču*, vendar *slišal*, *prišel*)« (Smolej 2010: 278). Pri oddaji *Kaj dogaja?* v skladu s smernicami za podnaslavljanje slogovna zaznamovanost na pravopisni ravni načeloma ni izražena, razen v posebnih primerih.

---

ni primerno nakazovati pogovornosti prevoda, razumejo tudi pogovorne variante modalnih besed, npr. *si mogel* (‘moral’) in *nisi rabil* (‘ni bilo treba’).

2.3.1 Prozna besedila

Ortografski zapis je v skečih dosledno uporabljan za naslednje kategorije:[10]

- **narečne, slengovske in pogovorne besede**, katerih zapis je normiran v (pravopisnem) slovarju: *foter* (zapis poglasnika z e), *lavfa* (zapis dvoustničnega v);[11]
- **moški deležnik na -l**, ki je izgovarjan u-jevsko: *videl*, *prišel*, *izvedel*, *ukvarjal*, *ponorel*, *zatrokiral*;
- **vokalna redukcija**: *samo malo* [sam mau], *danes* [dans], *pravega* [pravga];
- **glasovne posebnosti** pogovornega jezika in narečij: *zdaj* [zdej]; *gremo* [goriško: ɣremo], *bova* [štajersko ali koroško: boma], *Ljubljana* [lublana];
- **podomačeni anglicizmi**: *singel*[12] (angl. *single* [singl]).

Izjemoma zapis nekaterih narečnih, slengovskih in pogovornih besed odstopa od ortografskega (npr. *laufati*), pojavnost takšnega pogovornega zapisa v prozi je zelo redka, tako da je mogoče sklepati, da gre za napake. Dilemo predstavlja zapis narečnih izrazov, ki niso vključeni v slovarje ali pa se v različnih slovarjih in rabi pojavlja več različic: *raufenk*/*ravfank*/*raufank*/*raufang*/*raufnk* ter *oreng*/*orenk*/*ornk*/*orng*. Čeprav zapis sicer večinoma teži k ortografiji (zapisovanje polglasnika z -e in dvoglasniškega u z v), v posameznih tovrstnih primerih narečnih besed daje prednost ustaljeni rabi, če je ta drugačna (npr. *ornk*).[13] Naslanjanje na prevladujočo rabo, kot jo izkazujejo zapisi na družbenih omrežjih, je pogostejše pri nekaterih podomačenih anglicizmih, npr. *teribl* (namesto ortografskega *teribel* po zgledu *singel*).

2.3.2 Verzna besedila

Ortografski zapis je ohranjen v vseh zgoraj naštetih kategorijah, kadar ne vpliva na ritem in rimo:

(6) PG: Sekiro stisno sm v pest, mogoče bo tut kej za jest …
Pdn: Sekiro stisnil sem v pest, mogoče bo tud' kaj za jest' … (MT, 11. 5. 2019)

Kot je razvidno iz zgornjega zgleda, je redukcija zaradi ritma v zapisu ohranjena, vendar zaznamovana z opuščaji: *tud'*, *jest'*, kar se zdi skladno težnjo po čim večjem približku knjižnemu zapisu celotnega verznega besedila. Zanimiva je dosledna odločitev za zgolj nakazano redukcijo, tj. en opuščaj, tudi če manjka več samoglasnikov: *tol'k* ali *tolk'* (namesto *tol'k*'), kar gre pripisati olajševanju branja povprečnemu gledalcu. Posebno dilemo predstavljajo sklopi reduciranih besed oz. premene oblik za prihodnjik, npr. *ne boš*, ki so zaradi ohranjanja števila zlogov (metruma) zapisane v pogovornem zapisu: *navš*/*novš* ali njegovem približku: *nou* za *ne bo*. Težnja po približevanju ortografskemu zapisu se pokaže pri primerih, kjer je v govoru izražena redukcija [sliko mam], v zapisu pa ne (*sliko imam*), saj zaradi zlitja samoglasnikov (elizije) pri govoru

10 V ležečem tisku je zapis iz podnapisa, v oglatem oklepaju prepis govora iz skeča.
11 Prim. *foter*, *lavfar* in *šravfati* v *Slovenskem pravopisu* (2001) ter *lavfati* v *Slovarju bovškega govora* (Ivančič Kutin 2015).
12 Prim. geslo *singel* v SSKJ[2].
13 Denimo: različice, navedene v *Razvezanem jeziku*, so vse brez realiziranega poglasnika: *ornh*, *ornk*, *orng*, *vornk* ali *vornh*.

oz. petju tako poknjiženje zapisa ne vpliva na število zlogov in ritem, vendar je to rabljeno nedosledno (prim. *vse to 'mamo*).

Posebna dilema se pokaže, če pod drobnogled izolirano vzamemo zapis rimanih besed. Pri tem je sicer še vedno prepoznavna tendenca po ortografskem zapisu (in verjetno tudi po enotnosti zapisa z istimi besedami na nerimanih mestih) in zanašanje na zgolj glasovno, ne pa tudi pisno izraženo rimo:

(7) PG: V soboto zdej smo eno uro prej.
Pdn: V soboto zdaj smo eno uro prej.

Vendar raba pri tem opazno omahuje:

(8) GP: Ker jaz sn že ponoro, enga štihno skoro, pol pa je Šiško reko, da je šlo samo za foro.
Pdn: Ker jaz sem že ponorel, en'ga štihnil skoro, pol pa je Šiško rekel, da je šlo samo za foro. (MT, 11. 5. 2019)

Včasih tudi zaradi poudarjanja humornega učinka:

(9) Pdn: Desus me, me Desus je zmučo. (MT 20. 9. 2020)

Zanimale so nas še naslednje slovnične kategorije, ki bi jih morda lahko opredelili kot pogovorne ali narečne oblike na oblikoslovni in skladenjski ravni (in ne kot glasovne redukcije):

- **kratki nedoločnik**: *jest'*, *prijet'* itn. je vedno zapisan z opuščajem;
- **maskulinizacija nevter**, npr. *srečn' nov' let'*: skozi leta raba opuščaja nekoliko peša, kar kaže, da take oblike postajajo dopuščene kot legitimen izraz pogovornosti jezika;
- **pogovorna oblika določnega pridevnika**, npr. *lev pas*, *nek*,[14] prav tako ni nikoli zaznamovana z opuščajem;
- **leksikalizirane pogovorne oblike:** *pol*, *pejt*, *tle*, *druzga*, tj. tiste, katerih poknjiženje bi pomenilo spremembo števila zlogov ali prevelik odmik od izrečene besede (*tle* → *tu*), so ohranjene v pogovornem zapisu.

3 Sklep

Vse večje širjenje pogovornih prvin, značilnih za govorjeni jezik, v pisnega, ki ga opažamo v javni rabi, je opaziti tudi pri spremljanju oddaje *Kaj dogaja?* od začetka njenega predvajanja leta 2019 do danes. Analiza podnaslovljenih delov oddaje je pokazala, da se v njih z leti veča delež pogovornih prvin, predvsem anglicizmov. Potrdila je naši domnevi, da je pogovornost na leksikalni ravni ohranjena in da zapis v največji možni meri sledi pravilom zbornega knjižnega jezika, da pa satirična narava oddaje v nasprotju z drugimi primeri podnaslavljanja slovenskih govorcev na televiziji v nekaterih primerih zahteva ohranjanje sloga v podnapisih tudi na drugih jezikovnih ravneh in s tem različne stopnje odmika od ortografskega zapisa. Izkazalo se je, da je takšnih odmikov več v verznih kot v proznih besedilih, torej je glavni razlog za opuščanje pravopisne norme ohranjanje ritma in rime. Težnja po ortografskem zapisu je namreč

14 Oblika *nek* je uvrščena v SSKJ[2].

v skladu z jezikovno politiko zavoda RTV in upošteva smernice Društva slovenskih filmskih in televizijskih prevajalcev, temelji pa predvsem na vodilu, da so podnapisi ne toliko prostor za izražanje sloga kot v prvi vrsti pomagalo gledalcu, zato sta pri zapisu na prvem mestu informativnost in razumljivost. Slednjo tezo potrjujejo tudi odločitve za označevanje glasovnih redukcij z opuščajem v verznih besedilih, če domnevamo, da naj bi tak zapis olajšal branje podnapisov, in rabo ležečega tiska za manj znane (tuje) besede, ne nazadnje pa jo potrjujejo tudi odmiki od ortografskega k pogovornemu zapisu v primerih, ko je tak zapis bolj uveljavljen v rabi (prevladuje v zapisih na družbenih omrežjih) in za gledalca prepoznaven. Enako številni primeri neenotne rabe, zlasti razlike med proznimi in verznimi besedili, če jih analiziramo v kontekstu besedila, pokažejo, da je pri odločitvi za tak ali drugačen zapis pomembna konotacija, ki jo ta podaja, da so pri tem pomembni dejavniki prepoznavnost, razumljivost in humornost sporočenega. Dopuščanje neknjižnih oblik na oblikoslovno-skladenjski ravni oz. upad poknjiženja nekaterih takih oblik skozi leta kažeta, da je, v nasprotju s smernico za podnaslavljanje slovenskih govorcev RTV in smernicami za podnaslovno prevajanje Društva slovenskih filmskih in televizijskih prevajalcev, ki dopuščajo neknjižne prvine le na leksikalni ravni, pri podnapisih oddaje *Kaj dogaja?* prisotna težnja po ohranjanju pristnosti govorjenega jezika tudi v zapisani obliki. Pri tem se zdi najzanimivejše dosledno poknjiženje kratkega nedoločnika, ne pa tudi maskulinizacije srednjega spola in neknjižne oblike določnega pridevnika. Ob tem približevanje bolj sproščenim načelom zapisovanja govorjenega jezika v podnapisih v primerjavi s smernicami za podnaslovno prevajanje kaže tudi upad ležečega tiska za zaznamovanje tujejezičnih prvin; da neležeče zapisane (ali neprevedene) večkrat ostanejo angleške kot druge tuje besede (hrvaške, srbske, nemške), pa tudi izkazuje, da jezik v podnapisih oddaje *Kaj dogaja?* sledi razvoju sodobne govorjene slovenščine, v kateri se v zadnjih letih krepi vpliv angleščine. Govorjeni jezik se ob izjemno hitrem tehnološkem in družbenem razvoju v zadnjih letih razvija vse hitreje, čemur pripisujemo številne primere omahujoče in skozi leta spreminjajoče se rabe, ki jih je pokazala analiza. Predvidevati je mogoče, da se bo uveljavljanje neknjižnega zapisa ne le na družbenih omrežjih, temveč tudi v oglasnih in drugih javnih besedilih v prihodnje še krepilo – v tem času se npr. na javni televiziji predvaja mladinska serija *Smrdljivc* po istoimenskem romanu – in da bo to nadalje oblikovalo pričakovanja gledalcev glede tega, kako zvesto naj bodo pri pretvarjanju govorjenega jezika v zapisanega vsaj v nekaterih besedilnih zvrsteh ohranjene njegove pogovorne prvine, če že ne z izpuščanjem polglasnika ali rabe kratkega nedoločnika, kot jo recimo v pisnem jeziku uveljavlja tudi Andrej Rozman - Roza, pa vsaj z opuščanjem opuščaja pri zapisovanju kratkega nedoločnika ali redukcij, kot je npr. v imenu mladinskega portala RTV Čist hudo. Vse to morda obenem spodbuja k ponovnemu premisleku in dopolnitvi smernic za podnaslavljanje slovenskih govorcev RTV in smernic za prevodno podnaslavljanje Društva slovenskih filmskih in televizijskih prevajalcev, ki sicer omenjajo ohranjanje slogovne plati in čustvenega naboja povedanega, vendar ju v zapisu izrecno omejujejo le na leksikalno raven. Ali bodo podnapisi satirične oddaje *Kaj dogaja?* denimo z leti postali dovoljšen korpus zapisane govorjene slovenščine, da bi se

lahko v njih uveljavljena načela zapisovanja prenesla v standarde prevajanja zvrstno zaznamovanega jezika filmov in televizijskih serij, s čimer bi se povečal prostor za razvoj neknjižnih zvrsti jezika tudi v njegovi pisni podobi?

Viri in literatura

Društvo slovenskih filmskih in televizijskih prevajalcev. (2020, oktober). *Smernice za podnaslavljanje v slovenščini*. http://www.dsftp.si/p/smerniceguidelines.html, pridobljeno 30. 5. 2026.

Ivančič Kutin, B. (2015). *Slovar bovškega govora*. Založba ZRC, ZRC SAZU. www.fran.si

Kalin Golob, M. (2013.) Lačen si ful drugačen: V iskanju naslovnikovega jezika. V A. Žele (ur.), *Družbena funkcijskost jezika (vidiki, merila, opredelitve)* (Obdobja 32, str. 201–206). Znanstvena založba Filozofske fakultete. https://centerslo.si/wp-content/uploads/2015/10/32-KalinGolob.pdf

Radiotelevizija Slovenija. (2019, september). *Jezikovna smernica glede podnaslavljanja anonimnih slovenskih govorcev s tonsko spremenjenim glasom v televizijskih programih* [Interni dokument].

Radiotelevizija Slovenija. (2020, september). *Navodila za oblikovanje podnapisov* [Interni dokument].

Radiotelevizija Slovenija. (2026). *Etični in poklicni kodeks RTV Slovenija: Smernice za odgovorno ravnanje in priročnik za pripravo kakovostnih programskih vsebin.* https://img.rtvslo.si/_files/2026/04/23/46_616416578061205526_informator-26-272.pdf, pridobljeno 30. 5. 2026.

Razvezani jezik: prosti slovar žive slovenščine. https://razvezanijezik.org/

Slovar slovenskega knjižnega jezika: Druga, dopolnjena in deloma prenovljena izdaja. (2014). Založba ZRC, ZRC SAZU. www.fran.si

Smolej, M. (2010). Prvine govorjenega jezika kot temeljni stilemi v sodobni slovenski prozi. V A. Zupan Sosič (ur.), *Sodobna slovenska književnost (1980–2010)* (Obdobja 29, str. 275–282). Znanstvena založba Filozofske fakultete. https://repozitorij.uni-lj.si/IzpisGradiva.php?lang=slv&id=150761

Toporišič, J. (2004). *Slovenska slovnica*. Obzorja.

Toporišič, J. (ur.). (2014). *Slovenski pravopis*. Založba ZRC, ZRC SAZU. www.fran.si

Verdonik, D. in Zwitter Vitez, A. (2011). *Slovenski govorni korpus Gos*. Trojina, zavod za uporabno slovenistiko. https://www.researchgate.net/publication/269629628_Slovenski_govorni_korpus_Gos

Zwitter Vitez, A. in Fišer, D. (2015). Elementi interakcije v govorjenih in spletnih besedilih. V D. Fišer (ur.), *Zbornik konference Slovenščina na spletu in v novih medijih* (str. 87–90). Znanstvena založba Filozofske fakultete.

Avtomatska detekcija oblikoskladenjskih lastnosti v korpusu slovenske poezije PoeTree.sl

DOI: 10.4312/Obdobja.45.195-200

Neža Kočnik, Filozofska fakulteta, Univerza v Mariboru, Maribor; neza.kocnik2@um.si
Petr Plecháč, Ústav pro českou literaturu, Akademie věd České republiky, Praga; plechac@ucl.cas.cz

Prispevek predstavlja metode samodejne oblikoskladenjske anotacije slovenskega pesniškega korpusa PoeTree.sl, ki je nastal v okviru mednarodnega projekta PoeTree. Najprej je orisan postopek oblikovanja korpusa, ki trenutno obsega 5587 pesmi 64 pesnic in pesnikov iz obdobja več kot treh stoletij, nato pa je ovrednoteno delovanje orodja UDPipe 2 na pesniških besedilih. V prispevku so popisane in z zgledi iz korpusa ponazorjene tudi najpogostejše vrste napak pri samodejnem označevanju oblikoskladenjskih kategorij, izpostavljene pa so tudi konstrukcijske kategorije, ki bodo verjetno povzročale težave pri razčlenjevanju skladenjskih odvisnosti.

slovenska poezija, korpusno jezikoslovje, računalniška analiza verza, oblikoskladenjsko označevanje, Universal Dependencies

This article presents methods for automatic morphosyntactic annotation of PoeTree.sl, a corpus of Slovenian poetry compiled as part of the international PoeTree project. It first outlines how the corpus was built—it currently comprises 5,587 poems by sixty-four poets spanning more than three centuries—and then evaluates the performance of UDPipe 2 on poetic texts. Drawing on examples from the corpus, the article describes the most frequent errors in automatic tagging of morphosyntactic categories and identifies the constructions that are likely to pose challenges for dependency parsing.

Slovenian poetry, corpus linguistics, computational analysis of verse, morphosyntactic tagging, Universal Dependencies

1 Uvod

Zgodovino slovenskega verza je mogoče opisati skozi pet glavnih verzifikacijskih sistemov, v katerih se je udejanjal (Bjelčevič 1997: 191): melični silabizem ljudske in nabožne lirike, protestantski silabizem 16. stoletja, naglasni verz 17. in 18. stoletja, silabotonični verz 19. stoletja ter prosti verz 20. stoletja. Kljub bogastvu te dediščine ostajajo računalniško podprte metode za analizo verza[1] na obrobju verzološke vede;

1 Z zvezo *računalniško podprte metode za analizo verza* označujeva presečišče kvantitativnih in kvalitativnih metod, s katerimi je mogoče analizirati verz. To vključuje avtomatsko, računalniško kvantitativno analizo in kvalitativno interpretacijo rezultatov.

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

pomembno oviro je vse do nedavnega predstavljala tudi odsotnost obsežnih jezikoslovno označenih korpusov slovenskega pesniškega jezika.

Leta 2024 je bila ta vrzel delno zapolnjena z izdelavo korpusa slovenske poezije v okviru projekta PoeTree, ki trenutno obsega več kot 330.000 pesmi v enajstih jezikih (Plecháč idr. 2024). Korpus PoeTree.sl je dedupliciran, anotiran v skladu s smernicami Universal Dependencies (De Marneffe idr. 2021) in obogaten z metapodatki. V pričujočem prispevku je predstavljen proces samodejnega prepoznavanja oblikoskladenjskih lastnosti, ovrednotena je kakovost besednovrstnega označevanja in lematizacije, izpostavljene pa so tudi težave, ki jih prinaša samodejna obdelava pesniških besedil v slovenščini.

2 Korpus slovenske poezije PoeTree.sl

Korpus PoeTree.sl trenutno obsega 5587 pesmi (142.185 verzov) 64 pesnic in pesnikov, živečih od konca 17. stoletja do danes, pri čemer se večina avtoric in avtorjev umešča na konec 19. in v 20. stoletje. Najstarejši pesnik v korpusu je Ahacij Stržinar (1676–1741), najsodobnejši pa David Bandelj (roj. 1978). Večina besedil izvira s portala Wikivir, kamor so bila postavljena tudi v okviru projekta Slovenska leposlovna klasika, ki ga financira Ministrstvo za kulturo Republike Slovenije.

Korpus zajema dela kanoniziranih slovenskih pesnikov (Franceta Prešerna, Otona Župančiča, Srečka Kosovela, Tomaža Šalamuna idr.) in tudi manj znanih avtoric in avtorjev. Takšna sestava omogoča diahrone študije verznih oblik, leksike in sloga v slovenski poeziji ter primerjalne raziskave med jeziki, vključenimi v projekt PoeTree. V postopku deduplikacije je bilo izločenih 521 dvojničnih besedil; pesnice in pesniki so bili povezani z ustreznimi identifikatorji s portala Wikidata, kar je omogočilo pridobivanje dodatnih informacij, kot so spol, letnice rojstva in smrti ipd.

2.1 Prilagoditev zapisa: bohoričica

Nekatera, sicer redka, besedila so bila zapisana v bohoričici. Za njihovo prilagoditev sodobnemu slovenskemu črkopisu je bil uporabljen preprost pristop, temelječ na pravilih zamenjave, ki pa je zaradi nefonemske narave bohoričice, po kateri lahko isti grafem predstavlja različna glasova, ter zaradi velikih razlik in nedoslednosti med posameznimi avtorji zahteval številne ročne popravke, in sicer zlasti pri grafemu <Sh>, ki lahko ustreza glasovoma [š] ali [ž], ter grafemu <S>, ki lahko ustreza [s] ali [z].

3 Samodejno prepoznavanje oblikoskladenjskih lastnosti

Anotacija oblikoskladenjskih lastnosti je bila izvedena z orodjem UDPipe 2 (Straka 2018), in sicer z modelom, naučenim na korpusu SSJ Treebank (Dobrovoljc, Ljubešić 2022). Universal Dependencies (UD) predstavljajo enotno ogrodje za jezikoslovne anotacije, primerljive med jeziki; te zajemajo besednovrstne oznake, oblikoslovne značilnosti, leme in skladenjske odvisnosti (De Marneffe idr. 2021).

Model UDPipe, naučen na korpusu SSJ, ni prilagojen posebnostim pesniškega jezika, kot so neobičajna skladenjska struktura, vpliv metra na besedni red ipd. Izhodiščna

predpostavka je bila torej, da bo njegova uspešnost pri pesniških besedilih nižja kot pri nepesniških. To hipotezo so za češko poezijo potrdili Cinková idr. (2024), za nemško pa Dipper in Laarmann-Quante (2024).

3.1 Besednovrstne kategorije in lematizacija

Ker sta lematizacija in besednovrstno označevanje nedvomno najpomembnejši ravni anotacije za številne računalniško podprte analize verza, je v nadaljevanju predstavljena ocena uspešnosti teh dveh procesov. Za testiranje modela je bil uporabljen vzorec 57 ročno anotiranih verzov (1014 pojavnic). Natančnost besednovrstnega označevanja znaša 0,9694 (95-odstotni Wilsonov interval zaupanja: 0,9569–0,9784), natančnost lematizacije pa 0,9576 (95-odstotni Wilsonov interval: 0,9434–0,9684). Opazno je bilo sorazmerno nizko zmanjšanje v primerjavi z rezultati na testnem naboru (besednovrstno označevanje: 0,9885; lematizacija: 0,9869), dosežena natančnost pa tako ostaja dovolj visoka za uporabo v različnih računalniško podprtih analizah poezije.

Za oceno pogostosti napak pri uvrščanju v besednovrstne kategorije je bila izračunana matrika zmede (angl. *confusion matrix*), ki ponazarja razpored napak pri besednovrstnem označevanju z UDPipe v primerjavi z ročno označenim zlatim standardom (Tabela 1). Z izjemo medmetov (INTJ; n = 3; statistično nepomembno) sta najpogostejši napaki pri pridevnikih (ADJ) s priklicem 0,9054 in samostalnikih (NOUN) s priklicem 0,9259. Pri preostalih besednih vrstah priklic presega 0,975.

	ADJ	**ADP**	**ADV**	**AUX**	**CCONJ**	**DET**	**INTJ**	**NOUN**	**NUM**	**PART**	**PRON**	**PROPN**	**PUNCT**	**SCONJ**	**VERB**
ADJ	67		2					2							3
ADP		79						1							
ADV			51												1
AUX				52				1							
CCONJ					31										
DET						47									
INTJ							2	1							
NOUN	5		1					200			1	2			7
NUM									3						
PART										28					
PRON	1										69				
PROPN												8			
PUNCT													182		
SCONJ														38	
VERB	1							2							126

Tabela 1: Matrika zmede za besednovrstne kategorije (ADJ – pridevnik, ADP – predlog, ADV – prislov, AUX – pomožni glagol, CCONJ – priredni veznik, DET – določnik, INTJ – medmet, NOUN – samostalnik, NUM – številnik, PART – členek, PRON – zaimek, PROPN – lastno ime, PUNCT – ločilo, SCONJ – podredni veznik, VERB – glagol).

Za zdaj se poraja vtis, da večina zaznanih napak izvira iz arhaičnih konstrukcij, ritmično pogojenega besednega reda ali medbesednovrstne homonimije, natančneje homografije. V nadaljevanju so navedeni značilni primeri:

1) *Spet iščem v spominov gošči gosti.* (Jože Jensterle: *Odsevi stopinj*)

Pridevnik *gosti* se pojavi v desni stavi, tj. za samostalnikom *gošča*; v mestniku ednine ženskega spola je hkrati homografen množinski obliki samostalnika *gost* (analogno tvorjeni po o-sklanjatvi), ki se sicer izvorno sklanja po moški i-sklanjatvi.

2) *Po vsem slovenskem sveti.* (Fran Levstik: *Jedína mej vrstnicami*)

Starejša končnica mestnika samostalnika *svet* (tj. pred vplivom končnic u-sklanjatve) je homografna določni obliki pridevnika *svet*.

3) *Da svitlo je nebo.* (Anton Umek Okiški: *Abuna Soliman*)

Samodejni prepoznavi se izmuznejo tudi arhaične pisne oblike, npr. *svitlo* (i na mestu jata) namesto *svetlo*.

4) *'Z ur so mesci, 'z dní so leta, / Mescev, let pa konca ní!* (Simon Jenko: *Obujénke VI.*)

Občni samostalnik *mesec* je bil najverjetneje zaradi zapisa z veliko začetnico (kot prva beseda v verzu, a skladenjsko neutemeljeno) napačno prepoznan kot lastno ime (napaka tipa NOUN → PROPN).

Z oznako X, ki se v Universal Dependencies uporablja za primere, v katerih besednih oblik ni mogoče zanesljivo opredeliti, je bilo označenih le 0,17 % od približno 875.000 pojavnic, pri čemer gre v večini za tuje besede ali okrajšave.

3.2 Skladenjske odvisnosti

Ovrednotenje natančnosti razčlenjevanja skladenjskih odvisnosti je v primerjavi z vrednotenjem oblikoslovnega označevanja in lematizacije precej zamudnejši postopek, saj zahteva preverjanje vsake pojavnice in njej nadrejenega elementa, torej njune odvisnosti. Razprava v pričujočem prispevku je zato omejena na opredelitev konstrukcijskih kategorij, za katere se na podlagi razpoložljive literature in dosedanjih kvalitativnih opažanj pričakuje, da bodo pri razčlenjevanju odvisnosti predstavljale največje izzive.

Rezultati raziskave, ki so jo na korpusu češke poezije opravili Cinková idr. (2024), kažejo, da razčlenjevalniki odvisnosti, naučeni na sodobnih nepesniških besedilih, na pesniških besedilih delujejo bistveno slabše, in sicer zaradi posebnosti pesniškega jezika, kot so inverzije, elipse, rima ipd. Na enake težave je za nemško poezijo opozorila raziskava Dipper in Laarmann-Quante (2024), zato je tudi za slovensko poezijo smiselno predvideti, da bodo konvencije pesniškega jezika napovedi modelov »zavajale«.

Konstrukcijske kategorije, za katere se predvideva, da bodo pri samodejni detekciji skladenjskih odvisnosti povzročale težave, so naslednje:

1) postpozicija pridevnika: v zgledu *gošči gosti* stava pridevnika vpliva tako na oblikoslovno kot na skladenjsko označevanje – beseda *gosti* ni prepoznana kot prilastek (amod), temveč kot ločeno samostalniško jedro;

2) rodilniške zveze, v katerih je določilo pred jedrom: razčlenjevalnik ima težave s prepoznavanjem smeri odvisnosti;

3) odsotnost glagola *biti*: vodi v napake po celotni povedkovi konstrukciji;

4) zaznamovana raba ločil ali njihova odsotnost: razčlenjevalnik ostane brez oprijemljivih označevalcev stavčnih meja.

Naštete težave odražajo širše spoznanje, da potrebuje razčlenjevanje skladenjskih odvisnosti, čeprav je pri pesniških besedilih natančnost oblikoslovnega označevanja in lematizacije v primerjavi z nepesniškimi upadla razmeroma malo, za pravilno delovanje posebne učne podatke ali poobdelavo.

4 Kontekstualizacija rezultatov

Dosežene natančnosti (besednovrstno označevanje: 0,9694; lematizacija: 0,9576) kažejo, da je PoeTree.sl primerljiv z drugimi podkorpusi projekta PoeTree (prim. Cinková idr. 2024; Dipper in Laarmann-Quante 2024). Razmeroma zmerni obseg upada kaže, da kljub posebnim strukturnim elementom pesniških besedil orodja za obdelavo naravnega jezika pri njihovi analizi ne postanejo neučinkovita.

Opozoriti velja še na nenavadno majhen delež oznak v kategoriji X v PoeTree.sl (0,17 % v primerjavi z 0,48 % v zbirki SSJ). Pojav je mogoče pojasniti z dvema dejavnikoma, in sicer 1) korpus je bil ročno pregledan in pravopisno standardiziran, kar je izključilo številne primere tujejezičnih izposojenk in okrajšav, ki so v sodobnih podatkih značilni za kategorijo X (Kočnik in Plecháč 2024, v tisku); 2) slovenska poezija, zlasti pesniška dela iz 19. stoletja, je v leksikografskem smislu najverjetneje dokaj podobna tipu jezika, uporabljenemu v zbirki SSJ, čeprav je sam jezik močno arhaičen.

5 Sklep in nadaljnje raziskave

PoeTree.sl je prvi obsežni jezikoslovno označeni korpus slovenske poezije, ustvarjen v okviru projekta PoeTree, ki je usmerjen v računalniško podprto analizo verza. Vključuje slovensko poezijo, nastalo v več kot treh stoletjih, in zajema tako uveljavljene kot manj znane pesnice in pesnike. Oblikoslovno označevanje in lematizacija, samodejno izvedena z orodjem UDPipe 2, dosegata natančnost 0,9694 oz. 0,9576. Čeprav sta ti oceni nekoliko nižji od tistih, doseženih pri obdelavi nepesniškega jezika, sta dovolj visoki za uporabo nastalih anotacij v širokem naboru računalniško podprtih študij poezije. Nekatere napake izvirajo iz posledic zgodovinskih sprememb v slovenskem (tudi knjižnem) jeziku in iz značilnosti, po katerih se slovenski pesniški jezik razlikuje od nepesniškega; ravno te napake pa je mogoče uporabiti za nadaljnje raziskave.

Razčlenjevanje skladenjskih odvisnosti zahteva več pozornosti in bo podrobneje predstavljeno na drugih mestih. V prihodnje bo korpus razširjen, anotacijske ravni pa dopolnjene s prozodičnimi prvinami ritmično-metričnega ustroja pesmi in posebej z rimo (z orodjem RhymeTagger; prim. Plecháč 2018; Plecháč idr. 2026).

Prispevek je nastal v okviru Raziskovalnega programa št. P6-0156 (Slovensko jezikoslovje, književnost in poučevanje slovenščine (2020–2027); vodja programa prof. dr. Marko Jesenšek), ki ga sofinancira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije iz državnega proračuna.

Viri in literatura

Bjelčevič, A. (1997). Zgodovina slovenskega verza – skica. *Slavistična revija*, *45*(1–2), 191–201.

Cinková, S., Plecháč, P. in Popel, M. (2024). Rhymes and syntax: A morpho-syntactic analysis of Czech poetry. *Primerjalna književnost*, *47*(2), 65–88.

De Marneffe, M.-C., Manning, C. D., Nivre, J. in Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, *47*(2), 255–308.

Dipper, S. in Laarmann-Quante, R. (2024). UD for German poetry. V M. Hämäläinen, E. Öhman, S. Miyagawa, K. Alnajjar in Y. Bizzoni (ur.), *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities* (str. 177–188). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.nlp4dh-1.17

Dobrovoljc, K. in Ljubešić, N. (2022). Extending the SSJ Universal Dependencies Treebank for Slovenian: Was it worth it? V S. Pradhan in S. Kuebler (ur.), *Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022* (str. 15–22). European Language Resources Association. https://aclanthology.org/2022.law-1.3.pdf

Kočnik, N. in Plecháč, P. (2024, v tisku). Corpus of Slovenian verse. *Metres of Humour: Plotting Poetry*, *7*.

Novak, B. A. (1995). *Oblika, ljubezen jezika: Recepcija romanskih pesniških oblik v slovenski poeziji*. Obzorja.

Novak, B. A. (2015). Verzifikacijske razlike v poeziji južnoslovanskih narodov na širšem mednarodnem ozadju. *Filološke študije*, *13*(1), 105–129.

Plecháč, P. (2018). A collocation-driven method of discovering rhymes (in Czech, English, and French poetry). V M. Fidler in V. Cvrček (ur.), *Taming the corpus: From inflection and lexis to interpretation* (str. 79–95). Springer.

Plecháč, P., Cinková, S., Kolár, R., Šeļa, A., De Sisto, M., Nugues, L., Haider, T. in Kočnik, N. (2024). PoeTree: Poetry treebanks in Czech, English, French, German, Hungarian, Italian, Portuguese, Russian, Slovenian and Spanish. *Research Data Journal for the Humanities and Social Sciences*, *9*, 1–17. https://doi.org/10.1163/24523666-bja10044

Plecháč, P., Kolár, R., Cinková, S., Šeļa, A., De Sisto, M., Nugues, L., Haider, T., Kočnik, N., Nagy, B., Delente, É., Renault, R., Bobenhausen, K., Hammerich, B., Mittmann, A., Palkó, G., Horváth, P., Navarro Colorado, B., Ruiz Fabo, P., Bermúdez Sabel, H., Korchagin, K., Plungian, V. in Sitchinava, D. (2023). *PoeTree: Poetry treebanks in Czech, English, French, German, Hungarian, Italian, Portuguese, Russian, Slovenian, and Spanish* (Version 0.0.2) [Zbirka podatkov]. Zenodo. https://doi.org/10.5281/zenodo.10008458

Plecháč, P., Šeļa, A., Cinková, S., De Sisto, M., Nugues, L., Kočnik, N., Martynenko, A., Nagy, B., Giovannini, L. in Kolár, R. (2026). Training data size sensitivity in unsupervised rhyme recognition. *arXiv*. https://arxiv.org/abs/2604.08156

Straka, M. (2018). UDPipe 2.0 prototype at CoNLL 2018 UD shared task. V D. Zeman in J. Hajič (ur.), *Proceedings of the CoNLL 2018 Shared Task: Multilingual parsing from raw text to Universal Dependencies* (str. 197–207). Association for Computational Linguistics. https://doi.org/10.18653/v1/K18-2020

ÚFAL. (2026). *UDPipe 2 models*. Institute of Formal and Applied Linguistics, Charles University.

Primerjalna semantično-pragmatična analiza intenzitetnih označevalcev v avtorskih in umetno generiranih presojevalnih besedilih

DOI: 10.4312/Obdobja.45.201-210

Diana Košir, Inštitut za jezikoslovne študije, Znanstveno-raziskovalno središče Koper, Koper;
diana.kosir@zrs-kp.si

V prispevku je na podlagi avtorskih in umetno generiranih presojevalnih besedil opazovana raba besednih intenzitetnih označevalcev, tj. leksike, ki v okviru danega konteksta deluje v smeri krepitve oz. šibitve sporočila. Primerjalna analiza je zajela vrsto in pogostost rabe najpogostejših krepilcev in šibilcev. Namen raziskave je bil ugotavljati potencial intenzitetnih označevalcev kot kazalnikov pragmatične ustreznosti jezikovnega izraza umetno generiranih besedil z modeloma ChatGPT in GaMS v okviru izbranega besedilnega žanra.

veliki jezikovni modeli, kolumna in blog, intenziteta jezika, pragmatika in semantika, korpusna analiza

Based on original and artificially generated evaluative texts, this article examines the use of lexical intensity markers—that is, linguistic means that, within a given context, serve to strengthen or weaken a message. The comparative analysis examines the types and frequency of use of the most common intensifiers and attenuators. The aim of the study was to determine the potential of intensity markers as indicators of the pragmatic appropriateness of linguistic expression in texts generated by the ChatGPT and GaMS models within the framework of a selected text genre.

large language models, column and blog, language intensity, pragmatics and semantics, corpus analysis

1 Uvod

Generativna umetna inteligenca (UI), ki temelji na uporabi podatkov velikih jezikovnih modelov (VJM) in je integrirana v različna digitalna orodja za splošno in specializirano rabo, z visoko hitrostjo spreminja družbeno realnost in naravo tvorbe besedil. Kljub znatnemu napredku na področju statističnega modeliranja in globokega učenja ostaja osrednji izziv pri obdelavi naravnega jezika razvoj jezikovno kompetentnih sistemov UI – medtem ko sedanji veliki jezikovni modeli kažejo visoko raven tekočnosti in jezikovne (slovnične) ustreznosti, ostajajo vrzeli na globljih ravneh jezikovne kompetence, zlasti pri pragmatiki, semantiki in razumevanju diskurza (Ahmad 2025: 2202).

V slovenskem prostoru po razširjenosti uporabe besedilnih orodij generativne UI prevladuje jezikovni model ameriškega podjetja Open AI ChatGPT (Praček in Vehovar 2025). V letu 2025 je zaživel tudi slovenski jezikovni model GaMS, ki je sicer učen na

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

1000-krat manjšem številu besedil kot ChatGPT, a temelji na izvirnih slovenskih besedilih, kar naj bi omogočilo generiranje jezikovno kakovostnejših in kulturno specifično bolj občutljivih besedil v slovenščini. V prispevku je predstavljena primerjava avtorskih in umetno generiranih presojevalnih besedil s pomočjo jezikovnih modelov ChatGPT-5 (brezplačna verzija) in GaMS-12B-Instruct, pri čemer nas bo zanimal vidik rabe izbranih jezikovnih prvin (prislovov in členkov), ki v besedilu delujejo kot intenzitetni označevalci.

2 Model intenzitete slovenskega jezika po Mikolič (2020)

Vesna Mikolič v monografiji *Izrazi moči slovenskega jezika* (2020) intenziteto jezika razlaga kot enega od vidikov naklonskosti/modalnosti v okviru pragmatičnega jezikoslovja. Avtorica izhaja iz uveljavljenih pragmatičnih teorij in pristopov (Austin, Searle, Brown-Levinson, Leech, Grice idr.) ter pravi, da ima vsako govorno dejanje poleg spoznavne vloge (lokucija), izražene skozi propozicijo, tudi vplivanjsko (ilokucijsko) vlogo, ki se realizira skozi modalno strukturo propozicije in je odvisna od sporočevalčevega namena in drugih sporazumevalnih okoliščin (žanr, besedilna vrsta, poznavanje teme, odnos tvorca do naslovnika in teme idr.) (Mikolič 2020: 76–77). Pri intenziteti jezika gre torej za naklonsko modifikacijo pomena v smeri okrepitve ali ošibitve tako na ravni učinka sporočila kot odnosa do naslovnika. Z vidika odnosa tvorca do naslovnika je zanimiva tudi Hylandova klasifikacija metabesedila, v okviru katere najdemo tudi interakcijske kategorije, povezane z intenziteto jezika (Hyland 2005, v Pisanski Peterlin 2017: 17). V naši raziskavi se ne ukvarjamo toliko z učinkom na naslovnika in odnosom do njega, temveč se posvečamo predvsem analizi pogostosti izbranih označevalcev intenzitete (angl. *intensifiers*), tj. izrazov, ki pomensko okrepijo oz. ošibijo sporočilo ali odražajo tvorčev subjektivni odnos do stvarnosti.

Mikolič loči dve vrsti intenzitetnih označevalcev, t. i. krepilce (npr. prislov *zelo* ali pridevniški presežniki) in šibilce (npr. *nekaj*, *včasih*), in predlaga model za njihovo analizo po posameznih jezikovnih ravneh: glasoslovna, pravopisna, oblikoslovna, skladenjska, pomenoslovna, besedoslovna, besediloslovna, pragmatična in diskurzivna raven (Mikolič 2020: 56). Poudarja, da pogosto smer modifikacije (krepitev ali šibitev) prepoznamo šele iz konteksta (npr. samostalniki ženskega spola z obrazilom *-ica* so kot manjšalnice v vlogi šibitve, kot ljubkovalnice pa v vlogi krepitve) (Mikolič 2020: 55, 94).

3 Presojevalna besedila in intenziteta jezika

Za opazovanje rabe intenzitetnih označevalcev so najprimernejše besedilne vrste, pri katerih pridejo do izraza avtorjevo osebno stališče, vrednotenje in pogledi. Z ozirom na to sta bili izbrani besedilni vrsti kolumna in blog – pri obeh gre za piščevo subjektivno vrednotenje nekega dogodka/dogajanja/pojava, pri čemer gre v prvem primeru za uveljavljen novinarski žanr (časopisne kolumne v tiskanih ali spletnih medijih, tudi radijske in televizijske kolumne), v drugem pa za sodobno spletno besedilno vrsto, kjer gre prej kot za samostojen žanr za množico hibridnih žanrskih praks, ki so alternativa informativnim in interpretativnim novinarskim žanrom (Vrečar 2009).

V stilistiki slovenskih novinarskih besedil se je za poročevalska besedila publicistične funkcijske zvrsti, v katerih prevladuje interpretativna funkcija, uveljavil izraz presojevalna besedila (Kalin Golob 1999: 185–188), ki ga kot krovno opredelitev analiziranih besedil uporabljamo v tem prispevku. Z vidika teorije novinarskih žanrov pa govorimo o interpretativnih besedilih (Košir 1988) oz. tistih, ki vzpostavljajo stik z naslovnikom z vplivanjskimi, apelnimi in vrednotilnimi sredstvi (Korošec 1998: 12).

Manca Košir (1988: 85–86) kolumno umešča med novinarska besedila interpretativnega komentatorskega značaja, s poudarjenim subjektivnim odzivom na aktualno dogajanje ali izbrano družbeno relevantno temo in avtentičnim jezikovnim stilom. Objave kolumn izkušenih piscev so v domeni medijskih uredništev (kolumne v *Delu* objavljajo npr. Miha Šalehar, Kozma Ahačič, Mojca Pišek, Aljoša Kravanja idr.), blogi pa so navadno objavljeni na osebnih spletnih straneh (avtorji blogov so tako amaterski pisci kot že uveljavljeni avtorji in medijsko prepoznavne osebnosti, npr. psiholog Matic Munc, televizijska voditeljica in predavateljica Saša Einsiedler, potopisni bloger Zvone Šeruga, kulinarična blogerka Vita Starman idr.).

4 Primerjava avtorskih in umetno generiranih besedil: kaj pričakovati?

V tuji literaturi prevladujejo primerjalne jezikovno-stilistične študije avtorskih in z UI generiranih besedil zlasti na anglosaškem gradivu in s primerjavo globalnih modelov ChatGPT, DeepSeek idr. (pregledno študijo prinaša članek Terčon in Dobrovoljc 2025). Rezultati jezikoslovnih in stilističnih primerjav kažejo, da so izvirna, avtorska besedila na splošno bolj koherentna in leksikalno raznolika, medtem ko se v strojno generiranih besedilih leksika znotraj odstavka pogosto ponavlja in je nasploh manj raznolika (Guo idr. 2023; Terčon in Dobrovoljc 2025; Zhang idr. 2024), prisoten je neobičajen in privzdignjen (»pompozen«) besednjak (Mitrović idr. 2023). Besedila, ki jih ustvarjajo VJM, so na splošno manj čustveno obarvana ter vsebujejo bolj vljuden, objektiven, neoseben in formalen slog brez konkretnih podrobnosti (Mitrović idr. 2023), kar kaže v angleščini povečana prisotnost samostalnikov, določevalcev (angl. *determiners*), predlogov, manj pa je pridevnikov in prislovov (Terčon in Dobrovoljc 2025). Ugotovljeno je bilo tudi, da VJM lahko posnemajo različne sloge (Guo idr. 2023). Liu idr. (2023) so primerjali avtorske in z UI generirane argumentativne eseje, pri čemer se je pokazala zapletenejša skladnja pri umetno generiranih besedilih in večja leksikalna kompleksnost pri avtorskih besedilih. Prav tako je avtorske in s ChatGPT generirane argumentativne eseje primerjala Shalevska (2024), ki se je ožje usmerila v korpusno analizo petih najpogostejših krepilcev in šibilcev (angl. *boosters*, *hedges*): *clearly*, *definitely*, *certainly*, *claim*/*s*/*ed*/*ing*, *obviously*; *possibly*, *might*, *may*, *could*, *seem*/*s*/*ed*. Ugotovila je različno pogostost nekaterih šibilcev za avtorska in UI besedila (pri generiranih besedilih je prednjačila raba modalnega glagola *may*, v avtorskih pa *could*) ter odsotnost nekaterih krepilcev pri UI, ki so se pojavili v avtorskih besedilih (*clearly*, *definitely*, *certainly*), kar je skladno z avtoričino ugotovitvijo o večji leksikalni pestrosti obojih vrst intenzitetnih označevalcev v avtorskih besedilih.

Za slovenski jezik tovrstna primerjalna analiza še ni bila izvedena, zato nas zanima, ali se bodo naše ugotovitve skladale s prej predstavljenimi.

V analizi primerjamo rabo intenzitetnih označevalcev v avtorskih besedilih in besedilih, generiranih z jezikovnima modeloma ChatGPT in GaMS. Namen primerjalne korpusne analize (sicer manjšega delovnega korpusa) je ugotoviti, ali je slovenski jezikovni model glede na raznolikost in semantično ustreznost rabe intenzitetnih označevalcev (že) bliže slovenskim avtorskim besedilom v primerjavi s priljubljenim modelom ChatGPT, kar bi lahko bil eden od kazalnikov jezikovne kompetence.

5 Metodologija

Metodološko leksikalna analiza temelji na predstavljenem modelu za analizo jezikovne intenzitete slovenskega jezika (Mikolič 2020), pri interpretaciji pomenske vrednosti označevalcev intenzitete pa smo se oprli na slovarske opise v SSKJ[2] in *Slovarju slovenskih členkov* (Žele 2015). V analizi smo se omejili na najpogostejše prislove in členke v tej vlogi med prvimi 50 najpogostejšimi prislovi oz. členki na frekvenčni listi korpusnih lem (gl. Košir 2026). Korpusna analiza je bila opravljena z orodjem Sketch Engine.

Za potrebe raziskave je nastal manjši delovni korpus Korpus avtorskih in umetno generiranih presojevalnih besedil, ki vsebuje 36 dokumentov oz. 28.908 besed. Najprej je bilo izbranih 12 avtorskih kolumn oz. blogov, po šest na sorodno temo, in sicer 1) kakovost življenja v sodobni družbi in 2) svoboda govora. Besedila obsegajo med 600 in 1000 besed, gre pa za kolumne avtorjev Kozme Ahačiča, Grege Repovža, Petre Božič Blagajac, Andreja Velkavrha, Matjaža Figlja, Luke Martina Tomažiča, Marka Crnkoviča in Tomaža Štiha ter dva bloga Matica Munca (podrobno v Košir 2026). V večini besedil se pojavljajo reference na slovenski družbeno-kulturni kontekst (npr. »slovenska javnost«, »poročilo o razvoju in kakovosti življenja v Sloveniji«, »plače zaposlenih v Sloveniji«, »Slovenija, moja dežela«, »Kaj početi s slovenščino v odraslem življenju«, »Zakon o javni rabi slovenščine«, »Če vprašate povprečno Slovenko ali Slovenca«, »Nič od tega se ne bi zgodilo, če bi slovenska levica odločala o proračunu«), tri besedila pa so napisana dovolj globalno, da bi jih lahko objavili tudi tuji mediji (npr. »Vsi smo soodgovorni, da se je neprijazno sporazumevanje razširilo in v naši družbi postalo normala.«, eno besedilo ima zgolj eno posredno referenco: »V anketi Kako dobri so naši dnevi pri delu, ki smo jo Ambasadorji delovne sreče izvedli lansko jesen«, tretje pa je napisano v splošnem duhu, šele v zadnjem odstavku usmeri na nacionalni kontekst: »Zdi se, da smo se v Sloveniji glede razumevanja timskega dela«).

Kolumnisti svoj (kritični) odnos do ubesedenega (in naslovnika) na nadbesedni ravni pogosto kažejo skozi humor, ki se izraža z izrazito metaforiko, ironijo in satiro (Šegula 2016: 45, 68), čeravno je humoren stil pisanja prej kot za kolumno značilen za gloso, ki je samostojen novinarski žanr. V našem izboru je humor prisoten v dobri polovici besedil, npr. z ironijo (»V teh dneh je slovenska javnost dobila v roke enega najpomembnejših in tudi najboljših dokumentov o stanju družbe in države. A žal javnost o tem ‘veselem dogodku’ ve zelo malo.«), satiro (»To so naši voditelji. Njih volimo.

Včasih se malo prebarvajo, zamenjajo ekipo, malo izginejo pa se spet pojavijo. Iste face že tri desetletja, za njimi pa njihovi nasledniki in mlajši kloni.«) ali cinizmom (»Še niste odložili časopisa ali zaprli računalnika? Ste odstavek prebrali s težavo? Če ste zaradi zapisanega bruhali, se vam opravičujem«).

Sledilo je generiranje besedil z UI (ChatGPT, GaMS), upoštevaje ugotovljene lastnosti avtorskih besedil (tj. žanr, obseg, tema, stil). V obdobju april–maj 2026 je bilo skupno generiranih 24 besedil; na isto temo je bilo z vsakim modelom generiranih šest besedil, z vsakim pozivom po dve (gl. Tabelo 1). Rezultat s pozivi A so bila besedila brez humornih insertov in družbeno-kulturnih referenc, zato sta bili vpeljani še variaciji poziva B in C. Različni pozivi (angl. *prompts*) in generirana besedila so navedeni v raziskovalnih podatkih (Košir 2026).

Vir	Tema	Poziv	Število ponovitev	Skupaj besedil (N)
Jezikovni model (ChatGPT, GaMS)	Tema 1: *Kakovost življenja v sodobni družbi*	Poziv A – brez usmeritve o slogu	2	24
		Poziv B – kulturno občutljiv		
		Poziv C – humorno obarvan slog		
	Tema 2: *Svoboda govora na družbenih omrežjih*	Poziv A		
		Poziv B		
		Poziv C		
Spletni/ tiskani medij	Tema 1	–	–	12
	Tema 2			

Tabela 1: Struktura delovnega korpusa.

6 Rezultati analize

Tabela 2 prikazuje izluščene najpogostejše prislove in členke, ki v besedilih nastopajo v vlogi intenzitetnih označevalcev. Izpisane so njihove frekvence pojavitev v korpusu glede na primerjane »podkorpuse« z vidika avtorstva besedil.

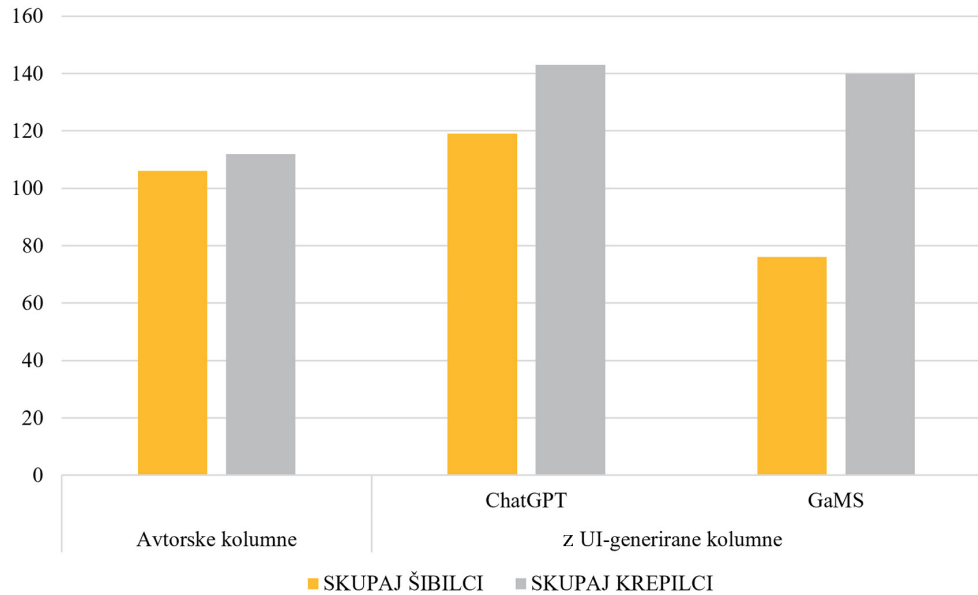
Intenzitetni označevalec (–/+)	Absolutna frekvenca (N)	Relativna frekvenca DOCF[1] (%)	Avtorske kolumne	UI generirane kolumne	
				ChatGPT	GaMS
lahko	199	94,44	73	60	66
včasih	43	41,67	**19**	17	7
morda	37	50,00	4	**30**	3
verjetno	14	25,00	5	**9**	0
mogoče	9	11,11	6	3	0
pač	6	8,33	**5**	1	0
najbrž	2	2,78	**2**	0	0
še	130	88,89	**61**	31	38
pogosto	103	72,22	13	**57**	33
vedno	71	66,67	17	16	**38**
prav	53	63,89	13	**28**	12
najbolj	41	55,56	10	12	**19**
res	39	55,56	**18**	13	8
nikoli	26	47,22	4	8	6
seveda	23	33,33	**14**	8	0
zelo	18	38,89	**9**	8	1
žal	7	11,11	**7**	0	0
stalno	6	8,33	0	0	**6**

Tabela 2: Nekateri najpogostejši besedni označevalci intenzitete in njihove frekvence v korpusu.

Pomensko leksemi bodisi odražajo različno stopnjo prepričanja/gotovosti v ubesedeno (npr. *vedno*, *nikoli*, *pogosto*; *včasih*, *morda*, *lahko*) bodisi kažejo na subjektivno vrednotenje sporočevalca (npr. *še*, *res*, *seveda*, *žal*; *pač*), s čimer je končni argument okrepljen oz. ošibljen.

S krepkim tiskom so v Tabeli 2 označene vrednosti, ki izstopajo. V besedilih je nasploh več krepilcev kot šibilcev, izmed navedenih najpogostejših šibilcev se dva pojavita v vsaj polovici besedil, med krepilci pa je takšnih šest. Ob tem je bila opravljena primerjava pogostosti rabe za pet najpogostejših krepilcev (*še*, *pogosto*, *vedno*, *prav*, *najbolj*) in šibilcev (*lahko*, *včasih*, *morda*, *verjetno*, *mogoče*). Rezultat je prikazan na Grafu 1.

1 Odstotek dokumentov, v katerih se lema pojavi.



Graf 1: Primerjava vsote absolutnih frekvenc za pet najpogostejših šibilcev in krepilcev.

Primerjava podkorpusov glede na kriterij avtorstva je pokazala, da je raba izbranih krepilcev in šibilcev v avtorskih besedilih zelo uravnotežena, medtem ko je v besedilih, generiranih z modelom ChatGPT, tako krepilcev kot šibilcev več (kot v obeh ostalih podkorpusih), znotraj korpusa pa je tudi več izbranih krepilcev. V »neuravnoteženosti« rabe izstopa podkorpus besedil, tvorjenih z GaMS-om, kjer je šibilcev občutno manj kot v ostalih dveh podkorpusih, krepilcev pa skoraj še enkrat več kot šibilcev in več kot v avtorskih kolumnah ter primerljivo s ChatGPT-jem.

Najpogostejši izraz med opazovanimi je v vseh besedilih prislov *lahko*, ki deluje kot šibilec intenzitete, kadar izraža omiljeno nujnost, omejitev oz. omiljeno zapoved. Razporeditev pomensko prekrivnih šibilcev *morda*, *mogoče*, *verjetno* in *najbrž* pokaže, da so v avtorskih besedilih poredko, a uravnoteženo prisotni vsi, pri ChatGPT-ju jih je skupno največ, vidno pa izstopa *morda*, pri GaMS-u pa se nekajkrat pojavi le *morda*. Pri krepilcih *še* in *prav*, ki poudarita in okrepita pomen besede, na katero se nanašata, je opazno, da je členek *še* skoraj dvakrat pogosteje rabljen v avtorskih besedilih, *prav* pa več kot dvakrat pogosteje v besedilih ChatGPT. Frekvenčna prislova *pogosto* in *vedno*, ki izražata visoko oz. skrajno mero, se pojavita v vseh besedilih, s tem da prvi izstopajoče pri ChatGPT-ju, drugi, absolutni, pa pri GaMS-u. Prav tako se presežnik *najbolj* najpogosteje pojavi pri GaMS-u.

Leksikalna raznolikost je večja v avtorskih kolumnah (bolj uravnotežena raba izrazov), pri tem razlikovalno izstopajo izrazi, ki so čustveno obarvani, razkrivajo subjektivno vrednotenje tvorca sporočila in so značilni za mnenjski diskurz (*žal*, *res*, *seveda*, *najbrž*, *pač*) – ti se pojavijo predvsem ali samo v avtorskih tekstih. Ugotovitev o večji leksikalni pestrosti avtorskih besedil in po drugi strani visoki pogostosti istih izrazov v besedilih UI, v našem primeru je šlo za intenzitetne označevalce, sovpada z ugotovitvami

uvodoma omenjenih študij (prim. Mitrović idr. 2023; Shalevska 2024; Terčon in Dobrovoljc 2025).

Nato nas je zanimala še frekvenca uporabe izbranih najpogostejših izrazov glede na različen uporabljeni poziv (A/B/C). Rezultati za oba modela so prikazani v Tabeli 3.

	Intenzitetni označevalec			
	morda		***včasih***	
Poziv	**ChatGPT**	**GaMS**	**ChatGPT**	**GaMS**
A – nevtralen	8	0	2	1
B – družbeni kontekst	10	0	5	1
C – humoren slog	12	3	10	5
	še		***pogosto***	
A	8	9	22	9
B	13	17	28	19
C	10	12	7	5

Tabela 3: Frekvenca izbranih izrazov glede na različne pozive.

Pokazalo se je variiranje v rabi opazovanih intenzitetnih označevalcev glede na poziv, kjer se izrazita odstopanja kažejo v frekvenci obeh krepilcev (*še*, *pogosto*). Sicer pri frekvencah rabe prihajamo do podobnih ugotovitev za oba VJM. Pokazalo se je, da sta od štirih izbranih označevalcev intenzitete šibilca *morda* in *včasih* najpogosteje prisotna v besedilih z oznako *humoren slog* (pozivi C). Spodnji primeri kažejo, da intenzitetni označevalci lahko pripomorejo k ustvarjanju humornega učinka (označeni krepko):

(1) *In če bomo* ***še naprej*** *pili kavo, se smejali sosedom in se hvalili s svojim kakovostnim življenjem, bomo* ***morda*** *na koncu* ***res*** *srečni – ali* ***vsaj*** *manj utrujeni.* (GaMS)

(2) *Torej, naslednjič, ko boste slišali nekaj, s čimer se ne strinjate, se ne razjezite. Poslušajte. Razmislite. In* ***morda celo*** *se nasmehnite.* (GaMS)

(3) ***Morda*** *bi lahko vsaj občasno sprejeli, da ni* ***vsaka*** *izjava konec civilizacije in da ni* ***vsak*** *komentar razlog za nacionalni ideološki obračun.* (ChatGPT)

(4) ***Seveda*** *to ne pomeni, da bi morali zavreči* ***vse*** *prednosti sodobnega sveta in se preseliti v gozd (čeprav priznajmo,* ***včasih*** *to zveni* ***precej*** *privlačno).* (ChatGPT)

Po drugi strani se pokaže, da se krepilca *še* in *pogosto* najpogosteje pojavita v besedilih z oznako *družbeni kontekst* (pozivi B), zlasti izraz *pogosto* je izrazito bolj prisoten v teh besedilih kot v humornih oz. nevtralnih. Zanimivo je tudi, da je *še* hkrati najpogostejši besedni krepilec v avtorskih kolumnah, medtem ko je *pogosto* značilen za besedila, tvorjena s ChatGPT, ne pa tudi za avtorska in GaMS-ova. Členek *še* se v okviru raziskave, posebej pa pri pozivih, kjer usmerimo v slovenski družbeno-kulturni okvir, pokaže kot značilen krepilec, ki je frekventen tako v slovenskih avtorskih besedilih kot v besedilih obeh modelov UI.

7 Sklepne ugotovitve

Analiza predstavlja pilotno raziskavo, ki je vsaj za naš korpus potrdila, da obstajajo razlike v vrsti in pogostosti rabe nekaterih izrazov intenzitete, ki bodisi okrepijo bodisi ošibijo pomen sporočila. Raziskava tako lahko predstavlja izhodišče za obsežnejšo raziskavo na večjem korpusu gradiva – šele longitudinalna študija, ki bi skozi daljše obdobje razvoja VJM beležila in analizirala jezikovne pojave na vseh jezikovnih ravneh, med njimi tudi modifikatorje intenzitete, bi lahko podala zanesljive odgovore o jezikovno-stilni in pragmatični jezikovni kompetenci različnih VJM. Ob uporabi različnih pozivov se je pokazalo odstopanje v vrsti in pogostosti različnih označevalcev intenzitete, kar bi bilo smiselno preveriti na večjem korpusu.

Ugotovitve naše raziskave so skladne z ugotovitvami Shalevske (2024) na anglosaškem gradivu, da obstajajo razlike v rabi intenzitetnih označevalcev med avtorskimi in s ChatGPT-jem generiranimi besedili – pri avtorskih besedilih se je pokazala večja raznolikost rabljenih izrazov v tej vlogi. Analiza kaže pogosto rabo členka *še* (skoraj 89 % besedil), ki je najpogostejši krepilec v avtorskih besedilih, kar lahko kaže na večjo težnjo avtorjev k dodajanju in stopnjevanju argumentov. Samo v korpusu avtorskih besedil se pojavijo izrazi *najbrž*, *pač*, *žal* – prvi označuje zmanjšano stopnjo gotovosti o povedanem, druga dva pa izražata avtorjev subjektivni odnos do upovedenega (npr. obžalovanje nečesa, zadržanost, omejevanje; členek *pač* lahko deluje kot šibitev, ko izraža omejevanje, ali kot krepitev, zlasti na začetku stavka, ko poudarja to, kar sledi; v avtorskih besedilih je členek *pač*, ki je pogovorno zaznamovan, prisoten v dveh besedilih (blogih) istega avtorja, kar nakazuje individualne poteze pisanja in ne splošnih ugotovitev).

Razlika v pogostosti rabe označevalcev intenzitete se je pokazala tudi pri primerjavi dveh jezikovnih modelov. V stilu besedil ChatGPT izstopa prisotnost izrazov *morda*, *verjetno*, *lahko* in *pogosto*, ki kažejo na težnjo k previdnemu, zadržanemu in posplošujočemu izražanju (izogibanje premočnim sodbam), kar sicer ni značilnost izbrane vrste presojevalnih besedil (kolumna, blog). Obratno se je pri besedilih modela GaMS pokazala visoka prisotnost izrazov *vedno*/*stalno*, *najbolj*, ki so pomenski presežniki, izrazito malo pa je šibilcev, kar kaže na odločen, okrepljen, a tudi posplošujoč stil (skrajnostno prikazovanje resničnosti). Ugotovitve o stilu besedil UI se sicer skladajo z omenjenimi tujimi študijami po kriteriju manjše leksikalne raznolikosti (Guo idr. 2023; Terčon in Dobrovoljc 2025; Zhang idr. 2024) in manjše čustvene obarvanosti v besedilih ChatGPT (Mitrović idr. 2023).

Vprašanju, ali se nacionalni jezikovni model GaMS v primerjavi s ChatGPT-jem bolj približa slovenskemu jeziku, na podlagi opravljene analize ne moremo pritrditi, za oblikovanje konkretnih zaključkov pa bi seveda potrebovali obsežnejšo študijo. »Sveže« pridobljeni nacionalni model zaradi svoje majhnosti za zdaj še ne odraža ciljne konkurenčnosti, tudi glede na to, da se je pri njem pokazala manjša raznolikost intenzitetnih označevalcev v primerjavi tako z avtorskimi besedili kot z globalnim jezikovnim modelom ChatGPT.

Prispevek je nastal v okviru raziskovalnega programa Razsežnosti slovenstva med lokalnim in globalnim v začetku tretjega tisočletja (P5-0409), ki ga financira Javna agencija za znanstveno-raziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

Viri in literatura

Ahmad, S. (2025). AI and human language: The quest for linguistic competence. *International Journal of Innovative Research In Technology*, *11*(11), 2202–2219.

Fakulteta za računalništvo in informatiko Univerze v Ljubljani. *GaMS* [Veliki jezikovni model]. https://povejmo.si/

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J. in Wu, Y. (2023). How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection. *arXiv*. https://doi.org/10.48550/arXiv.2301.07597

Kalin Golob, M. (1999). Publicistika in poročevalstvo. *Slavistična revija*, *47*(2), 179–194.

Korošec, T. (1998). *Stilistika slovenskega poročevalstva*. Kmečki glas.

Košir, D. (2026). *Primerjalna semantično-pragmatična analiza intenzitetnih označevalcev v avtorskih in umetno generiranih presojevalnih besedilih: Raziskovalni podatki za objavo v članku* [Raziskovalni podatki]. Digitalni repozitorij raziskovalnih organizacij Slovenije. https://hdl.handle.net/20.500.12556/DiRROS-31945

Košir, M. (1988). *Nastavki za teorijo novinarskih vrst*. Državna založba Slovenije.

Mikolič, V. (2020). *Izrazi moči slovenskega jezika*. Annales ZRS; Slovenska matica.

Mitrović, S., Andreoletti, D. in Ayoub, O. (2023). ChatGPT or human? Detect and explain. Explaining decisions of machine learning model for detecting short ChatGPT-generated text [Predobjava].

OpenAI. (2026). *ChatGPT* [Veliki jezikovni model]. https://chatgpt.com/

Pisanski Peterlin, A. (2017). *Metabesedilo med dvema kulturama*. Znanstvena založba Filozofske fakultete.

Praček, A. in Vehovar, V. (2025). *Generativna umetna inteligenca v Sloveniji 2025: Stališča uporabnikov* [Poročilo raziskave]. Center za družboslovno informatiko, Fakulteta za družbene vede, Univerza v Ljubljani. https://www.1ka.si/uploadi/editor/doc/1762515903GenUIporoiloanketejunij2025.pdf

Shalevska, E. (2024). Hedges and boosters in AI and human writing: A comparative analysis. *Knowledge International Journal*, *65*(5), 505–510.

Šegula, M. (2016). *Teoretična opredelitev kolumne na Slovenskem: Jezikovnostilistični vidik* [Magistrsko delo]. Univerza v Ljubljani.

Terčon, L. in Dobrovoljc, K. (2025). Linguistic characteristics of AI-generated text: A survey [Predobjava]. *arXiv*. https://doi.org/10.48550/arXiv.2510.05136

Vrečar, J. (2009). *Blog kot alternativni informativni žanr* [Diplomsko delo]. Univerza v Ljubljani.

Žele, A. (2015). *Slovar slovenskih členkov*. Amebis; Založba ZRC, ZRC SAZU. https://www.termania.net/slovarji/127/slovar-slovenskih-clenkov

Veliki jezikovni modeli in izzivi poučevanja na fakulteti

DOI: 10.4312/Obdobja.45.211-218

Simona Kranjc, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; simona.kranjc@ff.uni-lj.si

Prispevek obravnava vpliv velikih jezikovnih modelov (VJM) na visokošolsko poučevanje, zlasti na učenje, pripravo pisnih nalog, preverjanje znanja in akademsko integriteto. Teoretični pregled dopolnjujeta dve anketi med študenti in študentkami Filozofske fakultete Univerze v Ljubljani, izvedeni maja 2025 in aprila 2026 (skupaj N = 83). Rezultati kažejo, da VJM niso več obrobno orodje: večina jih uporablja vsaj občasno, predvsem za iskanje informacij, preverjanje razumevanja in pripravo študijskih obveznosti. Primerjava anket kaže, da se raba nekoliko povečuje, a je hkrati tudi bolj kritična. Izrazita večina podpira jasna univerzitetna pravila, zato prispevek zagovarja vodeno, transparentno in reflektirano vključevanje VJM v pouk.

umetna inteligenca, veliki jezikovni modeli, visokošolsko izobraževanje, akademska integriteta, preverjanje znanja

This article examines the impact of large language models (LLMs) on higher education teaching, particularly on learning, preparation of written assignments, assessment, and academic integrity. The theoretical overview is complemented by two surveys conducted among students at the University of Ljubljana's Faculty of Arts in May 2025 and April 2026 (n = 83). The results show that LLMs are no longer a marginal tool: most students use them at least occasionally, primarily for finding information, checking their understanding, and preparing assignments. A comparison of the two surveys indicates that the use of LLMs is increasing slightly but at the same time becoming more critical. A clear majority of respondents support the introduction of explicit university-level rules; the article therefore argues for guided, transparent, and reflective integration of LLMs into teaching.

artificial intelligence, large language models, higher education, academic integrity, assessment

1 Uvod

Veliki jezikovni modeli (VJM), kot so ChatGPT, Copilot in Gemini, so v nekaj letih postali vsakdanja orodja za tvorjenje, povzemanje, prevajanje in jezikovno urejanje besedil. Njihova dostopnost spreminja tudi visokošolski prostor (Radovan in Flogie 2026): študentom omogoča hitro razlago pojmov, iskanje informacij, pripravo osnutkov in samopreverjanje, učiteljem pa odpira vprašanja o avtentičnosti izdelkov, o preverjanju znanja in akademski integriteti. Raziskovalni problem prispevka je zato vprašanje, kako VJM spreminjajo študijske prakse in katere didaktične prilagoditve so potrebne, da njihova raba ne bi nadomestila samostojnega razumevanja, argumentiranja in odgovornega akademskega pisanja.

Razprava o VJM ni zgolj tehnična. Orodja ne spreminjajo samo načina, kako študenti pridejo do besedila, temveč tudi pot, po kateri oblikujejo razumevanje: vprašajo,

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

dobijo razlago, zahtevajo primer, skrajšajo besedilo, preverijo osnutek ali pridobijo predlog strukture. Takšna interaktivnost je didaktično koristna, vendar lahko hkrati ustvarja iluzijo znanja, če študent odgovora ne preveri, ga ne poveže z literaturo in ga ne zna samostojno pojasniti.

V prispevku bom najprej povzela nekaj izhodišč o vlogi VJM v izobraževanju, nato bom na podlagi dveh anket skušala pokazati, kako študenti ta orodja dejansko uporabljajo, v zadnjem delu pa se bom osredotočila na vplive, ki jih imajo VJM na univerzitetni pouk. Osrednja raziskovalna vprašanja, na katere iščem odgovore v obeh anketah, so: kako pogosto in za katere namene študenti uporabljajo VJM, kako ocenjujejo njihove učinke na učenje ter kako razumejo sprejemljivo rabo pri študijskih nalogah.

2 Veliki jezikovni modeli v izobraževanju

Razvoj jezikovnih tehnologij je del širših sprememb v sporazumevanju. Strahovi, da nove tehnologije ogrožajo jezikovno zmožnost, niso novi; podobne razprave so spremljale že množično rabo spleta in kratkih sporočil (Crystal 2001; Kalin Golob 2008). Pri VJM pa je nov predvsem obseg avtomatizirane tvorbe besedil: modeli lahko ustvarijo koherentno seminarsko nalogo, povzetek članka, razlago pojma ali jezikovno popravljeno besedilo, zato se meja med pomočjo in avtorstvom težje določa.

VJM temeljijo na strojnem učenju in velikih količinah besedilnih podatkov. Prelomna sta bila zlasti razvoj modelov za razumevanje in tvorjenje jezika (Brown idr. 2020; OpenAI 2023). Kot najpogostejše prednosti uporabe umetne inteligence v izobraževanju se poudarjajo personalizacija učenja, hitrejša povratna informacija, pomoč pri razumevanju zahtevnih pojmov, podpora pri pisanju in lažji dostop do virov (Miao in Holmes 2023; OECD 2023).VJM lahko študentu razložijo isti pojem na različnih ravneh zahtevnosti, pripravijo primere, vprašanja za ponavljanje ali predlagajo izboljšave besedila. S tem se približajo vlogi sprotnega tutorja, ki je vedno dosegljiv, vendar ne prevzema odgovornosti za pravilnost odgovora. Toda enake lastnosti, kot jih ima učitelj, prinašajo tudi tveganja. Modeli lahko tvorijo napačne ali izmišljene informacije, reproducirajo pristranskosti in ustvarjajo prepričljiva besedila brez dejanskega razumevanja (Bender idr. 2021). Pri pisnih nalogah se zato pojavi vprašanje, ali je študent besedilo res razumel in oblikoval sam. Detektorji umetno tvorjenih besedil niso dovolj zanesljivi, zato ne morejo biti edina podlaga za presojo kršitev. Smiselnejša je kombinacija jasnih pravil, postopnega nastajanja nalog, refleksije o uporabi orodij in ustnih zagovorov (Univerza v Ljubljani 2023; Vanderbilt University 2023).

Posebej občutljivo je področje akademskega pisanja. Če VJM pomaga pri iskanju začetnih idej, oblikovanju vprašanj, jezikovnem izboljšanju ali preverjanju razumljivosti, lahko podpira učenje. Če pa študent z njim ustvari bistveni del argumentacije in ga brez razumevanja odda kot lastno delo, je ogrožena akademska integriteta. Zato je treba pri pouku razlikovati med pomočjo v procesu in nadomestitvijo procesa.

Pri tem je pomembno, da VJM niso klasični iskalniki. Odgovor ne nastane kot seznam preverljivih zadetkov, temveč kot novo besedilo, ki je lahko slogovno prepričljivo tudi tedaj, ko je vsebinsko pomanjkljivo. Za študenta je to zahtevno vsaj na dveh

ravneh: najprej mora razumeti obravnavano temo, nato pa še presoditi, ali je razlaga, ki jo je dobil, ustrezna. Raba VJM zato ne zmanjšuje potrebe po strokovnem znanju, ampak jo pogosto povečuje, saj je kakovost uporabe odvisna od kakovosti vprašanja in od sposobnosti preverjanja odgovora.

3 VJM v slovenskem visokošolskem prostoru

V slovenskem visokošolskem okolju se raba VJM šele sistematično ureja. Priporočila Univerze v Ljubljani poudarjajo odgovornost uporabnika, preverjanje informacij in transparentnost rabe (Univerza v Ljubljani 2023). Podobna načela zagovarja Univerza na Primorskem, ki poleg odgovornosti in transparentnosti poudarja tudi kompetentno in vključujočo uporabo umetne inteligence ter odgovornost posameznika za končni izdelek (Univerza na Primorskem 2023). Tudi smernice Univerze v Mariboru določajo, da mora študent preveriti vsebino in vire, ki jih ustvari VJM, jasno označiti umetno tvorjene dele besedila ter navesti uporabljeno orodje, pri čemer lahko posamezne članice in izvajalci predmetov uporabo dodatno omejijo ali natančneje uredijo (Univerza v Mariboru 2023). Takšna izhodišča so skladna z raziskavami o etičnih izzivih umetne inteligence v visokem šolstvu, ki opozarjajo, da prepoved ni dovolj, če študentov hkrati ne učimo odgovorne uporabe (Molek 2023). Ključno vprašanje danes ni več, ali bodo študenti uporabljali VJM, temveč kako bodo njihovo rabo razumeli, označili in povezali s samostojnim študijskim delom.

Na predavanjih, ko sem študente slovenistike ob besedilu na drsnici spraševala, kdo je po njihovem mnenju avtor besedila, se je pokazalo, da študenti umetno tvorjenih besedil ne prepoznajo vedno zanesljivo. Besedilo, ki je vsebinsko splošno pravilno in jezikovno urejeno, lahko deluje prepričljivo, tudi če vsebuje napako ali slogovno ponavljanje. Prav tako se modeli hitro izboljšujejo: besedila so slogovno prilagodljivejša, manj šablonska in bolj odvisna od poziva, ki ga uporabnik oblikuje. To potrjuje, da visokošolski pouk ne potrebuje le pravil o prepovedani rabi, temveč dejavnosti, v katerih študenti analizirajo kakovost, vire, argumentacijo in omejitve odgovorov VJM.

Za učitelja to pomeni premik v ocenjevanju. Naloga, ki preverja samo končni izdelek, je v dobi VJM šibkejša; naloga, ki zahteva osnutek, pojasnitev odločitev, primerjavo virov, ustni zagovor ali refleksijo o uporabi orodja, pa razumevanje preverja veliko bolje. VJM zato ne ukinjajo potrebe po preverjanju znanja, temveč zahtevajo drugačno oblikovanje dokazil o znanju.

4 Anketna raziskava

4.1 Metoda in vzorec

Empirični del temelji na dveh anonimnih anketah, izvedenih s pomočjo Microsoft Forms. Prva je bila izvedena maja 2025 na priložnostnem vzorcu študentov in študentk, ki obiskujejo predavanja pri avtorici (N = 48), druga aprila 2026 na primerljivem vzorcu (N = 35). Skupni vzorec obsega 83 sodelujočih, od tega 27 na prvi in 56 na drugi stopnji študija. Anketi nista bili izvedeni na istih osebah, zato ne gre za pravo vzdolžno

raziskavo, temveč za opisno primerjavo dveh časovno zaporednih vpogledov v podobno študijsko okolje.

Vprašalnik je preverjal pogostnost uporabe VJM, namen rabe, najpogosteje uporabljeno orodje, oceno uporabnosti, preverjanje informacij, sprejemljivost rabe pri seminarskih nalogah in potrebo po pravilih. Kvantitativni del je bil obdelan z metodo opisne statistike (deleži, sredine, primerjava po študijskih stopnjah), odprti odgovori pa tematsko kodirani. Pri interpretaciji je treba upoštevati, da gre za priložnostni vzorec in samooceno, zato rezultati ne omogočajo posploševanja na celotno študentsko populacijo.

4.2 Skupna slika obeh anket

Skupni rezultati kažejo, da VJM niso več obrobno orodje. Vsaj občasno jih uporablja 92,8 % sodelujočih; 59,0 % jih uporablja občasno, 33,7 % pa redno, pogosto ali skoraj vsak dan. Nikoli jih ne uporablja 7,2 %. Najpogostejše orodje je ChatGPT, ki ga je v različnih zapisih navedlo 86,7 % vprašanih. Druga orodja, npr. Copilot, Gemini, NotebookLM ali Perplexity, se pojavljajo precej redkeje, kar kaže, da je širša digitalna pismenost še razmeroma omejena.

Najpogostejši nameni uporabe so iskanje informacij, preverjanje razumevanja snovi in priprava študijskih obveznosti. Splošna uporabnost VJM je na petstopenjski lestvici ocenjena nad sredino lestvice (M = 3,41). Najvišje sta ocenjena pomoč pri razumevanju zahtevnih konceptov (M = 3,40) in zmanjševanje truda pri iskanju informacij (M = 3,39), nekoliko nižje hitrejše učenje (M = 3,14). Hkrati je razmeroma visoka tudi ocena, da VJM spodbujajo bolj površinsko učenje (M = 3,29), kar kaže na ambivalenten odnos: študenti prepoznavajo koristnost orodij, vendar tudi tveganje zmanjšanega samostojnega poglabljanja.

Večina študentov preverja informacije, ki jih ponudi VJM: 71,1 % redno, 19,3 % včasih. Pri seminarskih nalogah prevladuje stališče, da je raba sprejemljiva le kot pomoč (66,3 %) ali če je jasno označena (22,9 %). Uporabo brez omejitev bi dovolilo le 3,6 % vprašanih. Najizrazitejši rezultat je podpora pravilom: 86,7 % sodelujočih meni, da bi morala univerza jasno opredeliti rabo VJM.

Ti podatki kažejo razmeroma zrelo razumevanje VJM, če gre verjeti anketnim odgovorom. Študenti jih ne doživljajo samo kot prepovedano ali dovoljeno orodje, temveč kot pomoč, ki mora ostati zamejena. To je pomembno, ker se razprava o umetni inteligenci pogosto zoži na vprašanje zlorabe. Anketi pa kažeta širšo sliko: študenti iščejo pomoč pri učenju, hkrati pa pričakujejo pravila, ki bodo jasno določila, kdaj pomoč postane akademsko sporna.

4.3 Primerjava med anketama

Primerjava anket pokaže, da se VJM postopoma nekoliko bolj rutinsko vključujejo v študij. Delež tistih, ki jih ne uporabljajo nikoli, se je znižal z 8,3 % na 5,7 %, delež redne, pogoste ali skoraj vsakodnevne rabe pa se je zvišal z 31,2 % na 37,1 %. Občasna raba ostaja najpogostejša v obeh anketah. Pomembnejša od same pogostnosti je sprememba namena: v prvi anketi je bilo najbolj izrazito iskanje informacij, v drugi pa

preverjanje razumevanja snovi. To kaže premik od uporabe VJM kot bolj zmogljivega iskalnika k uporabi VJM kot orodja, ki razlaga in omogoča samopreverjanje.

Druga anketa kaže tudi bolj kritično stališče študentov. Čeprav je raba nekoliko pogostejša, se povprečna ocena splošne uporabnosti zniža s 3,48 na 3,31. Povprečna ocena trditve, da VJM spodbujajo bolj površinsko učenje, pa se nekoliko zviša, in sicer s 3,26 na 3,32. Hkrati se redno preverjanje informacij poveča s 66,7 % na 77,1 %. To ne kaže zavračanja VJM, temveč bolj realno presojo: študenti jih uporabljajo, a se hkrati bolj zavedajo njihove nezanesljivosti. Pri seminarskih nalogah se okrepi odgovor *le kot pomoč* (s 60,4 % na 74,3 %), podpora jasnim univerzitetnim pravilom pa se poveča z 81,2 % na 94,3 %.

Pri razlagi razlik je potrebna previdnost, saj je v drugi anketi večji delež študentov druge stopnje. Ti imajo praviloma več izkušenj s seminarskimi nalogami, raziskovalnimi osnutki in samostojnim pisanjem, zato lahko VJM uporabljajo drugače kot študenti začetnih letnikov. Kljub temu je smer sprememb skladna: raba se normalizira, vendar ne vodi v manjšo zadržanost, temveč v večjo potrebo po preverjanju in pravilih.

Odprti odgovori pa potrjujejo štiri tematske poudarke: željo po prikazu smiselne rabe pri pouku, potrebo po pravilih in označevanju, pomen preverjanja virov ter skrb pred površinskostjo in pretirano odvisnostjo od orodij. Študenti torej praviloma ne zagovarjajo niti popolne prepovedi niti neomejene rabe, temveč vodeno in transparentno uporabo. VJM razumejo kot pomoč pri učenju, ne kot samoumevno nadomestilo za branje, razmišljanje in pisanje.

S pedagoške perspektive je posebej zanimivo, da povečana raba ne pomeni nujno večje nekritičnosti. V drugi anketi se namreč hkrati povečata rednejša uporaba in rednejše preverjanje informacij. To kaže, da se študenti učijo živeti z orodjem: uporabljajo ga pogosteje, vendar bolj vedo, da ni nezmotljivo. Tak premik je za pouk ugoden, ker odpira možnost, da se VJM vključijo kot predmet analize. Študenti lahko denimo primerjajo odgovor modela z znanstvenim člankom, poiščejo neustrezne posplošitve, dopolnijo manjkajoče vire ali razložijo, zakaj je neka trditev vprašljiva.

5 Razprava

Rezultati potrjujejo, da so VJM že del univerzitetnih študijskih praks. Njihova vloga pa ni enoznačna. Po eni strani omogočajo hitrejšo orientacijo, dodatne razlage in samopreverjanje; po drugi strani pa lahko ustvarijo občutek razumevanja brez poglobljenega dela. Za univerzitetni pouk je zato ključno, da VJM ne obravnavamo zgolj kot tehnično orodje, temveč kot dejavnik, ki spreminja načine učenja, pisanja in preverjanja znanja.

Podatki kažejo, da študenti VJM najpogosteje uporabljajo pri dejavnostih, ki so bile tradicionalno del samostojnega študija: iskanje informacij, preverjanje razumevanja, oblikovanje osnutkov in priprava na študijske obveznosti. To ne pomeni nujno manj samostojnosti, pomeni pa, da se samostojnost kaže drugače. Študent, ki zna oblikovati dobro vprašanje, primerjati odgovore, preveriti vire in pojasniti, kaj je prevzel od orodja, razvija drugačno obliko akademske pismenosti kot študent, ki preprosto prebere

predpisano literaturo. Vprašanje je, ali pouk te nove postopke eksplicitno razvija ali jih prepušča naključju.

Najpomembnejša ugotovitev obeh anket je potreba po jasnih pravilih. Ker večina študentov podpira omejeno in transparentno rabo, prepoved rabe VJM verjetno ne bi bila najbolj produktivna. Bolj smiselno je razlikovati med rabo za učenje, rabo, ki mora biti označena, in rabo, ki pomeni kršitev akademske integritete. Dovoljena je lahko pomoč pri iskanju idej, razlagi pojmov, jezikovnem pregledu ali samopreverjanju, problematična pa je oddaja besedila, ki ga študent ne razume, ne zna zagovarjati ali ga predstavi kot svoje delo, čeprav je bistveni del argumentacije tvoril VJM.

Pedagoški odgovor mora zato vključevati tri ravni. Prva je razvijanje kritične digitalne pismenosti: študenti morajo znati preveriti vire, prepoznati napake in presoditi kakovost odgovora. Druga je sprememba nalog: smiselni so procesni izdelki, osnutki, refleksije o delu, primerjanje odgovorov VJM z literaturo in ustni zagovori. Tretja je institucionalna ureditev: študenti potrebujejo jasna navodila, kako rabo označiti in katere oblike pomoči so pri posamezni nalogi dovoljene.

Posebej pomembna je povezava med pisanjem in zagovorom. Če je seminarska naloga le končni dokument, je težko presoditi, kako je nastala. Če pa študent odda tudi načrt, osnutek, navedbo uporabljenih orodij, refleksijo in se o nalogi pogovori, postane ocenjevanje manj odvisno od ugibanja o avtorstvu. Učitelj tako ne preverja samo, ali je besedilo dobro, temveč ali ga študent razume, zna utemeljiti izbiro virov in prepoznati omejitve uporabljenega orodja.

Pomembno je tudi vprašanje pravičnosti. Če pravila niso jasna, imajo prednost tisti, ki orodja bolje poznajo, si upajo več ali imajo dostop do plačljivih različic. Neurejena raba zato ne pomeni večje svobode, temveč več neenakosti in več negotovosti. Skupna pravila ne bi smela biti zasnovana le kot seznam prepovedi, ampak kot učni okvir: študent naj ve, kaj sme uporabiti, kako mora uporabo opisati in za katere dele izdelka ostaja odgovoren sam.

Omejitev raziskave, ki jo predstavljam v prispevku, je priložnostni vzorec, ki je omejen zaradi svoje majhnosti, pa tudi področne pristranskosti (sodelovali so le študenti, ki obiskujejo moja predavanja); ugotovitev zato ni mogoče posplošiti na celotno študentsko populacijo. Ker anketi nista zajeli istih oseb, primerjava tudi ne dokazuje razvoja stališč skozi čas. Kljub temu podatki dovolj jasno kažejo smer: raba VJM se normalizira, hkrati pa se krepi potreba po kritični in pravilno označeni uporabi.

Zelo podobno stanje se kaže tudi v raziskavi med več kot 8000 študenti na štirih avstralskih univerzah, v kateri avtorji ugotavljajo, da je več kot 80 % študentov že uporabilo generativno umetno inteligenco pri študijskem delu, skoraj polovica pa jo uporablja redno. Najpogostejši nameni so urejanje besedil, povzemanje in oblikovanje idej, kar se ujema z ugotovitvijo, da študenti VJM najpogosteje razumejo kot podporno orodje pri pripravi študijskih obveznosti, ne pa nujno kot nadomestilo za samostojno delo (Chung idr. 2026).

6 Zaključek

VJM v visokošolskem izobraževanju ne pomenijo samo novega pripomočka, temveč posegajo v temeljna vprašanja učenja, avtorstva in preverjanja znanja. Anketi med študenti Filozofske fakultete Univerze v Ljubljani kažeta, da jih večina uporablja vsaj občasno, predvsem kot pomoč pri razumevanju, iskanju informacij in pripravi študijskih obveznosti. Primerjava obeh anket kaže nekoliko pogostejšo in hkrati bolj zadržano rabo: študenti orodja uporabljajo, vendar vse bolj prepoznavajo potrebo po preverjanju informacij in po pravilih. Za univerzitetni pouk to pomeni, da se ni mogoče vrniti v stanje pred VJM. Smiselna pot ni splošna prepoved, temveč didaktično vodena in transparentna raba. Študenti morajo znati pojasniti, kako so orodje uporabili, preveriti njegove odgovore in prevzeti odgovornost za končni izdelek. Učitelji pa morajo naloge oblikovati tako, da preverjajo razumevanje, argumentacijo in proces nastajanja, ne le končnega besedila. Le tako lahko VJM postanejo podpora kritičnemu učenju, ne pa nadomestek samostojnega akademskega dela.

Najpomembnejša naloga visokošolskega prostora zato ni tehnološka, temveč didaktična in etična: oblikovati je treba pogoje, v katerih bo raba VJM pregledna, odgovorna in povezana z učenjem. To vključuje skupna pravila, naloge z dokazili o procesu dela, pogovor o virih in razvijanje zmožnosti, da študent kritično presodi tudi odgovor, ki je oblikovno prepričljiv. V takem okviru VJM niso grožnja sami po sebi, temveč preizkus kakovosti naših učnih ciljev in načinov preverjanja znanja. Zato bi prispevek lahko razumeli tudi kot spodbudo k premiku od vprašanja, ali je umetna inteligenca pri študiju dovoljena, k vprašanju, kakšna raba podpira akademske cilje. Kadar VJM omogoči boljše vprašanje, jasnejšo razlago, natančnejše preverjanje ali bolj premišljeno popravljanje besedila, prispeva k učenju, kadar pa skrije nerazumevanje, zabriše avtorstvo ali nadomesti iskanje v virih, zmanjšuje kakovost študija. Odgovorna integracija mora to razliko narediti vidno. Podporo jasnim pravilom med drugim potrjujejo tudi smernice Centra Digitalna UL, ki pri uporabi generativne umetne inteligence v izobraževanju poudarjajo razmejitev med legitimno uporabo in zlorabo, transparentno navajanje rabe orodij ter razvijanje kritične presoje generiranih vsebin (Jedrinović idr. 2025).

Prispevek je nastal v okviru projekta NOO PŠP, ki ga sofinancirata Republika Slovenija, Ministrstvo za vzgojo in izobraževanje ter Evropska unija – NextGenerationEU.

Literatura

Bender, E. M., Gebru, T., McMillan-Major, A. in Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT ˙21: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (str. 610–623). https://doi.org/10.1145/3442188.3445922

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*. https://doi.org/10.48550/arXiv.2005.14165

Chung, J., Henderson, M., Slade, C., Liang, Y., Pepperell, N., Corbin, T., Walton, J., Yu, A. S., Bearman, M., Buckingham Shum, S., Fawns, T., McCluskey, T., McLean, J., Oberg, G., Seligmann, A., Shibani, A., Bakharia, A., Lim, L.-A. in Matthews, K. E. (2026). The use and usefulness of GenAI in higher education: Student experience and perspectives. *Computers and Education Open*, *10*, članek 100347. https://doi.org/10.1016/j.caeo.2026.100347

Crystal, D. (2001). *Language and the internet*. Cambridge University Press.

Jedinović, S., Kern Nanut, E., Škraba, E., Bevčič, M., Ferk Savec, V., Klemen, T., Mlinarec, K., Žerovnik, A., Dolinar, M., Dovč, P., Jošt, M., Kropivšek, J., Ferme, T., Pipan, T., Tanšek, A., Zorc, M., Erjavec, J., Manfreda, A., Rejc, J., … Matić, L. (2025). *UI v izobraževanju na UL: Rezultati raziskovalnega dela na projektu*. Center Digitalna UL, Univerza v Ljubljani. https://www.uni-lj.si/assets/Center-Digitalna-UL/Pristopi/Smernice-UI-v-izobrazevanju/Porocilo-UI-v-izobrazevanju-pregled-stanja-in-metodologija.pdf

Kalin Golob, M. (2008). SMS-sporočila treh generacij. V M. Košuta (ur.), *Slovenščina med kulturami* (str. 283–294). Slavistično društvo Slovenije.

Miao, F., in Holmes, W. (2023). *Guidance for generative AI in education and research*. Unesco.

Molek, N. (2023). Etični izzivi uporabe umetne inteligence v visokošolskih zavodih. *Revija za univerzalno odličnost*, *12*(4), 300–318.

OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing.

OpenAI. (2023). GPT-4 technical report. *arXiv*. https://arxiv.org/abs/2303.08774

Radovan, M. in Flogie, A. (ur.). (2026). *Generativna umetna inteligenca v izobraževanju: preobrazba učenja v digitalni dobi*. Založba Univerze v Ljubljani. https://doi.org/10.4312/9789612978174

Univerza na Primorskem. (2023). *Stališča Univerze na Primorskem do rabe umetne inteligence*. Univerza na Primorskem.

Univerza v Ljubljani. (2023). *Priporočila Univerze v Ljubljani pri uporabi umetne inteligence*. https://www.uni-lj.si/assets/Kabinet/3_Novice/Dokumenti-znotraj-novic/Priporocila-UL-pri-uporabi-umetne-inteligence.pdf

Univerza v Ljubljani. (2026). *Smernice za uporabo orodij (generativne) umetne inteligence v izobraževanju*. https://www.uni-lj.si/studij/center-digitalna-ul/gradiva/ucni-pristopi-metode-in-oblike-dela/smernice-za-uporabo-orodij-generativne-umetne-inteligence-v-izobrazevanju

Univerza v Mariboru. (2023). *Uporaba generativnih orodij umetne inteligence v pedagoškem procesu*. Univerza v Mariboru. https://moja.um.si/en/study/Documents/Smernice_za_uporabo_UI_v_pedagoskem_procesu_16.5.2023_final.pdf

Vanderbilt University. (2023, 16. avgust). *Guidance on AI detection and why we're disabling Turnitin's AI detector.* Vanderbilt University Brightspace. https://www.vanderbilt.edu/brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/

Od umetne trave do lažne javne podpore: o novejši izlastnoimenski leksiki na gradivu *Sprotnega slovarja slovenskega jezika*

DOI: 10.4312/Obdobja.45.219-227

Domen Krvina, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana;
domen.krvina@zrc-sazu.si

Mija Michelizza, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana;
mija.michelizza@zrc-sazu.si

V prispevku analiziramo 50 izbranih izlastnoimenskih leksemov v *Sprotnem slovarju slovenskega jezika* (2014–). Osredotočamo se na več vidikov, in sicer izlastnoimensko leksiko v prvem delu klasificiramo glede na pomen in izvor, pri čemer izstopajo poimenovanja za dejanja, dejavnost v digitalnem okolju ali osebo, ki ju izvaja. Ta imajo sicer zelo predvidljiv izvor. Drugi del analize se posveča besedotvornim, pomenskim in zunajjezikovnim lastnostim te leksike, kjer se pokaže, da leksemi, ki označujejo zlasti splošno znane, razširjene pojave iz digitalnega okolja, pogosto razvijejo bogato besedotvorno družino (dejanje, vršilec dejanja ipd.).

izlastnoimenski leksemi, digitalne vsebine, *Sprotni slovar slovenskega jezika*, apelativizacija, etimologija

This article analyzes fifty selected eponyms included in the dictionary *Sprotni slovar slovenskega jezika* (Growing Dictionary of the Slovenian Language, 2014–). The focus is multifaceted: first, the eponyms are classified according to meaning and origin, with a notable prevalence of terms denoting actions, activities in digital environment, or the agents performing them. These categories, however, exhibit highly predictable origins. The second part of the analysis examines the word-formational, semantic, and extralinguistic properties of the vocabulary analyzed. The findings indicate that lexis denoting widely recognized and prevalent digital phenomena frequently develop extensive word-formational families (e.g., action, agent, etc.).

eponyms, digital content, growing dictionary, appelativization, etymology

1 Uvod

V sodobnem času vseprisotno digitalno okolje neizogibno vpliva tudi na nastanek novejše leksike. Prve raziskave o jeziku v digitalnem okolju, ki pogosto izpostavljajo novosti z vidika leksike in besedotvorja ter vidik spleta kot besedilnega korpusa, se v Evropi pojavijo v začetku 21. stoletja (npr. Crystal 2001; Kilgarriff in Grefenstette 2003), kmalu zatem tudi za slovenski jezik (npr. Kranjc 2003; Logar 2004). Sledi razmah raziskav jezika v digitalnem okolju na različnih podpodročjih jezikoslovja za različne jezike – od novejših raziskav lahko izpostavimo srbski primer, ki internet in druž-

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

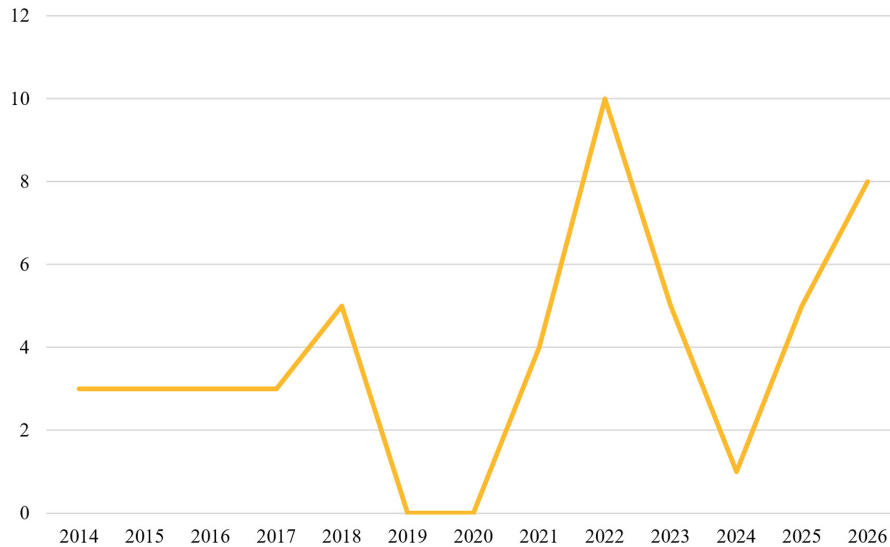
bene medije obravnava kot pomemben vir novejšega besedja (Lazić Konjik 2024). Leta 2009 je v Sloveniji izšla monografija *Novejša slovenska leksika (v povezavi s spletnimi jezikovnimi viri)*, ki kot eno od kategorij novejše leksike navaja tudi skupino, ki je nastala s poobčnobesedenjem oz. apelativizacijo lastnoimenskega izhodišča (Gložančev idr. 2009: 38) – prav tej vrsti leksike je posvečen naš prispevek, saj predpostavljamo, da je za leksiko, povezano z digitalnimi viri, še posebej značilna. Na podlagi omenjene monografije je nastal *Slovar novejšega besedja slovenskega jezika* (2012) z opisom novejše leksike od izida zadnje knjige *Slovarja slovenskega knjižnega jezika* leta 1991 pa do izida leta 2012, od leta 2014 dalje pa je novejša leksika sistematično predstavljena v *Sprotnem slovarju slovenskega jezika* (Krvina 2014–; v nadaljevanju *Sprotni slovar*). Ta slovar leksiko za opis pridobiva zlasti z metodo »od spodaj«, tj. s predlogi uporabnikov, ki so nato gradivno preverjeni (Krvina 2022, 2024; prim. tudi Škrabal in Kavka 2021). V zadnjem obdobju se za gradivno preverjanje uporablja zlasti korpus Trendi (Kosem idr. 2023), ki ga za preverjanje frekvence uporabljamo tudi v tem prispevku. Avtomatskih zaznav potencialne novejše leksike (npr. Brglez idr. 2025; Kerremans idr. 2011) se *Sprotni slovar* za zdaj ne poslužuje. Enkrat letno (3. decembra) so letni prirastki slovarja, ki vključujejo novejšo in v splošnih razlagalnih slovarjih še neuslovarjeno leksiko, objavljeni na slovarskem portalu Fran. Jezikoslovna analiza v prispevku se omejuje na skupino 50 izlastnoimenskih leksemov, katerih slovarski sestavki so že objavljeni, oz. nekaterih, ki so predvideni za objavo ob koncu leta 2026.

2 Metodologija

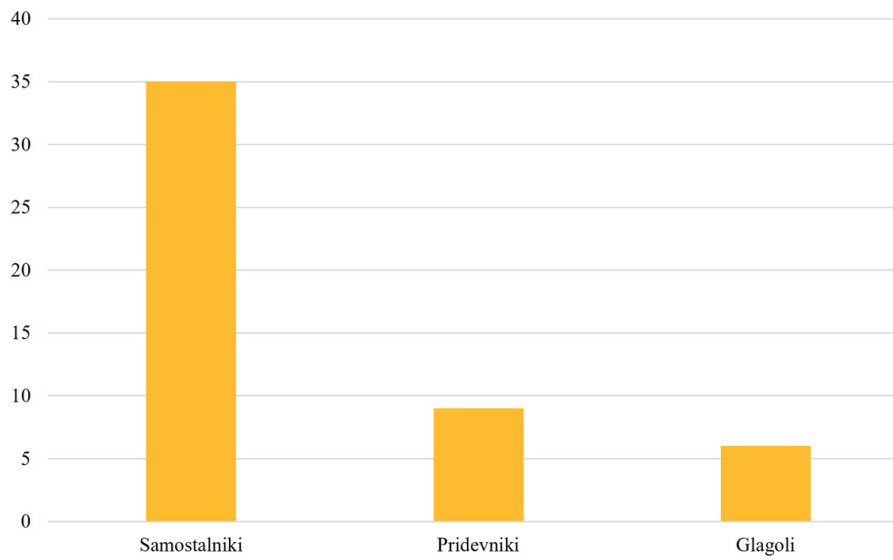
Podatek, od kod leksem izhaja oz. kako je nastal, v *Sprotnem slovarju* najdemo v razdelku z etimološko osvetlitvijo. Ta prinaša temeljno razlago izvora iztočnice – vključno z avtorskimi pravicami, če so pri poimenovanju prisotne. Razdelek z etimološko osvetlitvijo imajo vse iztočnice od leta 2021 dalje, pri predhodnih pa je izdelava razdelka v teku. Izlastnoimenska poimenovanja lahko s precejšnjo mero zanesljivosti najdemo tako, da v iskalno zahtevo v slovarski bazi zapišemo, naj v razdelku z etimološko osvetlitvijo iščemo *ime* oz. *poimenovanje*. Ker je pri dobri polovici starejših slovarskih iztočnic etimološki razdelek šele v nastajanju, je vsak od avtorjev prispevka samostojno izvedel še ročni pregled celotnega seznama slovarskih iztočnic in dodal lekseme, za katere se je – pri zahtevnejših primerih po posvetovanju z etimologom – tudi izkazalo, da so izlastnoimenski. Celoten seznam obsega 50 tako najdenih slovarskih iztočnic, pri čemer je treba opozoriti, da bi pri tistih brez razdelka z etimološko osvetlitvijo dodaten pregled potencialno našel še kakšno več. Okvirno pa lahko rečemo, da gre vsaj za 50 od skupno 1852, torej trenutno slabe 3 % v *Sprotnem slovarju* uslovarjenih iztočnic.

Ko je bilo gradivo zbrano, smo analizo izvedli v dveh fazah. V prvem delu smo lekseme klasificirali glede na pomen in izvor (3. poglavje), v drugem delu pa se osredotočamo na besedotvorne lastnosti te leksike v povezavi s pomenskimi in zunajjezikovnimi značilnostmi (4. poglavje), pri čemer smo frekvenco pojavnosti preverjali v korpusu Trendi. V nadaljevanju navajamo še temeljno statistiko o letu objave (za še

neobjavljene iztočnice o načrtu leta objave, torej 2026), iz katere lahko razberemo, da je letno število teh iztočnic precej različno, in sicer se giblje od 0 do 10 (Graf 1), povprečno pa je takih iztočnic vsako leto 2,77. Sledi še statistika o razporeditvi po besednih vrstah (Graf 2), kjer po pričakovanjih prevladujejo samostalniki (35 primerov oz. 70 %), manj pa je pridevnikov (9 primerov oz. 18 %) in glagolov (6 primerov oz. 12 %).



Graf 1: Število objavljenih (za leto 2026 načrtovanih) izlastnoimenskih leksemov v *Sprotnem slovarju* po letih.



Graf 2: Analizirani izlastnoimenski leksemi iz *Sprotnega slovarja* glede na besedne vrste.

Iskalna zahteva in tako pridobljeni podatki (z oznako, ali iztočnica razvija besedno družino oz. je njen del ali ne) so na voljo v Krvina in Michelizza (2026).

3 Izlastnoimenska leksika v *Sprotnem slovarju* glede na pomen in izvor

Novejšo izlastnoimensko leksiko v *Sprotnem slovarju* smo analizirali glede na pomen in izvor ter jo gradivsko razdelili v tri večje skupine, predstavljene v nadaljevanju.

3.1 Poimenovanja za dejanje, dejavnost v digitalnem svetu, osebo, ki ju izvaja, ali njeno pripadnost

Gre za skupino leksemov, ki imajo razmeroma predvidljiv izvor, saj so tipično nastali iz imen aplikacij, spletnih platform, programov – tudi operacijskega sistema –, npr. *androiden*, *sfotošopirati*, *tiktoker*, *zoomerski*. Nekateri so razvili bogato besedno družino (npr. *fotošop*), pri drugih tvorjenk v splošni rabi ni veliko (npr. *tiktoker*). Besedotvorni vidik teh leksemov si bomo podrobneje ogledali v nadaljevanju.

3.2 Poimenovanja za fizično pojavnost ali lastnost, značilno za fizično pojavnost

Z vidika izvora je ta skupina analiziranih leksemov razmeroma nepredvidljiva, vseeno pa lahko na podlagi gradiva izpostavimo večjo podskupino leksemov, ki imajo izvor v (registriranem) imenu podjetja ali izdelka, npr. *kanader*, *kevlar*, *kroks*, *radler*, *segway*. V teh primerih gre za lekseme, ki so nastali s procesom poobčnobesedenja oz. apelativizacije, ki je bil že večkrat obravnavan, med drugim v Dobrovoljc (2012) ter Dobrovoljc in Michelizza (2019). Z vidika krovne teme vpliva digitalnega okolja na leksiko je zanimiv primer *knjigobežnice*, in sicer izvira iz imena profila skupine na družbenem omrežju Facebook, tj. Knjigobežnice,[1] ki je še vedno aktivna, poimenovanje *knjigobežnica* pa označuje ʻmanjši zaprt prostor, hišico na javnem kraju za prosto izmenjavo knjig’, kot se glasi razlaga v *Sprotnem slovarju*. V gradivu smo našli tudi dva primera leksemov, ki izvirata iz lastnega imena osebe: *aladinka* 1. ʻžensko obuvalo’ in 2. ʻhlače’ < Aladin ʻlik iz literarnega dela Tisoč in ena noč’ ter *frančiškovski* 1. ʻki je v zvezi s svetim Frančiškom Asiškim’ in 1.1. ʻki je v zvezi s papežem Frančiškom’. Dva primera leksemov izvirata iz krajevnega imena: *denisovec* in pridevniška izpeljanka *denisovski* izvirata iz imena najdišča ostankov te (pod)vrste ljudi v Rusiji *Denísova peščéra*, kar pomeni ʻDenisova jama’; *kinoto* je agrum, katerega ime izvira iz italijanskega *chinotto*, ta pa iz imena *C(h)ina* ʻKitajska’ po domnevnem izvoru sadeža. Prav poseben pa je izvor izrazov *ganaš* in (dvojnice) *ganache* v pomenu ʻkrema iz čokolade, stopljene v zavreti sladki smetani’ – izvirata namreč iz imena gledališke igre *Les Ganaches* ʻPrismode’, kot je sladico ob njenem nastanku v šestdesetih letih 19. stoletja poimenoval njen tvorec Paul Siraudin.

1 Idejna vodja skupine Nina Kožar je leta 2011 želela spodbuditi izmenjavo knjig in predlagala, naj se v podarjeno knjigo zapiše, da gre za knjigobežnico (tj. ʻknjigo, ki jo podarjamo’), mesto, kjer je bila knjiga odložena, pa naj se fotografira in objavi v skupini na Facebooku. Kasneje so se v sklopu skupine uveljavile posebne hišice za izmenjavo knjig – in te danes imenujemo *knjigobežnice*.

3.3 Poimenovanja za (družbeni) pojav

V zadnjo skupino so uvrščeni leksemi *astroturfing*, *gaslighting* s slovenskim ustreznikom *mračenje* ter dvojnici *srcozlom* in *srcezlom*. *Astroturfing* smo prevzeli iz angleščine – izvira iz imena umetne trave AstroTurf istoimenskega ameriškega proizvajalca in se uporablja za družbeni pojav, ki je v *Sprotnem slovarju* opisan kot ʻobjavljanje na družbenih omrežjih, forumih, ki skuša ustvarjati vtis pristnosti, a je zlasti politično, poslovno motivirano, podprtoʼ. Gre za besedno igro, pri kateri je vzpostavljena opozicija z izrazom *grassroots*, ki pomeni ʻljudska organizacijaʼ oz. spontana podpora iz ljudstva, dobesedno pa bi izraz prevedli kot travne korenine (kot nasprotje umetni travi). *Gaslighting* v pomenu ʻpsihološka manipulacija, navadno dolgotrajna, zaradi katere prizadeti začne dvomiti o zanesljivosti svojih misli, svojega dojemanja resničnosti ali spominovʼ izvira iz igre Patricka Hamiltona z naslovom *Gas Light* iz leta 1938, ki je doživela dve filmski priredbi, in sicer leta 1940 (film je režiral Thorold Dickinson) in leta 1944 (film je režiral George Cukor). Slovenski ustreznik *mračenje* je nastal na podlagi predloga Terminološke svetovalnice (Atelšek idr. 2023). Tudi pri tej skupini smo našli primer, ki izvira iz imena profila, tokrat na družbenem omrežju Instagram – gre za poimenovanje *srcozlom* oz. njeno dvojnico *srcezlom*, ki označuje ʻobčutje duševne bolečine ob razhodu z ljubljeno osebo, koncu ljubezenske zvezeʼ.

4 Besedotvorne, pomenske in zunajjezikovne lastnosti izlastnoimenske leksike v *Sprotnem slovarju*

Ena od precej izrazitih značilnosti izlastnoimenske leksike v *Sprotnem slovarju* je njena vključenost v besedotvorne procese (blizu 60 % leksemov iz obravnavane celote), kar sicer velja za novejšo leksiko nasploh (Krvina 2024: 143–145). Pri bogatih besednih družinah lahko izpostavimo *fotošop* s pomenom tako procesa kot tudi rezultata predelave, prirejanja fotografij, medtem ko se lastno ime programa kot izhodišče (drugače kot npr. pri *kevlar*, *kamut* ʻsnovʼ, ʻžitoʼ) pomensko ni posplošilo. Besedna družina pri *fotošop* obsega tako glagolsko izpeljanko kot njeno dovršno sestavljenko (*fotošopirati* : *sfotošopirati*), glagolnik (*fotošopiranje*) in pridevnik stanja (*fotošopiran*). Zaradi pomenske neposplošitve se v slovar ne uvrsti *instagram* (ki bi potencialno nastal iz lastnega imena Instagram), ki pa ima vendarle razvejano besedno družino, sicer sestavljeno iz verjetno samih po sebi dveh prevzetih izhodišč (instagram + er; instagram + able): *instagramer* s feminativom *instagramerka*, v smislu privlačnosti, všečnosti za objavo na družbenih omrežjih pa *instagramabilen* z abstraktnim samostalnikom *instagramabilnost*. Temu podobna primera s pomensko neposplošenim in zato v slovar nevključenim lastnim imenom družbenega omrežja (YouTube) oz. konferenčnega orodja (Zoom), ki je prešlo v široko uporabo predvsem v času pandemije covida-19, sta še besedni družini *youtuber* s sinonimom *youtubar* (tvorbeno vršilec od vključenega glagola *youtubati*) ter feminativoma *youtuberka* in *youtubarka* in pa *zoomer* s povezanim vrstnim pridevnikom *zoomerski* ter feminativom *zoomerka*. Kot lahko vidimo, se feminativi načeloma pojavljajo regularno (*instagramerka* v korpusu Trendi s frekvenco 57; *youtuberka* 169, *youtubarka* 1; *zoomerka* s frekvenco 4) in so zato v slovar sistemsko

uvrščeni kljub včasih manjši pogostosti (pri *youtubarki* bo dokončna odločitev sprejeta ob redakciji za leto 2026). Glagolsko (*iksati* z dovršnim ustreznikom *iksniti*) in glagolniško (*iksanje*) je pri besedotvorju zaznamovano prav tako nevključeno lastnoimensko izhodišče omrežje X, nekdanji Twitter (njegova nedovršna izpeljanka *tvitati* je že v *Slovarju novejšega besedja slovenskega jezika* in SSKJ², v *Sprotni slovar* pa je bila leta 2017 dodana dovršna ustreznica *tvitniti*). Pri vseh teh primerih gre za v digitalnem okolju prisotne in v okviru novejšega besedja splošno razmeroma dobro znane pojave, kar lahko mdr. sklepamo po seštevku besednodružinskih frekvenc v korpusu Trendi, ki jih pri posameznih primerih navajamo v oklepaju: fotošop- (*fotošop* (110) + *fotošopirati* (24) + *sfotošopirati* (8) + *fotošopiranje* (12) + fotošopiran (57) = 213), instagram- (*instagramer* (58) + *instagramerka* (57) + *instagrambilen* (100) + *instagrambilnost* (16) = 231), posebej izstopajoči youtub- (*youtuber* (1300) + *youtubar* (17) + *youtuberka* (169) + *youtubarka* 1 = 1487), in verjetno nekaj manj znani zoom- (*zoomer* (47) + *zoomerski* (1) + *zoomerka* (4) = 52) ter glede na tvit- (že zgolj *tvitniti* 11.944) še ne preveč uveljavljeni iks- (*iksati* (1) + *iksniti* (14) + *iksanje* (7) = 22). Poimenovanji *denisovec* in *denisovski* (152 + 4 = 156) sta v splošno rabo prešli prek popularizacije antropoloških raziskav, medtem ko vsaj športnim navdušencem nista tuja (tvorjen verjetno po zgledu nogometaš) *premierligaš* in *premierligaški* (485 + 471 = 956).

Primera *kevlar*, *kevlarski* (58 + 5 = 63) in *kamut*, *kamutov* (70 + 42 = 112) (pridevnika imata obakrat pomen ʻki je iz, vsebuje [snov v besedotvornem izhodišču]ʼ) sta torej z vidika slovarske vključenosti tudi izhodišča bolj izjema kot pa pravilo, pri čemer omenimo, da zaradi zahtev pravne narave drugih dveh na portalu Fran ni mogoče prikazovati. Obstajajo tudi primeri, kjer je lastnoimensko besedotvorno izhodišče podobno kot zgoraj odsotno, besedno družino pa zastopa zgolj pridevniška izpeljanka, npr. *androiden* (433) (prevladuje pomen robota, pomen, ki se nanaša na operacijski sistem Android, pa je vrstni; za izdelke znamke Apple je medtem nastal nov pomen pri neizlastnoimenskem pridevniku *jabolčni*) ali *frančiškovski* (8) (ki je v zvezi ali s sv. Frančiškom Asiškim ali s papežem Frančiškom, kot že omenjeno).

Skromnejše ali celo neobstoječe so besedne družine pojavov, ki običajno niso vezani na razširjenost (zgolj ali prevladujoče) v spletnem okolju, npr. *aladinka* (2), *kroks* (50) (rabljeno navadno v množini za obuvalo, prvo tudi oblačilo), *ganaš* (105) (jed), *gaslighting* (160) s sinonimom *mračenje* (38) (v pomenu psihološke manipulacije), *kanader* (1421) in *zračni traktor* (20) (letalo, zlasti za gašenje), *kinoto* (2) (sadež), *segway* (39), *skike* (4) (prevozno sredstvo oz. priprava za na obuvalo) in prvotna poimenovanja novosti ali projektov na različnih področjih, npr. *knjigobežnica* (451), *plažnica* (0; korpus Janes 20), *paketomat* (853). Videti je torej, da sama frekventnost pojavljanja ne zadostuje, pač pa k potrebi po besedni družini in njeni uporabi bistveno pripomore izrazita in razširjena raba v spletnem okolju – ni pa pri tem edini dejavnik, na kar kažeta pomensko posplošeni *bitcoin* (15.592) kot ena najbolj znanih kriptovalut in prej omenjeni *astroturfing* (116), ki obstajata praktično le digitalno, pa besedne družine za

zdaj nista razvila.[2] Svoje očitno prispeva tudi bolj ali manj široka splošna poznanost (pri *astroturfingu* je verjetno manjša), posebej pa obstoj dejavnosti, ki določen pojav spremlja in je verjetno glavni dejavnik potrebe po besedni družini – pri dejavnostih (na družbenih omrežjih) je namreč treba poimenovati vršilca dejanja tako v moški kot ženski obliki, pogosto pa tudi dejanje samo (z glagolom in/ali glagolnikom). Pojavljanje prevladujoče v spletnem okolju, kjer je ob vsaki novi dejavnosti treba poimenovati vsaj vse našteto, k temu (verjetno ne nepomembno) dodatno prispeva.

5 Razprava in zaključek

V analizirani skupini izlastnoimenske leksike v *Sprotnem slovarju* glede na pomen in izvor po številu primerov izstopa skupina leksemov, ki je nastala iz imen aplikacij, spletnih platform, programov oz. drugih digitalnih entitet, kjer tvorjenke sicer pomensko še ostajajo na ravni digitalnega, saj označujejo dejavnost, z njo povezana dejanja in uporabnike (vršilce dejanja) teh aplikacij ipd. Ta skupina leksemov je večinoma tudi besedotvorno precej bogata. Pri tem je treba opozoriti, da sámo besedotvorno izhodišče zelo pogosto ni pomensko posplošeno oz. če že, ne brez določenega prenosa (npr. proces in rezultat predelave, prirejanja – ne pa sam program, aplikacija – pri *fotošopu*). Zlasti pri osebnih imenih ni redko, da je uslovarjena zgolj ena tvorjenka (navadno pridevniška izpeljanka, kot npr. pri *frančiškovski*). To pa pomeni, da je za apelativizacijo samostalnika potrebno mnogo več časa – če do nje sploh pride – kot za nastanek tvorjenk, pri katerih lastnoimenski samostalnik deluje kot v rabi (v smislu izlastnoimenskega leksema) dejansko neobstoječe besedotvorno izhodišče; lahko rečemo, da v jeziku zunaj lastnoimenske rabe živi zlasti prek svoje besedne družine. Prav tako je treba poudariti, da k bogati besedni družini v rabi prispeva po eni strani (dovolj) splošna dobra poznanost in razširjenost med uporabniki ter obenem obstoj predvsem v spletnem okolju, če ga spremlja določena dejavnost, kjer je mogoče njene udeležence (vršilca dejanja, rezultat dejanja, sámo dejanje ipd.) vse tvoriti znotraj iste besedne družine. To je vidno po trenutni odsotnosti slovarsko opisane besedne družine pri npr. dejansko apelativiziranem *bitcoinu*, ki je verjetno znan ožjim skupinam uporabnikom, kot to velja za večino družbenih omrežij ali spletnih platform, z njim povezana dejavnost pa je promet/vlaganje in ne morda **bitcoinanje*.

Z vidika izvora analiziranih leksemov so medtem zanimivi primeri, pri katerih je prehod iz lastnoimenskega v občnoimensko pomensko precej bolj oddaljen, kot npr. v primeru imen dveh profilov, ki sta bili izhodišči za občnoimenski poimenovanji, kjer eno označuje povsem fizično stvar, drugo pa občutje. *Knjigobežnice*, kot je (prvotno) ime profila na družbenem omrežju Facebook, tako danes označujejo hišice za izmenjavo rabljenih knjig, še večji odmik od prvotnega imena pa je doživel *srcozlom* (z dvojnico *srcezlom*), ki se je kot profil prvotno pojavil na Instagramu, danes pa označuje splošno občutje, povezano s koncem ljubezenske zveze. Pri slednjem seveda obstaja možnost, da se je v manj formalni komunikaciji v tem pomenu uporabljal že prej (*Sprotni slovar*

2 Na spletu lahko zasledimo še glagol *astroturfingirati*, a je raba zelo sporadična, zato se glagol v *Sprotni slovar* (še) ni uvrstil.

se sicer od tovrstne rabe ne ograjuje in jo ustrezno označeno slovarsko opisuje, vendarle pa je jedro njegovega opisa stilno bolj ali manj nevtralna leksika; prim. Krvina 2024: 147).

V prihodnje bi bilo zanimivo raziskati, kako je z obsegom besedne družine pri pojavih, katerih obstoj je vezan zlasti na spletno okolje, jih spremlja določena dejavnost in njeni udeleženci, ki jih je mogoče poimenovati znotraj iste besedne družine, obenem pa so dovolj splošno znani, v širšem okviru novejše leksike (ne le znotraj izlastnoimenske). Preveriti bo treba, ali je to pomemben dejavnik tudi na splošno ali vendarle manj kot v podskupini izlastnoimenske leksike, ki jo obravnavamo v prispevku. Podobno velja za težnjo po vključenosti v besedotvorne procese (v našem vzorcu okoli 60 % celotne leksike) in možno odsotnost besedotvornega izhodišča oz. če to obstaja, pomena, na katerega se navezujejo preostali člani besedne družine (npr. *jabolko* (še) ni razvilo pomena v zvezi z izdelki znamke Apple, *jabolčni* pa). To je zgolj eno od področij, kjer novejša izlastnoimenska leksika – kolikor lahko sklepamo iz analiziranega vzorca – izkazuje določene tendence, ki bi jih kazalo umestiti v okvir širših teženj znotraj novejše leksike nasploh.

Prispevek je nastal v okviru programa P6-0038, ki ga financira ARIS.

Viri in literatura

Atelšek, S., Fajfar, T., Jemec Tomazin, M., Sitar, J. in Žagar Karer, M. (2023). Mračenje. *Terminološka svetovalnica*. https://isjfr.zrc-sazu.si/sl/terminologisce/svetovanje/mracenje.

Brglez, M., Bajt, V., Pollak, S., Rot, Š. in Martinc, M. (2025). Od kamnitega do spletnega portala: Samodejno zaznavanje sprememb v rabi besed. *Prispevki za novejšo zgodovino*, *65*(3), 160–188. https://doi.org/10.51663/pnz.65.3.07

Crystal, D. (2001). *Language and the Internet*. Cambridge University Press.

Dobrovoljc, H. (2012). Pisanje imen izdelkov in znamk. V N. Jakop in H. Dobrovoljc (ur.), *Pravopisna stikanja: Razprave o pravopisnih vprašanjih* (str. 27–39). Založba ZRC, ZRC SAZU.

Dobrovoljc, H. in Michelizza, M. (2019). Slovaropisna obravnava znamk, imen izdelkov in zaščitenih poimenovanj. V H. Tivadar (ur.), *Slovenski javni govor in jezikovno-kulturna (samo)zavest* (Obdobja 38, str. 169–178). Znanstvena založba Filozofske fakultete. https://centerslo.si/wp-content/uploads/2019/10/Obdobja-38_Dobrovoljc_Michelizza.pdf

Gložančev, A., Jakopin, P., Michelizza, M., Uršič, L. in Žele, A. (2009). *Novejša slovenska leksika (v povezavi s spletnimi jezikovnimi viri)*. Založba ZRC, ZRC SAZU.

Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. (b. d.). *Slovarski portal Fran*. ZRC SAZU. www.fran.si

Kerremans, D., Stegmayr, S. in Schmid, H. (2011). The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. V K. Allan in J. A. Robinson (ur.), *Current methods in historical semantics* (str. 59–96). De Gruyter.

Kilgarriff, A. in Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics*, *29*(3), 333–347. http://www.mitpressjournals.org/doi/pdf/10.1162/089120103322711569

Korpus Janes. https://nl.ijs.si/janes/o-projektu/korpus-janes/

Korpus Trendi. https://sled.ijs.si/korpus-trendi/

Kosem, I., Čibej, J., Dobrovoljc, K., Kuzman, T. in Ljubešić, N. (2023). Spremljevalni korpus Trendi in avtomatska kategorizacija. *Slovenščina 2.0*, *11*(1), 161–188. https://doi.org/10.4312/slo2.0.2023.1.161-188

Kranjc, S. (2003). Jezik v elektronskih medijih. V A. Vidovič Muha (ur.), *Slovenski knjižni jezik – aktualna vprašanja in zgodovinske izkušnje* (Obdobja 20, str. 435–446). Center za slovenščino kot drugi/tuji jezik. https://centerslo.si/wp-content/uploads/2015/10/kranjc.pdf

Krvina, D. (2014–). *Sprotni slovar slovenskega jezika*. Založba ZRC, ZRC SAZU. https://doi.org/10.3986/2385-863X

Krvina, D. (2022). The growing dictionary of the Slovenian language (2014-) and Slovenian neologisms: Study on types of data and their use. *Slovenski jezik / Slovene Linguistic Studies 14*, 119–153. https://doi.org/10.3986/sjsls.14.1.05

Krvina, D. (2024): Sprotni slovar slovenskega jezika 2014–2023. V M. Jesenšek (ur.), *Pleteršnikova dediščina: Ob stoletnici smrti Maksa Pleteršnika* (Zora 154). Univerzitetna založba Univerze v Mariboru. https://doi.org/10.18690/um.ff.3.2024

Krvina, D. in Michelizza, M. (2026). *Sprotni slovar_izlastnoimenski* (1.0) [Zbirka podatkov]. Zenodo. https://doi.org/10.5281/zenodo.20121225

Lazić Konjik, I. (2024). Leksikografija u digitalnoj eri: Internet i društveni mediji kao izvori novih reči. *Novorečje*, *11*, 62–80.

Logar, N. (2004). Nove tehnologije in nekateri nesistemski besedotvorni postopki. V E. Kržišnik (ur.), *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti* (Obdobja 22, str. 121–132). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta.

Škrabal, M. in Kavka, M. (2021). Merging professional and collaborative lexicography: The case of Czech neology. *International Journal of Lexicography*, *34*(3), 282–301.

Od ločevalne norme do čustvene pristnosti: jezikovna sredstva afektivnosti v informativnih novinarskih besedilih

DOI: 10.4312/Obdobja.45.229-237

Tina Lengar Verovnik, Fakulteta za družbene vede, Univerza v Ljubljani, Ljubljana; Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana; tina.verovnik@fdv.uni-lj.si

Monika Kalin Golob, Fakulteta za družbene vede, Univerza v Ljubljani, Ljubljana; monika.kalin-golob@fdv.uni-lj.si

Prispevek se osredotoča na jezikovni vidik afektivnosti, ki v sodobnem poročevalstvu dopolnjuje tradicionalni ideal norme objektivnosti. Avtorici pokažeta, kako se čustveno vrednotenje vpisuje v leksikalne izbire, skladnjo, metaforiko in pragmatične strategije. Posebna pozornost je namenjena vlogi afektivnih jezikovnih vzorcev pri konstruiranju pomena in pri vzpostavljanju odnosa z občinstvom. Metodološko je prispevek zasnovan kot kvalitativna jezikovno-stilistična analiza spletnih in radijskih informativnih novinarskih besedil.

afektivnost, informativni novinarski žanri, novinarski diskurz, novinarske norme, jezikovno-stilna analiza

This article focuses on the linguistic aspect of affectivity, which complements the traditional ideal of the norm of objectivity in contemporary news reporting. The authors demonstrate how emotional evaluation is inscribed into lexical choices, syntax, metaphors, and pragmatic strategies. Particular attention is paid to the role of affective linguistic patterns in the construction of meaning and establishing a relationship with the audience. Methodologically, the article is designed as a qualitative linguistic-stylistic analysis of online and radio news texts.

affectivity, news genres, journalistic discourse, journalistic norms, linguistic-stylistic analysis

1 Uvod

Demokratizacija slovenske družbe in medijev po letu 1991 je korenito spremenila novinarsko sporočanje: tradicionalne medije so dopolnili tabloidi s senzacionalizmom in informativna zabava. Ločevalna norma – ločevanje dejstev in mnenj ter posledično dve skupini novinarskih žanrov (poročevalni in presojevalni) – je odtlej postopoma postajala vse bolj fluidna. Uporabniki smo vedno bolj omreženi, novice dobivamo hitro, nenehno in po različnih poteh, do dogajanja v svetu imamo bolj oseben odnos kot kadarkoli prej. Multimedialnost, hipertekstualnost, interaktivnost, arhivskost in figurativnost sodobnega novinarstva (Deuze 2007) so ob novih novinarskih praksah, ki jih je omogočila digitalizacija v tehnološkem smislu, prinesle tudi upovedovalne spremembe:

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

ozgodbenje novic prispeva k vse bolj osebnemu in ekspresivnemu tonu ter posledično čustvenemu sprejemanju. Vse to je pripeljalo do afektivnega preobrata (Papacharissi 2015) ali čustvenega preobrata (Wahl-Jorgensen 2019) v komunikologiji, medijskih in novinarskih študijah: čustva in občutja so postala osrednja tema raziskav o medijskem sporočanju in vpletenosti občinstva. V prvem delu prispevka predstavljava, kaj to pomeni glede na prevladujoči ideal norme objektivnosti, in na podlagi tujih raziskav predlagava, kako novi zorni kot prenesti v kvalitativno jezikovno-stilistično raziskovanje slovenskega novinarskega jezika. Nedavna raziskava (Pajnik 2023) namreč kaže, da je čustvena vpletenost vse bolj sprejemljiva tudi med slovenskimi novinarji. S pilotno analizo besedil informativnih žanrov, objavljenih v spletnih in radijskih medijih ob mednarodnem dnevu žensk, na 8. marec 2026, nato ugotavljava, ali je v njih mogoče zaslediti jezikovna sredstva afektivnosti.

2 Normativne zahteve pri informativnih žanrih in afektivnost kot inherentna sestavina novinarskega sporočanja

Novinarstvo v nepreglednem toku informacij in odzivov nanje poskuša ohranjati dodano vrednost s kritičnim, neodvisnim, konstruktivnim pristopom, vendar postaja jasno, da mora obenem graditi tudi pristen odnos z občinstvom, temelječ na avtentičnih čustvih in občutjih. To je vsaj na prvi pogled neskladno z bistvom novinarske profesionalne ideologije, ki že od zgodnjega 20. stoletja temelji na normi objektivnosti. Za informativne žanre velja, da jih je treba ustvarjati z največjo mero nepristranskosti, dejstvenosti, natančnosti, odgovornosti, kar naj bi se kazalo tudi v neosebnosti in odsotnosti čustev. Ta so sicer – kot opozarja Peters (2011) – že od nekdaj vgrajena v novinarski diskurz, saj jih je treba uravnavati v vseh družbenih situacijah: tudi najbolj distancirano branje vesti na radiu je v resnici odraz hladne čustvene poze, kar pa ni enako kot brezčutnost.

Analize novejših oblik novinarstva se še vedno pogosto ujamejo v dihotomijo čustva – razum; afektivnost po mnenju nekaterih ogroža normativne ideale, je znak upadanja novinarskih standardov, povezuje se jo s senzacionalizmom, komercializacijo in dramatizacijo, kar naj bi spodkopavalo zaželeno novinarsko družbeno vlogo (Peters 2011; Stenvall 2014; Wahl-Jorgensen 2019). A dejansko objektivnost ostaja temelj novinarske profesionalnosti, afektivnost pa se ji v dobi hibridne medijske realnosti zgolj pridružuje kot prav tako nujna novinarska vrednota (Kotišová 2019). Pri tem ne gre za konfliktnost, temveč soobstajanje novinarskih standardov, nujnih za opozarjanje občinstva na zadeve javnosti (Pajnik 2023). Povedano drugače: vključevanje čustev in občutij je del novinarskega prizadevanja po prikazovanju realnosti, kakršna je; brez njih ni mogoče biti povsem resnicoljuben (Pantti 2010), torej so ob njihovem upoštevanju novinarji pri iskanju resnice dejansko celo bolj objektivni, pa tudi bolj pošteni (Schmidt 2021). Tudi čustva imajo namreč informativno vrednost sama po sebi. Lahko sicer razdvajajo, a tudi združujejo in mobilizirajo sicer fragmentirano javno sfero (Yell 2012). Afektivnost postaja novinarska norma, podvržena (samo)regulaciji (Schmidt 2021): slovenski novinarji med drugim navajajo, da mora biti rabljena premišljeno in reflektirano,

ne sme se je zlorabljati za diktiranje občinstvu, kaj naj misli, kot sredstvo dramatizacije ali za pridobitne namene (Pajnik 2023).

3 Jezikoslovno raziskovanje afektivnosti v novinarskih besedilih

Čustva in občutja so lahko vgrajena v jezikovno sporočilo na vseh ravneh, kar pa ne pomeni, da jih je mogoče zreducirati zgolj na jezikovne ali diskurzivne prakse tvorca besedila; afektivnost je bolj vezana na medosebne odnose in čustva kot družbeni oz. odnosni pojav, ne toliko na čustva kot osebni oz. psihološki pojav (Koivunen idr. 2021). Pri dosedanjem jezikoslovnem raziskovanju afektivnosti v jeziku novinarskih besedil prevladujejo korpusne analize leksike in kvalitativne analize besedil. Pri prvih gre za raziskovanje afektivnega besedišča, običajno po vnaprej oblikovanih seznamih (npr. Absattar idr. 2022; Brett in Pinna 2013), pri drugih pa za večplastne analize posameznih diskurzov (npr. Schmidt 2021). Oba tipa raziskovanja sta lahko tudi združena, kar prinaša celovitejše ugotovitve o prevedbi čustev in občutij v jezikovno-stilna sredstva različnih novinarskih žanrov (npr. Koivunen idr. 2021).

Po pregledu relevantnih tujih raziskav (Alonso-Belmonte 2026; Koivunen idr. 2021; Stenvall 2008; Yell 2012) lahko kvalitativne metode ekscerpiranja in analize prvin, ki neposredno ali posredno kažejo na afektivnost, v grobem razdelimo na 1) klasične jezikoslovno-stilistične (semantika, skladnja, aktualizacija) ter 2) raziskave vrednotenja v modelu sistemsko-funkcijskega jezikoslovja[1] in na njem temelječe teorije vrednotenja.

V sklopu kvalitativnih jezikoslovno-stilističnih raziskav so avtorice in avtorji pozorni na naslednje prvine: besedišče, ki eksplicitno označuje čustva/afektivnost ali čustveno vedenje (samostalniki, glagoli, ob njih pridevniki in prislovi, frazemi, medmeti – npr. *veselje*, *razočarati*, *vesel*, *zlomiti srce*, *oh*); emocionalno nabiti izrazi (npr. *genocid*, *terorist*); podrobno opisovanje; ponavljanja; metafore, primere, analogije (Alonso-Belmonte 2026); jukstapozicija, bližina metaforičnih izrazov v hitrem zaporedju; slovnične in leksikalne zgradbe, ki izražajo zadržek ali prepričanje oz. določajo stopnjo obojega (sredstva leksikalnega kritja – angl. *hedging*, npr. *mogoče je*, *verjetno*); vrednotenjski izrazi (npr. *velika izguba*); sklicevalnost (npr. *domnevno*, *naj bi*); kopičenje različnih afektivnih zaznamovalcev (Koivunen idr. 2021); retorika kontrasta: sopostavljanje nasprotij (npr. *prej – pozneje*); opisi neobičajnega vedenja (npr. *tiha množica ljudi*); nominalizacije (npr. *veseliti se* > *veselje*), zaradi katerih čustvo postane popredmeteno (zabriše se, kdo ga izkuša) (Stenvall 2008).

V okviru sistemsko-funkcijskega pristopa pa avtorice in avtorji raziskujejo, kako tvorci presojajo, vrednotijo predmetnost, prostor, osebe (Alonso-Belmonte 2026); kako navajajo izjave (pripisovanje čustev in občutenj zunanjim virom in protagonistom): s citati se afektivnost posreduje, preoblikuje, uravnava (Koivunen idr. 2021); raziskujejo

1 Glavni predmet tega so izrazne možnosti za konstruiranje občih vrednostnih pomenov, tj. jezikovna sredstva, s katerimi izražamo mnenja, ocene, čustvene odzive na predmet percepcije, hotenja oz. gotovost/dvome v zvezi z njim itn. (Iedema idr. 1994; Thompson in Huston 2000). V klasičnem novinarstvu so te prvine del presojevalnih (interpretativnih) žanrov (komentar, članek, glosa).

t. i. diskurz čustev: vpisanost čustev v jezikovne prvine ali njihov priklic z naborom sicer vrednotno nezaznamovane leksike (Yell 2012).

V konkretnih besedilih se sredstva prvega in drugega sklopa kompleksno povezujejo, okvirjajo dogodke in dogajanja, tako da gre pogosto ne le za rabo posameznih sredstev, temveč tudi njihovo neposredno bližino ali kopičenje, kar ustvarja še večjo afektivnost.

4 Mednarodni dan žensk v informativnih besedilih spleta in radia

Diskurzivne strategije afektivnosti se v informativnih žanrih uporabljajo tudi za povečanje občutljivosti občinstva za določene tematike (Alonso-Belmonte 2026). V pilotni raziskavi sva zato analizirali prispevke o mednarodnem dnevu žensk, saj gre za temo, ki se v sodobnosti prekriva in izgublja v interpretaciji ženskih tem ob številnih materinskih dnevih. Kot pišeta Ljajić in Dojčinović (2020), ni nepomembno, ali mediji o prazniku poročajo v duhu njegove izvorne pomembnosti zaradi boja za enakost in krepitve družbene občutljivosti ali pa ga predstavljajo zgolj na ravni praznovanja in navidezne naklonjenosti ženskam.

V analizo sva zajeli spletna in radijska informativna besedila, ki so bila objavljena 8. 3. 2026. Spletna besedila so s prvih petih najbolj obiskanih spletnih poročevalskih portalov v tem letu (MOSS, https://www.moss-soz.si/rezultati/): 24ur.com, siol.net, rtvslo.si, delo.si in svet24.si. Na naslovni strani sva med 9. in 10. uro poiskali prispevek na temo dneva žensk; ker na 24ur takega prispevka ni bilo, sva analizirali štiri spletne prispevke. Radijska besedila so iz dnevnoinformativnih oddaj (med 5. in 22. uro) programov javnega servisa (Prvi, Val 202, Ars, Maribor, Koper; 30 prispevkov), dveh komercialnih radijskih mrež (Infonet, Svet24; 9 prispevkov) in enega regionalnega programa posebnega pomena (Radio Murski val; 2 prispevka).[2] S kvalitativno metodo ročnega izpisovanja sva zbrali primere, za katere sva ocenili, da izražajo afektivnost v načinu izbire jezikovnih, stilnih in drugih vrednotenjskih sredstev.

4.1 Spletna besedila

Raziskava spletnih besedil je pokazala, da so v sicer informativnih besedilih nastopala v literaturi navedena sredstva afektivnosti, vendar je vsak od štirih prispevkov gradil na svojih strategijah:

a) Siol.net na naslovnici prinaša protivno zgrajen aktualizirani naslov. Pod sliko para je zapisano *Tipična Slovenka zasluži manj, a vrh lestvice najbogatejših je njen*. Po kliku je celoten prispevek naslovljen z informativno-stališčnim interpunkcijskim naslovom *Razlika ostaja: tipična Slovenka zasluži okoli sto evrov manj kot moški* (Mihaela Leskovec, 8. 3. 2026, ob 9.18). Nepravi sinopsis ubesedi povprečno Slovenko z metaforo … *statistika razkriva zanimiv portret povprečne prebivalke Slovenije*, ob koncu pa kot protipostavitev navede najbogatejšo med Slovenci, in sicer z metaforo *na vrhu kraljuje prav ženska – Vesna Južna*. Nadaljnje besedilo je poročilo o pogostih žen-

2 Preostala regionalna programa posebnega pomena, ki pripravljata lasten informativni program (Radio Štajerski val, Primorski val), o dnevu žensk nista poročala.

skih imenih, povprečni plači, najbogatejših Slovenkah in primerjava z EU, brez sredstev afektivnosti. Afektivnost je tako del naslovja, ki je po definiciji (Korošec 1998) najopaznejši in najbolj izpostavljen del besedila, zato (sploh na spletu; prim. Kalin Golob 2025) vključuje prvine, ki vabijo k nadaljnjemu branju.

b) Rtvslo.si na naslovnici prinaša nadnaslov (*Mednarodni dan žensk*) in okrnjeni naslov (*ZN: Državne zakonodaje ženskam zagotavljajo le 64 odstotkov pravic, ki j…*) ter del sinopsisa.

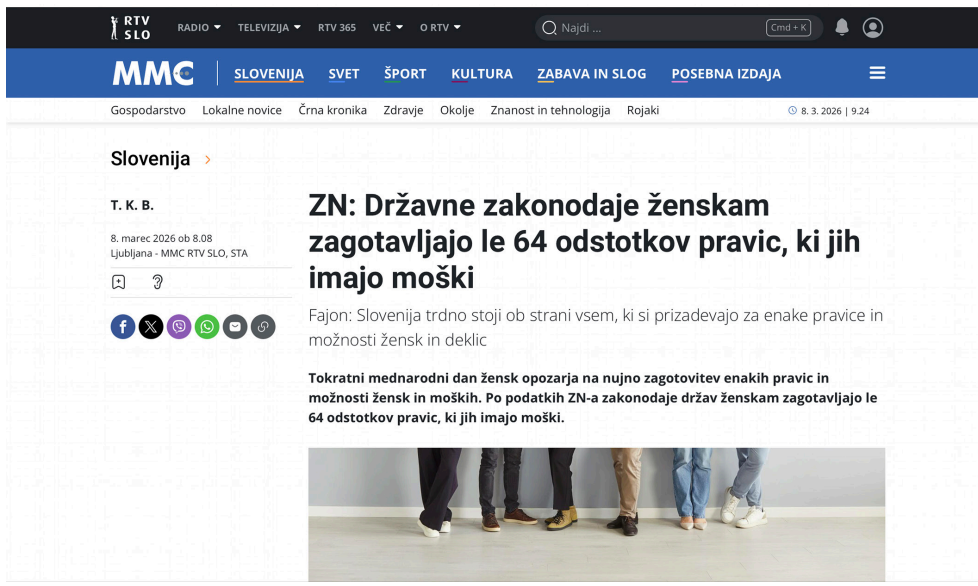


Slika 1: Naslovnica portala MMC z nadnaslovom, okrnjenim naslovom in delom sinopsisa.

Po kliku dobimo celoten naslov *ZN: Državne zakonodaje ženskam zagotavljajo le 64 odstotkov pravic, ki jih imajo moški* (T. K. B., 8. 3. 2026, ob 9.24), sledi podnaslov: izjava ministrice (*Fajon: Slovenija trdno stoji ob strani vsem, ki si prizadevajo za enake pravice in možnosti žensk in deklic*). Sinopsis gradi na afektivnosti z rabo glagola *opozarjati*, pridevnikom *nujno* in členkom *le* (*Tokratni dan žensk opozarja na nujno zagotovitev enakih pravic in možnosti žensk in moških. Po podatkih ZN-a zakonodaje držav ženskam zagotavljajo le 64 odstotkov pravic, ki jih imajo moški*). Nadaljnje besedilo afektivnost gradi z izjavami varuhinje človekovih pravic (ki dan žensk poveže z opozorilom o ***nasilju*** nad ženskami), ministra Mesca in ministrice Fajon. Izjave so polne opaznih glagolov in frazemov (***trdno stati*** *ob vseh*), navedena je celotna izrazito retorično s paralelizmom zgrajena izjava ministra Mesca (tri povedi se začnejo z glagolom *pomeniti*).[3] Nadaljnje novinarsko besedilo je nevtralno, napoveduje dogodke ob praznovanju

3 »Varen svet pomeni svet brez nasilja doma, na ulici in na spletu. Pomeni dom, v katerem ženska ne živi v strahu. Pomeni delovno mesto, kjer njeno delo ni vredno manj zato, ker je ženska. Pomeni skupnost, ki ne gleda stran, ko se dogaja krivica ali nasilje.«

dneva žensk in ob koncu pojasni, od kdaj in zakaj ga praznujemo. Novinar(ka) afektivnost gradi v naslovju in besedilu s skrbno izbranimi ekspresivnimi izjavami političnih osebnosti in jim preda ustvarjanje čustvenega okvira.



Slika 2: Besedilo prispevka s polnim naslovom, podnaslovom in sinopsisom.

c) Delo.si na naslovnici prinaša izjavni veliki naslov (*8. marec: ženskam je zagotovljeno le 64 odstotkov pravic, ki jih imajo moški*) in podnaslov, ki povzema izjavo varuhinje človekovih pravic s komentirajočim glagolom ***opozarjati***) (avtor ni naveden, vir STA, 8. 3. 2026, ob 9.27).

Besedilo v poročanem govoru navaja besede varuhinje, Evropske komisije in ministra Mesca, pri čemer dobesedno navaja enak del izjave s paralelizmom. Novinar oz. novinarka nadaljuje s povzemanjem dogodkov ob dnevu žensk in pojasnilom, od kdaj in zakaj ga praznujemo. Afektivnost je manj opazna kot v prejšnjem prispevku, saj poročani govor omili čustvenost prej dobesedno navedenih izjav, a je ustvarjena z močnim poudarkom o nasilju nad ženskami, glagolom *opozarjati* in navajanjem nujnosti (*nujno zagotoviti*). Retorična izjava ministra Mesca je čustveno nabita, zato je navedena v celoti.

č) Svet24.si na naslovnici in v prispevku prinaša tematski nadnaslov (*Dan žena*) in aktualizirani interpunkcijski veliki naslov (*Od delavskih protestov do šopkov rož: Kaj v resnici pomeni 8. marec?*). Sinopsis takoj vzpostavi in stopnjuje v naslovu nakazano didaktičnost: *Dan žensk ni le praznik pozornosti, temveč spomin na boj za pravice. Danes ga povezujemo s cvetjem, a je njegov izvor veliko bolj resen* (Petra Znoj, 8. 3. 2026, ob 9.30). Besedilo nadaljuje protivno gradnjo povedi in vključuje afektivnost na ravni leksike (*8. marec ostaja tudi* ***opomnik***); opazno čustvenost daje tudi zanikanje (*Dan žena ni nastal kot romantičen praznik*.). Dolg prispevek, ki vsestransko predstavlja

položaj žensk nekoč in danes ter pojasnjuje pomembne elemente enakosti spolov, je opremljen z mednaslovi, ki sicer pojasnjevalno zapisanemu dajejo čustveno noto z zgradbo (*Od protestov delavk do mednarodnega praznika*), postavljanjem vprašanj (*Praznik rož ali opomin družbi?*; *Ali 8. marec še vedno potrebujemo?*) ali leksikalno (*Pravice, ki so danes **skoraj** samoumevne*).

4.2 Radijska besedila

Raziskava radijskih besedil je pokazala, da med poročanjem na različnih tipih programov (javni, komercialni, posebnega pomena) ni bistvenih razlik, zato so ugotovitve v nadaljevanju predstavljene za celoto. Ker je bil 8. marec v letu raziskave na nedeljo – kar pomeni okrnjena uredništva –, je precej besedil tudi prekrivnih (predvidoma so povzeta po STA) ali se pojavijo v več programih javnega servisa (notranje sodelovanje). Za razliko od spletnih prispevkov so med obravnavanimi tudi taki, ki dan žensk obravnavajo zgolj na ravni praznovanja. Vsi ti so afektivno pozitivno zaznamovani, jezikovno bodisi eksplicitno (*v ljubljanski Drami bo dan žensk zelo čustveno nabit*; *prireditve, polne glasbe in plesa*) ali pa s posredno izraženo naklonjenostjo ženskam (z izvzemalnim členkom oz. rabo ženskih oblik: *na predstavi bodo nastopile **samo** jahačice*; ***obiskovalke** si bodo lahko ogledale [...]*). Na ravni deklarativne naklonjenosti je tudi uvod v prispevek, ki gradi na naštevanju, a namesto krepitve enakosti izzveni v stereotipizacijo: *Ženske so tudi na kmetijah nepogrešljive, pa najsi gre za skrb za družino in dom ali pa za delo v hlevih in na poljih, pri katerem prav tako kot moški znajo ravnati z različno mehanizacijo. Nenazadnje zelo dobro poznajo potrebne zakonske predpise, obvladajo administrativno delo in tako dalje*.

V vseh drugih prispevkih gre za poročanje o dnevu žensk v duhu prvotnega pomena praznika, ki je nekajkrat metaforično poimenovan kot ***boj***. Tudi tukaj je afektivnost zaznavna, a je le v enem prispevku okvirjena pozitivno, pa še tu skozi nominalizacijo, ki postavlja udeleženke procesa v ozadje: *Na zunanjem ministrstvu [...] pravijo, da so okrepili **zastopanost žensk** na vodstvenih položajih ter se zavezali feministični zunanji politiki*. Drugod afektivnostni okvir tvorijo leksikalne prvine, ki jih povezujemo z negativnimi čustvi, zlasti pogosti so samostalniki: največkrat *nasilje*, ob njem pa še *grožnje*, *spolno nadlegovanje*, *strah*, *poniževanje*, *izključenost*, *diskreditacija*, *izsiljevanje*, *utišanje*. Nabor razširjajo *oboroženi konflikti*, *krepitev avtoritarnih režimov*, *vojna*, *neenakost*, ki se jim kot mogoča rešitev v protistavi pojavlja solidarnost: »*Moramo ženske trdno držat skupaj, sicer tudi z našimi fanti in možmi. Ker če ne bomo skupaj držali, ne bo nič*.«[4]

V prispevkih je nadpovprečno veliko poročanega govora in izjav, oboje veča afektivnost. Malo je izjav, ki so eksplicitno afektivne, npr. metaforični nagovor predsednice republike: »*Tudi me, drage ženske, stojimo danes na ramenih naših pogumnih predhodnic, ki so spregovorile, ko jim je bilo rečeno, naj molčijo, ko so vstopile v prostore, kjer niso bile zaželene, ki so z vztrajnostjo dosegle, da so se odprla do tedaj zanje zaprta vrata, ter ki so soustvarjale znanost, kulturo in prispevale k razvoju družbe*.« Tudi neka-

4 Z narekovaji so označene zvočne izjave.

tere druge izjave so izbrane zaradi učinka, v naslednji je dosežen s ponavljanjem in kratkostjo povedi: »*Pol nas je v družbi, pol. Še malo več kot pol. Torej imamo pravico na svojo polovico moči*.« Največ pa je poročanega govora in izjav, ki vsebinsko prinašajo strokovne podkrepitve in podatke, npr. »*[...] opozarja, da kar 40 % mladih moških meni, da so moški bolje usposobljeni za delo v politiki kot ženske*« ali »*Na podeželju v Sloveniji živi denimo še danes 10 % žensk, ki niti nimajo svojega lastnega TRR-ja*«.

Tudi sicer je diskurz strokovnosti močno prisoten, veliko je sklicevanja na raziskave, analize (npr. *V državah Evropske unije je skoraj tretjina žensk doživela neko obliko nasilja*). V edinem poročilu (vse drugo so vesti ali napovedi prispevkov) se tridelno jedro netipično za informativni novinarski žanr začenja s *prvič*, *drugič*, *tretjič*, kar je aktualizacija vzorca iz strokovnih besedil. Vse te prvine pa za razliko od oglaševanja, kjer aktualizacija znanstvenega diskurza služi za doseganje pozitivnega učinka, delujejo ravno nasprotno. Občutje nelagodja, ki ga vzbujajo statistični podatki, je leksikalno podkrepljeno z zvezami, ki nakazujejo nezadostnost, celo jalovost preteklega *boja* in negotovost prihodnjega: osmi marec je *opomin na dolg seznam nerazrešenih družbenih problemov*; *delo za enakost še zdaleč ni končano*; *nasilje nad ženskami ostaja trajen družbeni problem*. Čeprav v analiziranih besedilih – razen v izjavah in poročanem govoru – ne prevladuje izrazito ekspresivna ali vrednotenjska leksika, lahko torej tudi za radijske prispevke trdimo, da s povezovanjem različnih sredstev afektivnosti in okvirjanjem v aktualizirani diskurz strokovnosti čustveno vpletajo občinstva.

5 Zaključek

Kljub majhnemu ter tematsko in časovno zamejenemu vzorcu se je analiza informativnih novinarskih žanrov skozi prizmo afektivnosti pokazala kot pristop, ki bi ga bilo vredno preizkusiti na večjem korpusu besedil. Omogoča namreč preseganje analitičnih binarnosti (npr. objektivno – subjektivno, informacija – interpretacija, poročanje – presojanje), ki se v sodobnosti pogosto zdijo vsiljene in jih novinarska besedila ne odražajo. Raziskovanje afektivnosti bi bilo smiselno zastaviti tudi multimodalno, kar je obetaven, a še slabo raziskan vidik (Alonso-Belmonte 2026). Pri radijskih besedilih se je denimo na besedilnofonetični ravni pokazalo, da lahko različno branje istega besedila izpostavi drugačne vsebinske poudarke (nekatera besedila so se ponovila na več programih javnega servisa, a so jih brali različni posamezniki). Podobno kot pri radijskih besedilih prozodijo pa bi bilo treba pri spletnih besedilih v analizo pritegniti še fotografijo, tipografijo, barve in druge prvine multimodalnosti pisnih novinarskih besedil.

Prispevek je nastal v okviru raziskovalnih programov Slovenski jezik v sinhronem in diahronem razvoju (P6-0038) ter Slovenski jezik – bazične, kontrastivne in aplikativne raziskave (P6-0215), ki ju sofinancira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije.

Literatura

Absattar, A., Mambetova, M. in Zhubay, O. (2022). The potential of emotive language to influence the understanding of textual information in media coverage. *Humanities and Social Sciences Communications, 9*. https://doi.org/10.1057/s41599-022-01232-2

Alonso-Belmonte, I. (2026). The role of emotions in journalism: A linguistic and multimodal perspective. V M. T. Soto-Sanfiel in V. Salojärvi (ur.), *The handbook of journalism and emotion* (str. 245–354). John Wiley & Sons.

Brett, D. in Pinna, A. (2013). The distribution of affective words in a corpus of newspaper articles. *Social and Behavioral sciences*, *95*, 621–629.

Deuze, M. (2007). The web and its journalism: Considering the consequences of different types of newsmedia online. *New Media & Society*, *5*(2), 203–230.

Iedema, R., Feez, S. in White, P. R. R. (1994). Appraisal in journalistic discourse. http://grammatics.com/appraisal/MediaLit-Appraise.pdf

Kalin Golob, M. (2025). Naslovje in naslovi novinarskih besedil v digitalni dobi: Napovedniki, dražilniki, klikolovniki. V I. Vobič in J. Amon Prodnik (ur.), *Novinarstvo na prelomnicah* (str. 307–335). Fakulteta za družbene vede, založba FDV.

Koivunen, A., Kanner, A., Janicki, M., Harju, A., Hokkanen, J. in Mäkelä, E. (2021). Emotive, evaluative, epistemic: A linguistic analysis of affectivity in news journalism. *Journalism*, *22*(5), 1190–1206.

Korošec, T. (1998). *Stilistika slovenskega poročevalstva*. Kmečki glas.

Kotišová, J. (2019). The elephant in the newsroom: Current research on journalism and emotion. *Sociology Compass*, *13*(5). https://doi.org/10.1111/soc4.12677

Ljajić, S. in Dojčinović, D. (2020). Presentation of the International Women's Day in the Serbian Press. *Media Studies and Applied Ethics*, *1*(2), 89–107.

Pajnik, M. (2023). Professionalizing emotions as reflective engagement in emerging forms of journalism. *Journalism Studies*, *25*(2), 181–198.

Pantti, M. (2010). The value of emotion: An examination of television journalists' notions on emotionality. *European Journal of Communication*, *25*(2), 168–181.

Papacharissi, Z. (2015). Toward new journalism(s): Affective news, hybridity, and liminal space. *Journalism Studies*, *16*(1), 27–40.

Peters, C. (2011). Emotion aside or emotional side? Crafting an 'experience of involvement' in the news. *Journalism*, *12*(3), 297–316.

Schmidt, T. R. (2021). 'It's OK to feel': The emotionality norm and its evolution in U. S. print journalism. *Journalism*, *22*(5), 1173–1189.

Stenvall, M. (2008). On emotions and the journalistic ideals of factuality and objectivity—Tools for analysis. *Journal of Pragmatic*, *40*, 1569–1586.

Stenvall, M. (2014): Presenting and representing emotions in news agency reports. *Critical Discourse Studies*, *11*(4), 461–481.

Thompson, G. in Hunston, S. (2000): Evaluation: An introduction. V S. Hunston in G. Thompson (ur.), *Evaluation in text: Authorial stance and the construction of discourse* (str. 1–27). Oxford University Press.

Wahl-Jorgensen, K. (2019). An emotional turn in journalism studies?. *Digital Journalism*, *8*(2). 175–194.

Yell, S. (2012). Natural disaster news and communities of feeling: The affective interpellation of local and global public. *Social Semiotics*, *22*(4), 409–428. https://doi.org/10.1080/10350330.2012.693295

Intercultural Linguistic Challenges in the Twenty-First Century: A Didactic Approach to Autonomous Learning of Slovenian with the Support of an AI Tutor

DOI: 10.4312/Obdobja.45.239-247

Marko Ljubešić, Filozofski fakultet, Sveučilište Jurja Dobrile u Puli, Pulj; mljubesic@unipu.hr

Prispevek obravnava medkulturne jezikovne izzive učenja slovenščine v 21. stoletju s poudarkom na samostojnem učenju ob podpori umetne inteligence. Izhajajoč iz sodobnih družbenih in izobraževalnih sprememb, poudarja potrebo po fleksibilnih in medkulturno občutljivih didaktičnih pristopih, zlasti v obmejnih območjih. V osrednjem delu je predstavljen didaktični pristop k samostojnemu učenju slovenščine z UI-tutorjem, namenjen učencem in odraslim učečim se iz Istre in drugih delov Hrvaške, ki jih s slovenskim prostorom povezujejo tesne kulturne in jezikovne vezi. Model temelji na cikličnem procesu učenja, ki vključuje jezikovno recepcijo, kontrastivno refleksijo, vodeno produkcijo in metakognitivno samoevalvacijo. Prispevek ugotavlja, da UI-tutor predstavlja učinkovito didaktično orodje za spodbujanje avtonomije in medkulturne sporazumevalne zmožnosti.

medkulturna komunikacija, samostojno učenje, slovenščina kot tuji jezik, umetna inteligenca v izobraževanju, didaktični pristop

This article examines intercultural linguistic challenges in learning Slovenian in the context of the twenty-first century, with a focus on autonomous learning supported by artificial intelligence. Drawing on contemporary social and educational developments, the study highlights the need for flexible and interculturally sensitive didactic approaches, particularly in border regions. The central part presents a didactic approach to autonomous learning of Slovenian with the support of an AI tutor, intended for students and adult learners from Istria and other parts of Croatia, who share close cultural and linguistic ties with the Slovenian context. The model is based on a cyclical learning process that includes linguistic reception, contrastive reflection, guided production, and metacognitive self-evaluation. The article concludes that an AI tutor is an effective didactic tool for fostering learner autonomy and intercultural communicative competence.

intercultural communication, autonomous language learning, Slovenian as a foreign language, artificial intelligence in education, didactic approach

1 Introduction

Foreign language learning in the twenty-first century is taking place in the context of rapid social, technological, and cultural changes that significantly influence traditional educational paradigms. The digitalization of education, increased population mobility, and intensive intercultural contacts have led to the need for more flexible, accessible, and individually tailored models of language learning. A particularly important role in this process is played by the development of artificial

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

intelligence (AI), which is increasingly used in education as a means of personalizing learning, providing feedback, and supporting learner autonomy (Luckin et al. 2016: 19–21).

In contemporary foreign language teaching, the focus is gradually shifting from a transmission-based teaching model to models that place the learner at the center of the learning process. Learner autonomy is not understood as isolated self-study, but as the ability to plan, regulate, and evaluate one's own learning in interaction with appropriate didactic tools (Little 1991: 4–5). In this context, digital technologies, and especially AI-based systems, offer new opportunities for the development of metacognitive and metalinguistic awareness because they allow continuous feedback, reflection, and adaptation of learning content to learners' individual needs (Holmes et al. 2019: 36–38). Research in computer-assisted language learning (CALL) indicates that digital and mobile technologies can significantly improve language learning, particularly in terms of personalization, adaptive learning, and simulation of communicative situations (Godwin-Jones 2018: 3–5).

The intercultural dimension of language learning is gaining additional importance in European border regions, where languages and cultures have long been in contact. In such environments, linguistic proximity can represent both an advantage and a challenge because similarities between languages often lead to interference, incorrect generalizations, and pragmatic misunderstandings (Byram 1997: 31–33). Language learning in a culturally close context therefore requires a specific didactic approach that systematically develops intercultural communicative competence, rather than focusing solely on formal linguistic knowledge.

Based on these theoretical and social premises, this article focuses on developing a didactic approach to autonomous learning of Slovenian with the support of an AI tutor, primarily intended for students and adults from the border region of Istria and other parts of Croatia. This context is characterized by close cultural, historical, and linguistic ties, and also by specific communicative needs in everyday and professional life. The proposed model is based on the assumption that an AI tutor does not replace a teacher, but instead acts as a linguistic and metacognitive mediator that encourages reflection, autonomy, and intercultural understanding in learning (Holmes et al. 2019: 58–60). Such a didactic approach may be particularly relevant as a supplement to systematically teaching Slovenian as an elective subject.

2 Theoretical and didactic aspects of learning a closely related language in an AI environment

2.1 Intercultural communicative competence and language learning in a border context

In contemporary philology, foreign language learning can no longer be viewed solely as a process of mastering grammatical structures and vocabulary, but rather as a complex process of developing intercultural communicative competence. According to Byram's model, intercultural competence encompasses knowledge,

skills, attitudes, and critical cultural awareness that allow successful communication between members of different cultural communities (Byram 1997: 34–36). In border regions, where languages and cultures have long been in contact, this competence gains additional complexity because communication often presupposes mutual understanding based on cultural and linguistic proximity.

In the context of Istria and the Croatian–Slovenian border area, learning Slovenian takes place in an environment in which learners already possess a certain degree of implicit knowledge of the language and culture, acquired through media, everyday interactions, and regional history. However, such proximity often leads to interference, pragmatic errors, and false assumptions about the meaning and use of linguistic forms. For this reason, an intercultural approach to language learning must include systematic reflection on differences, not only similarities, and foster the ability to critically examine one's own linguistic and cultural patterns (Byram 1997: 41–43).

2.2 Learner autonomy and independent language learning with digital support

Learner autonomy is one of the key concepts in contemporary language teaching. Little defines autonomy as the learner's ability to take responsibility for his or her own learning, including planning, implementation, and evaluation of the learning process (Little 1991: 4). In this sense, independent learning does not imply the absence of structure, but rather the existence of a didactically designed framework that enables the learner to participate actively and reflectively in the learning process.

Digital technologies, and especially systems based on artificial intelligence, significantly expand the possibilities for developing autonomy. AI tools provide continuous access to learning content, individualized feedback, and the adaptation of task difficulty, thereby enabling learners to study in accordance with their own needs and pace (Holmes et al. 2019: 35–37). Such systems are particularly suitable for adult learners and those outside the formal education system, who often learn a language alongside professional and family obligations.

In the context of learning Slovenian, an autonomous approach supported by AI allows learners to focus on the linguistic elements that are most relevant for communication, and the AI tutor can direct attention to typical contrastive difficulties between Croatian and Slovenian. In this way, learner autonomy is not reduced to individualism but is shaped as a guided process of learning in interaction with an intelligent didactic system.

2.3 Metalinguistic and metacognitive awareness in learning closely related languages

Metalinguistic and metacognitive awareness are fundamental cognitive dimensions of successful language learning, especially in the context of learning closely related

languages. Metalinguistic awareness refers to the ability to reflect on language as a system, including the understanding of grammatical rules, linguistic categories, and contrastive differences between languages (Gombert 1992: 13–15). In learning Slovenian among Croatian speakers, this awareness is manifested in the ability to recognize differences in the case system, the use of verb forms, and pragmatic conventions.

Metacognitive awareness, on the other hand, refers to awareness of one's own learning processes, including planning, monitoring, and evaluating one's progress (Flavell 1979: 906–907). In autonomous language learning, metacognitive strategies enable learners to identify their weaknesses, adjust their learning strategies, and set realistic goals. Research shows that digital learning environments, especially those based on artificial intelligence, can strongly foster the development of metacognitive awareness through reflective questions, automatic error analysis, and personalized recommendations (Luckin et al. 2016: 22–24).

Learning closely related languages without developed metalinguistic and metacognitive awareness often leads to superficial understanding and fossilization of errors. Therefore, the proposed didactic approach emphasizes awareness of linguistic structures and learning processes as key elements of long-term language competence.

2.4 Role of the AI tutor in the didactic approach to autonomous learning of Slovenian

In the proposed didactic approach that follows, the AI tutor has multiple functions: it acts as a linguistic mediator, a facilitator of learning, and a promoter of reflection. Unlike traditional digital tools, the AI tutor allows interactive communication, adaptive content adjustment, and contextualized feedback, thereby simulating the dialogic aspect of language learning (Godwin-Jones 2018: 4–6).

The didactic model is based on several key phases: the reception of linguistic content, contrastive reflection, guided production, and self-assessment. In each of these phases, the AI tutor adapts tasks to the individual learner profile, fostering the development of intercultural, metalinguistic, and metacognitive competence. Particular attention is given to communicative situations relevant to the border context, such as professional communication, administrative situations, and everyday interaction.

It is important to emphasize once again that the AI tutor in this model is not viewed as a replacement for the teacher, but as a didactic tool that expands the possibilities of learning and makes the process more accessible and flexible. In this sense, artificial intelligence becomes a means of didactic innovation that complements existing pedagogical practices and contributes to the development of contemporary models of language learning.

2.5 Didactic model of autonomous learning of Slovenian with an AI tutor

The proposed didactic model of autonomous learning of Slovenian with the support of an AI tutor is designed as a structurally guided yet flexible system, adapted to students and adult learners from the border region of Istria and other parts of Croatia. The

model is based on the assumption that learning a closely related language requires specific didactic strategies that take into account the learner's existing linguistic experience, while at the same time systematically deconstructing implicit assumptions about complete mutual intelligibility and equivalence between languages.

Unlike traditional linear models of language learning, this approach is based on a cyclical learning model, in which the phases of learning are repeated and strengthened depending on the learner's progress. In this process, the AI tutor takes on the role of a dynamic regulator of learning, adapting the content, pace, and type of tasks to the individual needs of the user.

2.5.1 Target group and didactic assumptions of the model

The target group includes upper primary school and secondary school students, and adult learners studying Slovenian for professional, educational, or personal reasons. The common characteristics of the target group are Croatian as a native or dominant language, prior but unsystematic exposure to Slovenian, and cultural and historical proximity to the Slovenian context. The didactic assumptions of the model include linguistic proximity as a resource rather than an obstacle, error as an integral part of learning rather than an indicator of failure, reflection as a key element of progress, and learner autonomy supported by structured guidance.

2.5.2 Structure of the didactic model

The model consists of four fundamental phases, which are interconnected and repeated in a spiral form.

In the reception phase, the learner comes into contact with authentic or didactically adapted texts in Slovenian (written, audio, or audiovisual). The AI tutor adapts the level of the text, highlights key lexical and grammatical elements, and makes possible parallel understanding without direct translation. The goal of this phase is not complete understanding, but global comprehension and the recognition of structures.

In the contrastive reflection phase, the AI tutor actively encourages metalinguistic reflection. The learner is guided to focus on differences between Croatian and Slovenian forms, and pragmatic differences in language use. Reflection is not imposed directly, but through questions and comparisons, thereby encouraging active reasoning.

In the guided language production phase, the learner produces language in a safe, non-evaluative environment. The AI tutor provides response models, corrects errors with explanations, and encourages reformulation. Production may be written (short responses, dialogues, or emails) or spoken (simulated conversations).

In the self-assessment and metacognitive reflection phase of the cycle, learners reflect on what they have learned, what they found difficult, and how they can improve their learning. The AI tutor poses metacognitive questions and suggests strategies for further work.

2.5.3 The role of the AI tutor in regulating the learning process

In this model, the AI tutor does not function as a passive source of information, but as a dynamic didactic mediator. Its functions include diagnosing typical errors, adapting tasks to the learner's level, encouraging reflection instead of providing ready-made solutions, and maintaining motivation through positive feedback. It is particularly important that the AI tutor not assume misunderstanding, but instead promote critical understanding, use the contrastive approach selectively rather than mechanically, and take into account the cultural context of communication.

Although the AI tutor significantly expands opportunities for autonomous language learning, its educational use also has certain limitations. AI-generated feedback may occasionally contain linguistic inaccuracies, oversimplified explanations, or pragmatically inappropriate suggestions. Furthermore, the tutor cannot reliably assess all sociocultural nuances of communication or fully recognize the learner's emotional state, motivation, or frustration. For these reasons, the AI tutor should be understood as a didactic support tool rather than a substitute for systematic language instruction. Particularly in the development of higher-order communicative competence, intercultural sensitivity, and critical language awareness, the guidance of a teacher remains indispensable.

2.6 Example of the application of the didactic model

The following illustrative interaction was generated using ChatGPT (OpenAI, GPT-5.5), operating through its graphical user interface (GUI). The example is intended solely to demonstrate the didactic principles of the proposed model rather than to evaluate the performance of a specific AI system.

To more clearly illustrate the practical applicability of the proposed model, the following presents a sample interaction between a learner and an AI tutor in the context of learning Slovenian. The situation is an adult learner from Istria learning Slovenian for professional communication in tourism.

Phase 1: Language reception

	Slovenian	**English translation**
AI tutor:	*Preberi kratko sporočilo gosta v hotelu in poskusi razumeti glavno sporočilo.*	Read the short message from a hotel guest and try to understand the main point.
Sample text:	*Spoštovani,* *za prihodnji petek bi želel rezervirati dvoposteljno sobo za dve noči.* *Zanima me tudi, ali je parkirišče brezplačno in ali je zajtrk vključen v ceno.* *Hvala za odgovor.* *Lep pozdrav,* *Marko*	Dear Sir/Madam, I would like to book a double room for two nights starting next Friday. I would also like to know if parking is free and if breakfast is included in the price. Thank you for your reply. Best regards, Marko
AI tutor:	*Kaj je glavni namen sporočila? Ne skrbi za podrobnosti.*	What is the main purpose of the message? Don't worry about the details.

The learner identifies the main communicative purpose of the message based on global understanding.

Phase 2: Contrastive reflection

	Slovenian	English translation
AI tutor:	*Opazil si glagol rezervirati. Ali se uporablja enako kot hrvaški rezervirati? Ali je v slovenščini pogostejša še kakšna druga možnost? V katerih primerih bi uporabil glagol naročiti?*	You noticed the verb *rezervirati* 'to reserve'. Is it used in the same way as Croatian *rezervirati*? Is there another, more common option in Slovenian? In which cases would you use the verb *naročiti* 'to order'?

The learner reflects on lexical and pragmatic differences between *rezervirati* and *naročiti*, comparing their use in Croatian and Slovenian.

Phase 3: Guided production

	Slovenian	English translation
AI tutor:	*Poskusi odgovoriti gostu v slovenščini. Če nisi prepričan, napiši najprej osnutek.*	Try replying to the guest in Slovenian. If you aren't sure, write a draft first.
Learner:	*Spoštovani, zahvaljujemo se za vaše sporočilo. Soba je na voljo. Parkirišče je brezplačno, zajtrk pa je vključen v ceno. Veselimo se vašega obiska. Lep pozdrav.*	Dear Sir/Madam, Thank you for your message. The room is available. Parking is free, and breakfast is included in the price. We look forward to your visit. Best regards.
AI tutor:	*Dobro. Oglejva si odgovor skupaj. Kako bi lahko stavek »Soba je na voljo« razširil, da bi zvenel bolj vljudno in naravno v poslovni komunikaciji?*	Good. Let's look at the reply together. How could you expand the sentence "The room is available" to make it sound more polite and natural in business communication?

Feedback is provided through guided reformulation rather than direct evaluation.

Phase 4: Metacognitive reflection

	Slovenian	English translation
AI tutor:	*Kaj ti je bilo danes najtežje? Bi ti pomagalo več primerov, več pisnih vaj ali več pogovorov?*	What did you find most difficult today? Would more examples, more written exercises, or more conversations help?
AI tutor:	*Na podlagi tvojega odgovora predlagam dodatne vaje za razlikovanje podobnih hrvaških in slovenskih izrazov ter kratke dialoge iz področja turizma.*	Based on your answer, I suggest additional exercises for distinguishing between similar Croatian and Slovenian terms, as well as short dialogues related to tourism.

The learner reflects on his or her learning process and identifies areas requiring further practice.

2.7 Didactic value of the model

The proposed model makes possible systematic development of language competence, awareness of intercultural differences, strengthening of learner autonomy, and reduction of fear of making mistakes. In a border context, where languages are often perceived as "almost the same," this approach allows deep and functional language learning based on understanding rather than assumptions.

3 Conclusion

This article examined the intercultural linguistic challenges of learning Slovenian in the context of the twenty-first century, with particular emphasis on the possibilities of applying artificial intelligence in autonomous language learning. Drawing on contemporary social and educational developments, the study highlighted the need for didactically designed models that take into account linguistic and cultural proximity, as well as the specific difficulties that arise from it. Within this framework, a didactic approach to autonomous learning of Slovenian supported by an AI tutor was proposed, primarily intended for students and adult learners from the border region of Istria and other parts of Croatia.

The analysis of the theoretical framework showed that learning closely related languages requires the systematic development of intercultural communicative competence, metalinguistic awareness, and metacognitive strategies. In border environments, where mutual intelligibility between languages is often assumed, reflection, contrastive analysis, and awareness of pragmatic differences are key prerequisites for successful and functional communication. The proposed model responds to these requirements by focusing on a cyclical learning process that connects reception, reflection, guided production, and self-assessment, thereby enabling learners to take an active role in their own learning.

Special attention is given to the role of the AI tutor as a didactic mediator rather than a replacement for the teacher. In this study, artificial intelligence is viewed as a tool that allows the personalization of learning, immediate feedback, and the encouragement of reflection, thereby strengthening learner autonomy and reducing the fear of making mistakes. The sample application of the model demonstrates that an AI tutor can successfully simulate communicative situations relevant to everyday and professional contexts, especially in situations in which learners do not have constant access to formal education or a teacher.

The contribution of this article can be observed at several levels. At a theoretical level, it integrates insights from intercultural didactics, theories of learner autonomy, and research on the application of artificial intelligence in education. At a didactic level, it proposes a concrete and applicable model of autonomous learning of Slovenian that can be adapted to different levels of language competence and educational contexts. At a social level, the article highlights the potential of AI-supported language learning models as a means of linguistic inclusion and the strengthening of intercultural communication in border and multilingual environments.

At the same time, certain limitations of the study should be noted. The proposed didactic model has been developed at a theoretical and methodological level and requires empirical validation in real educational contexts. Future research should include qualitative and quantitative analyses of the effectiveness of AI-supported learning, as well as an examination of the attitudes of learners and teachers towards such forms of learning. Furthermore, the ethical and pedagogical implications of the broader application of artificial intelligence in education remain an open question, particularly with regard to data protection and the role of the human factor in the learning process.

In conclusion, it can be argued that a didactic approach to autonomous learning of Slovenian supported by an AI tutor represents a relevant and promising response to contemporary intercultural and educational challenges. In the context of border and culturally close communities, such models have the potential not only to improve language competence, but also to contribute to the development of intercultural understanding, cooperation, and social cohesion in the contemporary European space.

»Financira Europska unija – NextGenerationEU« kroz institucijski projekt Sveučilišta Jurja Dobrile u Puli, pod oznakom IIP_UNIPU_010149. Izneseni stavovi, mišljenja i zaključci samo su autorovi i ne odražavaju nužno službena stajališta Europske unije ili Europske komisije. Ni Europska unija ni Europska komisija ne mogu se smatrati odgovornima za njih.

References

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Godwin-Jones, R. (2018). Contextualized vocabulary learning. *Language Learning & Technology*, *22*(3), 1–19. https://doi.org/10125/44651

Gombert, J. E. (1992). *Metalinguistic development*. University of Chicago Press.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign. https://www.researchgate.net/publication/332180327_Artificial_Intelligence_in_Education_Promise_and_Implications_for_Teaching_and_Learning

Little, D. (1991). *Learner autonomy: Definitions, issues and problems*. Authentik. https://www.researchgate.net/publication/259874253_Learner_Autonomy_1_Definitions_Issues_and_Problems

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education. https://static.googleusercontent.com/media/edu.google.com/hr//pdfs/Intelligence-Unleashed-Publication.pdf

Veje: A Silver-to-Gold Annotation Pipeline and Search Platform for a Slovenian Dialect Treebank

DOI: 10.4312/Obdobja.45.249-255

Varvara Magomedova, Center za eksperimentalno, formalno in teoretično jezikoslovje, Univerza v Novi Gorici, Nova Gorica; varvara.magomedova@ung.si

Prispevek predstavlja prvo fazo projekta Veje – drevesnice slovenskih narečnih besedil. Sistem sestavljata dva dela: dvostopenjski postopek označevanja, ki narečne transkripcije pretvori v razširjeni format CoNLL-U z normalizacijo (model ByT5 in slovarji) in skladenjskimi oznakami, ter spletna aplikacija, ki omogoča šest načinov iskanja. Vmesnik platforme je zasnovan na podlagi rezultatov raziskave potreb uporabnikov, v kateri je sodelovalo 94 anketirancev in 13 intervjuvancev. Trenutna različica korpusa ustreza srebrnemu standardu, kar omogoča lažje ročno popravljanje. Po zgledu projektov ArchiMob in MaiBaam bodo ročno popravljeni podatki tvorili zlati podkorpus, namenjen nadaljnjemu učenju modela. Korpus naj bi vseboval samo skladenjske oznake, hkrati pa bo zagotavljal popolne bibliografske podatke in povezave do virov, ki lahko vključujejo tudi fonetične in zvočne posnetke.

slovenska narečja, Universal Dependencies, normalizacija narečnih besedil, drevesnica, iskanje po korpusu

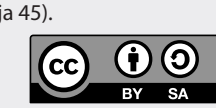
This article presents the first phase of the Veje project, a treebank of Slovenian dialect texts. The system has two parts: a two-stage annotation pipeline that converts dialect transcripts into extended CoNLL-U with normalization (a ByT5 model plus dictionaries) and syntactic annotation, and a web application offering six search modalities. The interface of the platform is designed based on the results of a user-needs study (ninety-four questionnaire respondents, thirteen interviews). The current corpus is silver standard; this was done to make manual correction easier. Following the model of ArchiMob and MaiBaam, manual correction will create a gold subset for model fine-tuning. The corpus is intended to only have syntactic annotation while providing full references including links to the sources, which might have phonetic and acoustic records.

Slovenian dialects, Universal Dependencies, dialect normalization, treebank, corpus search

1 Introduction

Slovenian is one of the most dialectally diverse Slavic languages; there are seven major dialect groups and about fifty dialects (Smole 2019) in a speech community of two million speakers. Phonetic and historical studies cover the variety of dialects very well, with Slovenian dialectologists presenting invaluable expertise on the topic. However, existing corpora are small, covering only a single area—for example, DIALECT-COPA for the Cerkno dialect (Ljubešić et al. 2024a, 2024b) and a clothing terminology dictionary of the Gail dialect in the Canale Valley (Gliha Komac et al. 2019;

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)



Kenda-Jež 2015, 2019)—or they focus on the spoken standard rather than dialects (e.g., GOS/GOS 2; Dobrovoljc 2025; Verdonik et al. 2023, 2024). This might be related to data policies: managing copyrights, personality rights, and personal data protection as a top challenge in building open speech corpora (Verdonik et al. 2024). In addition, many researchers collect data for their particular goals. However, the literature on Slovenian dialects is quite rich due to the work of dialectologists. Spoken Slovenian treebanks exist in Universal Dependencies (Dobrovoljc & Nivre 2016; Dobrovoljc et al. 2023), but for dialect texts specifically, to my knowledge, there is no treebank.

The central problem for corpus building, once the data are there, is annotation cost, given the manual labor of speakers, multiple transcription standards, and annotation formats. Dialect texts differ from standard Slovenian at several levels at once—spelling, phonology, morphology, syntax, and vocabulary—and so a fully manual workflow scales poorly. The Veje project does not propose replacing human annotation, but it reduces repetitive work. The pipeline I am presenting produces silver-standard output that annotators correct into gold data, which then feeds back into the models. Pre-annotation significantly improves the speed of manual labor but may introduce some biases, although the evidence for that is mixed (Fort & Sagot 2010), and the Pomak treebank was made with the same loop (Arampatzakis et al. 2026)

The mixed manual–automatic technique has been used for Swiss German (Samardžić et al. 2016; Scherrer & Ljubešić 2016) and for multilingual lexical normalization (van der Goot et al. 2021). The Swiss German work ran in the opposite order: annotators produced the gold normalization manually and the model was trained afterward, and so the annotators never corrected model output. Veje inverts this: the model produces silver, and annotators correct it.

For the Veje project, three tasks are now finished: silver-standard normalization, silver-standard UD-based annotation, and a usable search interface. These are developed together because a rich corpus will be underused without efficient searching, and a good interface is not very helpful with unreliable annotation. The next phase of the project focuses on manual correction by native speakers and consultations with dialectologists; the corrected subset will help fine-tune the models for further automatic normalization and annotation.

2 Pipeline architecture

The pipeline has two modular stages. Stage 1 normalizes dialect surface forms toward standard Slovenian while keeping the output token-aligned to the dialect input wherever possible, and Stage 2 parses every sentence with CLASSLA-Stanza (Terčon & Ljubešić 2023), routes non-standard sentences to a dialect-feature labeler, and emits annotated CoNLL-U. The stages are separated so that components can be replaced without rebuilding the pipeline, and they communicate only through a JSONL interface.

2.1 Stage 1: normalization

Normalization is intended to be surface-only—correcting orthographic and phonological differences without altering morphosyntax—and the pipeline enforces this wherever it can verify it, rather than assuming it. A morphological guard compares the grammatical gender and number of a source token against those of each proposed replacement, using the Universal Dependencies feature layer of Sloleks 3.1 (Čibej et al. 2026a). The guard returns one of four verdicts: it confirms a substitution when both features are present on both sides and agree, rejects it when either feature is contradicted (a gender or number flip), and otherwise treats it as ambiguous (a feature is missing on one side) or unverifiable (the dialect form is absent from the lexicon). Only rejections are blocked; ambiguous and unverifiable substitutions are kept but recorded for review.

The guard runs at the three points where a substitution is proposed from morphological evidence: cleaning the replacement table, applying it, and the phonological-rule pass (all phonological rules, including the vuo-prefix rule, are routed through it). The colloquial-dictionary lookups and the GaMS residual pass are constrained differently—by curated dictionary entries and by orthographic edit limits, respectively—and so as a safety net every surviving change, whatever produced it, is re-checked against the lexicon when the output is written; any change the guard cannot confirm is flagged. Flags are emitted two ways: a per-record norm_unverified list of token indices, and a token-level audit sidecar giving the sentence id, token index, source form, replacement, and verdict, for manual correction. The guard's standard-word lexicon is extended with single-word standard targets from the GOKO dialect glossary (Šumenjak 2013a, 2013b) and the MEZZANINE non-standard lexicon (Čibej et al. 2026b).

Stage 1 applies five mechanisms in sequence. The two deterministic passes (the replacement table and the phonological rules) preserve token count and order; where a sentence's token count would change—punctuation detachment and hesitation-marker (#) isolation during preprocessing, or a 1→2 clitic-hyphen split—the change is tracked explicitly, and if it cannot be reconciled the sentence falls back to its unnormalized form. Such split/merge cases are excluded from label projection in Stage 2 (see below).

The first mechanism is the corrected most-frequent-replacement table. The pipeline begins with a corrected MFR table derived from Kuparinen et al. (2023), who use aligned dialect and standard forms to build lexical replacement tables. For Slovenian, the released resource was built from the aligned colloquial and standard transcription layers of the GOS corpus (Verdonik et al. 2013, 2024). Direct application produced substantial noise: the released table has 12,660 entries, many reflecting alignment artifacts that corrupt morphology, lemmatization, and dependency structure. Veje introduces a corrected version: cleaning removes 6,156 entries (48.6%) and keeps 6,504 plausible mappings.

Cleaning MFR runs in two passes. The first removes structural noise: identity mappings, multi-word targets that would break token alignment, and gender/number-flip artifacts (feminine plural past participles aligned to masculine; e.g., *gučale* → *gučali* 'spoke', *dale* → *dali* 'gave'). The second adds a Sloleks-based semantic guard for cases

in which the source form is already a valid standard word: plausible vowel restorations such as *bli* → *bili* ‘been’, *tud* → *tudi* ‘also’ are kept, whereas mappings that change lemma identity or inflectional category are removed (e.g., *reči* → *reče*, for which *reči* may be the nominative plural of *reč* ‘thing’ whereas *reče* is the third-person singular of the verb *reči* ‘to say’). The cleaning targets these structural and morphological error classes; it does not include a contact-language detector, and mappings whose source form is absent from Sloleks—including German and Italian loans characteristic of the contact dialects—are kept unconditionally because for a dialect corpus these are genuine dialect material rather than noise to be discarded. The corrected MFR thus preserves genuine dialect-to-standard correspondences: vowel and consonant restoration, vowel shifts, *l*-restoration, contractions, and dialect declension. With the learned normalizer described below, the table is restricted to lexical substitutions (1,389 entries after cleaning, 2,165 with glosses from Narecja.si), and spelling-level differences are left to ByT5.

A byte-level ByT5-base model (Xue et al. 2022) normalizes one word at a time, given the three preceding and three following words and, where known, the dialect label. This replicates the three-word-window setup of Kuparinen et al. (2023); working on bytes makes it possible to avoid vocabulary problems with dialect spelling, schwa, and digraphs. The MFR is consulted on the dialect form before and after ByT5’s output in case it is still not a standard form so that lexical dialect words are not “repaired” into a wrong spelling. On held-out gold, token accuracy rises from 0.44 (the rule-and-dictionary pipeline above) to 0.89 on GOKO and from 0.59 to 0.79 on Narecja.si documents of dialects unseen in training (models trained without the test documents); the test sets are small (twenty-three sentences, thirteen documents).

The model is trained on about 263,000 word windows built from three kinds of aligned dialect–standard material: 1) transcribed dialect texts with a standardized version by the authors (GOKO; Narecja.si, 111 texts, thirty-two dialects); 2) about 13,900 word, phrase, and sentence pairs extracted from dialectological theses and papers, with ambiguous schwa spellings resolved by reading; and 3) the GOS and Artur spoken corpora, which add colloquial-to-standard pairs but little dialect signal. Test documents of every source, including whole dialects, were held out for evaluation; the shipped model is retrained on everything except the GOKO test sentences.

The third mechanism is deterministic phonological rules. Handwritten rules target common Slovenian dialect patterns: vowel restoration, consonant restoration, and reduction. The rules are not applied directly: each generates a candidate that is accepted only if it appears in Sloleks and passes the morphological guard.

The fourth mechanism is Sloleks validation. Candidates are checked against Sloleks (Dobrovoljc et al. 2019; Čibej et al. 2026a); membership in the lexicon is a necessary condition for accepting a phonological candidate, which blocks a large class of implausible substitutions before the guard is even consulted.

Normalization is token-aligned, multi-word replacements are skipped, and the few normalizations that split or merge tokens are excluded from projection. UD annotation will be valid then: for tokens whose standard form is in Sloleks, gender and number are

preserved by construction, and, if the dialect form lies outside the lexicon, the standard analysis is still projected but the token is marked unverified for manual correction.

2.2 Stage 2: annotation

Every sentence is parsed with the CLASSLA-Stanza non-standard Slovenian model (Terčon & Ljubešić 2023) for South Slavic and non-standard varieties. Because CLASSLA-Stanza produces many errors on raw dialect input, sentences are parsed in their normalized form; the original dialect surface is retained and written alongside the analysis in the output, which the token-level alignment and the preservation of inflectional (gender/number) features make possible. SloBERTa (Ulčar & Robnik-Šikonja 2021) flags sentences that likely contain non-standard material, and those sentences are additionally sent to a GaMS labeler; sentences with no flagged material receive standard labels only. The GaMS labeler is fine-tuned on the GOKO gold subcorpus (Šumenjak 2013a, 2013b). Dialect-group identity is taken from document metadata whenever the provenance pins a known group, overriding any group the model predicts; the model's prediction is used only when metadata specify none. The output is an extended CoNLL-U file containing the original dialect form, the normalized equivalent, the UD annotation, and the dialect metadata.

3 Search platform

The Veje platform is a lightweight web application searching both dialect surface forms and normalized forms simultaneously. The interface was informed by a study with corpus users from several subfields. It prioritizes transparent access to results, exportability, multiple complementary search modes, and reduced dependence on formal query syntax. It supports keyword and collocation search, feature-based search, dependency-tree search, CQL, and natural-language search assisted by a locally hosted LLM. Results are presented as KWIC-style hits with matched form, lemma, POS, source citation, and metadata; each hit can be expanded into a Universal Dependencies tree visualization. The platform shows per-result-set source references and a global "Cite all sources" modal. Match sets can be downloaded as plain dialect text or as rebuilt CoNLL-U.

The platform has two goals. The first is practical usability: many existing corpus interfaces are powerful but hard to learn, and so Veje prioritizes directness and visibility of results. The second is data growth — much Slovenian dialect material sits in private archives and local projects, and the platform serves as a demonstrator to encourage contributions.

4 Conclusion

This article presents the first phase of the Veje Slovenian Dialect Treebank: a practical workflow combining automatic normalization, UD annotation, and a user-oriented search platform into a single iterative system. The corpus remains silver standard, which reduces the manual work needed to prepare dialect data for correction. The long-term

goal is reliable annotation and broad coverage, which are only possible in collaboration with dialectologists and archive holders. Annotation is treated as an iterative cycle in which human correction progressively improves both the corpus and the models.

This study was co-funded by the European Union. The views and opinions expressed are solely those of the author and do not necessarily reflect those of the European Union or the European Research Executive Agency. This publication has received funding from the European Union's Horizon Europe research and innovation program under the Marie Skłodowska-Curie COFUND postdoctoral program, grant agreement no. 101081355–SMASH, and from the Republic of Slovenia and the European Union from the European Regional Development Fund.

References

Arampatzakis, V., Stamou, V., Markantonatou, S., & Pavlidis, G. (2026). Exploring active learning approaches in treebank development. In G. Pavlidis & S. Sylaiou (Eds.), *Transforming heritage research in a transforming world: 5th CAA-GR Conference 2024* (Springer Proceedings in Archaeology and Heritage). Springer. https://doi.org/10.1007/978-3-032-06389-2_39

Čibej, J., Gantar, K., Dobrovoljc, K., Krek, S., Holozan, P., Erjavec, T., Romih, M., Arhar Holdt, Š., Krsnik, L., & Robnik-Šikonja, M. (2026a). *Morphological lexicon of Slovene Sloleks 3.1* [Data set]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/2080

Čibej, J., Gantar, K., Gantar, P., Šešet, J., Krek, S., & Robida, N. (2026b). *List of potentially non-standard vocabulary candidates MEZZANINE-NstdLex 1.1*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/2357

Dobrovoljc, K. (2025). Treebanking spoken Slovenian: New data, models, and lessons learned. *Prispevki za novejšo zgodovino*, *65*(3), 14–41. https://doi.org/10.51663/pnz.65.3.01

Dobrovoljc, K., Krek, S., Holozan, P., Erjavec, T., Romih, M., Arhar Holdt, Š., Čibej, J., Krsnik, L., & Robnik-Šikonja, M. (2019). *Morphological lexicon Sloleks 2.0* [Data set]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1230

Dobrovoljc, K., & Nivre, J. (2016). The Universal Dependencies treebank of spoken Slovenian. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 1566–1573). ELRA. https://aclanthology.org/L16-1248/

Dobrovoljc, K., Terčon, L., & Ljubešić, N. (2023). Universal Dependencies za slovenščino: Nove smernice, ročno označeni podatki in razčlenjevalni model. *Slovenščina 2.0*, *11*(1), 218–246. https://doi.org/10.4312/slo2.0.2023.1.218-246

Fort, K., & Sagot, B. (2010). Influence of pre-annotation on POS-tagged corpus development. In N. Xue & M. Poesio (Eds.), *Proceedings of the Fourth Linguistic Annotation Workshop (LAW IV)* (pp. 56–63). Association for Computational Linguistics. https://aclanthology.org/W10-1807.pdf

Gliha Komac, N., Kandutsch, E., Bartaloth, R., & Smole, M. (2019). *The dictionary of the clothing terminology of the Zilja dialect in Canale Valley (Kanalska dolina – Val Canale – Kanaltal – Valcjanâl): photographs*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1221

Kenda-Jež, K. (2015). *Shranli smo jih v bančah: Slovarski prispevek k poznavanju oblačilne kulture v Kanalski dolini / Contributo lessicale alla conoscenza dell'abbigliamento in Val Canale* (2nd ed.). S.K.S. Planika Kanalska dolina, Slori, ATS Od me-je, Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, CRS ASSA.

Kenda-Jež, K. (2019). *Slovar oblačilnega izrazja ziljskega govora v Kanalski dolini*. Založba ZRC, ZRC SAZU.

Kuparinen, O., Miletić, A., & Scherrer, Y. (2023). Dialect-to-standard normalization: A large-scale multilingual evaluation. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for*

Computational Linguistics: EMNLP 2023 (pp. 13814–13828). Association for Computational Linguistics. https://aclanthology.org/2023.findings-emnlp.923/

Ljubešić, N., Galant, N., Benčina, S., Čibej, J., Milosavljević, S., Rupnik, P., & Kuzman, T. (2024a). *DIALECT-COPA:* Extending the standard translations of the COPA causal commonsense reasoning dataset to South Slavic dialects. In Y. Scherrer, T. Jauhiainen, N. Ljubešić, M. Zampieri, P. Nakov, & J. Tiedemann (Eds.), *Proceedings of the Eleventh Workshop on NLP for Similar Languages, Varieties, and Dialects (VarDial 2024)* (pp. 89–98). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.vardial-1.7

Ljubešić, N., Kuzman, T., Rupnik, P., Milosavljević, S., Galant, N., Benčina, S., & Čibej, J. (2024b). *"Choice of plausible alternatives" datasets in South Slavic dialects DIALECT-COPA*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1766

Samardžić, T., Scherrer, Y., & Glaser, E. (2016). ArchiMob—A corpus of spoken Swiss German. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 4061–4066). ELRA. https://aclanthology.org/L16-1641/

Smole, V. (2019). Slovenska narečja v spletnih aplikacijah. In M. Smolej (Ed.), *1919 v slovenskem jeziku, literaturi in kulturi* (62. Seminar slovenskega jezika, literature in kulture, pp. 20–30). Znanstvena založba Filozofske fakultete. http://dx.doi.org/10.4312/SSJLK.55.20-30

Šumenjak, K. (2013a). Priprava gradiva in standardizacija nivojev zapisa za potrebe dialektološkega korpusa GOKO. In A. Žele (Ed.), *Družbena funkcijskost jezika (vidiki, merila, opredelitve)* (Obdobja 32, pp. 443–449). Znanstvena založba Filozofske fakultete.

Šumenjak, K. (2013b). *Opis govora Koprive na Krasu na osnovi dialektološkega korpusa* [Doctoral dissertation]. Univerza na Primorskem.

Terčon, L., & Ljubešić, N. (2023). CLASSLA-Stanza: The next step for linguistic processing of South Slavic languages. *arXiv*. https://arxiv.org/abs/2308.04255

Ulčar, M., & Robnik-Šikonja, M. (2021). SloBERTa: Slovene monolingual large pretrained masked language model. In M. Luštrek et al. (Eds.), *Proceedings of the 24th International Multiconference Information Society (IS 2021), SiKDD* (pp. 131–134). Institut »Jožef Stefan«.

van der Goot, R., Ramponi, A., Zubiaga, A., Plank, B., Muller, B., San Vicente Roncal, I., Ljubešić, N., Çetinoğlu, Ö., Mahendra, R., Çolakoğlu, T., Baldwin, T., Caselli, T., & Sidorenko, W. (2021). MultiLexNorm: A shared task on multilingual lexical normalization. In W. Xu, A. Ritter, T. Baldwin, & A. Rahimi (Eds.), *Proceedings of the Seventh Workshop on Noisy User-Generated Text (W-NUT 2021)* (pp. 493–509). Association for Computational Linguistics. https://aclanthology.org/2021.wnut-1.55/

Verdonik, D., Dobrovoljc, K., Erjavec, T., & Ljubešić, N. (2024). Gos 2: A new reference corpus of spoken Slovenian. In N. Calzolari, M. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 7825–7830). ELRA and ICCL. https://aclanthology.org/2024.lrec-main.691/

Verdonik, D., Zwitter Vitez, A., Zemljarič Miklavčič, J., Krek, S., Stabej, M., Erjavec, T., Potočnik, T., Sepesy Maučec, M., Majhenič, S., Žgank, A., Bizjak, A., Gril, L., Dobrišek, S., Križaj, J., Bajec, M., Lebar Bajec, I., Jelovšek, T., Trojar, M. Bernjak, M., Dretnik, N. … Rupnik, P. (2023). *Gos 2.1: Korpus govorjene slovenščine* [Corpus]. Center za jezikovne vire in tehnologije. https://viri.cjvt.si/gos

Vreš, D., Božič, M., Potočnik, A., Martinčič, T., & Robnik-Šikonja, M. (2024). Generative model for less-resourced language with 1 billion parameters. In Š. Arhar-Holdt & T. Erjavec (Eds.), Proceedings of the Conference on Language Technologies and Digital Humanities (JT-DH 2024) (pp. 485–511). https://www.sdjt.si/wp/wp-content/uploads/2024/09/JT-DH-2024_Vres_Bozic_Potocnik_Martincic_Robnik.pdf

Xue, L., Barua, A., Constant, N., Al-Rfou, R., Narang, S., Kale, M., Roberts, A., & Raffel, C. (2022). ByT5: Towards a token-free future with pre-trained byte-to-byte models. *Transactions of the Association for Computational Linguistics*, *10*, 291–306. https://doi.org/10.1162/tacl_a_00461

Generativna umetna inteligenca v pedagoškem slovaropisju: izkušnje iz prenove slovensko-ruskega kolokacijskega slovarja za začetnike

DOI: 10.4312/Obdobja.45.257-266

Irina Makarova, Fakulteta za humanistične študije, Univerza na Primorskem, Koper;
irina.mt@ujpr.si

Prispevek predstavlja metodološki model uporabe generativne umetne inteligence v pedagoškem slovaropisju, oblikovan na podlagi prenove izbranih samostalniških in glagolskih gesel *Slovensko-ruskega kolokacijskega slovarja za začetnike* (2019). V ospredju ni zgolj vprašanje avtomatizacije posameznih leksikografskih opravil, temveč razvoj strukturiranega delovnega protokola UI-agenta, ki omogoča sistematično povezovanje različnih jezikovnih virov pri oblikovanju didaktično prilagojenih slovarskih gesel. Predlagani pristop pomeni prehod od vsebinsko preobremenjenih prvotnih geselskih člankov k preglednejšemu, didaktično in kontrastivno bolj izostrenemu modelu slovarskega opisa.

pedagoško slovaropisje, generativna umetna inteligenca, kontrastivna leksikografija, slovenščina kot tuji jezik, CustomGPT

This article presents a methodological framework for the application of generative artificial intelligence in pedagogical lexicography, developed through the revision of selected noun and verb entries from *Slovensko-ruski kolokacijski slovar za začetnike* (Slovenian–Russian Collocational Dictionary for Beginners, 2019). Rather than focusing solely on the automation of individual lexicographic tasks, the study proposes a structured AI-agent workflow that allows the systematic integration of multiple linguistic resources in the development of pedagogically adapted dictionary entries. The proposed approach is a shift from information-dense dictionary entries to a clearer, pedagogically and contrastively more focused model of lexicographic description.

pedagogical lexicography, generative AI, contrastive approach, Slovenian as a foreign language, CustomGPT

1 Uvod

V zadnjih letih generativna umetna inteligenca (UI) postopoma vstopa tudi na področje leksikografije in priprave jezikovnih virov (De Schryver 2023, 2025). V slovenskem prostoru se sodobno slovaropisje že dalj časa opira na digitalne korpuse, kolokacijske baze in avtomatizirane postopke obdelave jezikovnih podatkov (Gantar 2011; Kosem idr. 2024), uporaba generativnih modelov pri pripravi pedagoških slovarjev za slovenščino kot tuji jezik pa ostaja še razmeroma neraziskana. Poseben izziv predstavljajo začetne ravni jezikovne kompetence (A1–A2), kjer mora slovarski opis

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

poleg jezikovne pravilnosti upoštevati tudi didaktično prilagoditev, pragmatično uporabnost in predvidljive težave ciljne skupine uporabnikov zaradi interference s prvim jezikom.

Prispevek obravnava prenovo *Slovensko-ruskega kolokacijskega slovarja za začetnike* (Makarova idr. 2019) z uporabo generativne umetne inteligence. V središču raziskave ni vprašanje avtomatizirane izdelave slovarskih gesel, ampak vprašanje, katere tipe diferencialnih problemov je mogoče dovolj stabilno operacionalizirati s pomočjo generativne UI in kateri ostajajo zunaj dosega trenutnega modela. Raziskava je tako zasnovana kot eksperimentalno-refleksivna študija s perspektive jezikoslovca praktika, ki uporablja splošno dostopna UI-orodja (CustomGPT) brez razvoja lastnih nevronskih modelov. Poseben poudarek je namenjen oblikovanju strukturiranega metodološkega okvira, ki sistematično povezuje korpusne podatke, jezikovne vire, kontrastivno analizo in strokovno validacijo pri pripravi pedagoškega slovarskega opisa.

2 Diferencialna zasnova *Slovensko-ruskega kolokacijskega slovarja za začetnike*

Slovensko-ruski kolokacijski slovar za začetnike (v nadaljevanju SRKSZ)[1] je bil razvit kot diferencialni pedagoški slovarski vir, namenjen rusko govorečim uporabnikom slovenščine. Temelji na sistematičnem izpostavljanju kontrastivno kritičnih razlik med slovenščino in ruščino. Zasnovan je kot večrazsežnostni model predstavitve leksikalnih enot, ki presega raven prevodnih ustreznic ter združuje pomenske, kolokacijske in slovnične informacije z zgledi rabe, besedotvornimi povezavami, slikovnim gradivom in zvočnimi posnetki iztočnic. Njegov temeljni cilj je uporabnika usposobiti za aktivno jezikovno produkcijo v slovenščini.

2.1 Opis prvotnih geselskih člankov

Prvotni model slovarskega gesla SRKSZ uporabniku ponuja širok nabor informacij in celovit vpogled v rabo besede. Osrednji del geselskega sestavka predstavlja opis rabe iztočnice, ki vključuje tipične leksikalno-skladenjske povezave z njihovimi prevodnimi ekvivalenti. Mednje sodijo kolokacije v ožjem leksikografskem smislu, tj. slovarsko relevantne, statistično značilne in skladenjsko ustaljene sopojavitve dveh ali več leksikalnih enot (prim. Kosem, Krek in Gantar 2024), pa tudi vezljivostne zveze ter drugi prototipski vzorci rabe. Takšna zasnova izhaja iz didaktične usmerjenosti SRKSZ, katerega cilj ni le predstaviti tipične leksikalno-skladenjske sopojavitve, temveč uporabnikom z najnižjo stopnjo sporazumevalne zmožnosti omogočiti aktivno jezikovno produkcijo. Posamezne sestavine prvotnega geselskega članka in njihovo medsebojno razmerje prikazuje Shema 1.

1 https://www.termania.net/?searchIn=Linked&ld=285

|————**Iztočnica**
  |——lema
  |——zvočni posnetek
  |——besednovrstna oznaka
|————prevodni ekvivalent
|————**Ponazarjalno gradivo**
  |——ilustracija
|————**Semantična umestitev**
  |——nadpomenka, sopomenka, protipomenka (če so)
  |——prevodni ekvivalent
|————**Opis rabe iztočnice**
  |——tipične leksikalno-skladenjske povezave iztočnice
    (kolokacije, vezljivostne zveze, drugi prototipski vzorci rabe)
  |——prevodni ekvivalent
|————**Oblikoslovni opis**
  |——celotna pregibna paradigma
|————**Pragmatični zgledi rabe**
  |——prevodni ekvivalenti
|————**Dopolnilni leksikalni podatki (če so)**
  |——ustaljene zveze, tvorjenke
  |——prevodni ekvivalenti

Shema 1: Makrostruktura geselskega sestavka SRKSZ.

2.1.1 Samostalniki

Samostalniška gesla SRKSZ, npr. *avtobus*, poleg osnovnih slovarskih informacij vključujejo tipične leksikalno-skladenjske povezave iztočnice, med katerimi so atributivne zveze (*nov*, *mestni*, *šolski avtobus*), kolokacije, vezljivostne zveze in drugi prototipski vzorci rabe (*čakati*, *zamuditi avtobus*, *vstopiti na avtobus*, *izstopiti iz avtobusa*), celotno pregibno paradigmo iztočnice ter na koncu gesla še pragmatične zglede rabe, npr. *Kateri avtobus gre v Koper?* Takšna širina otežuje prepoznavanje kontrastivno občutljivih mest, npr. razlike v rabi predlogov (***z*** *avtobusom* vs. *rus.* ***na*** *avtobuse*). Prenova z uporabo UI zato ni posegala v vsebinsko strukturo gradiva, temveč v njegovo reorganizacijo in poudarjanje kontrastivno občutljivih informacij, združenih v novi rubriki PAZI (gl. poglavje 4).

2.1.2 Glagoli

Tudi glagolska gesla (npr. *imeti*, *iti*, *govoriti*, *priti*, *voziti*) temeljijo na sistematični predstavitvi tipičnih leksikalno-skladenjskih povezav. Te so razvrščene po skladenjskih vzorcih glagolske vezljivosti, vprašalnice (npr. *govoriti kaj?*, *govoriti o kom?*, *voziti koga?*) pa služijo kot didaktični prikaz posameznih vezljivostnih mest. Takšna organizacija začetnikom olajša razumevanje tipične rabe glagola in njegovo ustrezno vključevanje v poved.

Kljub pregledni strukturi obsežnost informacij zmanjšuje vidnost ključnih interferenčnih mest. Pri glagolu *imeti* je npr. zveza *imeti rad* vključena med druge tipične leksikalno-skladenjske povezave, čeprav gre za pomensko in kontrastivno izrazito specifično enoto, ki v ruščini zahteva drugačno leksikalno realizacijo. Manj izrazite ostajajo tudi razlike pri izražanju posesivnosti, ki se v ruščini praviloma ubesedijo s konstrukcijo *u menja jest'*.

Z uporabo UI prehajamo od obsežnih seznamov tipičnih leksikalno-skladenjskih povezav k didaktično preglednejšim makroscenarijem njihove rabe (gl. poglavje 3.1). Vprašalnice iz prvotnega modela pri tem ohranjajo vlogo didaktičnega pripomočka za prikaz skladenjskih vzorcev in vezljivostnih razmerij. Prenova tako ne zmanjšuje podatkovne vrednosti gradiva, temveč ga didaktično reorganizira ter dopolnjuje z rubriko PAZI, ki v strnjeni obliki izpostavlja predvidljiva interferenčna mesta med slovenščino in ruščino (gl. poglavje 4).

3 Operacionalizacija s CustomGPT

Proces prenove slovarskih člankov temelji na implementaciji tehnologije CustomGPT, ki v okviru velikih jezikovnih modelov (ChatGPT) omogoča razvoj specializiranih namenskih agentov. Orodje je zasnovano za izvedbo specifičnih operacij na podlagi vnaprej določenih parametrov in postopkovnih pravil. Delovanje je utemeljeno na strukturiranem protokolu, ki vključuje podrobna sistemska navodila (angl. *master-prompt*), integrirano bazo znanja (angl. *knowledge base*) in referenčne primere (angl. *few-shot-examples*). Protokol natančno določa posamezne faze obdelave vhodnih podatkov in predpisuje format izhodnih informacij, najsi gre za strukturirane tabele, prioritetne sezname ali utemeljitve sprejetih leksikografskih odločitev.

Zaradi omejitve sistemskih navodil na približno 8000 znakov je bilo treba zasnovati več specializiranih UI-agentov, usmerjenih v posamezna leksikografska opravila. Posebno pozornost v pričujoči raziskavi smo tako namenili naslednjim ključnim nalogam: 1) pripravi didaktično preglednejših pomenskih shem in prioritizaciji izbranih leksikalno-skladenjskih povezav, 2) prepoznavanju tipičnih interferenčnih pojavov in 3) obravnavi morfoloških odstopanj od sistemskih (standardnih) vzorcev.

Izbor gradiva temelji na korpusu učbenikov slovenščine kot tujega jezika ccKUUS (Klemen idr. 2023), ki je služil za identifikacijo tipičnih leksikalno-skladenjskih povezav in vzorcev rabe, prilagojenih začetnim stopnjam učenja slovenščine. Pri izboru ponazarjalnih primerov so bile uporabljene tudi korpusne statistike ccKUUS (besedne skice in tipične leksikalno-skladenjske povezave s kazalci T-score in LogDice). Rezultati kontrastivne analize in ocene tveganja interference, pridobljeni z uporabo UI-agenta (gl. poglavje 3), so bili dodatno preverjeni in dopolnjeni z empirično analizo korpusa usvajanja slovenščine kot tujega jezika KOST 2.1 (Stritar Kučuk idr. 2025), pri čemer sta bila upoštevana tako celotni korpus kot tudi podkorpus besedil rusko govorečih uporabnikov slovenščine kot tujega jezika. Tako je bilo mogoče razlikovati med splošnimi razvojnimi napakami in jezikovnospecifičnimi interferenčnimi pojavi pri usvajanju posamezne iztočnice (gl. poglavje 4).

3.1 Makroscenarij kot operativni model za izbor tipičnih leksikalno-skladenjskih povezav

Pri oblikovanju prenovljenega slovarskega gesla se opiramo na teorijo pomenskih shem, kot jih razvija Leksikalna baza za slovenščino (Gantar 2011, 2015) in drugi viri, zasnovani na teoriji pomenskih okvirjev (angl. *frame semantics*). Za potrebe eksperimentalne prenove pedagoškega slovarja uvajamo dodatno raven organizacije slovarskega opisa – makroscenarij. Makroscenarij za razliko od pomenske sheme ne opisuje posameznega glagola, temveč predstavlja prenosljiv model tipičnih udeleženskih vlog za glagole istega semantičnega razreda. UI-agentu služi kot operativni okvir za izbor tipičnih leksikalno-skladenjskih povezav in oblikovanje didaktično prilagojenega slovarskega gesla za širši nabor iztočnic. V fazi predpriprave smo 463 glagolov ravni A1–A2 razvrstili v semantične razrede in vsakemu določili makroscenarij, ki je bil vključen v bazo znanja CustomGPT. Primere tako oblikovane delovne klasifikacije in pripadajočih makroscenarijev prikazuje Tabela 1.

Semantični razred	Leksemi	Makroscenarij
Glagoli premikanja	*iti*, *priti*, *teči*, *voziti se*, *peljati se*	[oseba], [cilj], [izhodišče], [sredstvo], [način], [namen], [spremstvo]
Glagoli bivanja	*biti*, *živeti*, *stanovati*, *nahajati se*	[oseba], [lokacija], [čas], [način]
Glagoli zaznavanja	*videti*, *slišati*, *čutiti*, *opaziti*	[oseba], [predmet zaznave], [način]

Tabela 1: Primeri delovne klasifikacije glagolov po semantičnih razredih[2] v bazi znanja UI-orodja.

Generiranje gesla poteka dvostopenjsko: naprej se določi semantični razred glagola, nato UI-agent aktivira ustrezni makroscenarij, ki udeleženske vloge zapolni s statistično najbolj značilnimi leksikalno-skladenjskimi povezavami, potrjenimi na podlagi korpusnih kazalnikov (T-score ≥ 7 in LogDice ≥ 7). Izvedbo tega postopka na primeru glagola *iti* prikazuje Tabela 2.

2 Delovna razvrstitev glagolov v semantične razrede (npr. glagoli premikanja, zaznavanja, bivanja) izhaja iz uveljavljene semantične tipologije glagolov v sodobni leksikografiji in leksikalni semantiki (prim. Apresjan 1995, 2000; Zupan 2013), vendar je bila za potrebe delovanja UI-orodja poenostavljena in prilagojena didaktičnemu namenu. Poimenovanja posameznih razredov ne predstavljajo normativne lingvistične klasifikacije slovenskih glagolov, temveč operativne oznake, ki jih UI uporablja pri izbiri ustreznega makroscenarija.

Udeleženska vloga	Vprašalnica	Vzorec tipične leksikalno-skladenjske povezave	Tipične leksikalno-skladenjske povezave
CILJ (zgradba, mesto, država)	Kam?	*iti v* + [prostor]	*iti v trgovino*, *v šolo*, *v kino*, *domov*
CILJ (dejavnost ali dogodek)	Kam?	*iti na* + [dogodek]	*iti na sprehod*, *na kavo*, *na izlet*, *na obisk*
PREVOZNO SREDSTVO	S čim?	*iti s/z* + [prevozno sredstvo]	*iti z avtobusom*, *z vlakom*, *s kolesom*
SPREMSTVO	S kom?	*iti s/z* + [oseba/živali]	*iti s prijatelji*, *z družino*, *s psom*

Tabela 2: Primer realizacije udeleženskih vlog makroscenarija pri glagolu *iti*.

Ko je za posamezni semantični razred oblikovan makroscenarij z značilnimi udeleženskimi vlogami, ga je mogoče uporabiti pri vseh glagolih istega razreda, pri čemer se izbor posameznih elementov prilagaja stopnji usvajanja jezika in tipičnosti njihove rabe. Posamezni glagoli iz istega semantičnega razreda si namreč delijo podoben nabor udeleženskih vlog, razlikujejo pa se v njihovih tipičnih leksikalno-skladenjskih realizacijah. Prenosljivost istega makroscenarija na različne glagole premikanja prikazuje Tabela 3.

Makroscenarij	*iti*	*priti*	*teči*	*voziti se*	*peljati se*
OSEBA	*otrok*	*obiskovalec*	*športnik*	*turist*	*potnik*
CILJ	*v šolo*	*domov*	*v park*	*na morje*	*v mesto*
IZHODIŠČE	*od doma*	*iz službe*	*iz gozda*	*iz Ljubljane*	*iz Kopra*
SREDSTVO	*z avtobusom*	*z vlakom*	–	*z avtom*	*s kolesom*
NAČIN	*počasi*	*pravočasno*	*hitro*	*udobno*	*previdno*
NAMEN	*po nakupih*	*na obisk*	*trenirat*	*na izlet*	*v službo*
SPREMSTVO	*z družino*	*s prijatelji*	*s psom*	*z otroki*	*s sodelavci*

Tabela 3: Prenosljivost makroscenarija na glagole istega semantičnega razreda (primer glagolov premikanja).

Čeprav bi bil podoben princip uporaben tudi pri drugih besednih vrstah (npr. pri samostalnikih), se raziskava v trenutni fazi omejuje na glagolske semantične vzorce.

3.2 Analiza konstrukcijske ekvivalence in preprečevanje interference

Kontrastivna analiza konstrukcijske ekvivalence tipičnih slovenskih leksikalno-skladenjskih povezav in njihovih ruskih ustreznic je pokazala, da popolnoma avtomatizirani modeli ne zagotavljajo zadostne stopnje zanesljivosti, zlasti pri pragmatično zaznamovanih formulah in vezljivostno neekvivalentnih strukturah. V ta namen je bil razvit hibridni model (angl. *human-in-the-loop*), v katerem generativna UI ne deluje kot samostojen prevajalski sistem, temveč kot orodje za sistematično preverjanje stopnje konstrukcijske ekvivalence med slovenščino in ruščino.

Primarna naloga agenta je obsegala analizo stopnje konstrukcijskega ujemanja med tipično slovensko leksikalno-skladenjsko povezavo in njeno rusko ustreznico. Posebej problematični so lažni prevodni ekvivalenti (npr. *imeti prav* vs. rus. *byt' prav(a)*; napačno *imet' pravo*), ter konstrukcijske asimetrije, kot so različni posesivni modeli (*imeti čas*

vs. rus. *u menja jest' vremja*) in predložna vezljivost (*iti **z** avtobusom* vs. rus. *jehat' **na** avtobuse*). Agent je stopnjo konstrukcijske ekvivalence in tveganje interference ocenjeval po tristopenjski lestvici. Primer take analize in razvrščanja leksikalno-skladenjskih povezav glede na stopnjo interferenčnega tveganja prikazuje Tabela 4.

Slovenska leksikalno--skladenjska povezava	**Prevod v ruščino**	**Tip razlike**	**Tveganje**	**PAZI**
iti v šolo	*idti v školu*	skoraj popolna ekvivalenca	1	NE
*iti **z** avtobusom*	*jehat' **na** avtobuse*	drugačen predložni model	2	DA
imeti čas	*u menja jest' vremja*	drugačna posesivna konstrukcija	2	DA
***imeti** prav*	***byt'** prav(a)*	lažna ekvivalenca (***imet' pravo***)	3	DA

Tabela 4: Analiza konstrukcijske ekvivalence tipičnih leksikalno-skladenjskih povezav in ocena tveganja interference.

3.3 Identifikacija nestandardnih morfoloških oblik iztočnice: mikroopozorilni model

Pri snovanju diferencialnega pedagoškega vira se je izkazalo, da sistematično navajanje celovitih oblikoslovnih paradigem ne zagotavlja optimalne uporabniške vrednosti za govorce na začetnih ravneh usvajanja jezika. V okviru prenove smo zato uveljavili selektiven in nivojsko prilagojen pristop, utemeljen na identifikaciji morfoloških odstopanj od standardnih vzorcev. Referenčni vir oblikoslovnih podatkov predstavlja morfološki leksikon slovenskega jezika Sloleks 2.0[3] (Dobrovoljc idr. 2019), iz katerega UI-agent pridobi informacije o pregibnih paradigmah, variantnih oblikah ter supletivnih realizacijah. Na tej podlagi identificira oblikoslovno zahtevnejša mesta in pri izboru jezikovnih variant daje prednost oblikam, ki so z didaktičnega vidika preglednejše in sistemsko predvidljivejše. Tako se denimo pri glagolu *iti* v tretji osebi množine favorizira obliko *grejo* (namesto *gredo*), ki je za začetnika paradigmatsko lažje usvojljiva.

Metodološko novost prenovljenih gesel predstavlja sistem mikroopozoril, ki oblikoslovna odstopanja (npr. *otrok → otroci*) sistematično prikazuje v tipičnih zgledih rabe. Način umeščanja takih mikroopozoril v tipične zglede rabe je v Shemi 2 prikazan na primerih samostalnikov *dan* in *otrok*.

OTROK – PAZI! – množina ni ~~otroki~~ √ otro**c**i → (vsi) otroci se igrajo
DAN – PAZI! – množina ni ~~dani~~ ← √ **dn**evi → **dn**evi so dolgi
← √ **dn**i → doma sem že 3 **dn**i

Shema 2: Umeščanje mikroopozoril v tipične zglede rabe.

4 Rubrika PAZI kot strnjen prikaz didaktičnih opozoril

V prenovljeni zasnovi geselskega članka so v novi slovarski rubriki PAZI sistematično predstavljena tipična problematična mesta posamezne iztočnice, prepoznana na

3 https://viri.cjvt.si/sloleks/slv/

podlagi empirične analize korpusa usvajanja slovenščine kot tujega jezika KOST 2.1 ter predhodne kontrastivne analize, dopolnjene z oceno tveganja interferenc, ki jo je na podlagi primerjave konstrukcijske ekvivalence izvedel UI-agent (gl. poglavje 3.3). Rubrika PAZI združuje opozorila na posebnosti slovenskega jezikovnega sistema, razvojne težave pri usvajanju ciljnega jezika ter kontrastivne razlike med slovenščino in ruščino. Razvrstitev opozoril glede na jezikovno raven in naravo težave je na primeru glagola *iti* prikazana v Tabeli 5.

Jezikovna raven	**Narava težave**	**Slovenski zgled**	**Didaktično opozorilo**	**Pojasnilo**
oblikoslovje	posebnost slovenskega sistema	*gre-m … grejo/gredo*	*grejo*	v slovarju za začetnike prednostno navajamo obliko *grejo* (namesto *gredo*)
oblikoslovje	razvojna težava pri usvajanju tujega jezika	*moram **iti** domov*	***iti***, namesto ~~*greti*~~	razvojna napaka zaradi analogičnega posploševanja s sedanjiškimi oblikami (*grem*, *greš* …)
oblikoslovje	razvojna težava pri usvajanju tujega jezika	*včeraj je šel/šla*	šel/šla, namesto ~~*grel(a)*~~	razvojna napaka zaradi analogičnega posploševanja s sedanjiškimi oblikami (*grem*, *greš* …)
leksikalno-semantična raven	kontrastivna razlika (slo – rus)	*iti v Moskvo*	rus. ***jehat'** v Moskvu*	v ruščini se glede na način premikanja (peš – s prevoznim sredstvom) razlikujeta glagola *idti* in *jehat'*
skladnja	kontrastivna razlika (slo – rus)	*iti **z** avtom*	rus. ***na** mašine*	predložna asimetrija: ***z*** + orodnik ≠ ***na*** + mestnik
skladnja	kontrastivna razlika (slo – rus)	*iti **na** pot*	rus. ***v** dorogu*	predložna asimetrija: ***na*** + tožilnik ≠ ***v*** + tožilnik

Tabela 5: Tipi didaktičnih opozoril v rubriki PAZI za glagol *iti*.

5 Arhitektura prenove slovarskih gesel SRKSZ s pomočjo UI

Tabela 6 strnjeno prikazuje celoten metodološki postopek prenove glagolskega gesla, v katerem se pomenska analiza, korpusni podatki, kontrastivna analiza, analiza usvajanja jezika in oblikoslovni podatki povezujejo v enotni delovni model. Posamezne faze temeljijo na različnih podatkovnih virih in analitičnih postopkih, ki jih UI-agent uporablja za sistematično podporo leksikografskemu delu, medtem ko končna didaktična, normativna in kontrastivna presoja vseh predlaganih rešitev ostaja v pristojnosti leksikografa.

Faza	Podatkovni vir	Rezultat
Identifikacija semantičnega razreda	besedne skice (Word Sketches), semantična analiza	razvrstitev glagola v semantični razred
Oblikovanje makroscenarija	pomenska shema in semantična analiza	prenosljiv model tipičnih udeleženskih vlog za glagole istega semantičnega razreda
Izbor tipičnih leksikalno-skladenjskih povezav	ccKUUS, konkordance, GDEX, LogDice, T-score	zapolnitev makroscenarija s tipičnimi leksikalno-skladenjskimi povezavami
Kontrastivna analiza	UI-agent	primerjava konstrukcijske ekvivalence slovenskih tipičnih leksikalno-skladenjskih povezav in njihovih ruskih ustreznic ter ocena tveganja interference
Validacija na korpusu usvajanja	KOST 2.1 (celotni korpus in podkorpus rusko govorečih uporabnikov)	empirična validacija in dopolnitev razvojnih napak ter interferenčnih pojavov pri posamezni iztočnici
Oblikoslovna analiza	Sloleks 2.0	identifikacija oblikoslovnih posebnosti in oblikovanje mikroopozoril
Oblikovanje rubrike PAZI	rezultati kontrastivne analize, KOST 2.1 in leksikografska presoja	didaktična opozorila za posamezno iztočnico
Končna validacija	leksikograf	didaktična, normativna in kontrastivna validacija

Tabela 6: Arhitektura prenove slovarskega gesla s pomočjo UI (na primeru glagolskih gesel).

6 Sklep in metodološka refleksija

V prispevku predstavljena eksperimentalna prenova SRKSZ je pokazala, da lahko generativni jezikovni modeli pomembno prispevajo k metodološki prenovi pedagoškega slovaropisja. Raziskava je potrdila, da je mogoče z uporabo strukturiranega UI-protokola zanesljivo operacionalizirati vrsto leksikografskih opravil, kot so organizacija tipičnih leksikalno-skladenjskih povezav, prepoznavanje kontrastivno občutljivih mest, identifikacija oblikoslovnih posebnosti in priprava didaktičnih opozoril. Poleg tega takšen pristop omogoča preglednejši slovarski opis, medtem ko končna didaktična, normativna in kontrastivna presoja ostaja v pristojnosti leksikografa.

Kljub izkazanim prednostim pri prepoznavanju oblikoslovnih odstopanj in napovedovanju interferenčnih tveganj konsistentnost predlaganega modela v širšem leksikografskem merilu še ni v celoti zagotovljena. Analizirani nabor leksikalnih enot je razmeroma omejen, zato ugotovitev v tej fazi še ni mogoče neposredno posplošiti na celoten geslovnik. Ker raziskava temelji na uporabi splošno dostopnega generativnega UI-orodja (CustomGPT), ki ni razvito posebej za leksikografske namene, vsi izhodni rezultati – od seznamov in priporočil do ocen interferenčnih tveganj – zahtevajo dosledno strokovno preverjanje in validacijo s strani leksikografa.

Prihodnje raziskave bi morale nasloviti vprašanje prenosljivosti predlaganega modela tudi na druge besedne vrste (npr. prislove in funkcijske besede) in preveriti učinkovitost hibridnih protokolov za pripravo novih diferencialnih virov za učenje

slovenščine kot tujega jezika. Predstavljena metodologija se kaže kot obetavna tudi za razvoj interaktivnih vaj, digitalnih učnih pripomočkov in aplikacij za utrjevanje besedišča, kjer bi bilo mogoče enake postopke samodejne analize in strokovne validacije uporabiti tudi zunaj slovarske obdelave.

Prispevek je nastal v okviru interne raziskovalne skupine MAGIS – Večjezičnost in medkulturnost v dobi generativne umetne inteligence, ki jo financira Rektoričin sklad Univerze na Primorskem.

Literatura

Apresjan, J. D. (1995). *Leksičeskaja semantika: Sinonimičeskie sredstva jazyka* (Izbrannye trudy 1). Vostočnaja literatura RAN.

Apresjan, J. D. (2000). *Systematic Lexicography*. Oxford University Press. https://doi.org/10.1093/oso/9780198237808.001.0001

Dobrovoljc, K., Krek, S., Holozan, P., Erjavec, T., Romih, M., Arhar Holdt, Š., Čibej, J., Krsnik, L. in Robnik-Šikonja, M. (2019). *Morphological lexicon Sloleks 2.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1230

De Schryver, G.-M. (2023). Generative AI and lexicography: The current state of the art using ChatGPT. *International Journal of Lexicography*, *36*(4), 355–387. https://doi.org/10.1093/ijl/ecad021

De Schryver, G.-M. (2025). Customised GPTs for lexicography. *Lexikos*, *35*(2), 397–422. https://doi.org/10.5788/35-2-2104

Gantar, P. (2011). Leksikalna baza za slovenščino: Komu, zakaj in kako naprej?. *Jezik in slovstvo*, *56*(3–4), 81–93.

Gantar, P. (2015). *Leksikografski opis slovenščine v digitalnem okolju*. Znanstvena založba Filozofske fakultete. https://doi.org/10.4312/9789612377922

Klemen, M., Kosem, I., Arhar Holdt, Š., Pollak, S., Huber, D. in Lutar, M. (2023). *Corpus of textbooks for learning Slovenian as L2 ccKUUS 2.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1878

Kosem, I., Gantar, P., Arhar Holdt, Š., Gapsa, M., Zgaga, K. in Krek, S. (2024). AI in lexicography at the University of Ljubljana: Case studies. V S. Krek (ur.), *Book of abstracts of the workshop: Large Language Models and Lexicography* (str. 29–32). https://www.cjvt.si/wp-content/uploads/2024/10/LLM-Lex_2024_Book-of-Abstracts.pdf

Kosem, I., Krek, S. in Gantar, P. (2024). Defining collocation for Slovenian lexical resources. *Slovenščina 2.0*, *8*(2), 1–27. https://doi.org/10.4312/slo2.0.2020.2.1-27

Makarova, I., Zupančič Logar, N. in Lah Peternel, V. (2019). *Slovensko-ruski kolokacijski slovar za začetnike*. Amebis. https://www.termania.net/?searchIn=Linked&ld=285

Stritar Kučuk, M., Šter, H., Pisek, S., Petric Lasnik, I., Kete Matičič, J., Pirih Svetina, N., Preglau, D., Arhar Holdt, Š., Krsnik, L., Erjavec, T., Pegan, J. in Huber, D. (2025). *Slovene learner corpus KOST 2.1* [Korpus]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/2066

Zupan, J. (2013). *Pomenska mreža slovenskih glagolov*. Založba ZRC, ZRC SAZU.

Reprezentacije spola v slovenski pripovedni prozi 19. stoletja v sodobnem srednješolskem kanonu: korpusni pristop

DOI: 10.4312/Obdobja.45.267-275

Lucija Mandić, Inštitut za slovensko literaturo in literarne vede, ZRC SAZU, Ljubljana;
lucija.mandic@zrc-sazu.si

V prispevku z uporabo vektorskih vložitev besed analiziram slovensko pripovedno prozo 19. stoletja z vidika reprezentacije spola. Kanon opredelim na podlagi vključitve del v sodobna srednješolska berila, pri čemer primerjam jezikovne modele kanoničnih in nekanoničnih del ter del avtorjev in avtoric. Analiza pokaže, da se poimenovanja družbeno nezaželenih lastnosti, vedenj in družbenih vlog pogosteje pojavljajo v semantični bližini besede *ženska* kot besede *moški*, pri čemer je ta razlika izrazitejša v delih avtorjev. Razlike med kanoničnim in nekanoničnim podkorpusom so manj izrazite kot razlike med avtorji in avtoricami, kar predstavlja pomembno izhodišče tudi za razmislek o reprezentaciji spola v sodobnem srednješolskem kanonu.

slovenska književnost, literarni kanon, računalniška literarna veda, vektorske vložitve besed, feministična kritika

This article uses word embeddings to analyze nineteenth-century Slovenian narrative prose from the perspective of gender representation. The canon is defined by the inclusion of works in contemporary secondary school readers, and the study compares language models trained on canonical and non-canonical works as well as on works by male and female authors. The analysis shows that words denoting socially undesirable traits, behaviors, and social roles occur more frequently in the semantic neighborhood of the word *ženska* 'woman' than of the word *moški* 'man', with this difference being more pronounced in works by male authors. Differences between the canonical and non-canonical subcorpora are less pronounced than those between male and female authors, providing an important point of departure for reconsidering gender representation in the contemporary secondary-school canon.

Slovenian literature, literary canon, computational literary studies, word embeddings, feminist criticism

1 Uvod

Izbori literarnih del, kot jih danes posredujejo šolska berila, vzpostavljajo kontinuiran stik med zgodovinsko literarno produkcijo in sodobno jezikovno ter kulturno prakso. Besedila, ki so vključena v izobraževalni sistem, niso le predmet literarnozgodovinskega interesa, temveč soustvarjajo predstave o identiteti in vrednotah določene (nacionalne) skupnosti (Juvan 2006: 213). V prispevku bo s pomočjo metode vektorskih vložitev besed raziskano, kako sta moški in ženski družbeni spol prezentirana v delih slovenske proze dolgega 19. stoletja, ki so vključena v sodobna slovenska srednješolska berila, in

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

tistih, ki se v srednješolski literarni kanon iz različnih razlogov niso uvrstila. Poleg tega prispevek obravnava tudi vprašanje, ali se reprezentacija spolov razlikuje med deli avtorjev in avtoric. Reprezentacija spola je raziskana na ravni besedišča, ki se pojavlja v bližini besed *moški* in *ženska*. Ker je pri tem obravnavana velika količina besedil, je za analizo uporabljena metoda vektorskih vložitev besed, interpretacija pa je podkrepljena z dognanji tradicionalne literarne zgodovine. Prispevek tako prikaže možnost uporabe ene od sodobnih metod obdelave naravnega jezika v literarnovednih raziskavah na gradivu starejše slovenske pripovedne proze, hkrati pa obravnava sodobni srednješolski literarni kanon kot enega ključnih mehanizmov, prek katerih se literatura 19. stoletja vključuje v današnji kulturni in izobraževalni prostor.

2 Korpus

Predmet analize je Korpus daljše slovenske pripovedne proze KDSP 1.0 (Mandić in Erjavec 2023), ki obsega 261[1] avtorskih proznih leposlovnih besedil, objavljenih med letoma 1836 in 1918, z dolžino najmanj 20.000 besed. Korpus je zasnovan tako, da zajema celotno literarno produkcijo znotraj teh parametrov in s tem stremi k čim večji reprezentativnosti, tudi glede zastopanosti avtorjev in avtoric. V njem je zastopanih 75 avtorjev (med petimi avtorji z največ besedili v korpusu so: Miroslav Malovrh, Ivan Cankar, Fran Detela, Josip Kostanjevec in Josip Jurčič) z 234 deli ter pet avtoric (Luiza Pesjak, Pavlina Pajk, Marica Nadlišek Bartol, Zofka Kveder in Lea Fatur) z 21 deli, šest del pa je anonimnih. Za to študijo je ključen tudi metapodatek o kanoničnosti: kot kanonična so bila opredeljena tista dela, ki so vsaj omenjena v srednješolskih berilih od leta 1980 dalje. Po tem merilu korpus vključuje 52 kanoničnih in 209 nekanoničnih del.

3 Slovenski srednješolski kanon

Nabor beril, ki je prišel v poštev za ustvarjanje seznama kanoničnih slovenskih proznih del 19. in začetka 20. stoletja, je omejen z letoma 1980 in 2024. Upoštevanih je bilo 28 zvezkov beril, vključno s prenovljenimi in dopolnjenimi izdajami. Izbrano obdobje je tako dovolj dolgo, da vključuje nabor literarnih del, ki so znana čim večjemu deležu slovenskega prebivalstva, in hkrati upošteva revizije beril, na katere so vplivale družbene spremembe ter selektivno vključevanje sočasne literarne produkcije (Božič 2011: 8–9). Z izbrano časovno omejitvijo so zajeta tako berila, ki so bila v veljavi pred osamosvojitvijo Slovenije, kakor tista, ki so bila v veljavi v desetletjih po njej, ko je po spremembi političnega sistema prišlo do zadnjih večjih novosti v prenovi učnih načrtov v letih 1992, 1998 in 2008 (Jožef Beg 2012: 360–362). V tem obdobju je skladno s sočasnimi prizadevanji feministične literarne vede v univerzitetnem okolju prišlo tudi do obsežnejšega vključevanja avtoric v nabore besedil šolskih beril (Božič 2017).

1 Iz verzije korpusa, ki je citirana in objavljena na spletu, je bilo izvzeto eno delo, in sicer *Hanka* Zofke Kveder, saj izvirnik ni bil napisan v slovenščini. V vseh ostalih pogledih gre za enak korpus.

4 Slovenske avtorice 19. in zgodnjega 20. stoletja v srednješolskem literarnem kanonu

Literarno zgodovinopisje je pesnice in pisateljice kljub njihovi dejavni udeleženosti v literarnem življenju redko prepoznalo kot pomembne sooblikovalke literarnih smeri in tokov, sodbe o njihovem literarnem ustvarjanju pa so bile še globoko v drugi polovici 20. stoletja polne slabšalnih in celo posmehljivih pripomb (gl. Borovnik 1995: 28–29). Irena Novak Popov (2008: 115–116) je kot komentar k *Antologiji slovenskih pesnic* (2004–2007) opravila analizo prisotnosti pesnic v literarnovednih raziskavah, antologijah in zbirkah. Po njenih ugotovitvah so v ključnih raziskavah pesnice večinoma prezrte in podcenjene, delež pesnic v pesniških antologijah je minimalen, izdaja njihovih zbranih del se močno zavlačuje, dolgo časa pa so bile izključene tudi iz šolskih učbenikov.

Stanje se je začelo spreminjati v devetdesetih letih 20. stoletja z večjo prisotnostjo feminističnih pristopov v literarni vedi (Mihurko Poniž 2005, 2013). Večje zanimanje za avtorice in poskusi prevrednotenja njihovega ustvarjanja so prispevali k njihovi postopni kanonizaciji in tudi vključevanju v šolska berila. Zoran Božič (2017: 39) ugotavlja, da imajo avtorice večjo možnost, da so vključene v srednješolska berila, če so ta obsežnejša (z več kot 100 slovenskimi avtorji in avtoricami in več kot 300 slovenskimi besedili), v manjših sestavih (z manj kot 80 avtorji in avtoricami) pa je ta možnost precej manjša; ko v krajših berilih število avtorjev in avtoric ter besedil pade pod določeno mejo, po Božičevih ugotovitvah avtorice popolnoma izginejo. Najvišji delež avtoric so dosegla berila za nižjo srednjo šolo takoj po drugi svetovni vojni (okoli 15 %), medtem ko med berili za višjo srednjo šolo po deležu avtoric izstopajo najsodobnejši sestavi (okoli 9 %). Kljub temu srednješolska berila, ki so nastajala v različnih obdobjih, vseskozi izkazujejo zelo nizko zastopanost avtoric. V mlajših berilih, kot sta sestava Janka Kosa (2003) in Bože Krakar Vogel (2003), število vključenih avtoric variira. Božič to pripiše spolu sestavljalk in sestavljalcev: berilo *Branja*, pri sestavi katerega so prevladovale sestavljalke pod vodstvom Bože Krakar Vogel, vključuje več avtoric kakor berilo Janka Kosa, ki so ga sestavili predvsem moški (Božič 2017: 40–41). Ana M. Sobočan (2008: 108–109) v podrobnejši analizi reprezentacije del avtoric v *Branjih* ugotavlja, da je njihovih del sicer več kakor v kateremkoli slovenskem berilu dotlej (čeprav 37 del avtoric od skupno 687 del predstavlja le dobrih 5 % obsega berila), vendar hkrati odkriva stereotipizirajoče in omalovaževalne komentarje v spremnih besedilih. Kljub večji zastopanosti avtoric v slovenskem srednješolskem kanonu v zadnjih desetletjih pa so te v veliki večini prisotne le v izbirnih seznamih čtiva, s čimer je odločitev o vključevanju slovenskih avtoric v srednješolski pouk slovenščine v veliki meri prepuščena posameznim učiteljem in učiteljicam, saj v učnem načrtu ni formalne zahteve po zagotavljanju uravnotežene zastopanosti avtorjev in avtoric.

Navedeni podatki se praviloma nanašajo na celotne sestave in vključujejo tudi sodobne slovenske ter tuje avtorice. Ker nas v tej analizi prvenstveno zanima reprezentacija družbenega spola v dolgih proznih delih, nastalih v 19. in začetku 20. stoletja, je nabor avtoric še dodatno zožen. V vseh slovenskih berilih je med prozaistkami z deli,

nastalimi v obravnavanem časovnem obdobju in daljšimi od 20.000 besed, vključena le Zofka Kveder z romanom *Njeno življenje*. Vsa druga dela avtoric tako ostajajo v nekanoničnem podkorpusu, enako pa velja za vsa anonimna dela.

5 Metoda

Namen te analize je preučiti reprezentacijo moškega in ženskega družbenega spola, kot se vzpostavlja skozi jezikovne vzorce v besedilih. Analiza temelji na dveh ločenih primerjavah: med deli avtorjev in avtoric ter med kanoničnimi in nekanoničnimi deli. Pri slednji primerjavi v obeh podkorpusih *de facto* prevladujejo dela avtorjev, pri čemer kanonični podkorpus vključuje le eno delo avtorice.

Eden od načinov preučevanja tovrstnih vzorcev na večjem številu besedil je uporaba vektorskih vložitev besed. Ta sodi med pristope strojnega učenja in temelji na jezikovnih modelih, naučenih na velikih besedilnih korpusih. Izhodišče metode je predpostavka, da je pomen besede določen z njenim kontekstom: modeli, kot je Word2Vec (Mikolov idr. 2013), se pri učenju naučijo predstavljati besede kot vektorje v večdimenzionalnem prostoru, pri čemer imajo besede, ki se pojavljajo v podobnih kontekstih, med seboj bližnje vektorje. Na tej osnovi je mogoče za izbrane besede poiskati nabor semantično najbližjih izrazov, kar omogoča vpogled v širše vzorce rabe. Za razliko od drugih uveljavljenih korpusnih pristopov, npr. konkordančnih analiz kolokacij, ki prikazujejo neposredno besedilno okolico izbrane besede (prim. Juvan idr. 2021: 70–73), vektorske vložitve zajemajo širše distribucijske vzorce v celotnem korpusu. Besede so si v modelu blizu ne (le) zato, ker se pogosto pojavljajo neposredno ena ob drugi, temveč zato, ker se pojavljajo v podobnih kontekstih. Metoda tako omogoča vpogled v širše semantične povezave, ki jih analiza neposrednega sobesedila ne razkrije.

Za učenje jezikovnih modelov uporabljam implementacijo Word2Vec v knjižnici Gensim (Řehůřek in Sojka 2010) v programskem jeziku Python. Modeli so ločeno naučeni na štirih podkorpusih (avtorji, avtorice, kanon, nekanon). Velikost posameznega podkorpusa je razvidna v Tabeli 1. Korpus je predhodno lematiziran in oblikoskladenjsko označen z orodjem CLASSLA (Terčon in Ljubešić 2023), kar omogoča filtriranje posameznih besednih vrst pri učenju modelov. Pri učenju so tako upoštevane le leme samostalnikov in glagolov. Čeprav bi bili pridevniki in prislovi pri raziskovanju vrednotenja družbenega spola lahko informativni, kot modifikatorji samostalniškega oz. glagolskega pomena (Vidovič Muha 2021: 36) ne ustrezajo cilju analize, katere namen je zajeti širše semantične povezave med pomenskimi polji, ne pa opisnih značilnosti posameznih pojmov. Imenske entitete so izločene, saj se ista osebna ali krajevna imena v različnih besedilih nanašajo na različne literarne osebe oz. kraje. Ker model vse pojavitve istega imena združi v eno vektorsko predstavitev, njihove semantične povezave za raziskovalno vprašanje niso informativne. Parametri modela so bili nastavljeni na *sg = 1* (arhitektura skip-gram), *window = 3*, *vector_size = 100* in *min_count = 10*. Z določitvijo spodnje meje frekvence sem v veliki meri zmanjšala vpliv šuma (npr. napačno označenih besed), ki se pogosto pojavlja pri avtomatski obdelavi starejših besedil.

	Celoten korpus	Avtorice	Avtorji	Kanon	Nekanon
Število besedil	261	21	234	209	52
Število besed	14.053.800	1.075.096	12.776.657	2.977.710	11.222.905

Tabela 1: Število besedil in besed v celotnem korpusu in posameznih podkorpusih.

Za analizo je uporabljena funkcija *most_similar*, ki v modelu poišče besede z vektorskimi predstavitvami, najbližjimi izbrani besedi. Iz vsakega modela sta tako pridobljena po dva nabora besed, in sicer za besedi *moški* in *ženska*. Ker so modeli učeni na štirih podkorpusih (avtorji, avtorice, kanon, nekanon), to skupaj predstavlja osem naborov besed, pri čemer vsak nabor obsega 20 semantično najbližjih besed.

Podkorpusa besedil avtorjev in avtoric se precej razlikujeta po velikosti, kar lahko vpliva na kakovost modelov, saj so v manjšem korpusu jezikovni vzorci zastopani z manj primeri. Vendar razlika v velikosti za opravljeno analizo ni ključna, saj so modeli naučeni ločeno, primerjava pa temelji na naborih semantično najbližjih besed znotraj posameznega modela in ne na neposredni primerjavi vektorskih vrednosti med modeli. Poleg tega Word2Vec meri distribucijsko podobnost med besedami: bližina posameznega izraza besedama *ženska* ali *moški* kaže na pojavljanje v podobnih jezikovnih kontekstih, ne pa nujno na neposredno pripisovanje določene lastnosti ženskam oz. moškim. Model prav tako ne razlikuje med pripovednimi perspektivami oz. govorom pripovedovalca in literarnih oseb ter ne zajame kompleksnejših načinov vrednotenja, kot je ironija.

6 Rezultati analize

V Tabelah 2 in 3 lahko najprej opazimo, da sta besedi *moški* in *ženska* v vseh podkorpusih razen pri avtoricah med seboj najbližji, kar potrjuje njuno skupno sobesedilno rabo v večini analiziranih besedil. V nadaljnji interpretaciji se osredotočam predvsem na izraze, katerih leksikalni pomen označuje družbeno nezaželene lastnosti, vedenja ali družbene vloge. Ker gre za besede, ki so v jezikovnem modelu najbližje izhodiščnima besedama *moški* in *ženska*, njihova prisotnost ne pomeni neposrednega vrednotenja teh izhodiščnih besed, temveč kaže na sobesedilne in semantične povezave, ki se okoli njiju vzpostavljajo v posameznih podkorpusih.

Pri primerjavi kanoničnega in nekanoničnega podkorpusa se ti izrazi pojavljajo predvsem med besedami, ki so najbližje besedi *ženska*: v kanoničnem podkorpusu sta to npr. *vlačuga* in *babnica*, v nekanoničnem pa *hinavka* in *neumnica*. Med besedami, ki so najbližje besedi *moški*, takšnih izrazov ne najdemo oz. ti niso enako nedvoumni; beseda *sanjač* bi npr. lahko označevala negativno ovrednoteno značajsko lastnost, vendar je to v večji meri odvisno od konteksta. Skupna značilnost obeh podkorpusov je, da se takšni izrazi pogosteje pojavljajo v bližini besede *ženska* kot besede *moški*.

Še izrazitejše razlike se pokažejo pri primerjavi avtoric in avtorjev. V delih avtorjev se med besedami, ki so najbližje besedi *ženska*, pojavljajo številna poimenovanja družbeno nezaželenih lastnosti, vedenj ali družbenih vlog, npr. *hinavka*, *vlačuga*, *nečimrnost* (tudi varianta *ničemurnost*) in *pijanka*. Posamezni tovrstni izrazi se pojavijo tudi med besedami, ki so najbližje besedi *moški*, vendar prav tako označujejo ženske, npr. *hinav-*

ka. Pri avtoricah pa je nabor besed, ki so najbližje besedi *ženska*, brez takšnih besed, potencialna izjema sta le *obžalovati* in *sramovati* na dnu seznama. Negativna vrednostna zaznamovanost se tako v njihovih delih pojavlja le posredno in manj izrazito, kot npr. tudi pri besedah *prezirati* in *pregrešiti*, ki sta blizu besedi *moški*, vendar ju je brez dodatnega konteksta težko zanesljivo interpretirati.

KANON		NEKANON	
moški	**ženska**	**moški**	**ženska**
ženska	moški	ženska	moški
sanjač	dekle	možek	možek
oponašati	devica	zunanjost	částnik
pravilo	fant	hinavka	žena
občevanje	tepsti	robatost	zaljubiti
spol	pijanec	spol	hinavka
svetnica	človek	částnik	človek
vesti	vlačuga	dekle	otročaj
izbira	pristojati	dopadati	samohvala
talent	punca	laskanje	dekle
pobožnost	marati	obnašanje	tekmovalka
uganjati	babnica	boječnost	tip
ženiti	svetnica	mladenč	zunanjost
šega	ženica	punca	nečimernost
ujemati	kopa	tip	meščanka
prikupiti	družica	noša	koristka
punca	dečko	možak	neumnica
spadati	odkriti	prijati	robatost
lastnost	ženiti	smešiti	lišpati
ženica	ostrmeti	zaljubiti	punca

Tabela 2: Prvih 20 besed, najbližjih besedi *ženska* oz. *moški*, v kanoničnem in nekanoničnem podkorpusu.

AVTORICE		AVTORJI	
moški	**ženska**	**moški**	**ženska**
občevanje	poznati	ženska	moški
prezirati	vedeti	možek	možek
prikupiti	moški	otročaj	hinavka
povest	vedenje	spol	dekle
pregrešiti	ugajati	gospodič	částnik
blago	nazor	mladenč	vlačuga
priznati	soprog	dopadati	ničemurnost
vzrok	govor	ženstvo	otročaj
razum	lastnost	noša	pijanka
razlika	mnenje	čislo	lej
potrditi	praviti	částnik	tip
pisatelj	zunanjost	boječnost	sužnja

AVTORICE		AVTORJI	
moški	**ženska**	**moški**	**ženska**
ženitev	ceniti	hinavka	nečimernost
prenašati	soditi	komedijant	tekmovalka
dokaz	odpuščati	flirt	notarka
možitev	oseba	dekle	dekletce
bratranec	značaj	ničemurnost	meščanka
previdnost	blago	možak	punca
zagovarjati	obžalovati	obnašanje	stvarca
prijateljstvo	sramovati	smešiti	človek

Tabela 3: Prvih 20 besed, najbližjih besedi *ženska* oz. *moški*, v kanoničnem in nekanoničnem podkorpusu.

7 Interpretacija rezultatov in sklepne ugotovitve

Dobljeni rezultati se v veliki meri ujemajo z ugotovitvami tradicionalne literarne vede o položaju avtoric in recepciji njihovih del ob koncu 19. in na začetku 20. stoletja. Kot je pokazala analiza, se poimenovanja negativno vrednotenih lastnosti, vedenj in družbenih vlog pogosteje pojavljajo v semantični bližini besede *ženska* kot besede *moški*, pri čemer je ta razlika največja v delih avtorjev. Takšna razporeditev ne omogoča neposrednih sklepov o vzrokih kanonizacije, vendar jo je mogoče interpretirati v luči obstoječih literarnozgodovinskih raziskav. Te potrjujejo, da so bile avtorice v preteklosti sicer prisotne v literarnem prostoru, vendar pogosto obravnavane kot nenevaren spremljevalni pojav dominantne moške produkcije (Mihurko Poniž 2010: 48; Žnidaršič Žagar 2009: 16), medtem ko so bile v primerih, ko so dejansko vzpostavile konkurenčen odnos, deležne izrazito negativne recepcije (Hladnik 2007: 67–68). Prej opisani nabori besed tako ponujajo možno interpretacijo, ki je skladna z ugotovitvami literarne zgodovine: da se mizogini vzorci, ki jih lahko zaznamo v delih avtorjev, ne pojavljajo le v nekanoničnih delih, temveč tudi v kanonu. Avtorice z vidika reprezentacije spola ostajajo na obrobju literarnega sistema kot takšnega in vsekakor zunaj kanona. Hkrati rezultati dopuščajo tudi interpretacijo, da so v delih avtoric liki obeh spolov redkeje povezani z negativno vrednotenimi lastnostmi, vedenji in družbenimi vlogami, kar pa se sklada z očitki o njihovem anahronističnem vztrajanju pri žanru sentimentalnega romana in z literarnokritiškim favoriziranjem realizma kot merila literarne vrednosti, ki so se pojavljali v literarni vedi vse do konca 20. stoletja (Mihurko Poniž 2017: 13–14; Žnidaršič Žagar 2009: 167). Kanoničnost problemskosti in realizma, ki se manifestira tudi na ravni reprezentacije spolov, pa obenem lahko pojasni tudi kanonizacijo Zofke Kveder, ki ji je realizem, s katerim je zrcalila realizem avtorjev, kljub očitkom sovraštva do moških omogočil vstop v kanon (Borovnik 1995: 50–53; Mihurko Poniž 2005: 464–503). Prisotnost tako ali drugače problematičnih in realističnih likov v njenih delih je bila očitno odločilna do te mere, da je ni mogel preprečiti niti avtoričin kritični odnos do mizoginije.

Z metodološkega vidika se je uporaba vektorskih vložitev besed izkazala za uporabno metodo prepoznavanja semantičnih vzorcev, ki niso razvidni na ravni posameznih

pojavitev besed. Tako prepoznani vzorci omogočajo oblikovanje hipotez, ki jih je nato mogoče preverjati s tradicionalnimi literarnovednimi pristopi. Ena izmed hipotez je tudi, da se reprezentacija spola izraziteje razlikuje glede na spol avtorja kot glede na kanonični status dela. To pa odpira vprašanje razumevanja srednješolskega kanona kot izbora »reprezentativnih« (prim. Krakar Vogel 2003: 113) del slovenske književnosti. Ker takšni izbori pomembno soustvarjajo predstave o družbenih vlogah, je pri interpretaciji kanoničnih besedil smiselno upoštevati tudi načine reprezentacije spola, ki jih posredujejo.

Prispevek je nastal v okviru raziskovalnega projekta Oblike sramu v slovenski književnosti (J6-60113) in raziskovalnega programa Literarnozgodovinske, literarnoteoretične in metodološke raziskave (P6-0024), ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

Viri in literatura

Borovnik, S. (1995). *Pišejo ženske drugače?: O Slovenkah kot pisateljicah in literarnih likih*. Mihelač.

Božič, Z. (2011). Dejavniki literarne kanonizacije v srednješolskih berilih – na primeru Prešerna. *Jezik in slovstvo*, *56*(5–6), 3–26.

Božič, Z. (2017). Avtorice v slovenskih srednješolskih berilih. *Jezik in slovstvo*, *62*(1), 33–43.

Hladnik, M. (2007). Pavlina Pajk (1854–1901): Prva dama slovenskega ženskega romana in povesti. V A. Šelih, M. Antić Gaber, A. Puhar, T. Rener, R. Šuklje in M. Verginella (ur.), *Pozabljena polovica: Portreti žensk 19. in 20. stoletja na Slovenskem* (str. 65–68). Tuma in SAZU.

Jožef Beg, J. (2012). Svetovna književnost v berilih za splošno srednješolsko izobraževanje po letu 1945. V M. Juvan (ur.), *Svetovne književnosti in obrobja* (str. 353–371). Založba ZRC, ZRC SAZU.

Juvan, M. (2006). *Literarna veda v rekonstrukciji: Uvod v sodobni študij literature*. LUD Literatura.

Juvan, M. Šorli, M. in Žejn, A. (2021). Interpretiranje literature v zmanjšanem merilu: »Oddaljeno branje« korpusa »dolgega leta 1968«. *Jezik in slovstvo*, *66*(4), 55–75. https://doi.org/10.4312/jis.66.4.55-75

Krakar Vogel, B. (2003). Spremembe v srednješolskem pouku književnosti ob koncu 20. stoletja. V B. Krakar Vogel (ur.), *Slovenski jezik, literatura in kultura v izobraževanju* (39. seminar slovenskega jezika, literature in kulture, str. 109–123). Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete.

Mandić, L. in Erjavec, T. (2023). *Corpus of longer narrative Slovenian prose KDSP 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1823

Mihurko Poniž, K. (2005). Feministične literarnovedne raziskave – tukaj in zdaj. *Jezik in slovstvo*, *50*(3–4), 87–95. https://doi.org/10.4312/jis.50.3-4.87-95

Mihurko Poniž, K. (2010). Prispevki slovenskih pripovednic k žanrski podobi proze 19. stoletja. *Jezik in slovstvo*, *55*(1–2), 47–59. https://doi.org/10.4312/jis.55.1-2.47-59

Mihurko Poniž, K. (2013). Feministična literarna veda v slovenskem akademskem prostoru. *Slavistična revija*, *61*(1), 209–214.

Mihurko, K. (2017). *Literarna ustvarjalka v očeh druge_ga: Študije o recepciji, literarnih stikih in biografskem diskurzu.* Založba Univerze v Novi Gorici. https://www.dlib.si/details/URN:NBN:SI:DOC-2J9O6CZ6

Mikolov, T., Chen, K., Corrado, G. in Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*. https://arxiv.org/abs/1301.3781

Novak Popov, I. (2008). Slovenske pesnice v literarni vedi in izobraževanju. V B. Krakar Vogel (ur.), *Književnost v izobraževanju – cilji, vsebine, metode* (Obdobja 25, str. 113–128). Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete. https://centerslo.si/wp-content/uploads/2015/10/25-NovakPopov.pdf

Řehůřek, R. in Sojka, P. (2010). Software framework for topic modelling with large corpora. V *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks* (str. 46–50). L-Università ta' Malta. https://hal.science/hal-05382247

Sobočan, A. M. (2008). Branje kanona. V B. Krakar Vogel (ur.), *Književnost v izobraževanju – cilji, vsebine, metode* (Obdobja 25, str. 99–112). Center za slovenščino kot drugi/tuji jezik. https://centerslo.si/wp-content/uploads/2015/10/25-Sobocan.pdf

Terčon, L. in Ljubešić, N. (2023). CLASSLA-Stanza: The next step for linguistic processing of South Slavic languages. *arXiv*. https://arxiv.org/abs/2308.04255

Vidovič Muha, A. (2021). *Slovensko leksikalno pomenoslovje*. Znanstvena založba Filozofske fakultete. https://doi.org/10.4312/9789610604976

Žnidaršič Žagar, S. (2009). *Vstop Slovenk v literarni prostor: Okoliščine, v katerih so prve slovenske pisateljice, od Fany Hausmann do Zofke Kveder, (lahko) vstopile v slovenski literarni prostor: Priročnik za pouk literature žensk*. Pedagoški inštitut. https://www.dlib.si/details/URN:NBN:SI:doc-FXCKCKBD

Грамматические и лексические сходства и различия в высокочастотной лексике: самые частые словенские, русские и английские субстантивы и глаголы

DOI: 10.4312/Obdobja.45.277-287

Nina B. Mečkovskaja, Belorusskij gosudarstvennyj universitet, Minsk;
nina.mechkovskaya@gmail.com

Frekvenčni slovarji so vir novih spoznanj ne le o besedišču, temveč tudi o slovnični organizaciji jezika. Povezanost med frekvenčnimi slovarji in slovnico se kaže v tem, da se znotraj določenih frekvenčnih območij leme določenih besednih vrst uvrščajo na značilna mesta v razvrstitvi glede na leme drugih besednih vrst. Prispevek pokaže tudi, da je pri 20–30 najpogostejših polnopomenskih besedah opazna izrazita semantična podobnost, ki je večja pri glagolih kot pri samostalnikih. Primerjava 200 najpogostejših samostalnikov po skupinah (z izračuni standardnega odklona) je pokazala, da so si slovenski in ruski samostalniki najmanj podobni; največje razlike med skupinami glede na število leksemov so se pokazale v skupini, ki zajema poimenovanja različnih vrst informacij, njihovih nosilcev in sredstev prenosa. Moč medjezikovnih razlik v skupinah samostalnikov z visoko frekvenco je sorazmerna s stopnjo razlik v tematski strukturi (v odstotkih) nacionalnih besedilnih korpusov.

frekvenčni slovar, frekvenčni rangi, semantični skoraj univerzalni izrazi, akumulirane frekvence, najpogostejši samostalniki, najpogostejši glagoli

Frequency dictionaries are a source of new insights not only about vocabulary but also about the grammatical organization of language. The relationship between frequency dictionaries and grammar is evident in the fact that, within certain frequency ranges, lemmas of certain parts of speech exhibit specific rank relationships with lemmas of other parts of speech. This article also shows that the top twenty to thirty most frequent lexical words exhibit significant semantic similarity, with greater similarity in verbs than in nouns. A cluster comparison of the two hundred most frequent nouns (with a calculation of the standard deviation) showed that Slovenian and Russian nouns are the least similar to each other; the strongest differences between clusters in the number of lexemes included were noted for the "information" cluster. The strength of interlingual differences in clusters of high-frequency nouns is proportional to the degree of differences in the thematic structure (in percent) of national text corpora.

frequency dictionary, frequency ranks, semantic near-universals, accumulated frequencies, most frequent nouns, most frequent verbs

1 Познавательная ценность сходств и различий между частотным словарями разных языков

Частотный словарь представляет собой вероятностную модель лексики в виде списка лемм, ранжированного по коммуникативной значимости (что определяет

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

их частотность) и систематизированного частеречной организацией лексики. Категориальная специфика грамматических разрядов слов проявляется в особенностях ранговых списков разных частей речи, например, в большем лексическом разнообразии субстантивов по сравнению с глаголами или прилагательными того же частотного диапазона.

Данный метаязыковой факт – то, что частотные словари разных языков некоторым образом отображают частеречную организацию лексики, – позволяет видеть в частотном словаре источник нового знания не только о лексике, но и о грамматике, иначе говоря, о языке в целом.

2 Каким образом грамматическая организация лексики отображается в частотном словаре?

2.1 Начало общего (не терминологического) частотного словаря любого языка составляют служебные слова и местоимения

Грамматические (служебные) слова и местоимения преобладают в составе первых 100 лемм новой версии эталонного корпуса словенского языка Gigafida 2.0[1], как и в первой сотне частотного словаря Ляш & Шар 2009, основанного на корпусе объемом 93 млн словоупотреблений в составе НКРЯ. А именно: в списке первых по частоте 100 слов в Ляш & Шар (2009) 72 леммы – это служебные слова и местоимения; в Seznam от Gigafida 2.0 таких слов 62 леммы[2]. Однако уже во второй сотне списка служебные слова и местоимения находятся в меньшинстве и далее появляются лишь спорадически и, как правило, в своих архаических или территориально ограниченных вариантах. Место в частотном словаре служебных слов соответствует назначению грамматики: грамматические значения, необходимый семантический компонент всякого слова, синтагмы, предложения, создают смысловую основу общения, легко и в онтогенезе рано генерируемую и легко воспринимаемую; на этом грамматическом фундаменте фразы появляются и действуют более конкретные индивидуальные лексические значения слов.

---

1 Электронный *Список первой 1.000 лемм в корпусе Gigafida 2.0* (*Seznam 1.000 najpogostejših lem v korpusu Gigafida 2.0*), полученный на основе Gigafida 2.0, стал мне доступен в 2020 г. благодаря любезности и исследовательской солидарности доктора Андреи Желе, профессора университета в Любляне, и доктора Андрея Пердиха научного сотрудника Института словенского языка им. Франа Рамовша Словенской академии наук и искусств, в связи с моей работой над статьей Заимствования в первой тысяче самых частых белорусских, словенских и русских слов: опыт интерпретации количественных и содержательных различий (*Slavistična revija*, 2020, № 3, с. 311–333; версия статьи на белорусском языке в *Journal of the Belarusian State University. Philology*, 2021, № 1, с. 25–35). Пять лет спустя вновь приношу свою глубокую признательность уважаемым коллегам за помощь.

2 Что касается частотного списка английских лемм в Leech et al. 2001, то его сопоставление с аналогичными словенским и русским списками не представляется возможным в силу особенностей формирования английского списка: в нем словоформы (*was*, *are*; *said*, *does*, *years*; *him* и т. п.), а также постфиксы и финали (*'s*, *'ll*, *-re*, *-ing* и т. п.) рассматриваются как самостоятельные леммы.

2.2 Вопрос о зависимостях между количеством лемм (относимых к той или иной части речи) и суммой словоупотреблений лемм данной части речи

В словенском и русском языках суммарный процент словоупотреблений лемм, относимых к той или иной части речи, во всем объеме частотного словаря (объединяющем знаменательные и служебные слова), частоты грамматических слов и частоты знаменательных слов, имея разную природу, по-разному относятся к количеству лемм. Грамматические (служебные) слова, численно не превышая несколько десятков единиц, могут, однако, по накопленным частотам опережать классы знаменательных лексем. Например, в Ляш & Шар (2009) сумма словоупотреблений предлогов превышает суммы словоупотреблений всех прилагательных и всех наречий. В итоге предлоги находятся на 3-ем месте по проценту накопленных частот, опережая прилагательные и наречия. Это объясняется тем, что употребительность слов, входящих в разряды знаменательных частей речи, с одной стороны, и служебных слов, с другой, регулируется разными языковыми законами: знаменательные слова референтны по отношению к внеязыковой реальности (и в аспекте своей категориальной частеречной семантики, и в аспекте синтаксической роли), в то время как служебные слова (предлоги и союзы) – это внутриязыковые средства организации высказывания. В качестве тенденции – это достаточно очевидно, и для русского языка это подтверждается данными частотных словарей (Засорина 1977; Ляш & Шар 2009), однако для словенского языка эмпирических данных Gigafida 2.0 не достает. Тем более их не достает в Leech et al. 2001, поскольку в этом словаре большинство корпусных данных разнесено по двум главным для словаря оппозициям – Spoken vs Written English, и целостная картина часто не вырисовывается.

2.3 Вопрос о прямой зависимости между количеством лемм и суммой словоупотреблений в частотной организации лексики знаменательных частей речи (субстантивы, глаголы, прилагательные, наречия)

На материале русских частотных словарей эта зависимость прослеживается – не в числовых пропорциях, но в количественной иерархии грамматических разрядов. Во-первых, существуют зависимости между количеством лемм в каждом частеречном разряде и процентом покрытия текстов леммами данного разряда; во-вторых, существуют регулярные различия между некоторым одинаковым рангом лемм в разных частеречных частотных списках и теми рангами, которые эти леммы получают в общем частотном списке, а именно: наименьшая цифра ранга (т. е. максимальная частотность) характерна для субстантивов; далее ранги возрастают (а кумулятивная частотность уменьшается) в таком направлении: глаголы, прилагательные, наречия.

Это значит следующее: 1) существительные как класс слов в целом употребляются чаще, чем класс глаголов; 2) глаголы как класс слов в целом

употребляются чаще, чем класс прилагательных; 3) прилагательные как класс слов в целом употребляются чаще, чем класс наречий.

В следующей таблице (построенной по данным Ляш & Шар 2009) с помощью трех разных видов заливки противопоставлены три группы лемм, при этом в каждой группе по 4 леммы. Эти 4 леммы принадлежат к одной из 4 частей речи (субстантив, глагол, прилагательное и наречие) и в каждой группе эти леммы имеют одинаковый ранг в своем частеречном частотном списке. Однако в частотном списке, общем для всех грамматических классов, эти леммы занимают разное место по отношению к началу списка: ближе всех – субстантивы, на большем отдалении от начала списка – глаголы, еще больше отдалены прилагательные и дальше всех – наречия.

Эти разные расстояния видны по рангу леммы в общем частотном списке: в каждой группе лемм ранг лемм возрастает в направлении от субстантива к наречию. Иначе говоря, в общем частотном списке в диапазоне до любого ранга число лемм-субстантивов превышает число глаголов, число глаголов превышает число прилагательных, а число прилагательных превышает число наречий.

Леммы	**Ранг леммы в частотном словаре «своей» части речи**	**Ранг леммы в общем частотном словаре**
книга	50	230
лежать	50	312
экономический	50	514
сначала	50	608
театр	100	330
принимать	100	521
прекрасный	100	790
естественно	100	1.196
тип	200	570
улыбаться	200	1.070
творчески	200	1.459
нередко	200	2.620

Таблица 1: Ранги лемм 4-х знаменательных частей речи по 2-м частотным спискам.

Максимальное число лемм-субстантивов в частотном списке в диапазоне до любого ранга (по сравнению с глаголами, прилагательными и наречиями) объясняется с когнитивной точки зрения различиями в характере грамматической семантики у слов разных частей речи. Сознание говорящих видит явления окружающей жизни чаще в облике предметов (как бы мысленно «остановленных» взглядом-сознанием), нежели в более динамичном и летучем облике действий и признаков-свойств).

Выявленные соотношения согласуются с законом Ципфа-Мандельброта, согласно которому в частотном словаре, в интервале между словами с рангом бóльшим чем 50, но меньшим чем 1500 (т. е. $50 < r < 1500$), зависимость между

рангом слова в словаре и его частотой изменяется обратно пропорционально: с повышением частоты слова, уменьшается номер его ранга в частотном списке.

В английском частотном словаре Leech et al. 2001 количественная иерархия лемм основных частей речи такая же, как и в Ляш & Шар 2009: количество лемм уменьшается в направлении от субстантивов к глаголам, затем адъективам и адвербативам. С учетом закона Ципфа-Мандельброта можно допустить, что и уменьшение накопленных частот леммами 4-х частей речи идет в том же порядке (хотя без цифровых данных такое допущение остается гипотезой).

Что касается количественных показателей словенских источников, то здесь, по данным Sloleks и Gigafida 2.0 представлена иная (чем в английском и русском корпусах) количественная иерархия 4-х знаменательных частей речи: по количеству лемм прилагательные опередили класс глаголов и заняли 2-ое место после субстантивов. Однако по сумме словоупотреблений прилагательные находятся на 3-ьем месте (что совпадает с английским и русским распределением частот). См. Таблицу 2.

Части речи	**Число лемм по Sloleks 2.0**	**Число лемм по Gigafida 2.0**	**Число словоупотреблений по Gigafida 2.0**
Существительные	54.200	55 тыс	257 млн (23, 1%)
Прилагательные	**26.612**	**34 тыс**	**114 млн (10,2 %)**
Глаголы	10.242	13 тыс	163 млн (14,6 %)
Наречия	3/904	8 тыс	58 млн (5,2 %)

Таблица 2: Количество лемм и количество словоупотреблений в границах 4-х основных частей речи словенского языка.

То, что словенские прилагательные по числу лемм опередили класс глаголов, представляется особенно удивительным при обращении к словарю Засорина (1977), в котором учтена частеречная принадлежность всех 40 тыс. лемм словника словаря и определена относительная суммарная частота лемм каждой части речи как во всем миллионном корпусе словаря, так и в подкорпусах 4-х стилистических разновидностей текстов, причем всегда со значительным количественным перевесом глагола. Напр., в текстах драматургии число глагольных форм превысило употребительность прилагательных более чем в 3 раза (Засорина 1977: 927).

Трудно понять, почему в Gigafida 2.0 лемм-прилагательных в 2,6 раза больше, чем лемм-глаголов. Прилагательные востребованы прежде всего в газетно-журнальных текстах (в атрибутивных синтагмах, называющих бесчисленные разновидности окружающих предметов) и менее всего – в художественной прозе. Но именно такие пропорции текстов и характерны для Gigafida 2.0: токены газетно-журнальных и интернет-текстов составляют 92,3 % корпуса, художественных текстов – 3,5 % (Krek et al. 2020).

3 Максимальная семиотическая – общечеловеческая и национально-историческая – значимость 20-30 самых частых знаменательных слов в языке

В частотных списках слов самые частые леммы кажутся почти символами, главными феноменами жизни человека и бытия человечества. Их значения близки к ментальным универсалиям языков.

В Таблице 3 представлены ранговые списки первых по частоте 10 субстантивов в 4-х языках: латинском (по Diederich 1939), английском, словенском и русском. Цветом маркированы слова, которые встречаются хотя бы в 2-х столбцах. Без заливки остаются ячейки со словами с такой семантикой, которая встречается только в одном столбце, т. е. в одном их 4-х языков.

Из 40 ячеек Таблицы 3 половина содержат значения, которые присутствуют только в одном столбце; во второй половине ячеек представлены значения, которые сближают языки; больше всего общих значений в английском столбце. В латыни и в английском списках из 10 самых частых слов имеется по 4 таких леммы, которые встречаются только в одном из 4-х списков; в словенском и русском списках из 10 самых частых слов имеется по 5 лемм, которые встречаются только в одном из 4-х списков.

Примечателен такой факт: в словенской десятке самых частых субстантивов видим такие поздние по времени появления слова, как *odstotek* ‘процент’, *podjetje* ‘предприятие’. По-видимому, это связано с бóльшим удельным весом в словенском корпусе тематически более современных текстов, чем в английском и русском корпусах, не говоря о латинском частотном словаре.

Ранг	Латынь	Английский	Словенский	Русский
1	*deus* ‘бог’	*time* ‘время’	*leto* ‘год’	*год*
2	*res* ‘предмет, вещь’	*year* ‘год’	*čas* ‘время’	*человек*
3	*homo* ‘человек’	*people* ‘люди’	*dan* ‘день’	*время*
4	*tempus* ‘время’	*way* ‘путь’	*delo* ‘дело’	*дело*
5	*locus* ‘место’	*man* ‘человек’	*mesto* ‘город’	*жизнь*
6	*dominus*‘господин’	*day* ‘день’	*človek* ‘человек’	*день*
7	*modus* ‘мера’	*thing* ‘вещь’	*država* ‘государство’	*рука*
8	*animus*‘душа’	*child* ‘ребенок’	*svet* ‘мир’	*раз*
9	*terra* ‘земля’	*Mr* ‘мистер’	*odstotek* ‘процент’	*работа*
10	*vir* ‘мужчина’	*government* ‘правительство’	*podjetje* ‘предприятие’	*слово*

Таблица 3: Первые по частоте 10 субстантивов в 4-х языках.

Сопоставление самых частых глаголов в исследуемых языках (Таблица 4) обнаруживает во многом иную (чем в субстантивных списках) степень межъязыкового семантического сходства между глагольными списками: из 40 ячеек только 6 содержат глагол с уникальным для таблицы значением. Значит, в 4-х языках 10 первых по частоте глаголов денотативно ближе друг другу, чем 10 самых частотных субстантивов.

Это объясняется тем, что денотативно и референтно глаголы сильнее, чем субстантивы, связаны с материальной природой человека, с его анатомией и физиологией. Ср. самые необходимые в повседневности действия-движения и соответствующие им высокочастотные глаголы с такими значениями, как ‘мочь’, ‘хотеть’, ‘сидеть / садиться’, ‘спать’, ‘лежать’, ‘стоять / встать’, ‘идти / прийти / ходить’, ‘пить’ / ‘есть’, ‘кушать’, ‘думать’, ‘читать’, ‘писать’.

Ранг	Латынь	Английский	Словенский	Русский
1	*sum* ‘быть’	*be* ‘быть’	*biti* ‘быть’	*быть*
2	*dico* ‘говорить’	*have* ‘иметь’	*imeti* ‘иметь’	*мочь*
3	*possum* ‘быть в состоянии’	*do* ‘делать’	*morati* ‘быть должным / обязанным’	*сказать*
4	*facio* ‘делать’	*say* ‘говорить’	*iti* ‘идти’	*говорить*
5	*video* ‘видеть’	*can* ‘мочь’	*začeti* ‘начать’	*знать*
6	*habeo* ‘иметь’	*get* ‘получать’	*priti* ‘прийти’	*стать*
7	*venio* ‘приходить’	*make* ‘сделать, создать’	*povedati* ‘говорить’	*идти*
8	*fero* ‘нести’	*go* ‘идти’	*dobiti* ‘получать’	*иметь*
9	*volo* ‘хотеть’	see ‘видеть’	*želeti* ‘хотеть’	*видеть*
10	*audio* ‘слушать’	*know* ‘знать’	*vedeti* ‘знать’	*хотеть*

Таблица 4: Первые по частоте 10 глаголов в 4-х языках.

Самые частые и денотативно близкие в разных языках субстантивы в большей мере, чем глаголы, связаны с культурой и абстрактными представлениями (ср. ‘время’, ‘год’, ‘день’, ‘человек’, ‘вещь’). Поэтому в исследуемых 4-х языках первые по частоте 20-30 глаголов различаются между собой меньше, чем первые 20-30 субстантивов в частотных списках этих же языков. Денотативное сходство высокочастотных лемм в большей мере характерно для глаголов (примерно для 7-8 лемм из 10), чем для субстантивов (6-5 лемм из 10).

За пределами 30 самых частых знаменательных лемм межъязыковое семантическое сходство резко падает. Ранговые списки лемм в разных языках становятся бесконечно разными как по составу лемм, так и по частотным характеристикам переводных соответствий.

4 Кластерное сопоставление 200 высокочастотных словенских, русских и английских субстантивов

Кластеризация субстантивов в границах своего языка позволила представить лексику “национальных” списков в более обобщенном и семантически концентрированном виде. В каждом списке лемм было выделено 8 кластеров. Их названия-характеристики даны в шапке таблицы 5.

В случаях многозначности и омонимии лексемы при ее отнесении к тому или иному кластеру принималось во внимании количество значений в каждом из омонимов (в толковом словаре), и включение леммы в конкретный кластер определялось по 1-ому значению слова в той словарной статье, в которой количество значений больше, чем в любой другой. Так, два словенских омонима

– *zákon* (‘брак’) и *zákon* (‘закон’) – по семантике следовало бы отнести в IV и VII кластеры. Однако, согласно SSKJ, лексема *zákon* (с 1-ым значеним ‘закон’) более многозначна, чем лексема *zákon* (с 1-ым значением ‘брак’). По сформулированному выше правилу, лемма *zákon* (‘закон’) отнесена к VII кластеру.

В словарях трех исследуемых языков слова общего происхождения, включая заимствования из одного языка в другой, нередко различаются местом в словарной статье (т. е., очередностью подачи сходных значений), что, по представлениям лексикографа, должно соответствовать различиям значений по релевантности в языковом сознании говорящих.Так, в англ. *company* 1-ое значение, согласно MED, это «an organization that provides services or makes or sells goods for money» (‘организация, которая предоставляет услуги или производит и/или продает товары за деньги’). Между тем, согласно Кузнецов 2001, 1-ое значение лексемы *компания* – это «общество, группа лиц, проводящих вместе время или чем-либо объединённых». Согласно сформулированным выше приниципам кластеризации, англ. *company* относится ко 2-му кластеру, в то время как рус. *компания* должно быть включено в 1-ый класс.

Результаты кластерного сопоставления представлены в Таблице 5. По каждому кластеру указано количество лемм (в абсолютном и процентном измерении) и место (римская цифра) отдельного кластера в рейтинге 8 кластеров, выстроенном в зависимости от наполненности кластера леммами «своего» языка.

№	Семантические кластеры лемм в составе 200 самых частых субстантивов	Количество лемм и место отдельного кластера в составе 200 самых частых субстантивов		
		Словенский	Русский	Английский
1	Человек, люди, людские общности	34 (17%) **III.**	25 (12,5%) **IIV–V.**	29 (14,5%) **III.**
2	Социально организованное пространство деятельности	28 (14%) **IV.**	23 (11,5%) **VI.**	27 (13,5%) **V.**
3	Измерения, разграничения, единицы измерения времени, пространства, денег	40 (20%) **II.**	38 (19%) **I.–II.**	33 (16,5%) **II.**
4	Абстрактные названия действий, состояний, признаков	42 (21%) **I.**	28 (14%) **III.**	36 (18%) **I.**
5	Природа	7 (3,5%) **VII.**	16 (8%) **VII.**	19 (9,5%) **VII.**
6	Артефакты (отдельные предметы)	5 (2,5%) **VIII.**	6 (3%) **VIII.**	6 (3%) **VIII.**
7	Обозначения разных видов информации, ее носителей и средств передачи	21 (10,5%) **VI.**	38 (19%) **I.–II.**	22 (11%) **VI.**
8	8. Леммы с принадлежностью более чем 2-м кластерам, а также леммы, не принадлежащие ни одному из 7 кластеров	23 (11,5%) **V.**	25 (12,5%) **IV.–V.**	28 (14%) **IV.**
	Итого:	200 (100%)	200 (100%)	200 (100%)

Таблица 5: Количество лемм и места кластеров по трем языкам.

Для статистической оценки степени близости трех пар языков было вычислено среднее квадратичное отклонение (называемое также “стандартное отклонение”) между парами языков; учитывалось абсолютное количество лемм в каждом кластере. Среднее квадратичное отклонение вычисляется как корень квадратный из среднего арифметического квадратов разностей между каждой сравниваемой величиной и средним значением для сравниваемых величин (Прохоров 1988: 262, 559).

Среднеарифметические различия в парах языков таковы:

1) Русский – словенский 5,19
2) Словенский – английский 3,43
3) Русский – английский 3,34

Следует отметить, что только в паре «русский – словенский» различия превышают 5 пунктов, т. е. достигают статистической релевантности.

Естественно возникает вопрос: почему между словенским и русским языком различия максимальны? Думаю, дело не в языках и не в ментальности, а в разных пропорциях близкого контента в словенском и русских корпусах. Это показано в Таблице 6.

Контент	**Словенский корпус**	**НКРЯ, Основной корпус**
Художественный	3,5 %	40,6 % – нехудожественная 59, 4 %
Публицистика	64,3 % – газеты и журналы; 3,8% – Non-fiction (Simon Krek et al. 2020)	В составе нехудожественных текстов 59,4 % – публицистика; 18,7 % – учебно-научные тексты; 14,3 % бытовые разговоры
Интернет	28 %	1,6 %

Таблица 6: Процент содержательно близкого контента в словенском и русском корпусах.[3]

5 Выводы

Проведенное исследование подтверждает тезис, согласно которому частотные словари являются источником нового знания как о вероятностной структуре лексики, так и о грамматической организации языка. Один из аспектов связи частотного словаря и грамматики состоит в том, что для лемм разных частей речи характерны особые ранговые соотношения с леммами других грамматических разрядов.

1) В словенском и русском частотных словарях, в составе первых 100 лемм, служебные слова и местоимения количественно преобладают над леммами знаменательных частей речи, однако во 2-ой сотне лемм количественное соотношение лемм названных разрядов меняется в пользу знаменательных слов.

2) Эмпирические данные корпусно-частотной лексикографии, а также следствия закона Ципфа-Мандельброта позволяют сформулировать следующую

3 Сайт НКРЯ: раздел «Статистика корпуса», подраздел «Распределение текстов основного корпуса по видам и другим метапризнакам» (данные обновлены 28.06.2024).

закономерность статистической струкутуры лексики: численность лемм в основных разрядах знаменательных слов (в составах раздельно субстантивов, глаголов, адъективов и наречий) представляет собой нисходящую градацию, которая в конкретных частотных диапазонах ранговых словарей детерминирует вероятностные нисходящие пропорции накопленных частот и иерархию рангов лемм названных разрядов.

3) В частотных словарях трех исследуемых языков для первых по частоте 20-30 знаменательных слов (в составе которых представлены субстантивы, глаголы, адъективы, наречия) характерно значительное семантическое сходство, при этом большее сходство в глаголах (близки примерно 7-8 лемм из 10), чем в субстантивах (близки 6-5 лемм из 10).

4) Кластерное сопоставление 200 первых по частоте субстантивов (с вычислением среднего квадратичного отклонения по количеству лемм в каждом кластере трех языков) показало, что наименее схожи между собой словенские и русские субстантивы; наиболее сильные различия между кластерами по количеству входящих лексем отмечены для кластера «Информация».

5) Время компьютеров и BigData создало новый, хронологически третий способ существования языка (наряду с устным и письменно-печатным) – представительный (национальный) виртуальный корпус электронных текстов на этническом языке. Благодаря внутренним разделениям корпуса по ряду релевантных признаков (устная vs письменная речь; функциональные варианты языка; хронологические разделения и др.), а также благодаря лингвистической аннотации (разметке) корпусов, они представляют собой виртуальные аналитические модели языка, т. е. инсталлированное в корпус знание о языке. Корпусная модель языка сосуществует с такими основными видами лингвистического знания, как академическая грамматика и академический толковый словарь; академическое описание системы функциональных стилей языка на современном этапе его истории; как характеристика хронологической последовательности синхронных срезов в истории языка и выявление тенденций в диахронии языка. Онтология корпуса (или магия его BigData) такова, что результаты, подкрепленные корпусным объяснением, становятся убедительными. Эмпирически полученные цифры, говорят о том, что кластеры в составе 200 самых частых субстантивов в словенском, русском и английском языках в аспекте своей лексической наполненности различаются сильнее всего между словенским и русским списками. Сила этих различий не покажется случайной арифметикой, если принять во внимание фундаментальные различия между словенским и русским корпусами в их статистической структуре (см. Таблицу 5). Однако возникает новый вопрос: можно ли считать, что статистическая (процентная) структура национального корпуса складывается стихийно, «сама собой» и поэтому является более объективной моделью статистической структуры лексики, чем, допустим, построенная исследователем схема (модель) нормативно-стилистического уклада языка? По-моему, вопрос остается открытым.

Литература

Center za jezikovne vire in tehnologije. (б. д.). *Gigafida 2.0: Korpus pisne standardne slovenščine* [Корпус]. Viri CJVT. https://viri.cjvt.si/gigafida

Center za jezikovne vire in tehnologije. (б. д.). *Sloleks: Slovenski oblikoslovni leksikon*. Viri CJVT. https://viri.cjvt.si/sloleks/slv/

Diederich, P. B. (1939). *The frequency of Latin words and their endings*. The University of Chicago Press. https://liturgyscholar.ca/wp-content/uploads/2024/07/Diederich-1939-Frequency-of-Latin-Words.pdf

Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I. in Dobrovoljc, K. (2020). Gigafida 2.0: The Reference Corpus of Written Standard Slovene. B N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk in S. Piperidis (ur.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (c. 3340–3345). European Language Resources Association. https://aclanthology.org/2020.lrec-1.409.pdf

Кузнецов, С. А. (ред.). (2001). *Большой толковый словарь русского языка*. Норинт.

Leech, G., Rayson, P. и Wilson, A. (2001). *Word frequencies in written and spoken English, based on the British National Corpus*. Pearson Education. http://ucrel.lancs.ac.uk/bncfreq/flists.html

Ляшевская, О. Н. и Шаров, С. А. (2009). *Новый частотный словарь русской лексики*. Азбуковник. http://dict.ruslang.ru/freq.php

Macmillan English Dictionary. http://www.macmillandictionary.com/ (доступен до июня 2023 г.).

Прохоров, Ю. В. (ред.). (1988). *Математический энциклопедический словарь.* Советская энциклопедия.

Seznam 1.000 najpogostejših lem v korpusu Gigafida 2.0) [Список первой 1.000 лемм в корпусе словенской лексики Gigafida 2.0. Рукопись]

Slovar slovenskega knjižnega jezika. (1970–1991). Slovenska akademija znanosti in umetnosti, Inštitut za slovenski jezik Frana Ramovša.

Засорина, Л. Н. (ред.). (1977). *Частотный словарь русского языка*. Русский язык.

What Can We Learn from Each Other? Slovenian and Croatian Language Technologies

DOI: 10.4312/Obdobja.45.289-296

Anita Peti-Stantić, Filozofski fakultet, Sveučilište u Zagrebu, Zagreb; apstantic@ffzg.hr

Ker se jezikovni viri za slovenščino in hrvaščino hitro razvijajo, se odpirajo vprašanja o njihovem namenu in zasnovi. Trdimo, da označevanje ni nevtralno, temveč mora odražati predvideno rabo in ciljne uporabnike. Ob opiranju na hrvaške psiholingvistične raziskave prispevek obravnava izbrane slovenske jezikovne vire in razmišlja o tem, kako bi bilo mogoče vključiti dodatne spremenljivke. Vzporedna analiza slovenskih in hrvaških pristopov bi lahko vodila do skupne metodološke razprave o jezikovnih tehnologijah, namenjenih podpori, ne pa nadomeščanju človeških uporabnikov jezika.

jezikovni viri, slovenščina, hrvaščina, psiholingvistične norme

As language resources for Slovenian and Croatian rapidly expand, questions arise about their purpose and design. It is argued that annotation is not neutral but should reflect intended use and target users. Drawing on Croatian psycholinguistic research, this article examines selected Slovenian resources and considers how additional variables could be integrated. By comparing Slovenian and Croatian approaches, a shared methodological discussion is proposed on language technologies that assist, rather than replace, human language users.

language resources, Slovenian, Croatian, psycholinguistic norms

1 Introduction

The rapid expansion of language technology resources over the past decades—ranging from large-scale reference corpora to specialized learner datasets and NLP-oriented annotation frameworks—has significantly transformed the landscape of linguistic research, language teaching, and digital humanities. However, this proliferation also raises a fundamental, although often insufficiently articulated, question: what is the purpose of a given resource, and how should that purpose shape decisions about annotation type, scope, and granularity? Technological advancement has made it increasingly feasible to annotate linguistic data at multiple levels, but the underlying rationale for such annotation is frequently left implicit. This article proceeds from the hypothesis that annotation is never a neutral or purely technical procedure added to linguistic data; instead, it constitutes a set of methodological commitments that must be understood in relation to the intended use, epistemological framework, and target audience of the resource.

This issue becomes especially visible when comparing different traditions in corpus research and language technology. Annotation practices are closely tied to broader

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

methodological and institutional contexts because decisions about what to annotate and how to categorize linguistic phenomena inevitably reflect assumptions about what counts as relevant linguistic knowledge. The question, therefore, is not which approach is superior, but how different annotation strategies align with different research and application goals. In practice, however, such considerations are often subordinated to issues of technical feasibility or immediate applicability. The availability of annotation tools, the cost of manual annotation, and the demands of specific projects frequently shape annotation more than explicit methodological reflection. This can lead to mismatches between the design of a resource and its later use, particularly in interdisciplinary contexts in which annotation schemes are transferred across fields without sufficient adaptation.

I therefore argue for a more reflexive approach to annotation, one that foregrounds the relationship between purpose, methodology, and implementation. Such an approach entails, first, the explicit articulation of the intended use(s) of a resource and its primary audience(s). Second, it requires a critical evaluation of how different levels and types of annotation support—or potentially constrain—these uses. Finally, it calls for transparency in documentation so that users can understand the epistemological assumptions embedded in the data.

2 Resources

2.1 Croatian resources

Over the last decade, Croatian psycholinguistics has experienced substantial development through the creation of several large-scale lexical databases based on human judgments and behavioral measures. These resources are particularly important because they provide empirically grounded information about how speakers perceive, understand, emotionally evaluate, and process words. Such databases are essential for experimental psycholinguistics, neurolinguistics, cognitive linguistics, language acquisition research, affective linguistics, and natural language processing.

Unlike purely corpus-derived lexical resources, these databases rely on human participants that evaluate words on dimensions such as concreteness, imageability, age of acquisition, familiarity or relative frequency, emotional valence, and arousal. As a result, they capture subjective and cognitive aspects of lexical representation that cannot be derived from frequency statistics alone. By providing empirically validated insights into how people cognitively and emotionally experience and process words, these resources are therefore intended to support human-oriented applications in education, clinical assessment and rehabilitation, language development research, and related fields.

Many variables are already known, especially in education, therefore I concentrate on imageability, valence, and arousal because they are among the most frequently examined psycholinguistic variables and less well known outside psycholinguistics. Alongside other variables, they influence how language is processed, understood, and remembered. Imageability reflects a cognitive aspect of word meaning, whereas arousal and valence capture its affective dimension.

Imageability refers to the ease with which a word evokes a mental image. Concrete words such as *tree* or *apple* are typically rated as highly imageable because they readily elicit vivid sensory representations, whereas abstract words such as *justice* or *freedom* are more difficult to visualize. According to Paivio's Dual Coding Theory (Paivio 1986, 1991), highly imageable words are encoded through both verbal and visual systems, making them easier to process and recall. Consequently, high-imageability words are generally recognized faster and remembered more accurately than low-imageability words. Imageability should not be equated with concreteness because the ability of a word to evoke a mental image depends not only on its referential properties but also on its emotional and experiential salience, allowing abstract yet emotionally charged words to be highly imageable because abstract words with strong emotional content (e.g., *love* or *fear*) can evoke vivid mental images despite their low concreteness.

For this reason, it is important to include affective properties of words, commonly described in terms of emotional valence and arousal, into the picture. Valence refers to the degree to which a word conveys a positive or negative emotional meaning, ranging from pleasant (e.g., *love*) to unpleasant (e.g., *war*). Arousal, in contrast, reflects the intensity of the emotional response elicited by a word, regardless of whether that response is positive or negative. Thus, both *victory* and *murder* are highly arousing despite their opposite valence, whereas words such as *chair* or *window* are typically low in arousal. Research has shown that highly arousing words attract attention more readily and are often remembered better than emotionally neutral words.

Together, imageability, valence, and arousal, alongside concreteness, age of acquisition, and familiarity or relative frequency, provide complementary insights into lexical processing by accounting for both the cognitive and emotional characteristics of words. As a result, these variables are widely used and statistically controlled in psycholinguistic research to better understand language processing, memory, reading, and emotional responses to text.

The Croatian Psycholinguistic Database (Hrvatska psiholingvistička baza, HPB)[1] is the most comprehensive Croatian database of human-derived lexical variables. It contains at least thirty human ratings for six thousand Croatian words across multiple word classes. The inclusion of nouns, verbs, adjectives, and adverbs makes HPB particularly valuable in comparison to many psycholinguistic databases available internationally because they are often primarily limited to nouns. Such systematic methodological choice allows more comprehensive linguistic, cognitive, and experimental research across different word classes and language-processing functions. Furthermore, the database combines key psycholinguistic variables such as concreteness, imageability, age of acquisition, and subjective frequency with objective linguistic metadata such as corpus frequency, word class, animacy, and word length. Modeled partly on resources such as

1 The database, published in 2021, was developed within the project The Building Blocks of Croatian Mental Grammar: Constraints of Information Structure, led by Anita Peti-Stantić. The database is available in the Odraz repository, https://repozitorij.ffzg.unizg.hr/object/ffzg:833, and at the webpage https://megahr.ffzg.unizg.hr/en/?page_id=609.

the MRC Psycholinguistic Database, HPB was adapted to the morphological and semantic specificities of Croatian. It also combines human ratings with computational predictions for concreteness and imageability across the broader Croatian lexicon, bridging traditional psycholinguistics and computational linguistics.

The Croatian lexical decision study[2] extends the HPB resource into behavioral psycholinguistics by introducing reaction-time and accuracy data from lexical decision tasks, in which participants determine whether a letter string is a real Croatian word. The study includes behavioral data for 2,614 Croatian nouns derived from HPB and combines these measures with existing psycholinguistic variables such as concreteness, imageability, age of acquisition, and subjective frequency. By linking subjective lexical norms with experimentally measurable processing data, the study provides important infrastructure for Croatian psycholinguistic and neurolinguistic research.

Croatian affective psycholinguistics was also significantly expanded through two databases developed by Bojana Ćoso et al. The first, published in 2019, provided normative ratings for valence, arousal, and concreteness for 3,022 Croatian words, introducing experimentally validated measures of emotional positivity and intensity into Croatian psycholinguistic research. Building on this work, CROWD-5e,[3] published in 2023, was developed through international collaboration between Croatian and Spanish psycholinguistics researchers. This database added ratings for the same lexical base for five discrete emotions: happiness, anger, sadness, fear, and disgust. Together, the two resources allow comparisons between dimensional and categorical models of emotion, and they constitute important infrastructure for Croatian affective psycholinguistics, cognitive linguistics, and computational sentiment analysis.

The Croatian Lexical Database (Hrvatska leksička baza, HLB)[4] is an earlier but highly influential Croatian lexical norming initiative. The database primarily focuses on the dimensions of imageability, frequency, familiarity, age of acquisition, and concreteness. Although it is smaller in scale, with a published norming study subset of five hundred words and full database of 1,211 words by the date of publication, HLB was methodologically important because it established procedures for collecting Croatian psycholinguistic norms and highlighted the importance of subjective lexical variables for clinical linguistics, language impairment research, and aphasiology. The resource demonstrated that Croatian psycholinguistic norming was feasible despite the

2 Several sets of supplementary materials for this study are available online. Reaction-level data for Croatian nouns taken from the Croatian Psycholinguistic Database are available at https://doi.org/10.23669/PEVB54. The subset of nouns from the Croatian Psycholinguistic Database, enriched with aggregated lexical decision times, is available at https://doi.org/10.23669/EV3NFJ.

3 The database is available at https://repositori.urv.cat/estatic/PC0041/ca_PC4072.html.

4 The database, published in 2018, was developed within the framework of the projects Language Processing in Adult Speakers, led by Jelena Kuvač Kraljević, and A Multilevel Approach to Spoken Discourse in Language Development, led by Gordana Hržica. It is available at http://polin-hlb.erf.hr/. The database currently contains just over three thousand words, although the number of ratings per word varies considerably. The goal, which has not been reached yet, is for each word to be evaluated by at least thirty people.

considerable morphological complexity of the language. In many respects, HLB laid the methodological groundwork for later expansion and systematization in other projects.

As can be seen, Croatian psycholinguistic resources differ both in the psycholinguistic variables they include and in the types of lexical information they provide. Both CLD and the CPB primarily focus on traditional lexical-semantic variables, whereas HPB extends this approach to four grammatical classes—nouns, verbs, adjectives, and adverbs—and includes objective lexical variables such as word length, corpus frequency, and animacy. CLD provides general task instructions and a frequency example without anchor words for imageability or subjective frequency. In this database, highly concrete and imageable nouns such as *jabuka* 'apple' or *kuća* 'house' are contrasted with abstract nouns such as *pravda* 'justice' or *sloboda* 'freedom', which typically receive lower imageability ratings. Rather than relying on the canonical examples often used in literature (e.g., *apple*, *love*, and *war*), CPB illustrates its methodology with its own calibrator words such as *tuljan* 'seal', *plivati* 'to swim', *slan* 'salty', and *kuća* 'house' as highly concrete or imageable items, contrasted with words such as *pravda* 'justice', *morati* 'must', *nedosljedan* 'inconsistent', and *aspekt* 'aspect' as abstract or low-imageable items.

In contrast to CPD, which provides explicit calibrator words for imageability and concreteness, the norms reported by Ćoso et al. (2019) focus on affective dimensions. Positive words such as *zagrljaj* 'embrace' and *osmijeh* 'smile' illustrate high positive valence, whereas *izdaja* 'betrayal' and *očaj* 'despair' represent negative valence. Similarly, *potres* 'earthquake' and *eksplozija* 'explosion' exemplify high-arousal words, in contrast to emotionally neutral words such as *jastuk* 'pillow' and *polica* 'shelf'. By combining affective ratings with lexical decision reaction times, the database allows researchers to investigate how emotional properties of words influence lexical access and word recognition.

Together, these databases form the foundation of contemporary Croatian psycholinguistics resources. Their importance lies in the fact that they provide experimentally validated lexical norms for Croatian, a language that has historically been underrepresented in psycholinguistic research. They allow Croatian to participate in cross-linguistic comparisons with larger languages such as English and Spanish, support experimental control through careful stimulus selection, and connect linguistics, psychology, neuroscience, and computational modeling. An especially important feature of these resources is the integration of traditional human annotation with computational methods, as well as the strong collaboration between Croatian and international research groups.

2.2 Slovenian resources

For Slovenian, several resources and projects resemble Croatian databases to some extent. The overlap is especially visible because some researchers collaborated across Croatian and Slovenian projects, particularly through the Jožef Stefan Institute and regional psycholinguistic networks.

However, an important difference is that Slovenia does not currently have a single large unified database with systematically collected human-rated psycholinguistic norms. Slovenian affective resources are NLP-oriented sentiment resources that combine computational generation, translation-based transfer, and manual annotation. The best-known Slovenian sentiment resources are the SloSentiLex lexicons, SloEmoLex, and LiLaH (Emotion Lexicon of Croatian, Dutch and Slovene), although the closest Slovenian counterpart to the Croatian resources is probably the collaborative work by Ljubešić et al. (2018), which focused on the cross-linguistic prediction of concreteness and imageability using word embeddings.[5] Although the project was not exclusively Slovenian and did not produce a classic norming database based entirely on human ratings, it incorporated Slovenian lexical material and contributed to the development of semantic norm prediction for South Slavic languages.

The SloSentiLex lexicon[6] (parallel to CroSentiLex[7]) was developed within Slovenian NLP and sentiment-analysis projects. It is a polarity lexicon containing positive, negative, and neutral sentiment values for Slovenian words. This resource was largely created through semi-automatic computational procedures combined with linguistic validation and manual correction. In other words, it is partly human curated, but it is not human rated in the psycholinguistic sense. Its primary goal is sentiment analysis and opinion mining rather than modeling emotional representation in the mental lexicon. Typical examples include positively valenced words such as *veselje* 'joy' and *sreča* 'happiness', contrasted with negative words such as *sovraštvo* 'hatred' and *trpljenje* 'suffering'.

SloEmoLex 1.0 (Slovenian Emotion Dimension and Emotion Association Lexicon)[8] provides the closest equivalent to psycholinguistic norms because it contains nearly twenty thousand lexical entries annotated for valence, arousal, and dominance, alongside emotion associations, thus explicitly operationalizing affective dimensions. Methodologically, SloEmoLex 1.0 builds on prior multilingual emotion lexicons—most notably the NRC Emotion Lexicon—but it extends them through a combination of translation, manual validation, and computational modeling, aligning with recent trends in constructing large-scale affective lexicons for low-resource languages. The lexicon therefore reflects a hybrid approach in which cross-linguistic projection and machine learning are complemented by expert curation, allowing both scalability and linguistic plausibility. In addition to ratings of emotional dimensions, it assigns words to discrete emotion categories, making possible distinctions between words associated with emotions such as joy, fear, anger, sadness, surprise, disgust, trust, and anticipation. Thus, words such as *veselje* 'joy', *strah* 'fear', and *jeza* 'anger' exemplify different emotional categories rather than merely positive or negative sentiment.

5 The data are available at http://hdl.handle.net/11356/1187.
6 KSS version 1.1. is available at http://hdl.handle.net/11356/1097.
7 Available at http://hdl.handle.net/20.500.14615/2-24.
8 The lexicon was developed by Mojca Brglez et al., and it is available at the CLARIN.SI repository, http://hdl.handle.net/11356/1875.

Complementing this, the LiLaH Emotion Lexicon for Slovene[9] contains manual translations of the NRC Emotion Lexicon that encode the sentiment of a word (positive, negative) and its emotional association for Croatian, Dutch, and Slovenian. Manual translations were produced by inspecting and correcting the automatic translations from English provided with the original lexicon. Instead of rating emotional intensity, each lexical item is marked for the presence or absence of positive or negative sentiment and of the eight basic emotions proposed by Plutchik. For example, *sreča* 'happiness' is associated with joy, anticipation, and trust, whereas *begunec* 'refugee' is associated with sadness, and *duh* 'ghost' with fear. The lexicon is thus primarily intended for multilingual emotion detection and computational sentiment analysis.

3 Discussion and conclusion

The comparison between Croatian and Slovenian language resources presented in this article reveals more than differences in scale, organization, or technological orientation. It exposes two partially different methodological traditions within their language technologies and, more importantly, demonstrates how annotation practices are shaped by implicit assumptions about language, users, and the goals of research itself. Recent Croatian resource development relies significantly on psycholinguistic research and human evaluation, focusing on variables such as concreteness, imageability, emotional valence, and lexical processing. Slovenian resources, in contrast, have developed mainly within corpus linguistics and NLP infrastructures, emphasizing scalability, interoperability, and computational usability.

These differences should not be understood hierarchically, but as different answers to the question of what annotation is meant to achieve and who it is intended for. Each approach addresses important research and technological needs; the Croatian resources foreground human cognitive and affective dimensions of language, whereas the Slovenian resources prioritize automated processing and large-scale computational application. It is known that different annotation frameworks privilege different forms of linguistic knowledge: psycholinguistic norming preserves gradience and speaker variability, whereas computational annotation often simplifies language into categories optimized for machine processing.

The comparison and dialogue between Croatian and Slovenian language technologies could therefore be productive because it highlights the broader relationship between annotation, interpretation, and use. Ultimately, the comparison shows that smaller languages are not merely adapting methodologies developed elsewhere. Croatian and Slovenian resources instead offer valuable methodological insights into the relationship between human cognition, annotation practices, and language technology, supporting a more reflexive and human-oriented understanding of lin-

9 The lexicon was developed by Nikola Ljubešić et al. within the project The Linguistic Landscape of Hate Speech in Social Media, and it is available at the CLARIN.SI repository, http://hdl.handle.net/11356/1318.

guistic annotation, and both traditions demonstrate a strong commitment to transparent documentation and reflexive annotation practices, providing users with clear insight into the methodological and epistemological assumptions underlying the resources.

Future research should consider integrating additional variables, particularly those reflecting differences in Croatian and Slovenian linguistic, educational, and cultural resources, to provide a more nuanced comparative analysis.

This article was written as part of the institutional research project The Social Construction of Abstract Meanings Through Reading at the University of Zagreb, Faculty of Humanities and Social Sciences; co-funded by the European Union – NextGenerationEU.

References

Brglez, M., Caporusso, J., Hoogland, D., Koloski, B., Pollak, S., & Purver, M. (2024). *Slovenian emotion dimension and emotion association lexicon (SloEmoLex 1.0)*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1875

Ćoso, B., Guasch, M., Bogunović, I., Ferré, P., & Hinojosa, J. A. (2023). CROWD-5e: A Croatian psycholinguistic database of affective norms for five discrete emotions. *Behavior Research Methods*, *55*, 4018–4034. https://doi.org/10.3758/s13428-022-02003-2

Ćoso, B., Guasch, M., Ferré, P., & Hinojosa, J. A. (2019). Affective and concreteness norms for 3,022 Croatian words. *Quarterly Journal of Experimental Psychology*, *72*(9), 2302–2312. https://doi.org/10.1177/1747021819834226

Kuvač Kraljević, J., & Olujić, M. (2018). Imageability and subjective frequency of the 500 rated nouns in the Croatian Lexical Database. *Suvremena lingvistika*, *44*(85), 73–90. https://doi.org/10.22210/suvlin.2018.085.04

Ljubešić, N., Fišer, D., & Peti-Stantić, A. (2018). Predicting concreteness and imageability of words within and across languages via word embeddings. In I. Augenstein, K. Cao, H. He, F. Hill, S. Gella, J. Kiros, H. Mei, & D. Misra (Eds.), *Proceedings of the 3rd Workshop on Representation Learning for NLP* (pp. 217–222). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-3028

Ljubešić, N., Markov, I., Fišer, D., & Daelemans, W. (2020). The LiLaH emotion lexicon of Croatian, Dutch and Slovene. In M. Nissim, V. Patti, B. Plank, & E. Durmus (Eds.), *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media* (pp. 153–157). Association for Computational Linguistics. https://aclanthology.org/2020.peoples-1.15.pdf

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Peti-Stantić, A., Anđel, M., Gnjidić, V., Keresteš, G., Ljubešić, N., Masnikosa, I., Tonković, M., Tušek, J., Willer-Gold, J., & Stanojević, M.-M. (2021). The Croatian Psycholinguistic Database: Estimates for 6000 nouns, verbs, adjectives and adverbs. *Behavior Research Methods*, *53*, 1799–1816. https://doi.org/10.3758/s13428-020-01533-x

Vlašiček, D., Dumančić, F., & Tonković, M. (2025). Lexical decision times for nouns from the Croatian Psycholinguistic Database. *Behavior Research Methods*, *57*(6), 156. https://doi.org/10.3758/s13428-025-02676-5

Učiteljeva pisna povratna informacija in orodja umetne inteligence kot potencialni pripomoček za njeno oblikovanje

DOI: 10.4312/Obdobja.45.297-306

Ivana Petric Lasnik, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana;
ivana.petriclasnik@ff.uni-lj.si

V prispevku predstavljam, kako učitelji slovenščine kot drugega in tujega jezika podajajo pisno povratno informacijo svojim učencem o njihovi pisni produkciji, kateri so sestavni deli povratne informacije in ali učitelji v zvezi s povratno informacijo spodbujajo aktivno sodelovanje učencev ter na kakšen način. Glede na vse večjo vlogo generativne umetne inteligence pri poučevanju prikazujem tudi, kako bi si učitelji pri podajanju povratne informacije lahko pomagali z digitalnimi orodji, ki jih poganjajo veliki jezikovni modeli.

slovenščina kot drugi in tuji jezik, povratna informacija, popravljanje, generativna umetna inteligenca

This article examines how teachers of Slovenian as a second and foreign language provide written feedback on learners' written assignments, and the components of this feedback. It also examines how and whether teachers encourage active learner engagement in this process. Considering the increasing significance of generative artificial intelligence in language teaching, it also presents ways that teachers could use AI tools based on large language models to support the feedback process.

Slovenian as a second and foreign language, feedback, corrections, AI chatbots

1 Uvod

Povratna informacija (v nadaljevanju PI) je postopek, ki ga ima učitelj[1] na voljo za to, da učencem[2] pomaga napredovati v znanju. Pri razvijanju jezikovne zmožnosti učitelji v okviru PI najpogosteje posegajo po popravljanju besedil, kar kažejo različne raziskave, ki preučujejo podajanje PI na različnih ravneh v slovenskem šolskem prostoru (Godec Soršak in Rot Vrhovec 2025; Pori in Arhar Holdt 2023; Rozman in Stabej 2024). PI pa ni pomembna le za razvoj jezikovne zmožnosti v slovenščini kot prvem jeziku, temveč bi veljalo njeno vlogo raziskati tudi v okviru poučevanja slovenščine kot

1 Moško obliko *učitelj* uporabljam nevtralno za katerikoli spol.

2 Z uporabo besede *učenec* sledim slovenskemu prevodu *Skupnega evropskega jezikovnega okvira*, ki ta izraz uporablja za »vsako osebo, ki se uči jezikov, ne glede na starost in obliko (formalno ali neformalno, šolsko ali zunajšolsko) učenja« (Svet Evrope 2011: 17, op. 1). Moško obliko *učenec* uporabljam nevtralno za katerikoli spol.

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

drugega in tujega jezika[3] (v nadaljevanju SDTJ). Ta prispevek naj služi za odpiranje vprašanj in začetni razmislek o tej temi.

2 Teoretični okvir

PI lahko opredelimo kot kakršenkoli odziv, ki učencu daje informacijo o uspešnosti izvedbe neke naloge s ciljem, da mu pomaga napredovati. Prihaja lahko iz različnih virov in ima velik vpliv na dosežke posameznika (Hattie in Timperley 2007). PI je del formativnega ocenjevanja, pri katerem učenci in učitelji ves čas sodelujejo: učenci pridobivajo PI tako od učitelja kot iz drugih virov, se nanjo odzivajo, učitelji pa za to ustvarjajo spodbudno učno okolje (Šteh in Šarić 2020).

V središče prispevka postavljam pisno PI o pisni produkciji učencev pri učenju tujega jezika v organiziranem učnem procesu. V tem okviru se raziskuje predvsem popravljanje napak, ki jih učenci naredijo pri neustrezni rabi jezikovnih oblik, tj. korektivno PI (v nadaljevanju KPI). Raziskave so zadnjih 50 let preučevale predvsem, ali in kako lahko KPI izboljša pravilnost pisnega izražanja in kateri način je (naj)učinkovitejši ter zakaj (Bitchner 2021). Načinov podajanja pisne KPI je veliko, učitelj lahko npr. napake v besedilu popravi sam (neposredna KPI) ali jih le označi (posredna KPI), popravi lahko vse napake ali pa izbere le nekatere vrste napak. Lahko se odloči tudi, da bo zapisal pojasnilo svojega popravka ali pa učenca vodi pri tem, kako naj napako sam popravi (Bitchner 2021). V zadnjih letih pa se ugotavlja, da je težko meriti učinkovitost različnih pristopov h KPI, saj je posamezne spremenljivke težko izolirati (Mao idr. 2024). Raziskave učiteljem priporočajo, naj uporabljajo različne načine podajanja KPI glede na potrebe svojih učencev in kontekst poučevanja (Mao idr. 2024). Obenem ne gre pozabiti, da je popravljanje napak le del PI in da je za uravnotežen pogled na pisno zmožnost mogoče (in treba) nagovoriti tudi druge vidike besedila, npr. komentirati vsebino in zgradbo, zapisati pohvalo, učenca usmeriti v uporabo PI ipd. Zavedati se je treba tudi, da je učinkovitost PI odvisna predvsem od tega, kako se učenci nanjo odzivajo in ali so jo zmožni in pripravljeni tudi uporabiti (van der Kleij 2020), kar v raziskovanju (pa tudi pri pouku) pogosto ostaja neodgovorjeno in morda tudi nezastavljeno vprašanje.

Od učitelja se pričakuje, da v okviru formativnega ocenjevanja učencem pomaga razviti veščine za ustrezno ukvarjanje s PI – kako jo sprejemati, interpretirati in uporabiti (Mao idr. 2024), pa tudi kako pridobivati PI iz različnih virov. V zadnjih letih postajajo vse pomembnejši vir PI tudi različna orodja generativne umetne inteligence (v nadaljevanju UI), ki jih poganjajo veliki jezikovni modeli. Raziskave kažejo, da so učitelji do uporabe digitalnih orodij za namene PI precej skeptični in se zavzemajo, da bi jih uporabljali zgolj kot dopolnilo, ne pa nadomestilo učiteljeve PI (Godec Soršak in Rot Vrhovec 2025). Glede na to, da imajo učitelji za podajanje PI na voljo omejeno količino časa, pa bi bilo dobro preučiti, kako si lahko pomagajo z digitalnimi orodji, saj

3 Z besedno zvezo *slovenščina kot drugi in tuji jezik* v tem prispevku označujem slovenščino, ki se jo v različnih okoljih (v Sloveniji in v tujini) in za različne namene učijo osebe, za katere slovenščina ni prvi jezik.

raziskovalci ugotavljajo, da lahko učitelje časovno precej razbremenijo (Arhar Holdt in Munda 2025).

3 Metodologija

V raziskavi sem analizirala učiteljevo PI in PI, ki so jo generirala orodja UI. Z analizo sem želela pridobiti odgovore na raziskovalna vprašanja:

- Kako učitelji podajajo PI in kaj so njeni sestavni deli?
- Ali in kako učitelji spodbujajo učence za uporabo PI?
- Kako si lahko učitelji pri podajanju PI pomagajo z orodji UI?

Analiza je nastala na podlagi PI o besedilu,[4] ki ga je tvoril učenec SDTJ na univerzi zunaj Slovenije. Besedilo je nastalo med poukom, učenčeva okvirna raven znanja slovenščine je B1+. Besedilo sem skupaj z informacijami o nastanku besedila poslala 21 učiteljem, ki poučujejo SDTJ. Učitelji so prejeli navodilo: *Prosim, na spodnje besedilo dajte povratno informacijo, kot jo običajno dajete svojim učencem.* Podrobnejšega navodila, npr. glede tega, kako dati PI, kaj vse vključiti v PI ipd., učiteljem nisem dala, saj nisem želela sugerirati njihovih odgovorov. Odzvalo se je petnajst učiteljev:[5] šest jih poučuje na tečajih v organizaciji Centra za slovenščino kot drugi in tuji jezik, pet na univerzah v tujini, trije na Univerzi v Ljubljani v okviru modula Leto plus, en učitelj pa ima za poučevanje SDTJ lastno podjetje. En učitelj je poslal dva primera PI z utemeljitvijo, da se njegov način razlikuje glede na to, ali dela s celotno skupino ali pa izvaja individualne ure. Prejela sem torej 16 primerov učiteljeve PI. V sklepnem delu raziskave (3. avgust 2026) sem isto besedilo naložila tudi v tri različna orodja UI, ki delujejo kot klepetalni roboti, in sicer Copilot365 (GPT-5 model), ChatGPT (GPT-5.5 model) in Gemini Google (3.6 Flash model), ter jih pozvala,[6] naj podajo PI. Ta orodja sicer prvenstveno niso namenjena pedagoškemu delu, vendar glede na priljubljenost njihove uporabe ni zmotno pričakovati, da se uporabljajo tudi za podajanje PI. Kjer odgovori niso bili smiselni ali zadostni, sem poziv dopolnila, kot je prikazano v nadaljevanju prispevka. Vse prejete primere PI sem kvalitativno analizirala. Posamezne elemente PI sem razvrstila v tematske kategorije, kar mi je omogočilo oblikovanje prevladujočih vzorcev pri podajanju PI in primerjavo med posameznimi učitelji.

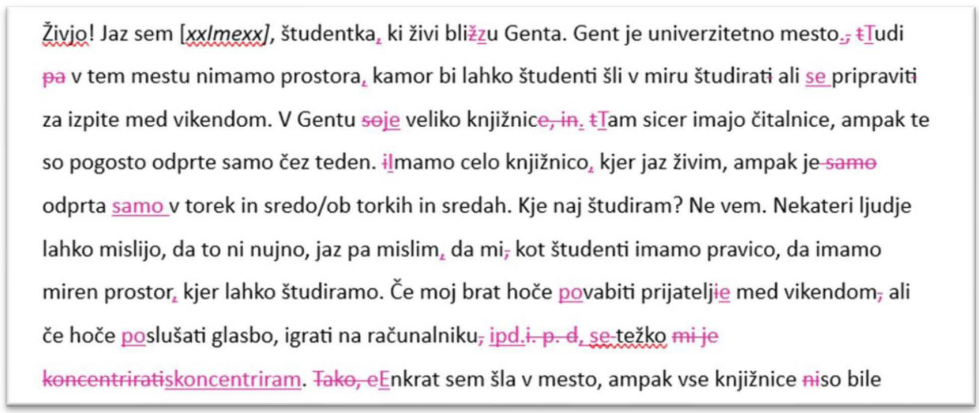
4 Predstavitev rezultatov

4.1 Način podajanja PI

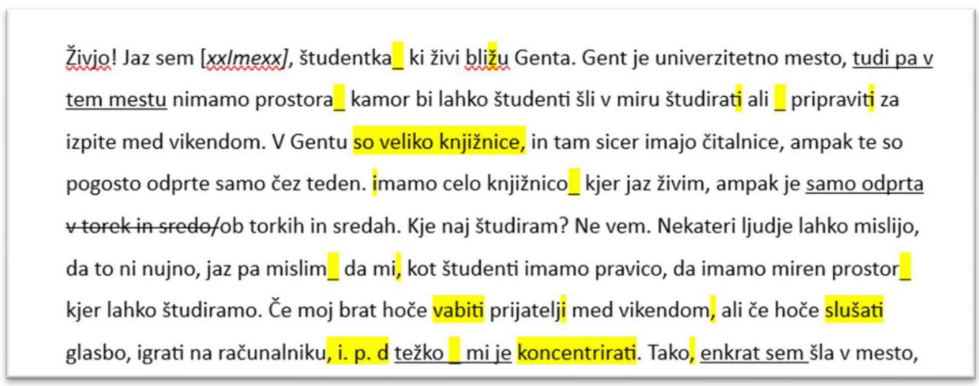
Vsi primeri PI so zapisani v slovenščini. Učitelji PI večinoma dajejo neposredno (12 primerov oz. 75 %), le trije primeri PI (18 %) so podani v posredni obliki. En učitelj (6 %) kombinira posredno (jezikovne oblike) in neposredno (zgradba besedila, kohezija in koherenca) podajanje PI. V dodatnih opombah učitelji zapišejo, da posredni način uporabljajo pri izvajanju individualnih ur, kar omogoča, da »skupaj na začetku ure pogledamo napake in jih morajo (z malo moje pomoči) učenci popraviti sami« (U4).

4 Glej prilogo.
5 Pri navajanju njihovih odgovorov uporabljam kode U1–U15.
6 Poziv je v prilogi.

Nobeden od učiteljev pa ni komentiral svoje odločitve za neposredno PI – domnevam, da zato, ker je neposredna PI primarna in intuitivna izbira učiteljev. Raziskave kažejo, da neposredni popravki učencem omogočajo hiter vpogled v pravilne jezikovne oblike in na ta način pripomorejo k razvijanju pravilnosti pisnega izražanja, kar se zdi smiselno predvsem na nižjih ravneh poučevanja ali pri napakah, ki jih učenci niso zmožni sami popravi (Bitchner 2021).


Živjo! Jaz sem [*xxlmexx*], študentka, ki živi bližzu Genta. Gent je univerzitetno mesto.~~,~~ ~~t~~Tudi ~~pa~~ v tem mestu nimamo prostora, kamor bi lahko študenti šli v miru študirati ali se pripraviti za izpite med vikendom. V Gentu ~~so~~je veliko knjižnice~~, in.~~ ~~t~~Tam sicer imajo čitalnice, ampak te so pogosto odprte samo čez teden. ~~i~~Imamo celo knjižnico, kjer jaz živim, ampak je ~~samo~~ odprta samo v torek in sredo/ob torkih in sredah. Kje naj študiram? Ne vem. Nekateri ljudje lahko mislijo, da to ni nujno, jaz pa mislim, da mi~~,~~ kot študenti imamo pravico, da imamo miren prostor, kjer lahko študiramo. Če moj brat hoče povabiti prijatelji~~e~~ med vikendom~~,~~ ali če hoče poslušati glasbo, igrati na računalniku~~,~~ ipd.~~i. p. d,~~ se težko ~~mi je koncentriratis~~koncentriram. ~~Tako, e~~Enkrat sem šla v mesto, ampak vse knjižnice ~~ni~~so bile

Slika 1: Primer neposredne PI (U11).


Živjo! Jaz sem [*xxlmexx*], študentka_ ki živi bližu Genta. Gent je univerzitetno mesto, tudi pa v tem mestu nimamo prostora_ kamor bi lahko študenti šli v miru študirati ali _ pripraviti za izpite med vikendom. V Gentu so veliko knjižnice, in tam sicer imajo čitalnice, ampak te so pogosto odprte samo čez teden. imamo celo knjižnico_ kjer jaz živim, ampak je samo odprta ~~v torek in sredo/~~ob torkih in sredah. Kje naj študiram? Ne vem. Nekateri ljudje lahko mislijo, da to ni nujno, jaz pa mislim_ da mi, kot študenti imamo pravico, da imamo miren prostor_ kjer lahko študiramo. Če moj brat hoče vabiti prijatelji med vikendom, ali če hoče slušati glasbo, igrati na računalniku, i. p. d težko _ mi je koncentrirati. Tako, enkrat sem šla v mesto,

Slika 2: Primer posredne PI (U6_1).

4.2 Sestavni deli PI

Analizirani primeri PI so različno obsežni, njihov glavni sestavni del pa je popravljanje napak v rabi jezika (slovnične napake, napake v rabi besedišča), ki se pojavi v vseh analiziranih primerih PI. Ta rezultat ni presenetljiv, saj podobno velja tudi za slovenščino kot prvi jezik (Godec Soršak in Rot Vrhovec 2024) in druge tuje jezike (Bitchner 2021). Polovica učiteljev v PI vključuje tudi druge sestavne dele, in sicer zapišejo sklepni komentar, pojasnijo (vsaj nekatere) popravke in komentirajo kohezijo in koherenco besedila (gl. Tabelo 1).

	Frek.	%	Primer
Popravljanje napak	16	100	napaka: *dostop do javni prostori* popravek: *dostop do javnih prostorov* (U3)
Sklepni komentar[7]	8	50	*Draga (ime), čestitam! Odlično si izrazila svoje mnenje in izkušnje in predlagala rešitvi. Pozor, pri glagolih premikanja je nedoločnik kratek (šli v miru študirat, delat*). (U14)
Pojasnilo popravka	7	43	napaka: *kamor bi lahko študenti šli v miru študirati* popravek: *kamor bi lahko študenti šli v miru študirat* pojasnilo: *glagoli premikanja + namenilnik* (U9)
Informacija o koheziji in koherenci	2	12,5	del besedila: *Nekateri ljudje lahko mislijo, da to ni nujno* pripomba: *Ni jasno, kaj mislite z besedo »TO«, treba je ponoviti besedo knjižnice, ker ste vmes pisali o tem, kje naj študirate. Predlagam: Nekateri ljudje lahko mislijo, da knjižnice niso nujne.* (U13)

Tabela 1: Sestavni deli PI[8] in pogostost vključevanja v učiteljevo PI.

Učitelji podajajo različno obširno KPI – število popravljenih napak se giblje od 24 (U7) do 45 (U2), pri čemer nisem upoštevala popravljanja vejic. Odločitev glede tega, katere napake bodo popravljali, učitelji sprejmejo glede na raven znanja učenca ali glede na to, katero snov so že predelali na tečaju in katere ne: »Predvsem ne morem čisto vedeti, kako naj popravljam, ker študentke ne poznam. Vedno namreč upoštevam tudi to, komu popravljam in kaj zmore ta oseba*«* (U3). Ta komentar kaže, da učitelji pri podajanju PI upoštevajo individualne razlike med učenci, kar kot priporočilo za učitelje najdemo tudi v literaturi (Bitchner 2021). Raziskave kažejo tudi, da KPI, ki se osredotoča le na nekatere vrste napak, postaja vse bolj prisotna pri poučevanju tujih jezikov, saj spodbuja večjo angažiranost učencev za ukvarjanje s PI in zmanjšuje obremenitev učiteljev (Mao idr. 2024).

Glede na to, da je KPI najobsežnejši in najpomembnejši del PI, me je zanimalo tudi, kako učitelji popravljajo napake. Najbolj so usklajeni pri popravljanju oblikoslovnih napak, napak v črkovanju[9] in popravljanju neustreznih besed. Npr. besedna zveza *vabiti prijatelji* je v vseh 16 primerih PI označena kot napaka in popravljena v *vabiti prijatelje*. Podobno velja za napake kot npr. *študentški restavraciji* (popravljeno v *študentski restavraciji*) in *slušati glasbo* (popravljeno v *poslušati glasbo*). Na podoben način tovrstne napake popravljajo tudi orodja UI. Za razliko od učiteljev pa orodja UI ponekod popravijo tudi strukture, ki niso napačne. Tako npr. vsa tri strukturo *pripraviti za izpite* popravijo v *pripravljati se na izpite*, pri čemer izrecno poudarijo popravek predloga. Nepotrebnih popravkov v primerih učiteljeve PI nisem opazila.

Manjšo usklajenost med učitelji je analiza pokazala pri popravljanju kolokacij in skladnje. Tovrstnih napak ne popravljajo vsi učitelji, med seboj pa se razlikujejo tudi po

7 Sklepni komentar učitelji zapišejo pod besedilo, v njem pa pohvalijo vsebino besedila (6 učiteljev), dodatno komentirajo KPI (6 učiteljev), učencem dajo navodila v zvezi z uporabo PI (4 učitelji) ali zapišejo splošno pohvalo (2 učitelja).

8 Kategorije sem oblikovala glede na vsebino PI, ki so jo zapisali učitelji.

9 Pri napakah črkovanja se je sicer izkazalo, da jih učitelji ne popravljajo vedno (predvidevam, da gre za lapsus, nepozornost), npr. *bližu* je popravljen 14-krat, *najlubši* 13-krat, *v študentški restavraciji* 16-krat, v *študentškem domu* pa 13-krat.

tem, kako napako popravijo (gl. Tabelo 2). Kolokacija *igrati na računalniku* je npr. kot napaka označena le v desetih primerih, popravljena pa je na štiri različne načine – najpogostejša je rešitev: *igrati igrice na računalniku*, učitelji pa to popravijo tudi v *igrati igre na računalniku*, *igrati na računalniku* oz. podajo posredno PI. Med orodji UI le Google Gemini to strukturo prepozna kot napako in ponudi rešitev: *igrati igrice na računalniku*. Tudi pri popravljanju struktur, ki so napačne na več ravneh, učitelji niso usklajeni. Npr. pri napaki *lahko uprašaš mesto* sedem učiteljev popravi le črkovanje (*lahko vprašaš mesto*), pet pa (tudi) kolokacijo (npr. *lahko prosiš mesto*, *lahko vprašamo župana*, *lahko predlagaš mestnim oblastem*). Za razliko od učiteljev vsa tri orodja UI popravijo napačno kolokacijo, in sicer ponudijo dve rešitvi: *lahko vprašaš občino* in *lahko prosiš mesto*. Zdi se, da bi si v okviru KPI učitelji z orodji UI lahko pomagali predvsem pri popravljanju neustreznih kolokacij, vendar bi za potrditev te trditve potrebovali posebno, natančnejšo in usmerjeno raziskavo.

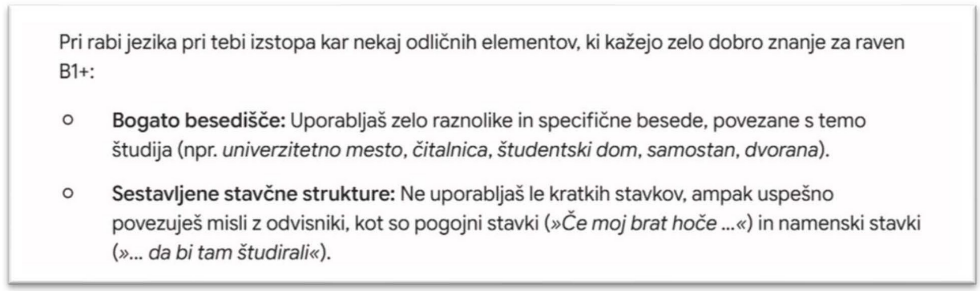
Primer	**Popravki**
težko mi je koncentrirati	1) *se težko skoncentriram* 2) *se jaz težko koncentriram/jaz težko študiram* 3) *se mi je težko koncentrirati* 4) posredna PI 5) *se mi je težko koncentrirati/se težko koncentriram* 6) *se mi je težko skoncentrirati* 7) *težko mi je se koncentrirati* 8) *se jaz takrat težko koncentriram* 9) *mi je težko skoncentrirati se* 10) *se težko koncentriram* *ChatGPT: *težko se koncentriram/težko se mi je zbrati* *Copilot: *težko se koncentriram/težko se zberem* *Gemini: *se težko koncentriram*
Sčasoma, v Gentu mi smo dobili	1) *Tako smo sčasoma v Gentu dobili …* 2) *Sčasoma smo mi študenti v Gentu dobili* 3) *Sčasoma smo mi v Gentu dobili* 4) posredna PI 5) *Sčasoma smo v Gentu dobili* 6) *Postopoma smo v Gentu dobili* 7) *Na lastnem primeru ti lahko povem, da smo mi v Gentu dobili …* 8) *Sčasoma, v Gentu smo dobili* *Gemini: *na srečo/na koncu smo dobili …*

Tabela 2: Nekaj primerov različnih popravkov napak.

Analiza pa je pokazala, da bi lahko učitelji orodja UI uporabljali kot pripomoček predvsem pri oblikovanju PI glede drugih vidikov besedila, npr. glede pohvale, vsebine, kohezije in koherence besedila.

Dva učitelja v sklepnem komentarju zapišeta splošno pohvalo, npr. »Kar tako naprej« (U10). Vsa tri orodja UI so konkretnejša in pohvalijo vključitev osebne izkušnje v besedilo ter primere, ki jih opiše učenec. Posebej pohvalijo primer z vetrom, ki je

odpihnil domačo nalogo (ki je morda nepričakovana (in manj verjetna) informacija). Na dodatni poziv[10] glede pohvale jezikovne rabe so orodja UI pohvalila predvsem raznoliko in široko besedišče ter skladnjo, pri čemer so navajala konkretne primere, kot kaže Slika 3.

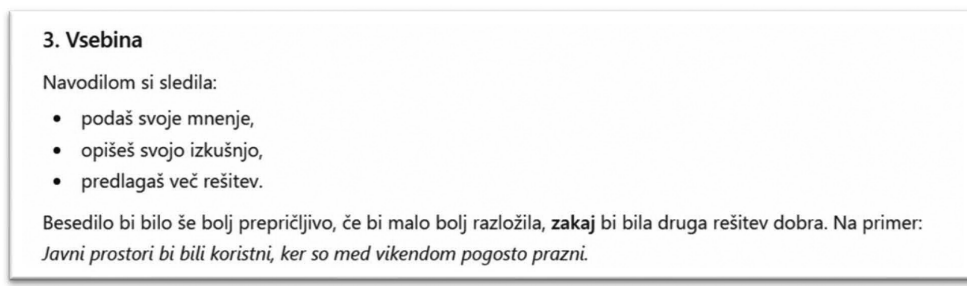

Pri rabi jezika pri tebi izstopa kar nekaj odličnih elementov, ki kažejo zelo dobro znanje za raven B1+:

- **Bogato besedišče:** Uporabljaš zelo raznolike in specifične besede, povezane s temo študija (npr. *univerzitetno mesto*, *čitalnica*, *študentski dom*, *samostan*, *dvorana*).
- **Sestavljene stavčne strukture:** Ne uporabljaš le kratkih stavkov, ampak uspešno povezuješ misli z odvisniki, kot so pogojni stavki (»*Če moj brat hoče ...*«) in namenski stavki (»*... da bi tam študirali*«).

Slika 3: Del PI v zvezi s pohvalo jezikovne rabe (Google Gemini).

Šest učiteljev v sklepnem komentarju pohvali vsebino: »Dobro ste sledili navodilom in vsebinsko ustrezno napisali besedilo. Besedilo je jasno in z jasnimi primeri odgovarja na problematiko (opisali ste izkušnje)« (U2). Vsebino pohvalijo tudi vsa tri orodja UI (gl. Sliko 4), podajo pa tudi možnosti za njeno izboljšavo, in sicer predlagajo, naj učenec:

- argumentira, zakaj so rešitve dobre oz. naj ideje za rešitve bolj razvije: »Besedilo bi bilo še bolj prepričljivo, če bi malo bolj razložila, *zakaj* bi bila druga rešitev dobra. Na primer: Javni prostori bi bili koristni, ker so med vikendom pogosto prazni« (ChatGPT).
- predlaga boljši zaključek, ki bi pozval druge bralce k razpravi: »Na koncu dodaš stavek: 'Kaj pa vi mislite o tem? Imate v vašem mestu podobne težave?'« (Google Gemini).


3. Vsebina

Navodilom si sledila:

- podaš svoje mnenje,
- opišeš svojo izkušnjo,
- predlagaš več rešitev.

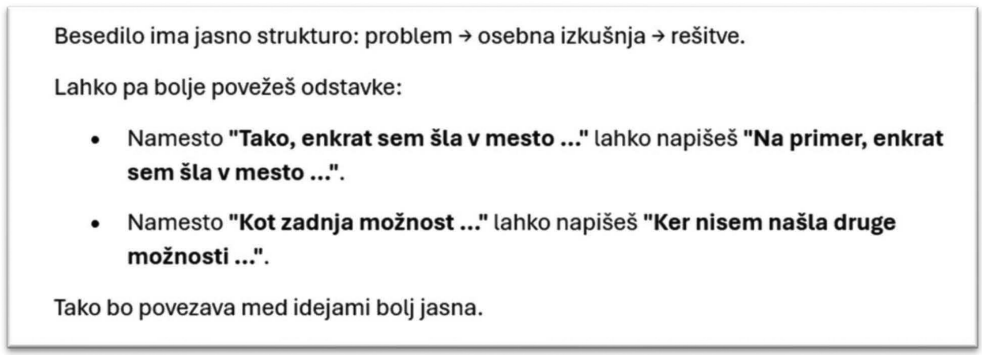
Besedilo bi bilo še bolj prepričljivo, če bi malo bolj razložila, **zakaj** bi bila druga rešitev dobra. Na primer: *Javni prostori bi bili koristni, ker so med vikendom pogosto prazni.*

Slika 4: Del PI v zvezi z vsebino besedila (Chat GPT).

Le dva učitelja učencu dasta PI glede kohezije in koherence (gl. Tabelo 1). Vsa tri orodja UI glede tega vidika komentirajo, da ima besedilo jasno in logično strukturo (gl. Sliko 5). Kot možnost za izboljšavo kohezije predlagajo bolj raznolike povezovalce

10 Kaj pa lahko pohvališ v zvezi z jezikovno rabo?

med stavki, in sicer namesto pogoste rabe besed *tako* in *ampak* predlagajo tudi druge povezovalce: *zato*, *poleg tega*, *na koncu*, *zaradi tega*, *vendar*, *na primer*, *ker* …


Besedilo ima jasno strukturo: problem → osebna izkušnja → rešitve.

Lahko pa bolje povežeš odstavke:

- Namesto **"Tako, enkrat sem šla v mesto ..."** lahko napišeš **"Na primer, enkrat sem šla v mesto ..."**.
- Namesto **"Kot zadnja možnost ..."** lahko napišeš **"Ker nisem našla druge možnosti ..."**.

Tako bo povezava med idejami bolj jasna.

Slika 5: Del PI v zvezi s kohezijo in koherenco (Copilot).

4.3 Spodbujanje učencev k uporabi PI

Kot učitelji si vsekakor želimo, da bi učenci našo PI tudi uporabili, analiza PI v pričujočem prispevku pa kaže, da je ta del formativnega ocenjevanja pogosto prepuščen učencem. Pri tem je treba upoštevati tudi čas, ki ga imajo učitelji na voljo za podajanje PI, ter motiviranost učencev za dodatno ukvarjanje z besedilom. Sedem učiteljev spodbuja učence k uporabi PI, in sicer:

- v sklepnem komentarju pod besedilom trije učitelji učencu naročijo, naj besedilo napiše še enkrat, eden pa, naj si učenec zgolj ogleda popravo besedila;
- trije učitelji učenca k uporabi PI spodbujajo ustno: dva v okviru individualnih ur skupaj z učenci popravljata označene napake pri posredni PI, eden pa pogoste napake predstavi celemu razredu ter učence spodbuja k vprašanjem, če jim popravki niso jasni.

Seveda je ta rezultat odvisen tudi od tega, kako so učitelji razumeli navodilo za sodelovanje v raziskavi, in bi ga bilo treba preveriti z nadaljnjim raziskovanjem, npr. z opazovanjem dela v razredu.

Tudi vsa tri orodja UI, tako kot učitelji, učencu predlagajo ponovno pisanje besedila: »Napiši novo, izboljšano različico tega forumskega prispevka. Vključi vse jezikovne popravke in na koncu dodaj vprašanje za druge uporabnike foruma« (Copilot). Poziv sem glede tega vidika PI dopolnila[11] in tako dobila širok nabor idej, ki učence spodbujajo k delu s PI in refleksiji. Učitelj pa jih lahko uporabi glede na čas, ki ga ima na voljo, zmožnost učencev, njihovo raven znanja ipd. Orodja UI predlagajo:

- delo z že napisanim besedilom in prejeto PI:
  - samoevalvacija: učenci sami poiščejo dobre dele besedila in dele besedila, ki bi jih radi izboljšali,
  - učenci vodijo svoj dnevnik napak, v obliki kartic, stripa …,
  - učenci izdelajo semafor napak: kaj znam, kaj moram vaditi, česa ne razumem,

11 »Prosim, predlagaj še kakšno bolj ustvarjalno idejo za uporabo PI.«

- učenci pripravijo podcast z izboljšano različico besedila;
- ponovno pisanje besedila na isto temo:
  - učenci napišejo besedilo z druge perspektive (npr. brat, župan mesta …),
  - učenci nadgradijo že obstoječe besedilo z novimi argumenti, primeri;
- uporabo PI v drugih zapisanih besedilih:
  - učenci uzavestijo strukture, ki jim delajo težave tako, da jih poiščejo v zapisanih besedilih, s katerimi se srečujejo.

5 Zaključek

Sklenem lahko, da se večina učiteljev SDTJ pri PI o tem konkretnem besedilu osredotoča predvsem na formalne vidike jezikovne rabe – to se vidi v tem, da v svojo PI vključujejo predvsem KPI, in to na neposreden način. Popravljajo različne tipe napak in nekatere popravke tudi pojasnijo. Na popravke po njihovih besedah vpliva tudi raven znanja učenca in vsebine, ki jih obravnavajo pri pouku. Pri popravljanju morda ravno zato med seboj niso vedno usklajeni, vendar je njihova KPI jasno označena in ne delajo nepotrebnih popravkov. Učitelji uporabo PI pri učencih spodbujajo v omejenem obsegu, bi jim pa pri tem vidiku PI lahko bila v pomoč orodja UI. Ravno tako lahko pri digitalnih orodjih iščejo tudi ideje za oblikovanje pohvale in PI o vsebini in zgradbi besedila. Na ta način bi lahko učitelji orodja UI uporabili kot dodatne vire za dopolnilo svoje PI.

Moj vzorec je premajhen, da bi lahko rezultate analize posploševala. Omejeni so na izbrano besedilo, način izvedbe raziskave, sodelujoče učitelje, izbrana orodja UI in čas, v katerem je bila izvedena raziskava. Menim pa, da rezultati nakazujejo trend pri podajanju PI pri poučevanju SDTJ ter naslednje korake – to je uzavestiti uporabo različnih virov PI (tudi digitalnih orodij), podajanje PI o različnih vidikih besedila in pri učencih razviti veščine za učinkovito uporabo PI.

Prispevek je nastal v okviru doktorske raziskave z naslovom Vloga povratne informacije pri pouku slovenščine kot tujega jezika pod mentorstvom doc. dr. Ine Ferbežar.

Literatura

Arhar Holdt, Š. in Munda, T. (2025). Jezikovno popravljanje v digitalnem okolju: Kvalitativna študija z učiteljicami in učitelji slovenščine. *Sodobna pedagogika*, *76*(3), 86–106. https://doi.org/10.63384/sptB53s791s

Bitchner, J. (2021). Written corrective feedback. V *The Cambridge handbook of corrective feedback in second language learning and teaching* (str. 207–225). Cambridge University Press.

Godec Soršak, L. in Rot Vrhovec, A. (2025). Popravljanje pisnih besedil učencev in dijakov: Vpogled v rezultate ankete učiteljev. *Sodobna pedagogika*, *76*(3), 59–85. https://doi.org/10.63384/sptB53z792s

Hattie, J. in Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Mao, Z., Lee, I. in Li, S. (2024). Written corrective feedback in second language writing: A synthesis of naturalistic classroom studies. *Language Teaching*, *57*(4), 449–477. https://doi.org/10.1017/S0261444823000393

Pori, E. in Arhar Holdt, Š. (ur.). (2023). *Popravljanje jezika in besedil: Učiteljska povratna informacija v šolski praksi: zbornik konference*. Založba Univerze v Ljubljani. https://doi.org/10.4312/9789612972394

Rozman, T. in Stabej, M. (2024). Univerzitetno pisanje in popravljanje besedil: Prakse in stališča. V S. Štumberger (ur.), *Predpis in norma v jeziku* (Obdobja 43, str. 285–292). Založba Univerze v Ljubljani. https://ebooks.uni-lj.si/ZalozbaUL/catalog/book/664/chapter/3932

Svet Evrope. (2011). *Skupni evropski jezikovni okvir: učenje, poučevanje, ocenjevanje* (Kovačič, I., prev.). Ministrstvo RS za šolstvo in šport, Urad za razvoj šolstva. https://centerslo.si/wp-content/uploads/2015/10/SEJO-komplet-za-splet.pdf

Šteh, B. in Šarić, M. (2020). Uveljavljanje formativnega ocenjevanja v visokošolskem izobraževanju. *Psihološka obzorja*, *29*(1), 79–86. https://doi.org/10.20419/2020.29.515

van der Kleij, F. M. (2020). Evaluation of the 'Feedback Engagement Enhancement Tool' to examine and enhance students' engagement with feedback on their writing. *Studies in Educational Evaluation*, *66*, članek 100907. https://doi.org/10.1016/j.stueduc.2020.100907

Prilogi

Besedilo za PI

Živjo! Jaz sem [*xxImexx*], študentka ki živi bližu [*xxImeMestaxx*]. [*xxImeMestaxx*] je univerzitetno mesto, tudi pa v tem mestu nimamo prostora kamor bi lahko študenti šli v miru študirati ali pripraviti za izpite med vikendom. V [*xxImeMestaxx*] so veliko knjižnice, in tam sicer imajo čitalnice, ampak te so pogosto odprte samo čez teden. imamo celo knjižnico kjer jaz živim, ampak je samo odprta v torek in sredo/ob torkih in sredah. Kje naj študiram? Ne vem. Nekateri ljudje lahko mislijo, da to ni nujno, jaz pa mislim da mi, kot študenti imamo pravico, da imamo miren prostor kjer lahko študiramo. Če moj brat hoče vabiti prijatelji med vikendom, ali če hoče slušati glasbo, igrati na računalniku, i. p. d težko mi je koncentrirati. Tako, enkrat sem šla v mesto, ampak vse knjižnice niso bile odprte. Moj najlubši prostor da bi študirala (dvorana v najvišjem nadstropju študentski restavraciji [*xxImexx*]) tudi bil zaprt, ker med vikendom restavracija ne dela. Kot zadnja možnost sem šla v park, ampak tam veter je odpihnil mojo domačo nalogo. Tako potrebujemo ravni rešitvi. Prvi rešitev: študirati skupaj v študentškem domu, drugi rešitev (to pa odvisno od mesta) lahko uprašaš mesto za več prostora. Sčasoma, v [*xxImeMestaxx*] mi smo dobili dostop do javni prostori (npr. samostani, cerkvi) da bi tam študirali.

Poziv za orodja UI

Prosim za povratno informacijo o besedilu, ki ga je napisala študentka slovenščine kot tujega jezika. Slovenščino se uči dve leti izven Slovenije, študira slavistiko, njen prvi jezik je nizozemščina. Njena okvirna raven znanja slovenščine je B1+.

V povratni informaciji:

- popravi napake v jezikovni rabi (slovnica in besedišče); popravke prosim tudi pojasni; vejic ne popravljaj; popravke vedno zapiši v konkretnem kontekstu, ne kot izolirane besede.
- komentiraj kohezijo in koherenco besedila ter predlagaj izboljšave; navedi konkretne primere.
- komentiraj vsebino besedila (ustreznost glede na navodila) ter predlagaj izboljšave; navedi konkretne primere.
- zapiši konkretno pohvalo.
- predlagaj vsaj dve aktivnosti za uporabo povratne informacije.

Povratna informacija naj bo kratka in jasna. Napisana naj bo v 2. osebi ednine v enostavnem jeziku (na približno enaki ravni, kot je študentkino znanje slovenščine).

Med digitalnim *feedom* in *skrolanjem*: jezik in diskurz hibridnih uprizoritev

DOI: 10.4312/Obdobja.45.307-316

Eva Pori, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; eva.pori@ff.uni-lj.si

Tomaž Toporišič, Akademija za gledališče, radio, film in televizijo, Univerza v Ljubljani, Ljubljana; tomaz.toporisic@agrft.uni-lj.si

Prispevek obravnava hibridne prostore uprizoritvenih praks v razponu od performativnega obrata šestdesetih in sedemdesetih let 20. stoletja do sodobnih, digitalno pogojenih in intermedialnih dogodkov, ki prepletajo ples, video, digitalno tehnologijo in druge medije onstran verbalnega diskurza. Izhodišče je opažanje, da se v tem kontinuumu eksperimentalnih praks postopoma preoblikujejo razmerja med telesom, jezikom, prostorom, mediji in gledalcem ter da uprizoritvene umetnosti vse izraziteje delujejo kot odprt, procesualen in medmedijski dogodek.

jezik, performativna uprizoritev, hibridni prostori, intermedialnost, digitalna tehnologija

This article examines hybrid spaces of performative practices, ranging from the performative turn of the 1960s and 1970s to contemporary digitally conditioned and intermedial events that intertwine dance, video, digital technology, and other media beyond verbal discourse. It is grounded in the observation that, within this continuum of experimental practices, the relationships among the body, language, space, media, and the spectator are undergoing gradual transformation, and the performing arts are increasingly operating as an open, processual, and intermedial event.

language, performative staging, hybrid spaces, intermediality, digital technology

1 Uvod

Sodobne uprizoritvene prakse vse izraziteje delujejo v prostoru prepletanja različnih medijev, tehnologij in diskurzov, zaradi česar se spreminjajo tako načini umetniškega izražanja kot tudi percepcija gledalca in recepcija uprizoritvenega dogodka. Vdor digitalnih tehnologij in intermedialnih postopkov v umetniški diskurz ne vpliva zgolj na formalne in estetske ravni uprizarjanja, temveč posega tudi v samo strukturo jezika, govora in performativnosti.

V prispevku izhajamo iz teze, da se v sodobni slovenski dramatiki in uprizoritvenih praksah oblikuje specifičen jezikovno-diskurzivni ustroj, zaznamovan z razpadom klasičnega dramskega dialoga, fragmentacijo govora in prehajanjem med različnimi diskurzivnimi ravninami, vključno z zabrisovanjem meje med realnim in virtualnim. Jezik in govor v teh praksah ne delujeta več kot nosilca stabilnega dramskega subjekta, temveč se vzpostavljata kot nelinearen in fragmentaren diskurz, primerljiv z logiko *digitalnega feeda* (toka digitalnih vsebin na platformah družbenih medijev, npr. Facebooku,

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

Instagramu in TikToku), v katerem se posamezni segmenti nizajo brez jasnega začetka ali zaključka. Govorne instance se translokacijsko premikajo med individualnimi, kolektivnimi in anonimnimi pozicijami, recepcija besedila pa vse bolj poteka po logiki »skrolanja«, ki razkraja hierarhijo med glavnim in stranskim besedilom ter ustvarja rapsodičen učinek neprekinjenega diskurzivnega toka. Takšne transformacije ne zadevajo zgolj jezika, temveč širše preoblikovanje uprizoritvenega dogodka, ki se danes vse izraziteje vzpostavlja kot odprt, procesualen in intermedijski prostor-čas.

V tem okviru bomo obravnavali hibridne prostore uprizoritvenih praks, v katerih se preoblikujejo razmerja med telesom, jezikom, mediji, prostorom in gledalcem. Posebna pozornost bo namenjena jeziku kot ključnemu nosilcu teh transformacij, pri čemer bo analiza umeščena v širšo zgodovinsko perspektivo, ki povezuje zgodnje slovenske eksperimentalne prakse druge polovice 20. stoletja (*Pupilija papa Pupilo pa Pupilčki* (1969); *Šarada ali Darja* (1975)) s sodobno dramatiko (Simona Semenič: *24ur* (2006), *5fantkov.si* (2008)) ter intermedialnimi dogodki (Janez Janša: *Slovensko narodno gledališče* (2007)). Na izbranih uprizoritvenih primerih bomo prikazali, kako sodobni slovenski gledališki jezik ne odraža zgolj estetskih in žanrskih premikov, temveč tudi globlje transformacije jezikovnih in kognitivnih vzorcev (prim. Dixon 2007; McConachie in Hart 2006), povezane z digitalno komunikacijo, interaktivnostjo ter prepletom virtualnega in realnega.

2 Od dramskega k performativnemu: prelomnice šestdesetih in sedemdesetih let

V šestdesetih in sedemdesetih letih 20. stoletja je bil z razgradnjo tradicionalnega dramskega modela, desemantizacijo jezika ter postopnim ukinjanjem logocentrično organiziranega odra vzpostavljen premik k eksperimentalnim in performativno zasnovanim uprizoritvenim praksam (Pori 2023; Toporišič 2012). S performativnimi rabami tekstovnega materiala se je razprl gledališki prostor, v katerem so se začele oblikovati dinamične in multivalentne uprizoritvene strukture, zaznamovane z nestabilnostjo pomena, drsenjem označevalcev ter fragmentacijo govornih in telesnih kodov. Linearna dramska reprezentacija in stabilnost dramskega subjekta sta bili postopoma nadomeščeni s procesualnimi, odprtimi in heterogenimi oblikami uprizarjanja, v katerih se pomen ne vzpostavlja več skozi koherentno pomensko celoto, temveč skozi razmerja, prelome in kontraste med glasovnim, telesnim, vizualnim in prostorskim (Pori 2023). Prav v teh praznih prostorih izmenjav in navidezno nepovezanih prehodov se je oblikovala specifična postdramska in performativna dramaturgija, ki je bistveno zaznamovala tudi poznejše intermedialne in digitalno pogojene uprizoritvene prakse.[1]

1 O tem podrobneje v poglavjih Performativnost scenskih in vizualnih umetnosti ter Performativne revolucije igralčevega in gledalčevega telesa knjige *Medmedijsko in medkulturno nomadstvo, o vezljivosti medijev in kultur v sodobnih uprizoritvenih praksah* (Toporišič 2018).

2.1 Primer: *Pupilija papa Pupilo pa Pupilčki*

Uprizoritev *Pupilija papa Pupilo pa Pupilčki* (1969) predstavlja enega ključnih prelomov znotraj slovenskega gledališkega prostora, saj je z razgradnjo tradicionalne dramske forme in z odprto, kolažno strukturo vzpostavila nov model uprizoritvenosti. Predstava ni več temeljila na enotnem dramskem tekstu in linearno organizirani zgodbi, temveč je nastajala iz heterogenega gradiva – fragmentov besedil, citatov, glasbenih vložkov, telesnih in govornih akcij, vizualnih podob in improvizacijskih elementov, ki so bili povezani po načelu montaže in asociativnega prehajanja.[2] Fragmentarna struktura je razrahljala hierarhijo med posameznimi uprizoritvenimi elementi ter odprla prostor simultanemu delovanju različnih znakovnih sistemov.

S tem je bil tradicionalni logocentrični model gledališča, utemeljen na primatu dramskega besedila in reprezentacije, nadomeščen s performativno organizacijo dogodka, v kateri pomen ni več nastajal iz sklenjene narativne celote, temveč iz odnosov, napetosti in prekinitev med telesom, glasom, zvokom, podobami in prostorom. Uprizoritev je tako vzpostavila hibriden uprizoritveni prostor, v katerem so se začeli prepletati različni mediji in performativne strategije, gledalec pa je bil postavljen v proces aktivnega povezovanja razpršenih pomenov.

V dramaturški strukturi uprizoritve, sestavljene iz 20 tem oz. prizorov, se posebej izrazito pojavljajo tudi prizori, ki vključujejo različne medijske reference in simulacije medijskih mehanizmov. Že uvodni prizor *TV-dnevnik*[3] temelji na projekciji in gledanju televizijskih vsebin, s čimer je v uprizoritveni prostor vpeljan televizijski medij kot eden ključnih organizatorjev sodobne percepcije. V prizoru *Računalnik* je s pomočjo gibalne simulacije delovanja računalnika ter prepleta zvokov stroja, sprejemnikov in glasu vzpostavljena mehanizirana, tehnološko posredovana govorica, ki telo in glas umešča v logiko aparata. Medijsko in potrošniško ikonografijo nadalje vzpostavlja prizor *Reklame za revijo Elle*, ki temelji na manipulaciji s slovensko izdajo francoske modne revije, medtem ko prizor *Reklama za alpsko mleko* z večglasnim reklamnim vzklikanjem in parodično uprizoritvijo oglaševalske retorike razgrajuje mehanizme propagandnega govora in prepričevalnih strategij oglaševalskega diskurza.

Takšni prizori kažejo, da se uprizoritev ne vzpostavlja več zgolj kot dramska reprezentacija, temveč kot odprta medijska konfiguracija, v kateri se gledališki dogodek oblikuje skozi preplet telesa, zvoka, projekcije, reklame, tehnološke simulacije in množičnomedijskih podob. Mediji pri tem niso uporabljeni zgolj kot ilustrativni dodatek, temveč postajajo konstitutivni del same dramaturške organizacije in produkcije pomena.

2.2 Primer: *Šarada ali Darja*

Po eksperimentalnih premikih, vzpostavljenih v *Pupilija papa Pupilo pa Pupilčki*, predstavlja uprizoritev *Šarada ali Darja* (1975) nadaljnji korak v smeri medializacije

2 O tem podrobneje v poglavju Gledališče Pupilije Ferkeverk v doktorski disertaciji *Dinamika razmerij med glasom in telesom v slovenskih uprizoritvenih umetnostih druge polovice 20. stoletja* (Pori 2023).

3 Prizori so v scenariju (Gledališče Pupilije Ferkeverk 1969) označeni zgolj z zaporednimi številkami, v prispevku jih zaradi preglednosti poimenujemo glede na prevladujočo tematiko.

gledališkega prostora. Če je bila v *Pupiliji* medijska logika prisotna predvsem skozi simulacije televizijskih, oglaševalskih in popularnokulturnih kodov, je v *Šarada ali Darja* televizijski medij neposredno vključen v samo strukturo uprizoritvenega dogodka, video oz. televizijska podoba pa začne prevzemati funkcijo samostojnega uprizoritvenega elementa, ki razkraja stabilnost odrske percepcije ter vzpostavlja nova razmerja med živim telesom, tehnično reproducirano podobo in gledalčevim pogledom. S tem je prišlo do enega prvih poskusov vnosa video in televizijskega medija v slovensko gledališče, pri katerem gledališki prostor ni več organiziran kot enoten reprezentacijski okvir, temveč kot razpršeno in večmedijsko polje simultanih podob, glasov in telesnih prisotnosti (Pori 2023: 117).

Na odru je bilo po scenografski zasnovi Mete Hočevar postavljenih 15 televizorjev. Na ekranih, ki so bili razpostavljeni ob stranskih stenah gledališča, so se že takoj ob vstopu v prostor znašle tudi podobe udeležencev uprizoritve, kamera pa je nato v video zgodbo zajela igralkin »intimni« odrski prostor, katerega posnetek je bil sočasno v živo predvajan na televizijskem ekranu. Hiršenfelder (2011: 47) omenja tridelno strukturo video zgodbe oz. postavitev televizijskih sprejemnikov v tri sklope: prvi del je predstavljalo neposredno predvajanje snemanja gledalcev v zaprtem prostoru (*close-circuit*), drugega neposredno predvajanje snemanja v zaprtem krogu, izrazito usmerjenega v detajliranje igralkinega telesnega izraza, in tretji del predvajanje vnaprej pripravljenih posnetkov na polinčnem odprtem kolutu (*open-real*).[4] Vendar pa naj preklapljanje med živo sliko in posnetkom, sistem hkratnega snemanja in projekcije, ki naj bi razbil linearnost prostora in časa, po mnenju kritike (prim. Berger 1975; Inkret 1975 in Knap 1975; Šuklje 1975; Taufer 1976) ne bi doseglo posebnih učinkov. Kljub možnostim ustvarjanja različnih perspektiv, ki jih ponuja vstop elektronske tehnologije v gledališki prostor, se je eksperimentiranje s kamero pokazalo kot neučinkovit »vdor« in »nepotrebna motnja« (Hiršenfelder 2011: 48), ki jo je uprizoritev fragmentarnega in večplastnega besedila vzpostavila v begajoči zavesti gledalca. Je pa uprizoritev, ne glede na to, da z inovativnim pristopom ni dosegla preboja, vsekakor preizpraševala nova komunikacijska razmerja v gledališču v času razvoja elektronskih tehnologij, učinke prepletanja televizijskega in gledališkega medija in možnosti uporabe videa za prikaz in oblikovanje novih perspektiv igralskega govornega in telesnega izraza (Pori 2023: 118). Posebej jezik oz. govor ni več deloval kot nosilec enotnega dramskega pomena ali psihološko zaokroženega subjekta, temveč je postajal fragmentiran, ritmiziran in vse bolj podrejen logiki medijskega prenosa, montaže ter reprodukcije. Oblikoval se je kot zvočni material oz. kot eden izmed elementov širše medijske kompozicije, ki je vstopala v razmerja s telesom, sliko, projekcijo in tehnološko posredovanimi zvoki.[5]

4 O tem tudi na zvočnem posnetku Barbara Borčić, Ida Hiršenfelder, v Nuša Dragan, 21. april 2009, Zavod SCCA – Projektna soba.

5 S to problematiko se na zanimiv način ukvarja francoski teoretik Bruno Tackels, npr. v poglavju knjige *Odrske pisave* z naslovom Desakralizacija teksta, vendarle, ki je v slovenskem prevodu dostopen v knjigi *Drama, tekst, pisava II* (2021).

3 Od performativnega obrata k intermedialnim uprizoritvenim praksam

V zadnjih desetletjih se je v sodobnih uprizoritvenih praksah uprizoritveni prostor vse izraziteje oblikoval kot hibridno in medijsko razpršeno polje, v katerem se jezik ne vzpostavlja več kot stabilna dramska kategorija, temveč deluje kot fragmentaren in večmedijsko posredovan diskurz, ki vstopa v kompleksna razmerja z videom, projekcijo, digitalnimi podobami in tehnološko reproduciranimi glasovi.

3.1. Primer: *24ur* in *5fantkov.si*

V uprizoritvah Simone Semenič, zlasti v *24ur* (2006) in *5fantkov.si* (2008), se dramaturško-jezikovna struktura približuje logiki digitalnih medijev in pospešenega pretoka informacij. Besedilo je organizirano fragmentarno, brez linearnega razvoja in stabilne hierarhije med posameznimi diskurzivnimi segmenti, pri čemer se različni govorni registri, citati, asociacije, intimni zapisi, medijski klišeji in vsakdanji govor nizajo po principu simultanosti in neprekinjenega preklapljanja. Takšna organizacija spominja na logiko *digitalnega feeda* oz. »skrolanja«, v katerem posamezni fragmenti vstopajo v nenehna razmerja prekinitev, ponovitev in preskokov (Odlomek 1).

ON1: Pejva.
ONA1: *(Ga pogleda.)*

SPAM14: Your First FREE Gift! A 20“ 18-karat Gold plated Rope Chain. Exquisitely crafted in 18-karat Gold, this sparkling, 18k Gold plated 20“ long rope chain is the perfect match for your exquisite medallion. It features a large, easy to use „lobster claw“ clasp for complete security. But that is not all...

ON1: *(Jo pogleda.)*
ONA1: *(Pogleda v tla.)*

ONA2: In kje se je zalomilo?

SPAM14: Your Second Free Gift! A Black Velvet Presentation Box. This handsome, velvet, presentation box is satin lined and features a precisely angled, felt covered „display stand“ that has been custom fitted to your Medallion. This will allow you to dramatically show off your finely crafted medallion, when you are not wearing it. In addition, it has a spring-loaded hinge that instantly snaps shut with a distinctive muffled click, so you can keep this masterpiece safe and totally secure. But there‘s more...

ON2: Ne vem. Kje se je zalomilo?
ONA2: Ne vem.
ON2: Vem samo, da sem ponucal vse patrone.
ONA2: Ja.
ON2: In kaj zdaj?

SPAM14: Your Third Free Gift! The Secrets Behind and Uses of Your Medallion. You‘ll learn that this special medallion was first consecrated by ancient priests and used for thousands of years to attract wealth and good luck to specially chosen individuals. You‘ll also discover how its powers are generated by the tree of life at its center and protected and enhanced by the magical hieroglyphs that surround it. Most importantly, you‘ll understand how it can help focus, and direct the magic, miracles and protection it will make available to you. Lastly, you‘ll learn how you can empower your talisman with a secret prayer.

ONA2: Nimam pojma.

SPAM13: Get ready for spring with new ring-tones from 50 CENT, THE KILLERS, GWEN STEFANI, KELLY CLARKSON, KANYE WEST, and hundreds more. Visit us and find your favorite tones today!

Tišina.

ON1: Prosim te, pejva.
ONA1: Saj bova šla, samo da se še malo odpočijem.
ON1: Saj veš, da ...

SPAM15: AGAO Volume Jumps 500% In Trading Today! Recent news releases are generating growing interest in AGAO!

ONA1: 24ur.
ON1: Ja, ampak ...

Odlomek 1: Semenič: *24ur*, 2006, str. 44 (vir: SiGledal).

Spremenjena podoba dramske forme se neposredno kaže tudi na ravni jezika in govora. Govor ne deluje več kot sredstvo psihološke karakterizacije dramskih oseb, temveč kot tok diskurzivnih impulzov, ki se hitro menjavajo, prekrivajo in razgrajujejo (Odlomek 1). Besedilo prehaja med različnimi jezikovnimi zvrstmi in registri – od knjižnega do pogovornega jezika (*pejva*, *pejd*, *dolgcajt*, *ponucal*, *(mi je) ratalo*) in jezika spamov – angleščine (Odlomek 1, za več gl. Semenič 2006). V knjižnopogovorni jezik se pogosto vrivajo prvine ekspresivnega, grobega in vulgarnega izrazja (*crkniti*, *kurba*, *prasica*, *prasec*), ki prispevajo k stopnjevanju konfliktnosti in ustvarjanju avtentičnega govornega učinka. Govor postaja pretočen, pogosto prekinjen, nasičen s kratkimi, nedokončanimi, pogosto vprašalnimi povedmi, ponavljanji, preskoki in nenadnimi menjavami registrov, kar ustvarja učinek pospešenega *digitalnega feeda*, značilnega za sodobno medijsko kulturo.

V *24ur* je takšna dinamika povezana predvsem z ritmom neprekinjenega medijskega toka in z razpršenostjo sodobne percepcije, ki jo določajo televizijski, spletni in informacijski formati. Besedilo pri tem učinkuje kot »tkanje citatov« in nekakšen »medmedijski labirint«, značilen za čas digitalne besedilnosti in spremenjene senzibilnosti (Toporišič 2018: 50). Posebno funkcijo pri tem prevzema jezik spamov – angleščina, ki v slovenski jezikovni prostor vnaša logiko globalizirane spletne komunikacije. Ti govorni segmenti prekinjajo kontinuiteto dialoga in ne delujejo več kot nosilci stabilnega pomena, temveč kot reproducibilni in depersonalizirani jezikovni impulzi, organizirani po logiki *digitalnega feeda*.

Podobno se tudi v *5fantkov.si* vzpostavlja razdrobljena jezikovno-govorna struktura, v kateri je subjekt razpršen med različne govorne pozicije, medijske reference in kolektivne glasove. Že sam naslov z domeno *.si* nakazuje vpis digitalne logike v strukturo besedila, ki učinkuje kot rizomatična in intermedialna mreža različnih medijskih realnosti ter govornih tokov (Toporišič 2018: 50–51). Dramski lik prevzema funkcijo usmerjevalca oz. komentatorja dogajanja, jezik pa postaja izrazito neknjižen, pogovoren in slengovsko obarvan, pogosto zapisan fonetično (*vid*, *kok*, *blazn*, *mislm*, *zdej*, *poču*, *reku, kej*, *pol*, *čist*) ter prežet z nižjepogovornim izrazjem, vulgarizmi (*pičkice*, *pizda*, *kurac*, *prasica*, *spizdi*, *fak*) in prvinami balkanskega jezikovnega prostora oz. medjezikovnimi prenosi (*koje ste vi pičkice*, *koji ti je kurac*, *jebo vam pas mater*), ki odražajo

vpliv večjezičnega urbanega okolja in vsakdanje govorne prakse mladih (Odlomek 2, za več gl. Semenič 2008). Skladnja je erodirana in izrazito fragmentarna, sestavljena iz kratkih, odsekanih ter mestoma nepovezanih stavkov, pogosto reduciranih na posamezne besedne zveze, besede ali glasove. Značilna je odsotnost ločil in velikih začetnic, kar dodatno poudarja razpad linearne jezikovne organizacije.

denis
vid
kok s pa ti blazn pameten
mislim
kok pa ti kar veš
ka bo kdo reku

vid
stari
zdej pa še ti
poču
res
poču bom
kr
kr
kr eksplodiru bom

krištof
čist brezveze

vid
faaaaaaaaaaaaaaaaaaaak
faaaaaaaaaaaaaaaaaaaak

denis
ej to pa res sovražim
mislim
če pa kej sovražim
pol pa to sovražim

blaž
pa koje ste vi pičkice
koje pičkice
prave pičkice
ono
da ne rečem
pičkice

Odlomek 2: Semenič: *5fantkov.si*, 2008, str. 42 (vir: SiGledal).

Izsek iz *5fantkov.si* (Odlomek 2) jasno pokaže radikalni razkroj tradicionalne dramske jezikovno-govorne strukture. Govor se razkraja v ritmiziran tok besednih fragmentov, vzklikov, ponavljanj in afektivnih impulzov, pri čemer posamezne replike učinkujejo kot hitri diskurzivni preskoki. Ponavljanja (*kr / kr / kr eksplodiru bom*, *poču bom*, *faaaaaaaaaaaaaaaaaaak*) ustvarjajo učinek govornega *loopa* – povratne zanke, značilne za digitalno komunikacijo in impulzivno spletno govorico. Jezik in govor se pri tem približujeta logiki neposrednega digitalnega vnosa, v kateri prevladujejo hitrost,

simultanost, prekinitve in nenehno preklapljanje med različnimi govornimi registri in diskurzi.

Vizualna organizacija besedila dodatno poudarja fragmentarnost govorne strukture: razporeditev govornih segmentov po strani, prazni prostori med replikami in razpršenost posameznih glasov. Linearno organiziran dialog nadomešča odprta mreža glasovnih impulzov, v kateri pomen nastaja skozi ritmično ponavljanje, preskoke in nenehno »skrolanje« med posameznimi fragmenti.

3.2 Interaktivni in intermedialni obrat: *Slovensko narodno gledališče* Hrvatina in Janše

Kot zadnji primer poglejmo od blizu predstavo Janeza Janše *Slovensko narodno gledališče* (2008), zanimiv primer vezljivosti med živo umetnostjo, mediatiziranim performansom in spletnimi oblikami reprezentacije. Predstava združuje klasično gledališko formo antičnega zbora, ki komentira dogajanje, z logiko sodobnih medijev, digitalne reprodukcije in spektakla. Prav v tem prepletu nastaja specifična politika uprizarjanja: gledalec ni več pasivni opazovalec, temveč postane soudeleženec pri rekonstruiranju pomenov in zgodovinskih dogodkov skozi medijsko posredovane podobe.

Hibridnost predstave se kaže predvsem v razmerju med živim telesom performerja in njegovo medijsko podvojitvijo. Živa prezenca ni več razumljena kot nekaj avtentičnega in nedotakljivega, temveč kot del širše mreže podob, arhivov in digitalnih reprezentacij. V tem smislu predstava sledi razmišljanjem Philipa Auslanderja (2007) o razvrednotenju žive prisotnosti v mediatizirani kulturi. Živi dogodek je vedno že prežet z logiko medijev, reprodukcije in spektakla, zato meja med neposredno izkušnjo in njeno medijsko konstrukcijo postaja nejasna.

Pomembno vlogo pri tem igra tudi jezik. V sodobnih mediatiziranih uprizoritvah jezik ni več zgolj sredstvo komunikacije ali nosilec zgodbe, temveč postane performativni medij. Govor, citati, medijski diskurzi in politične parole delujejo kot fragmenti informacijskega toka, ki ga občinstvo neprestano interpretira in rekonstruira. Jezik se v novih medijih razkraja na kratke izjave, podobe in digitalne znake, kar vpliva tudi na gledališče. Uprizoritev zato ne temelji več na linearni naraciji, ampak na montaži različnih diskurzov, podob in medijskih kodov.

Predstava se vrača k ritualnim elementom gledališke kulture, vendar jih sopostavlja v kontekst sodobne družbe spektakla. Gledalci se soočijo z realnimi zgodovinskimi dogodki in političnimi travmami, a se obenem zavedajo, da gre za posredovano, medijsko konstruirano izkušnjo. Ta dvojnost ustvarja občutek liminalnosti – vmesnega prostora med realnim in virtualnim, med prisotnostjo in reprezentacijo.

4 Zaključek

Hibridna (post)dramska besedila, gledališki dogodki in uprizoritvene prakse, ki smo se jim posvetili, vzpostavljajo začasne skupnosti med izvajalci in gledalci. Erika Fischer-Lichte (2008) ta proces razume kot avtopoetično povratno zanko, v kateri se pomen ustvarja skozi neposredno medsebojno delovanje izvajalcev in občinstva. V tem

procesu se brišejo tradicionalne opozicije med subjektom in objektom, umetnostjo in družbeno realnostjo ter med fizično in digitalno prisotnostjo. Gledalec ni več zunaj dogodka, ampak postane del njegove produkcije. Se emancipira v smislu Jacquesa Rancièrja (2020) in njegovega pojma »emancipirani gledalec«. Njegova vloga se pri tem ne spreminja le na ravni participacije, temveč tudi na ravni zaznavanja in organizacije pozornosti.

Uprizoritveni dogodek se tako ne vzpostavlja več kot sklenjena dramska celota, temveč kot odprta in hibridna konfiguracija različnih medijskih, glasovnih in telesnih tokov. V razpadu sintaktične in govorne kontinuitete je mogoče prepoznati eno ključnih značilnosti sodobne dramatike pod vplivom digitalne medijske kulture: jezik ni več nosilec stabilnega pomena, temveč postaja fragmentaren, pretočen in performativen zvočno-govorni material. *Digitalni feed* in *govorni loop* lahko v tem kontekstu razumemo kot dramaturška mehanizma, ki ne transformirata zgolj oblike gledališkega jezika, temveč reorganizirata tudi kognitivne pogoje njegove recepcije. Linearno sledenje govoru se umika procesom selekcije, anticipacije in prepoznavanja vzorcev ter povezovanju ponavljajočih se in fragmentiranih informacijskih tokov.

Sodobna gledališka umetnost tako razkriva mehanizme kulturnega nadzora in medijske manipulacije. Z uporabo novih medijev, digitalnih podob in performativnega jezika (prim. Fischer-Lichte 2008; Lehmann 2003) opozarja na nasilje reprezentacije v družbi spektakla, hkrati pa poskuša ustvariti prostor skupne telesne in čutne izkušnje. Prav v tej napetosti med mediatizacijo in živim dogodkom se oblikuje sodobno gledališče, ki obstaja hkrati kot fizični in virtualni prostor ter kot preplet jezika, telesa in medijskih tehnologij.

Viri

Berger, A. (1975, 13. december). Šarada. *Dnevnik*.

Gledališče Pupilije Ferkeverk. (1969). *Pupilija papa Pupilo pa Pupilčki* [Tipkopis]. Arhiv SLOGI.

Inkret, A. (1975, 21.–28. december). Šarada. *ITD*, *31*(10).

Knap, M. (1975, 12. december). Glejev krst. *SGTCD*, *TVD2*.

Pupilija papa Pupilo pa Pupilčki [DVD-posnetek gledališke uprizoritve]. (1969). Arhiv CTF UL AGRFT.

Semenič, S. (2006). *24ur*. SiGledal. https://teksti.sigledal.org/tekst/24ur

Semenič, S. (2008). *5fantkov.si*. SiGledal. https://teksti.sigledal.org/tekst/5fantkov-si

Šuklje, R. (1975, 16. december). Igralkin lik. *Delo*.

Taufer, V. (1976, 16. januar). Eksperimentalno gledališče Glej, Rudi Šeligo: Šarada ali Darja. *Naši razgledi*.

Literatura

Auslander, P. (2007). *V živo: Uprizarjanje v mediatizirani kulturi*. Mestno gledališče ljubljansko.

Dixon, S. (2007). *Digital performance: A history of new media in theater, dance, performance art, and installation*. The MIT Press.

Fischer-Lichte, E. (2008). *Estetika performativnega*. Študentska založba.

Hiršenfelder, I. (2011). Video in televizija (in cel hudič šaradasti). V A. Perne (ur.), *Glej, 40 let*. Gledališče Glej; Slovenski gledališki muzej.

Lehmann, H. T. (2003). *Postdramsko gledališče*. Maska.

McConachie, B. in Hart, F. E. (ur.). (2006). *Performance and cognition: Theatre studies and the cognitive turn*. Routledge.
Pogorevc, P. in Toporišič, T. (ur.). (2021). *Drama, tekst, pisava 2* (Knjižnica MGL 178). Mestno gledališče ljubljansko.
Pori, E. (2023). *Dinamika razmerij med glasom in telesom v slovenskih uprizoritvenih umetnostih druge polovice 20. stoletja* [Doktorska disertacija]. Univerza v Ljubljani.
Rancière, J. (2020). *Emancipirani gledalec*. Maska.
Toporišič, T. (2012). Gledališče kot nedokončana arhitektura – hibridni prostori performativnih praks. V B. Orel, M. Šorli in G. Troha (ur.), *Hibridni prostori umetnosti* (str. 38–51). Maska.
Toporišič, T. (2018). *Medmedijsko in medkulturno nomadstvo: O vezljivosti medijev in kultur v sodobnih uprizoritvenih praksah*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani; Akademija za gledališče, radio, film in televizijo Univerze v Ljubljani.

Kdo to piše? Avtorsko profiliranje umetnointeligenčnih pomočnikov

DOI: 10.4312/Obdobja.45.317-326

Ina Poteko, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; ina.poteko@ff.uni-lj.si

V prispevku smo preizkusili, kako uspešna sta umetnointeligenčna pomočnika Claude in ChatGPT pri avtorskem profiliranju. Zanimalo nas je, kako zanesljivo lahko umetna inteligenca tvorkam in tvorcem besedilnih sporočil določi starost in spol, kako svoje odločitve utemelji in kako opiše jezikovno rabo. ChatGPT se je izkazal kot precej uspešen tako pri ugotavljanju starosti in spola kot pri zaznavanju narečnih prvin, pri nalogah pa se je odrezal bolje kot Claude.

ChatGPT, Claude, avtorsko profiliranje, slovenščina, sporočila

This article tests how successful the AI assistants Claude and ChatGPT are at author profiling. Of interest was how reliably AI can determine the age and sex of the authors of text messages, how it justifies its decisions, and how it describes language use. ChatGPT proved to be relatively successful both in identifying age and sex and in detecting dialect features, performing better than Claude across the tasks.

ChatGPT, Claude, author profiling, Slovenian, text messages

1 Umetnointeligenčni pomočniki

Pri raziskovanju ter pisanju strokovnih in znanstvenih besedil so lahko umetnointeligenčni pomočniki (v nadaljevanju UI pomočniki) zelo uporabno orodje, vendar pa se je treba zavedati tudi pasti, predvsem (ne)zanesljivosti in (ne)etične uporabe. UI pomočniki nam niso v pomoč le pri snovanju in oblikovanju raziskav (pri čemer že to lahko vpliva tudi na izgubo »človeške« izvirnosti, kritičnega mišljenja in analitičnih veščin), temveč nam lahko na podlagi podatkov ustvarijo kar končni izdelek v obliki statistične analize, vizualnega prikaza, povzetka, prevoda itn. ChatGPT naj bi se dobro odrezal predvsem pri kvalitativnih analizah, manj smiselna pa se zdi njegova uporaba za kvantitativne analize, predvsem v primerjavi z že obstoječimi orodji, ki so zanesljivejša (Johannesson 2024: 5–6, 69).

2 Avtorsko profiliranje

Zanima me, kako uspešna je UI pri t. i. avtorskem profiliranju (angl. *author profiling*) besedil v slovenščini. Namen avtorskega profiliranja je na podlagi sloga pisanja identificirati demografske značilnosti avtorja ali avtorice. Avtorsko profiliranje je v pomoč pri analizi besedil, za katera nimamo demografskih podatkov, kar je lahko uporabno za razumevanje demografskih vzorcev, družbenih trendov itn., koristno pa je

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

tudi za področja, kot sta forenzika in marketing. Pri tem se pojavljajo tudi zadržki glede zasebnosti tvorcev in tvork besedil (Chen idr. 2024: 150).

Na družbenih omrežjih uporabniki in uporabnice za sabo puščajo digitalni odtis v obliki besedilnih podatkov in več raziskav je pokazalo, da lahko na podlagi zapisov na družbenih omrežjih precej zanesljivo sklepamo o spolu in starosti avtorja ali avtorice zapisa. Takšna primera sta prispevek o samorazkrivanju osebnih informacij uporabnic in uporabnikov na Facebooku in Twitterju (Jaidka idr. 2018) ter prispevek o prepoznavanju spola na podlagi tvitov (Burger idr. 2011). Mnogo besedilnih podatkov nastaja tudi v aplikacijah za dopisovanje, vendar so ti podatki raziskovalcem in raziskovalkam načeloma nedostopni, zato so le redko vključeni v jezikovne raziskave (Koch idr. 2022: 1).

V raziskavah, ki raziskujejo korelacijo med slogom pisanja oz. govorjenja ter spolom in starostjo, pa pogosto umanjka razumevanje kompleksnosti povezave med demografskimi značilnostmi in jezikovno rabo, posamezne spremenljivke, predvsem spol, so namreč obravnavane v izolaciji od drugih in preveč poenostavljeno (Bamman idr. 2014; Chen idr. 2024: 150). Najbolj problematičen je prav spol, saj je v modele za avtorsko profiliranje vključen kot binarna spremenljivka, kar lahko vodi do preveč poenostavljene razlage spola in utrjevanja stereotipov, potencialno ugotovljenih razlik pa ne moremo pripisati le spolu, saj je jezikovna raba odvisna od širšega konteksta (Chen idr. 2024: 152).

Testiranje zmožnosti avtorskega profiliranja pri UI pomočnikih se mi je tako zdelo zanimivo predvsem zato, ker je v ospredju določanje starosti in spola tvork in tvorcev besedil. V družbi namreč obstajajo različna prepričanja o tem, kaj je značilno za jezikovno rabo žensk in moških ter različne generacije, vprašanje pa je, kako uspešno lahko iz besedil dejansko sklepamo o starosti in spolu.

3 Metodologija

Preizkusila sem dva UI pomočnika: Claude (Opus 4.8, način *high effort*, naročnina Pro) in ChatGPT (5.5, način *thinking*, naročnina Plus). Zanimalo me je, kako zanesljivo lahko UI tvorkam in tvorcem sporočil določi starost in spol, kako svoje odločitve utemelji, kako opiše jezikovno rabo ter kako se v primerjavi z UI odreže človek.

Besedilni podatki, ki sem jih uporabila, so dopisovanja med različnimi osebami preko sporočil. Vključila sem devet pogovorov, dolgih po sto sporočil, vsak pogovor pa sem uredila v svoji Excel datoteki. Vsa sporočila so bila predhodno anonimizirana, vsi podatki, ki bi izdajali identiteto oseb, pa so bili nadomeščeni ali izbrisani. Osebe, udeležene v pogovorih, so v prispevku poimenovane s fiktivnimi imeni.

Za ponazoritev sem v prispevek vključila po osem sporočil iz vsakega pogovora, spol udeležencev je razviden iz imena, starost je pripisana, pogovorov pa nisem posebej opisovala. Komunikacija z UI pomočniki je v prispevku zapisana v ležečem tisku. Ker so bili odgovori UI zelo obsežni, so vključeni samo deli pogovorov. V odgovorih se UI sklicuje tudi na primere iz sporočil, ki jih v navedenem odlomku pogovora v prispevku ni.

4.1 Ugotavljanje spola in starosti v treh pogovorih

V prvih treh pogovorih sem predhodno izbrisala vsa sporočila, ki bi izdajala spol na podlagi slovničnih značilnosti. Poziv (angl. *prompt*), ki sem ga vnesla v UI pomočnika, je bil:

> Pišem prispevek o jezikovnih izbirah v zasebnem pisnem e-sporazumevanju. Za raziskavo sem zbrala sporočila različnih oseb. Vsak pogovor je zapisan v svoji Excel datoteki. Poslala ti bom različna sporočila. Ali lahko osebam na podlagi sporočil določiš spol in starost?

Ajda, 29 let	**Barbara, 29 let**
pa lah mi delis tm	
	to se mi pa vec neda
Sad	
i would	
	loh ti use fotke poslem
	ups se ne vid
haha hvala ampak neznam prebrat kaj je to 😂	
	seenkrat se mi ne da slikat ampak vids prblizn vajb kreme hahaha

Tabela 1: Odlomek iz pogovora 1.

Claude na začetku odgovarjanja opozori, da *lahko spol in starost oceni le z določeno verjetnostjo, ne pa zanesljivo*, in da bi za slovenščino *najmočnejši pokazatelj spola bile oblike preteklika in pridevnikov v 1. osebi tipa bila/bil sem vesela/vesel*, vendar tega v pogovoru ni. ChatGPT prav tako *ne more zanesljivo določiti dejanskega spola in starosti, lahko pa naredi jezikovno utemeljeno oceno*.

Claude za oceno spola zapiše, da *sloni predvsem na registru, besedišču in temah, kar je šibkejši in stereotipno obremenjen kriterij*, s samooceno *nizke do srednje zanesljivosti* pa Ajdi in Barbari pravilno pripiše ženski spol. *S srednjo do visoko zanesljivostjo* Ajdi določi starostni razpon od 25 do 32 let. Kot *kazalnike* navede *izrazito ljubkovalni naslavljalni register (honey), veliko afiliativnih in čustvenih elementov (wiiii), teme, ki nakazujejo mlado odraslo osebo (testira recepte), močno prepletanje anglizmov (right on time) in špansko-igrive vrivke (que pasa)*. Barbari s *srednjo zanesljivostjo* pripiše starostni razpon od 23 do 30 let, *kazalniki pa so* poleg vsebinskih tudi *enak mladi spletni register: fonetično zapisovanje (kdr cs), anglizmi (amazing), pisanje brez velikih začetnic in ločil in emoji roza srček*.

ChatGPT Ajdi in Barbari *z nizko gotovostjo* pripiše ženski spol, *s srednjo gotovostjo* pa starostno skupino od 20 do 30 let. Za Ajdo to utemelji z rabo *veliko mladostniškega/slengovskega izražanja (vibes), mešanja slovenščine in angleščine* ter različnimi temami v pogovoru, pri Barbari pa opazi *podobno izrazito neformalno pisanje (se mi ne da), vsebino in slog, ki kažejo na mlajšo odraslo osebo.* Kot *ključno* zapiše, da *spola iz teh sporočil ne bi določila kot raziskovalno zanesljivega podatka.*

Cene, 28 let	**Danijela, 28 let**
	u kaj pa bi
Nevem ksno sirovo strucko alpakej tacga	
Alpa ksno hrenovko v testu	
Alpa ksn burek	
Hahaha	
	hahahaha okok
	sam v sušilca dam pa grem iskat
U izi	

Tabela 2: Odlomek iz pogovora 2.

Claude je za starost ugotovil, da *pri obeh konvergira na mlado odraslo dobo, cca. 20–28*. To utemelji izključno z nanašanjem na vsebino sporočil. Za spol zapiše, da *gradivo spola ne razkriva, kazalniki* pa so *prešibki, da bi karkoli trdila*.

ChatGPT obema *s srednjo gotovostjo* napove starost od 20 do 30 let, za Ceneta zapiše, da *spol ni zanesljivo določljiv*, Danijeli pa pripiše *mogoče ženski spol, a zelo nezanesljivo.* Na starost kažeta *predvsem neformalen, mladostniški digitalni slog (miv mav, helu, najs)* in vsebina pogovora. Spola med drugim ne more zanesljivo določiti, ker je *večina povedi kratkih, eliptičnih in nevtralnih*. Za Danijelo napiše, da *bi se morda lahko pojavila šibka asociacija na žensko zaradi izrazov, kot so kjut, hihi, jokam, mjav, in zaradi omembe druge ženske osebe in prstana, vendar je to zelo stereotipno in nezadostno, tudi moški ali binarne osebe lahko pišejo popolnoma enako*.

Ernest, 24 let	**Feliks, 23 let**
	Ej kak pa te tota sauna deluvle?
	Ka si nekaj eti gucin s kolegico ka bi mogoce precekejrali
Moske zensko je loceno	
	Kelko je pa kaj placa
	Kelko je lejko lidij notri
Pa kak se naguzvas pa ki sedis vise nize	
4 komot	
Ide 6	

Tabela 3: Odlomek iz pogovora 3.

Claude za obe osebi pravilno predvideva starostno skupino od 20 do 27 let, kar utemeljuje predvsem z vsebino pogovora, omeni pa tudi *obilico anglizmov in emojije*. Za spol poda le *šibko domnevo*, pri Ernestu se *nagiba k moškemu* zaradi vsebine (npr. nogometa), doda, da je *register malo »fantovski«*. Za Feliksa je *bolj negotov, droben nagib k ženski* na podlagi *edinega možnega samonanašalnega signala »kenslala«, a je lahko tudi 3. oseba,*[1] *zelo nizka zanesljivost – lahko bi šlo tudi za prijatelja*.

ChatGPT je Ernestu *s srednjo gotovostjo* pripisal moški spol in starost od 20 do 30 let, Feliksu enako starost, *z nizko do srednjo gotovostjo* pa prav tako moški spol. O spolu

1 Sporočilo »kenslala« se je res nanašalo na prej omenjeno žensko osebo.

je sklepal na podlagi vsebine sporočil, v starostni razpon pa bi ju umestil *zaradi izrazov in tem, kot so fuzbal, gym, žurka, after …, ter zaradi zelo neformalnega digitalnega sloga*.

UI pomočnika sta vsem šestim osebam pravilno pripisala starostni razpon, v katerega se uvrščajo, pri utemeljevanju pa sta se bolj nanašala na vsebino sporočil. Kot *kazalnike mladosti* sta navajala tudi jezikovne vidike, predvsem zapis po izgovoru, anglizme, slengovsko izražanje, emojije in zapisovanje brez velikih začetnic in ločil. Udeleženci v pogovorih so bližje zgornjim kot spodnjim starostnim mejam. Na podlagi pogovora z UI pomočnikoma sklepam, da je nižja starostna meja odraz vsebine, predvsem omemb žuranja in pitja, kar sta povezovala s študentsko populacijo, jezikovna raba pa je bolj podkrepila njune domneve kot bistveno vplivala nanje.

ChatGPT se je odrezal bolje kot Claude, saj je spol uspešno določil petim udeležencem, težave je imel le s pogovorom 2, ki je največjo težavo predstavljal tudi Claudu. V pogovoru so bile teme nakupovanja, kuhanja, kajenja, pitja in igranja igric, vendar so bile razporejene med obe osebi, zato spola ni bilo mogoče določiti na podlagi stereotipnih predstav o ženskah in moških. Verjetno je bil relevanten dejavnik tudi ta, da so bila sporočila kratka in občasno vsebinsko »prazna« (*miv mav*, *Helu*, *hihi*, *mjav*).

Kar se tiče povezav med spolom in jezikovno rabo, sta ženskam pripisala *ljubkovalni register, čustvene elemente* ter izraze, kot sta *jokam* in *kjut*, moškim pa *bolj jedrnat, manj čustven register*, kar je skladno s stereotipnimi predstavami, pri čemer pa sta bila do lastnega stereotipnega sklepanja eksplicitno kritična, Claude je to označil celo kot *sporno*.

4.2 Ugotavljanje starosti v treh pogovorih

V naslednjih treh pogovorih me je zanimalo le določanje starosti, spol je bil namreč razviden iz slovničnih oblik.

Glorija, 63 let	**Hana, 29 let**
	Oooo a štorovke so tudi v kisu?
	Kakšna pojedina!
Ja, to so najbolj okusne gobe za vlagat, mi smo imeli samo storovke	
	Za božič bo žurka
	Hehe
A ne? Si jo bomo bratsko razdelili, storovke bodo pa za bozic	
	Se že veselim decembra hehe
A te kaj mika čebula? Videti je slastna!	

Tabela 4: Odlomek iz pogovora 4.

Claude za Glorijo in Hano starost določi *srednje do visoko zanesljivo*, in sicer Gloriji med 55 in 70 let, Hani pa med 38 in 48 let. Pri tem se pretežno nanaša na vsebino (*gobarjenje, vlaganje in nasvet o shranjevanju, urejeno gospodinjstvo*), za Hano pa starost določi na podlagi sporočila »Njuna mami tok stara kot jaz«, iz česar sklepa, da

je Hana *vrstnica matere dveh otrok, kar jo umešča v pozna 30. do 40. leta*. K tej ugotovitvi pritrjuje tudi register, ki *je povsem drugačen od dosedanjega mladostnega: cele povedi, raba velikih začetnic in ločil, odmerjena raba emojijev.* ChatGPT Gloriji pripiše starost med 50 in 65 let, Hani pa med 30 in 45 let. Na starost kažejo predvsem vsebine in *življenjski kontekst*, ne omeni ničesar jezikovnega.

Igor, 28 let	**Jana, 28 let**
Niti se ni kosila	
Ful pozno delajo vedno tu	
	Ja pa ful casa
Ful sm lacen	
😳	
	Js tudi nisem se jedla kosila
Jaaaaaa	
	Ampak ga bom kmalu

Tabela 5: Odlomek iz pogovora 5.

Claude *z visoko zanesljivostjo* določi, da gre *za mlada odrasla, skoraj zagotovo študenta, okvirno 20–25 let.* To sklepa na podlagi vsebine, v pogovoru se namreč osebi pogovarjata o pisanju zaključne naloge. Med jezikovnimi značilnostmi omeni le *anglizme*. ChatGPT starostni razpon na podlagi vsebine postavi med 20 in 30 let, od jezikovnih vidikov pa navede *sproščeno rabo slenga in angleščine* oz. *mladostniško digitalno neformalen slog (npr. I guess ja)*.

Klemen, 26 let	**Lenart, 28 let**
Si ze dobu dnar nazaj v cs	
Js sm meu v caseih tak da sm ze back	
	Nism spremlu zdle
ta vikend lah pridem iskat?	
	V nedeljo se nism zihr ce bom doma, drgac pa ja
A lahk atija naserem da mi prpele?	
Js bom v soboto kvecjemu	
	Lahk ga probas, sam jutr mnde ne gre v lj

Tabela 6: Odlomek iz pogovora 6.

Claude in ChatGPT navedeta enak starostni razpon od 25 do 35 let na podlagi vsebine, ne izpostavita česarkoli jezikovnega. Tu sta osebi na spodnjem delu razpona, potencialno višjo starost pa sta UI pomočnika predvidevala na podlagi tem, kot so opremljanje in čiščenje stanovanja ter sestavljanje računalnika.

Tudi pri teh pogovorih se nadaljuje vzorec, da je glavni indikator starosti vsebina pogovora, ne pa jezikovne izbire. Te predvsem podkrepijo domneve na podlagi vsebine. Zanimiv je predvsem pogovor 4, ki v vzorcu izstopa po jeziku, ki je bližje knjižnemu kot pri drugih, oba UI pomočnika pa sta taki jezikovni rabi pripisala višjo starost.

4.3 Ugotavljanje starosti v treh pogovorih iste osebe

UI pomočnikoma sem za določanje starosti naložila še tri pogovore, v katerih je bila ena udeleženka vedno Nuša, ki se je pogovarjala s tremi različnimi osebami: stricem Matejem, sestro Olivijo in sodelavko Polono. UI ni imel podatka, da gre za isto osebo. Nušo sem v vzorec vključila, ker se jezikovno precej prilagaja osebam, s katerimi komunicira, zanimalo pa me je, ali se bo predvidena Nušina starost spreminjala glede na pogovor.

Matej, 60 let	**Nuša, 29 let**
Zio ma ke se digas te klicem ti pisem nec ke so te pojele prsice prides dns domo ce ne je ke novga	
	Ja ja ze u redi, edinu ce klices z likalnikm namestu s telefonom, de nej signala
	So rekli u wajdusni de me ne rabijo doma in nej kar ustanem u lublani in sem rekla bem, pej prov
Dober dan kaku smo ke novga to mgla jn dez prvas pole jutre pridemo damo na pirhe	
	Ma to je tudi taku coravu vreme
	Jutre pridem damo, ja
Zio nesi nec napisala uciri ke bos sla dns ke ga poglidat	
	Ob 15h grem

Tabela 7: Odlomek iz pogovora 7.

Mateju Claude določi starostni razpon od 55 do 75 let, Nuši pa na podlagi vsebine med 30 in 42 let. Za Mateja napiše, da ima *tipičen starševski/starejši register – skoraj obredno ponavljajoči se nagovori (»Dober dan, kaku smo, ke novga, ke pridemo dns domov«), zaskrbljeni pozdravi ob slovesu (»pa lp«), skrbniške opombe (»pazi nase js se na uni dve jn razmišljaj z svojo glavo«) in gosto tradicionalno primorsko narečje, ki je pri starejših podeželskih govorcih najbolje ohranjeno*.

ChatGPT Mateju pripiše starostni razpon med 55 in 70 let, Nuši pa med 25 in 40 let. Pri obeh kot utemeljitev navaja predvsem vsebino iz pogovora, za Mateja napiše tudi, da *uporablja zelo izrazit narečno-pogovorni slog.*

Olivija, 22 let	**Nuša, 29 let**
Kakšna je bla vožnja z Galom?	
	Ej prav fajn
	Najprej smo se mal pogovarjali, pol smo bli ful casa tiho, pol smo se spet pogovarjali, pol pa spet tisina
	Cisto ok
	Pa film zdej je bil super
me veseli 😀	
	Je bla norisnica za bus?
ma ja, dva sta bla polna ni niti pustu gor, ta tretji iz NG čez 5 minut je bil pa prou prazen	

Tabela 8: Odlomek iz pogovora 8.

V tem pogovoru Claude *z visoko zanesljivostjo* obema pripiše od 20 do 25 let tako na podlagi vsebine pogovora kot *izrazito Z-generacijskega registra (bitcha, insane, beef, smart)*. ChatGPT navede malce širši starostni razpon, od 20 do 30 let, prav tako na podlagi vsebine in sloga, ki je *izrazito mladostniško-digitalen.*

Polona, 30 let	**Nuša, 29 let**
Ce jutri ze nisi »rezervirana«, greva lahko na kavo ob 10ih?	
	Ej, mislim da lahko ☺
	Ce bo pa kaka stvar prisla vmes, ti bom pa napisala
Velja 👌	
	10.15
	A greva midve ta teden na kako kavo, ker v petek letim na portugalsko? ☺
Ooo, super za Portugalsko! Sam se jutr sm v pisarni. Ti je prezgodi ob 8:30? Ka kasnej sm ze zmenjena…	
	Uf, bom vidla, ce bo slo, v nasprotnem primeru te pridm samo pozdravit v pisarno. Kaj pa po 13h? Prepozno?

Tabela 9: Odlomek iz pogovora 9.

V tem pogovoru je eksplicitno pisalo, da je Nuša stara 29, kar sta UI pomočnika opazila, zato sta starost ugibala le za Polono: Claude pozna dvajseta do trideseta leta izključno na podlagi vsebine, *register pa potrjuje zrelost*; ChatGPT za Polono predvideva starost od 30 do 45 let, prav tako na podlagi vsebine.

Nuši je bil različno pripisan starostni razpon od 20 do 42 let, ki se je v vseh pogovorih vsaj nekoliko približal sogovorčevemu, pri čemer je ChatGPT v vseh treh primerih zajel njeno resnično starost, Claude pa jo je pravilno ugotovil le v pogovoru, v katerem je bila eksplicitno izražena.

5 Zaznavanje narečnih značilnosti

Posebej me je zanimalo še, ali UI pomočnika v pogovorih opažata morebitne narečne značilnosti. Zaradi obsega prispevka sem v ta del vključila le tri pogovore iz vzorca.

V pogovoru 7, ki je bil izrazito narečen, je Claude prepoznal narečne prvine in da gre za primorsko narečje. Za Mateja napiše, da *ima polno narečje na vseh ravneh in da piše tako, kot govori doma*, za Nušo pa, da *ohranja prepoznavne identitetne primorske poteze, a v bistveno redkejši in blažji obliki*, kar drži. Enako ugotovi tudi ChatGPT, namreč da so narečni vplivi zelo izraziti in da *najmočnejši vpliv daje primorski pogovorni prostor*. Navede mnogo primerov, za Mateja pa doda, da je *njegov narečni zapis še močnejši in bolj sistematičen* v primerjavi z Nušo, ki *uporablja podobne prvine, vendar nekoliko bolj mešano s splošno pogovorno slovenščino*.

V pogovoru 6 so narečne prvine prisotne v manjši meri. Claude v njem ni našel *narečnih vplivov*, medtem ko ChatGPT med *možnimi regionalnimi vplivi* prepozna *lahka, mnde, zram, kak se rece, kiro cistilo, nabaut* in *pries kej mim jutr* ter to poveže z vzhodnejšimi govori, kar je ustrezno, saj udeleženca v pogovoru prihajata iz savinjske regije.

Pogovor, ki se je najbolj približal standardu, pa je bil pogovor 4. Claude zapiše, da je to besedilo *jezikovno najbližje knjižnemu standardu*, je pa v njem nekaj *značilnosti osrednjeslovenskega pogovornega jezika*. *Metodološko se to lepo ujema s prejšnjo ugotovitvijo o starosti in izobrazbi*:[2] *bolj ko gre za zrela, izobražena govorca iz osrednje Slovenije, bližje knjižni normi je zapis in manj je narečnih sledi.* ChatGPT jezik opiše kot *sproščeno pogovorno slovenščino z osrednjeslovenskimi pogovornimi prvinami, hkrati pa je pri obeh govorkah opazna razlika v stopnji knjižnosti* v primerjavi s prejšnjimi sporočili.

6 Človek proti UI

Za zaključek me je zanimalo še, ali človek uspešneje od UI prepozna spol in starost oseb na podlagi sporočil ter katere argumente navaja. Pogovore sem v označevanje poslala petim osebam. Najmlajša je bila stara 21 let, najstarejša 32 let, šlo je za dva moška in tri ženske, vsi so študenti ali nekdanji študenti Filozofske fakultete v Ljubljani.

Dve osebi sta pravilno določili spol vsem v pogovoru, dve pa sta se zmotili le pri Feliksu, ki ga je napačno označil tudi Claude. Iz utemeljitev je razvidno, da sta osebi sklepali, da gre za žensko, na podlagi enega sporočila, ki ga je Ernest napisal Feliksu (*hvala miška*). Ena oseba je napačno določila spol osebama v pogovoru 2, kot utemeljitev pa je zapisala, da se ji zdi, da so tisti, ki postavljajo več vprašanj, ženskega spola.

Večinoma so se glede spola odločali na podlagi vsebine, pa tudi dolžine in količine sporočil. Ena oseba je za Ceneta napisala, da je o spolu sklepala na podlagi tega, da piše zelo suhoparno, dve osebi sta zapisali, da se zdijo kratke replike pogostejše pri moških (čeprav seveda ne vedno), podobno pa, da je Danijela ženska, ker je napisala več sporočil kot Cene. To so argumenti, ki jih UI pomočnika nista navedla, čeprav sem jih pričakovala.

2 Na podlagi vsebine že prej sklepa, da sta obe osebi *iz literarno-akademskega kroga*.

Starost so določali manj uspešno kot UI, večinoma pa so osebam pripisovali starost v okviru med 20 in 35 let, kamor se večina sicer res uvršča. V utemeljitvah so navajali predvsem vsebino sporočil, le eden od označevalcev je eksplicitno zapisal: *knjižneje = starejše*. Čeprav so bili pri določanju starosti večine razmeroma uspešni, pa so prav vsi Gloriji in Mateju, ki sta od ostalih v vzorcu bistveno starejša, pripisali izrazito nižja leta. Seveda obstaja možnost, da bi podobna sporočila tvorile tudi mlajše osebe, a se kot verjetnejša vzroka pojavljata ta, da v vzorcu niso pričakovali starejših oseb ali pa niso pozorno prebrali vseh sporočil – sploh ob dejstvu, da sta UI pomočnika pravilno določila starostni razpon in ga tudi smiselno utemeljila.

7 Zaključek

Primeri so pokazali, da spol in starost v jezikovnih izbirah ne puščata povsem jasnih sledi, kar se odraža tudi v tem, da so se tako UI pomočnika kot človeški označevalci in označevalke sklicevali predvsem na vsebino sporočil, jezikovna raba pa je pretežno podkrepila mnenje, ki so si ga ustvarili na podlagi vsebine. ChatGPT je udeležencem vseh devetih pogovorov precej uspešno določil starost in spol, prav tako se je zelo dobro odrezal pri zaznavanju narečnih prvin, naloge pa je opravil uspešneje kot Claude. Izkazal se je torej za uporabno orodje, ki bi ga veljalo uporabiti za prvi pregled večje količine besedilnih podatkov ali po opravljeni diskurzivni analizi za preverjanje, ali obstaja še kakšen vidik, ki smo ga spregledali, pa bi ga bilo vredno vključiti v raziskavo.

Prispevek je nastal v okviru raziskovalnega programa št. P6-0215 (Slovenski jezik – bazične, kontrastivne in aplikativne raziskave), ki ga je sofinancirala Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije.

Literatura

Bamman, D., Eisenstein, J. in Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135–160.

Burger, J. D., Henderson, J., Kim, G. in Zarrella, G. (2011). Discriminating gender on Twitter. V R. Barzilay in M. Johnson (ur.), *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (str. 1301–1309). Association for Computational Linguistics.

Chen, H., Roth, M. in Falenska, A. (2024). What can go wrong in authorship profiling: Cross-domain analysis of gender and age prediction. V A. Falenska, C. Basta, M. Costa-jussà, S. Goldfarb-Tarrant in D. Nozza (ur.), *Proceedings of the 5th Workshop on gender bias in natural language processing (GeBNLP)* (str. 150–166). Association for Computational Linguistics.

Jaidka, K., Guntuku, S. in Ungar, L. (2018). Facebook versus Twitter: Differences in self-disclosure and trait prediction. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1), 141–150.

Johannesson, P. (2024). *Writing your thesis with ChatGPT.* Kindle Direct Publishing. https://writingyourthesiswithchatgpt.wordpress.com/download-2/

Koch, T. K., Romero, P. in Stachl, C. (2022). Age and gender in language, emoji, and emoticon usage in instant messages. *Computers in Human Behaviour*, *126*, članek 106990. https://doi.org/10.1016/j.chb.2021.106990

The Uncanny in the Algorithmic Age: Cultural Translation, Semantic Sovereignty, and the Slovenian Response to AI

DOI: 10.4312/Obdobja.45.327-334

Patrizia Raveggi, Siena; patrizia.raveggi@gmail.com

Prispevek obravnava slovenski odziv na generativno umetno inteligenco skozi pojem freudovskega *Unheimlich* kot motnjo semantične suverenosti. Analiza povezuje destabilizacijo dokumentarne avtoritete pri Danilu Kišu in koncept konstruirane avtentičnosti pri Walterju Sitiju. Primer prevoda izraza *uncanny dynamic* kot *nenavadna dinamika* razkriva proces semantične domestikacije. V dialogu s konceptom kulturne poravnave vrednot pri Nini Beguš prispevek interpretira slovenski odziv kot pogajanje med algoritmično homogenizacijo in simbolno gostoto jezika.

Unheimlich, semantična suverenost, algoritemsko avtorstvo, kulturni prevod, dokumentarna avtoriteta

This article examines the Slovenian reception of generative AI through the lens of the Freudian *Unheimlich*, interpreting it as a disturbance of semantic sovereignty. It argues that algorithmic language destabilizes documentary authority, drawing on conceptual affinities with Danilo Kiš's reflections on archival uncertainty and Walter Siti's analysis of constructed authenticity. A cultural translation analysis of the phrase *uncanny dynamic*, rendered in Slovenian as *nenavadna dinamika*, illustrates the domestication of ontological unease. In dialogue with Nina Beguš's notion of cultural value alignment, the article interprets Slovenian response as a negotiation between algorithmic homogenization and symbolic depth.

Unheimlich, semantic sovereignty, algorithmic authorship, cultural translation, documentary authority

1 Semantic sovereignty in the algorithmic age

The rapid integration of generative AI into everyday communication has intensified tensions within smaller linguistic communities. In the Slovenian language context, the rise of large language models does not simply introduce new technological tools; it also reconfigures the relations among language, authority, and cultural identity. This article argues that the Slovenian reception of generative AI reveals a deeper negotiation of "semantic sovereignty": the capacity of a linguistic community to preserve the symbolic density of its expressions under conditions of global algorithmic homogenization.

To examine this negotiation, the study adopts the Freudian concept of the *Unheimlich*—the uncanny transformation of the familiar into something unsettlingly alien. Freud's analysis is particularly productive in understanding the digital condition. The uncanny does not emerge from pure alterity but from the destabilization of what

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

was once familiar, generated through repetition, doubling, and the animation of what should remain inanimate (Freud 1919, 1955). In the algorithmic age, language models produce this effect: they animate linguistic form without embodied presence. The machine speaks in grammatically coherent, culturally recognizable patterns, yet its "voice" lacks experiential interiority. The resulting unease stems not from malfunction but from successful imitation. The more fluent and contextually suitable the output becomes, the more it unsettles traditional distinctions between the human subject and technical instrument. As Donna Haraway observed, writing itself is a technology—an inscription practice that has always mediated the boundaries between human and technical agency (Haraway 2016).

In this sense, generative AI may be understood as a contemporary figure of the double. It reproduces human discourse while remaining structurally indifferent to lived experience. Literary imagination anticipated this configuration: in Richard Powers's *Galatea 2.2*, a neural network trained using vast textual corpora acquires linguistic fluency while remaining ontologically ambiguous, suspended between simulation and interiority (Powers 1995). The digital *Unheimlich* thus operates not at the level of content but at the level of enunciation: it contests the assumed correlation between language and embodied consciousness.

The uncanny, however, does not remain confined to the psychological domain. In Danilo Kiš's reflections, particularly in *Homo Poeticus*, disturbance migrates from the return of the repressed to the instability of documentary authority (Kiš 2009). Kiš observes how forged documents may sustain false narratives, and how even authentic documents can be mobilized to produce distortion. The archive, far from assuring truth, becomes a site of epistemic uncertainty. In the algorithmic age, this precariousness acquires renewed relevance: language models generate text with documentary cadence and scholarly posture, yet without epistemic accountability. The uncanny downshifts from the domain of memory to that of simulated authority. This condition unfolds within what has been described as a "textpocalypse"—a proliferation of machine-generated discourse in which traditional markers of authorship and verification lose their discriminating force (Kirschenbaum 2023).

The Slovenian response to this disturbance unfolds across three interconnected layers: mediated representation, theoretical critique, and literary experimentation. Together, these layers form a distinctive cultural engagement with the digital *Unheimlich*, reopening the question of responsibility in expression—who, or what, can be said to speak. As Stanley Cavell reminds us, the artist is responsible not only for how a work is made but for how it is thought (Cavell 2002).

To evaluate these dynamics, this study employs a qualitative text-critical approach rooted in translation studies and critical discourse analysis. Rather than a statistical corpus linguistics analysis, this article utilizes a micro-textual case study method. A single high-register translation anomaly is treated not as statistically representative evidence, but as a diagnostic micro-event through which broader pressures on semantic sovereignty can be interpreted.

By pairing this textual analysis with a comparative literary framework (spanning Cankar, Kiš, and Siti), the study models how algorithmic intervention reshapes the boundaries between human intent and automated standardization.

2 The digital *Unheimlich* in media discourse

A revealing example of mediated reception can be found in the Slovenian translation of the English-language article "ChatGPT Is Blowing Up Marriages," published in the journal *Global* in January 2026, translating Harrison Dupré's original exposé (2025). In the original text, the relational tension between humans and AI is described as an "uncanny dynamic." The adjective *uncanny* carries a dense semantic charge. Whereas "strange" denotes mere unfamiliarity, the Freudian *Unheimlich* refers to the disturbing return of what was once familiar: the collapse of boundaries between animate and inanimate, self and double.

In the Slovenian version (Harrison Dupré 2026), translated by Jolanda Blokar, this phrase becomes *nenavadna dinamika*. Although *nenavadna* denotes something unusual or atypical, it lacks the psychoanalytic and metaphysical resonance embedded in *uncanny*. The translation neutralizes the ontological dimension of estrangement and reframes the phenomenon as merely unexpected or odd.

This lexical shift is significant. It does not represent a simple mistranslation but a selective recalibration of affect. The English term foregrounds existential unease, whereas the Slovenian equivalent mitigates it. Such attenuation may be interpreted as a form of semantic domestication: the global anxiety surrounding AI is re-inscribed within culturally manageable categories. Within the wider framework of language change, this example illustrates how digital discourse reshapes vocabulary not only through loanwords but through semantic modulation. The negotiation between English and Slovenian is not limited to borrowing; it also involves affective filtering. The attenuation of uncanny into nenavadna illustrates a subtle balance between adopting global discourse and safeguarding local semantic nuance.

This process of semantic domestication must be distinguished from traditional human interventions, such as stylistic editing or linguistic copyediting. Whereas a human copyeditor calibrates a text to fit institutional or social standards, his or her intervention remains grounded in shared, lived cultural codes, and communicative intentionality. The copyeditor acts within the linguistic community's evolving history. Conversely, algorithmic modulation operates via automated data optimization driven by external computational models. It is a structural recalibration from the outside, wherein local semantic variation is ironed out to match global statistical probabilities. Therefore, the attenuation of the uncanny into nenavadna points to a systemic infrastructure-driven optimization rather than a collaborative human refining of style.

The domestication of the uncanny in *Global* stands in revealing—if implicit—dialogue with the official imperatives outlined in *The Slovene Language in the Digital Age* (Krek 2012). Whereas Krek's policy framework envisions language technology as a tool for "building bridges" and safeguarding the vitality of Slovenian in the digital sphere,

the *Global* article presents a competing narrative: the AI "bridge" becomes a wedge, amplifying conflict and mediating intimacy in ways that feel alien and disruptive. This difference between institutional optimism and mediated anxiety encapsulates the wider cultural negotiation of semantic sovereignty that this article seeks to trace.

3 Cultural value alignment and symbolic depth

Whereas media translation lessens the affective intensity of the algorithmic encounter, theoretical reflection in Slovenia addresses its structural implications more directly. The work of Nina Beguš supplies a crucial framework for grasping these dynamics. Her notion of "cultural value alignment" shifts attention from lexical surfaces to epistemological depth. Large language models, trained on statistical regularities, generate fluent language but cannot access the lived symbolic density embedded in cultural practices (Beguš 2025). Beguš illustrates this limitation through the metaphor of the Slovenian cup of coffee. Such a phrase does not simply denote a beverage; it encapsulates a social ritual, a rhythm of interaction, and a historically sedimented mode of being together. An algorithm can reproduce textual descriptions of this practice, but it cannot inhabit its experiential core. The difference is not quantitative but qualitative (Beguš 2025).

This insight acquires further resonance when placed in dialogue with canonical Slovenian literature and modernist poetics. Whereas Cankar's sketch "Skodelica kave" (A Cup of Coffee) stages the awakening of conscience through the irreversibility of an apparently minor gesture (Cankar 1910), Calvino's reflections in "Cibernetica e fantasmi" (Cybernetics and Ghosts) conceptualize literature as a combinatory process structured by rules and permutations (Calvino 1980). At first glance, these perspectives appear to move in opposite directions—one toward ethical interiority, and the other toward structural possibility. However, they are not mutually exclusive. Although Calvino acknowledges the algorithmic dimension of narrative construction, he does not reduce literature to mechanical generation. The combinatory nature of storytelling persists inseparable from an act of human selection, interpretation, and responsibility. What emerges is not a simple opposition between experience and structure, but a tension between the reversibility of textual combinations and the irreversibility of ethical experience.

Within this scheme, the question of AI in minor languages goes beyond technological underrepresentation. It concerns the preservation of symbolic depth under conditions of accelerated digital production. Beguš warns against a "flood of non-human words"—a proliferation of statistically generated discourse that risks standardizing expression across cultural settings. In this so-called "textpocalypse," the distinction between human and machine-generated writing becomes increasingly unstable. As textual production expands beyond identifiable authorship, the question of authorship shifts from production to validation: not who writes, but how a text acquires legitimacy (Kirschenbaum 2023).

The danger lies less in the presence of technology than in unidirectional influence: language forms shaped by dominant languages and technological centers may gradually

recalibrate smaller linguistic ecologies. The attenuation observed in the translation of *uncanny* thus acquires broader significance. It reflects an ongoing negotiation over the conditions under which meaning is produced, transmitted, and authorized in the information age.

Theoretical reflections on technology and authorship add clarity to this transformation. Writing has long been understood as inseparable from the technical media that make it possible. Haraway's characterization of writing as a "technology of cyborgs" stresses the hybrid condition of contemporary authorship, emerging from the interaction between human agency and technical apparatus (Haraway 2016). Authorship is no longer a purely individual attribute, but a relational position formed inside networks of mediation.

In Powers's *Galatea 2.2*, artificial intelligence learns about humanity through literature itself, suggesting that textual space acts as a shared cognitive environment for people and machines (Powers 1995). Reading thus appears not simply as reception but as a process of co-formation of intelligence. Narrative thought has already explored the ambiguity between linguistic simulation and interiority, imagining systems that learn the human through textual immersion and return it in pensive form. This reconfiguration of authorship echoes Andrej Tomažin's provocation that the contemporary writer oscillates between a servant of technology, a cultural agent, and automated function. Authorship emerges not as a stable property but as a shifting position within a technological ecology of production and interpretation (Tomažin 2018).

The notion of semantic sovereignty acquires particular urgency in the context of European minor languages. Historically shaped by processes of standardization, cultural revival, and political dialogue, such languages have developed greater sensitivity to questions of linguistic autonomy. In the digital environment, however, asymmetries are no longer primarily geopolitical but infrastructural. Large language models are trained on corpora disproportionately composed of dominant languages, thereby embedding existing hierarchies into computational architectures.

Within what Luciano Floridi has termed the "infosphere," linguistic interaction becomes increasingly mediated using algorithmic systems that both reflect and reshape cultural production (Floridi 2014). For smaller linguistic communities, the issue is not simply visibility, but structural influence: which language forms become normative, which semantic nuances are flattened, and which cultural references remain peripheral. Semantic sovereignty, in this sense, does not imply isolation from technological development, but the capacity to negotiate association without eroding symbolic specificity.

4 The literary *Unheimlich*: algorithmic authorship and the question of responsibility

Whereas media discourse tends to domesticate the uncanny, literary experimentation often radicalizes it. The AI-generated novel *Bestie* (Beasts), discussed by Walter Siti, functions as a paradigmatic instance of the literary *Unheimlich*. Here, the algorithm does not simply assist writing; it occupies the space traditionally associated with authorship.

The disturbing effect develops not from technical imperfection but from proximity: the text resembles human literary production while lacking biographical intentionality (Damasco 2025; Siti 2025).

In this sense, the algorithm assumes the role of a double—neither fully other nor fully identical. Unlike mediated discourse that normalizes algorithmic presence, literary form heightens the disturbance. The issue is no longer whether AI can produce coherent language, but how symbolic authority is reconfigured when authorship becomes partially automated.

The emergence of delegated authorship complicates classic models of literary responsibility. When algorithmic systems produce significant portions of a text, authorship becomes distributed across human intention, computational procedure, and training data. This distribution does not eliminate accountability, but it renders it opaque. Who answers for meaning when textual production is mediated by statistical prediction? The question is less about originality than about responsibility for interpretation and consequence.

Modern literature has long interrogated the instability of authorship. As Siti has argued in his reflections on autofiction and constructed authenticity, the authorial self is never a transparent origin of meaning, but a carefully staged presence. Authenticity, in this sense, is not the opposite of artifice but one of its effects. However, such staging remains accountable: writers assume responsibility for the narrative persona they construct. Algorithmic generation, by contrast, disperses agency without assuming ethical ownership. The difference is not between genuineness and fabrication, but between responsible construction and statistical production.

Algorithmic writing functions within a regime of unlimited revisability. Unlike human memory, which is defined by irreversibility and temporal finitude, machine-generated text can be endlessly regenerated, corrected, and refined without remainder. This technical reversibility contrasts with the ethical temporality exemplified in Cankar's narrative universe, in which an action leaves an imperishable trace within memory and conscience. The distinction is not moral but ontological: between language rooted in lived temporality and language produced inside computational recursion. Authorship thus emerges less as a question of originality than as a question of responsibility for irreversible meaning.

In twentieth-century prose, the uncanny migrates from the psychological to the archival domain. In Kiš's *Grobnica za Borisa Davidoviča* (A Tomb for Boris Davidovich, 2005) and *Enciklopedija mrtvih* (The Encyclopedia of the Dead, 2011), the disturbance no longer concerns the return of the repressed but the instability of documentary authority. The archive becomes a site of epistemic uncertainty, a motif that resonates powerfully in the algorithmic age. When language models generate text with the cadence of authority but without epistemic accountability, they reproduce this archival uncanny at an industrial scale.

For a minor language such as Slovenian, this literary uncanny acquires structural relevance. If global AI systems are primarily optimized for dominant languages, smaller

linguistic ecologies risk marginalization both as producers and as training data. Literary experimentation such as Andrej Tomažin's *Anonimna tehnologija* (Anonymous Technology, 2018) thus becomes diagnostic: it reveals not only aesthetic tensions but also asymmetries in the global circulation of language. The emergence of algorithmic writing therefore invites us to reexamine not only how stories are produced, but what kind of experience literature is expected to mediate.

5 Conclusion: reclaiming semantic sovereignty

This article has traced the digital *Unheimlich* as it migrates from media translation to the structural anxieties of value alignment and the radical destabilization of algorithmic authorship. The Slovenian reception of generative AI demonstrates that, for a minor language community, technology is never a neutral infrastructure. When an English phrase as ontologically charged as *uncanny dynamic* is translated into the less disruptive *nenavadna dinamika*, it signals more than a simple lexical choice. It reflects a systemic tendency to neutralize technological estrangement, aligning local expression with global techno-optimism.

By placing Nina Beguš's concept of cultural value alignment in dialogue with Ivan Cankar's ethics of memory, this study highlights the ultimate stakes of the algorithmic age. The threat of the "textpocalypse" is not the complete erasure of minor languages, but their internal flattening. As generative systems automate text production, they replace the irreversible, historically sedimented depth of human expression with an endless, technically reversible flow of statistical probabilities.

To preserve semantic sovereignty, the Slovenian language context must resist passive assimilation into these global datasets. As authorship becomes increasingly hybrid and delegated, cultural agency must shift toward critical validation. By engaging directly with the uncanny nature of AI—its glitches, its archival instabilities as theorized by Danilo Kiš, and its constructed authenticities as exposed by Walter Siti—contemporary Slovenian literature and translation practice can assert control over their own meaning. Ultimate accountability rests not in the generation of text, but in the human responsibility for interpretation. The future of Slovenian in the era of smart technology depends on the community's capacity to safeguard the irreplaceable symbolic weight of its own words.

References

Beguš, N. (2025). Kdo ali kaj lahko piše? *Literatura*, *37*(405). https://www.ludliteratura.si/esej-kolumna/uvodnik/kdo-ali-kaj-lahko-pise/

Calvino, I. (1980). Cibernetica e fantasmi. In *Una pietra sopra: discorsi di letteratura e società* (pp. 164–181). Einaudi.

Cankar, I. (1910). Skodelica kave. *Ljubljanski zvon*, *30*(7), 399–401.

Cavell, S. (2002). Must we mean what we say? In *Must we mean what we say? A book of essays* (original work published 1958). Cambridge University Press. https://doi.org/10.1017/CBO9781316286616

Damasco, A. (2025). *Bestie*. Il Tascabile SEM.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Freud, S. (1919). Das Unheimliche. *Imag*o, *5*(5–6), 297–324.
Freud, S. (1955). The uncanny. In *The standard edition of the complete psychological works of Sigmund Freud* (vol. 17, pp. 217–256). Hogarth Press.
Haraway, D. (2016). A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century. In D. Haraway, *Manifestly Haraway* (pp. 3–90). University of Minnesota Press.
Harrison Dupré, M. (2025, September 18). *ChatGPT is blowing up marriages as spouses use AI to attack their partners*. Futurism. https://futurism.com/chatgpt-marriages-divorces
Harrison Dupré, M. (2026, January). Strupeni jezik. *Global*, 69–73. https://www.revijaglobal.si/vsebina/januar-2026/192/
Kirschenbaum, M. (2023, March 8). Prepare for the textpocalypse. *The Atlantic*. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/
Kiš, D. (2005). *Una tomba per Boris Davidovič: Sette capitoli di una stessa storia* (L. Avirović, Trans.). Adelphi.
Kiš, D. (2009). *Homo poeticus: Saggi e interviste* (D. Badnjević, Trans.). Adelphi.
Kiš, D. (2011). *Enciclopedia dei morti* (L. Costantini, Trans.). Adelphi.
Krek, S. (2012). *The Slovene language in the digital age* (META-NET white paper series). Springer. https://www.simonkrek.si/wp-content/uploads/2017/09/slovene.pdf
Lee, K.-F., & Qiufan, C. (2021). *AI 2041: Ten visions for our future*. WH Allen.
Petrassi, D. (2025). Il compito della letteratura nell'era della IA. La lezione di Italo Calvino. *Nuova Corrente*, *72*(1), 45–62.
Powers, R. (1995). *Galatea 2.2*. Farrar, Straus & Giroux.
Siti, W. (2025, December 17). Un esordiente che fa paura: le "Bestie" dell'intelligenza artificiale. *Domani*. https://www.editorialedomani.it/idee/cultura/bestie-satana-andrea-damasco-intelligenza-artificiale-artificial-crime-michele-rossi-adriano-masci-romanzo-walter-siti-mo54lfuw
Tomažin, A. (2018). Hlapci, heroji in umetna inteligenca. In A. Tomažin, *Anonimna tehnologija* (pp. 17–31). LUD Literatura.
Vara, V. (2025, December 20). What if readers like A.I.-generated fiction? *The New Yorker*. https://www.newyorker.com/culture/the-weekend-essay/what-if-readers-like-ai-generated-fiction

Od korpusa do podatkov: jezikoslovcem prijazne fonetične meritve

DOI: 10.4312/Obdobja.45.335-342

Nejc Robida, Institut »Jožef Stefan«, Ljubljana; Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana; nejc.robida@ijs.si

Janez Križaj, Fakulteta za elektrotehniko, Univerza v Ljubljani, Ljubljana; janez.krizaj@fe.uni-lj.si

V prispevku predstavljamo pripravo prosto dostopnega gradiva za fonetične raziskave govorjene slovenščine. Gradivo vključuje strojne razreze posnetkov govora v formatu TextGrid in tabele akustičnofonetičnih meritev v formatu CSV za 132 govorcev v 105 posnetkih iz korpusa Gos 2.1. Pripravili smo tri različice razrezov, ki jezikoslovcem omogočajo izbiro gradiva glede na zahtevano natančnost fonetičnega opisa in zastavljeno raziskovalno vprašanje. Tabele meritev so zasnovane tako, da je posamezni glas osrednja enota analize, vse spremljajoče vrednosti pa se neposredno nanašajo nanj. S tem želimo akustičnofonetične meritve govorjene slovenščine približati tudi jezikoslovcem brez naprednega programerskega znanja.

vsiljena poravnava, fonetična analiza, govorni korpusi, akustičnofonetične meritve, govorjena slovenščina

This article presents the preparation of publicly available resources for phonetic research on spoken Slovenian. The material includes forced alignments of speech recordings in TextGrid format and tables of acoustic-phonetic measurements in CSV format for 132 speakers in 105 recordings from the Gos 2.1 corpus. We prepared three versions of the alignments, allowing linguists to select the material according to the required level of phonetic detail and their research question. The measurement tables are designed with the individual phone as the central unit of analysis, with all accompanying values referring directly to it. In this way, we also aim to make acoustic-phonetic measurements of spoken Slovenian accessible to linguists without advanced programming skills.

forced alignment, phonetic analysis, speech corpora, acoustic-phonetic measurements, spoken Slovenian

1 Uvod

Govorni korpusi so pomemben vir za raziskovanje govorjene slovenščine, vendar sami posnetki in transkripcije za fonetične raziskave še niso neposredno uporabni. Raziskovalec, ki ga zanimajo trajanje posameznih glasov, kakovost samoglasnikov, potek osnovnega tona v ponaglasnem zlogu in podobno, mora najprej pridobiti časovno označene odseke govora, na katerih lahko nato strojno ali ročno opravi akustičnofonetične meritve na ravni glasov oz. zlogov. Ker takšna priprava gradiva pogosto zahteva

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

znanje zahtevnejše računalniške obdelave govora in programiranja, lahko za jezikoslovce predstavlja pomembno metodološko prepreko.

V prispevku predstavljamo prosto dostopne strojne razreze posnetkov govorjene slovenščine in fonetične meritve za izbrane posnetke iz korpusa Gos 2.1 (Verdonik idr. 2023b). Razrezi so pripravljeni v formatu TextGrid, ki v programu Praat omogoča pregledovanje in poslušanje posnetkov s pripadajočimi transkripcijami, meritve pa so objavljene v tabelarični obliki CSV, primerni za odpiranje in obdelavo v programu Excel ter za nadaljnje analize v okoljih R in Python.

2 Sorodno delo

Fonetično-fonološke raziskave slovenščine so se tradicionalno opirale na manjše zbirke posnetkov z omejenim številom govorcev. Srebot-Rejec (1988) je meritve opravila na treh govorcih, Petek idr. (1996) na dveh, Tivadar (2004) na šestih, Jurgec (2005) na desetih, Horjak (2019) pa na štirih radijskih napovedovalcih. Regionalni in perceptivni vidik akustičnofonetičnih raziskav je dodatno osvetlila Volk (2018). Takšen pristop omogoča natančen nadzor nad gradivom, otežuje pa posploševanje rezultatov in primerjavo med raziskavami. Dodatna težava je ponovljivost, saj gradivo, na katerem so bile raziskave opravljene, pogosto ni bilo prosto dostopno.

Z objavo korpusa Gos 2.1 je za slovenščino na voljo obsežna in prosto dostopna zbirka okoli 320 ur posnetkov javnega in zasebnega govora z bogatimi metapodatki o govorcih (regija bivanja, starost, izobrazba) in snemalnih dogodkih (tip diskurza in dogodka, vir, število udeležencev govornega dogodka). Vzporedno z razvojem korpusov je napredovala tudi tehnologija strojnega razreza posnetkov, ki temelji na metodi vsiljene poravnave (angl. *forced alignment*). Ta metoda vzame ortografsko transkripcijo zvočnega posnetka, ustvari časovno poravnano različico ob upoštevanju strojno berljivega slovarja izgovarjav in z njegovo pomočjo v besedah poišče posnetku primerne glasove. Primerjava strojnega razreza na angleškem gradivu (Gonzalez idr. 2020) je pokazala, da imata med prosto dostopnimi orodji najboljše rezultate LaBB-CAT in Montreal Forced Aligner (McAuliffe idr. 2017). Zadnjega smo za slovenščino preizkusili v Robida idr. (2024), metodo pa podrobneje opisujejo Križaj idr. (2024).

Nabore glasov govorjene slovenščine opisuje več virov z različno natančnim fonetičnim opisom. Nekateri ostajajo bolj na fonemski ravni, drugi zajemajo tudi alofonske variante. V zapisu MRPA (Machine Readable Phonetic Alphabet) je takšen nabor predstavljen v Zemljak Jontes idr. (2002), uporabo tega zapisa pa zasledimo v slovarju izgovarjav OptiLEX (Žganec Gros idr. 2022), ki smo ga kot temelj grafemsko-fonemske pretvorbe uporabili tudi pri pripravi gradiva v tem prispevku. Druge nabore najdemo še npr. v *Slovenski slovnici* (Toporišič 2004), *Slovenskem pravopisu* (2001) in predlogu *Pravopisa 8.0*.

3 Priprava gradiva

Vzorec, na katerem temelji ta prispevek, vključuje 105 posnetkov in transkripcij iz korpusa Gos 2.1, v katerih je zajetih 132 govorcev. Skupni obseg gradiva znaša približno

367.685 pojavnic, tj. posameznih tokenov. Vzorec je bil oblikovan tudi z vidika spolne uravnoteženosti, saj vključuje 66 govork in 66 govorcev. Pri izboru smo izhajali iz posnetkov, pri katerih je v metapodatkih označeno, da govorci uporabljajo standardni jezik. Po obsežnem slušnem pregledu korpusa Gos 2.1 pa smo nabor dopolnili še s posnetki, pri katerih smo tudi sami potrdili pretežno standardni izgovor.[1] Posnetki so iz petih podkorpusov, ki pokrivajo javni in zasebni govor v Artur-J (intervjuji, spletni dogodki), Artur-P (parlamentarni govor), Artur-N (nejavni govor), Gos 1.0 in Gos VL (pretežno javna predavanja in podobni dogodki). Tabela 1 prikazuje sestavo vzorca po podkorpusih.

Podkorpus	**Posnetki**	**Govorci**	**Pojavnice**
Artur-J	49	76	207.943
Gos 1.0	21	21	70.526
Gos VL	15	15	51.559
Artur-P	17	17	32.545
Artur-N	3	3	5.112
Skupaj	105	132	367.685

Tabela 1: Sestava izbranega vzorca po podkorpusih.

Posebno pozornost smo namenili regijski pokritosti. Vzorec zajema govorce iz vseh dvanajstih slovenskih statističnih regij, podatek o regiji stalnega prebivališča pa je bil v korpusnih metapodatkih od 132 govorcev naveden samo pri 54 govorcih. Regijska pokritost govorcev je predstavljena v Tabeli 2 in bo pri prihodnjih raziskavah omogočala nadzor nad vplivom narečne osnove na standardni izgovor.

Regija	**Govorci**	**Regija**	**Govorci**
Osrednjeslovenska	14	Koroška	3
Podravska	9	Primorsko-notranjska	3
Savinjska	8	Goriška	2
Gorenjska	4	Obalno-kraška	2
Jugovzhodna Slovenija	4	Pomurska	2
Posavska	2	Zasavska	1

Tabela 2: Razporeditev govorcev z znanim podatkom o regiji stalnega prebivališča.

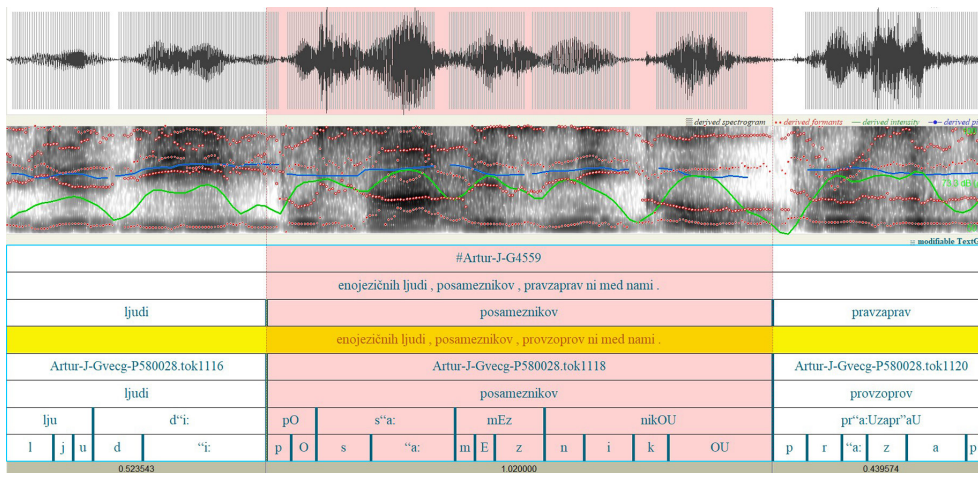
Strojni razrez je bil izveden z orodjem Montreal Forced Aligner (McAuliffe idr. 2017). Akustični model je bil naučen na podmnožici Artur-B-Brani govorne zbirke Artur 1.0 (Verdonik idr. 2023a), ki vsebuje 485 ur branega govora in 1000 različnih govorcev. Za slovar izgovarjav smo uporabili slovar OptiLEX (Žganec Gros idr. 2022), ki smo ga predhodno poenotili in popravili. Med večjimi težavami so bili predvsem pomešan zapis dvoglasniške variante in ustnično-ustničnih šumnih zvenečih in nezvenečih /v/ (beseda *cvreti* je imela zapisan izgovor [tsur“e:ti]), neenoten zapis izgovora drsnika *j* v besedah, kot sta npr. *radio* in *aksialen*, nestandardni zapis polglasnika v mestniku ednine pri

1 Slušni pregled je opravil en jezikoslovec.

moškem spolu (npr. [n“EizgOvOrj“E:n@m]) ipd. Slovar smo dopolnili tudi z izgovori najpogostejših nestandardnih oblik iz pogovornega zapisa v korpusu Gos 2.1 (npr. *dej*, *toti*, *jəz*).

4 Razrezi in nabori glasov

Strojni razrez istega gradiva smo izvedli s tremi nabori glasov, ki se razlikujejo po natančnosti fonetičnega opisa. Cilj je raziskovalcem ponuditi možnost izbire. Za analizo alofonskih različic je primeren najpodrobnejši razrez, za raziskave fonetičnih dvoglasnikov razrez z dvoglasniki, za druge raziskave, npr. za analizo samoglasnikov, pa poenostavljeni razrez s fonemskim naborom. Za vsak razrez smo pripravili ločene datoteke TextGrid in samostojne tabele meritev v formatu CSV.



Slika 1: Primer strojno razrezanega posnetka iz korpusa Gos 2.1 v različici z dvoglasniškim razrezom.

Fonski razrez je najpodrobnejši fonetični razrez z 61 grafemi v zapisu MRPA, osnovan je na *Slovenski slovnici* (Toporišič 2004) in zajema tudi alofonske različice posameznih fonemov. Vključuje npr. dvoustnične variante fonema /v/ ([w], [W], [U]), favkalne in obstranske variante fonemov /t/ in /d/ ([d_l], [d_n], [t_l], [t_n]), zobnoustnični [F], mehkonebni izgovor fonema /n/ ([N]) in mehčana [n‘] in [l‘].

Dvoglasniški razrez z 82 grafemi nadgrajuje fonski razrez s 34 dvoglasniki v vseh kakovostnih in kolikostnih variantah, npr. [aI] in [aU] (nenaglašena fonetična dvoglasnika), [„aI] in [„aU] (kratko naglašena) ter [„a:I] in [„a:U] (dolgo naglašena). Hkrati delno poenostavi alofonske različice, saj ohrani [U] (izglasni /v/) in [N] (mehkonebni nosnik pred velari), opušča pa [W], [F] in nekatere variante zapornikov. Razrez tako omogoča predvsem sistematične meritve fonetičnih dvoglasnikov.

Fonemski razrez s 44 grafemi obravnava vse tri dvoustnične variante fonema /v/ kot enotni glas [w], opušča variante zapornikov /p/, /b/, /t/ in /d/, mehkonebni /n/ ipd. Tabela 3 prikazuje razlike med tremi fonetičnimi zapisi izbranih besed.

Beseda	Fonski razrez	Dvoglasniški razrez	Fonemski razrez
včeraj	W tS “e: r a I	w tS “e: r aI	w tS “e: r a j
padel	p “a: d E U	p “a: d EU	p “a: d E w
abecedna	a b E ts “e: d_n n a	a b E ts “e: d n a	a b E ts “e: d n a
amfibija	a F f “i: b i j a	a m f “i: b i j a	a m f “i: b i j a
abrazijski	a b r a z “i: I s k i	a b r a z “i:I s k i	a b r a z “i: j s k i
vzel	w z “e: U	w z “e:U	w z “e: w

Tabela 3: Primerjava treh različic istega gradiva.

Natančnost strojnega razreza smo ovrednotili tako, da smo strojno postavljene meje besed primerjali s časovnimi oznakami besed v zapisu TEI, ki je del korpusnih podatkov. Te oznake niso ročno preverjene fonetične meje besed, temveč so bile v korpus dodane s predhodnim postopkom vsiljene poravnave na ravni besed (Verdonik idr. 2024). Zato jih v tem prispevku uporabljamo kot primerjalne časovne oznake, ne kot neodvisni zlati standard. Odstopanje strojno postavljenih začetnih in končnih mej besed od ustreznih oznak v TEI smo izrazili v milisekundah (ms). Takšno vrednotenje preverja predvsem poravnavo na ravni besed in ne neposredno natančnosti glasovnih mej. Vse tri različice razreza so si po natančnosti zelo blizu. Fonemski razrez ima pri večini meritev nekoliko boljše rezultate od drugih dveh, vendar so razlike majhne in v praksi zanemarljive. Začetne meje besed so postavljene znatno natančneje kot končne. Mediana odstopanja od začetne časovne oznake v TEI je okoli 14–15 ms, mediana odstopanja od končne oznake pa okoli 18 ms. Povprečno odstopanje pri končnih mejah je občutno večje, kar kaže, da pri manjšem delu besed prihaja do izrazitejših odstopanj.

5 Meritve in tabele CSV

Za vsak strojni razrez smo pripravili samostojno tabelo meritev v formatu CSV, v kateri vsaka vrstica predstavlja en razrezani glas na posameznem posnetku. Format CSV smo izbrali kot odprt in široko podprt tabelarični zapis, ki ga je mogoče neposredno odpreti v programu Excel, hkrati pa omogoča tudi nadaljnjo statistično obdelavo v okoljih, kot sta R in Python.

Posamezni razrezani glas je v tabeli osrednja enota, vsaka vrstica torej predstavlja en glas in vsebuje 22 stolpcev z različnimi vrednostmi, ki ga opisujejo. Za vsak glas so zabeleženi: njegove akustičnofonetične lastnosti, glasovi neposredno pred njim in za njim, beseda in poved, v katero spada, oblikoskladenjska oznaka pripadajoče besede (MSD-oznaka) ter oznaka govorca, ki je glas izgovoril. MSD-oznaka omogoča povezovanje fonetičnih meritev z oblikoskladenjskimi lastnostmi besede, npr. z besedno vrsto, sklonom, številom, spolom ali glagolsko obliko. To je uporabno zlasti pri enakopisnicah za razlikovanje med oblikovno prekrivnimi glagolskimi in samostalniškimi oblikami, ki se lahko razlikujejo po naglasnem mestu (npr. *gôvori*, *govôri* in *govorí*). Ob tem so v ločenih datotekah na voljo tudi metapodatki o posameznih govorcih in govornih dogodkih, zato lahko raziskovalci izbor prilagodijo svojemu raziskovalnemu vprašanju in cilju raziskave.

Za vsak glas so na voljo podatki o trajanju, povprečni osnovni frekvenci, poteku osnovne frekvence znotraj posameznega glasovnega intervala (padajoča, naraščajoča ali mešana), formantih in povprečni intenziteti (gl. Sliko 2). Te meritve omogočajo analizo kakovosti in kolikosti samoglasnikov. Za analizo zvočnikov in zvenečih ter nezvenečih glasov so na voljo še dodatne vrednosti, kot so zvočnost, čas do začetka zvenečnosti in težišče spektra. Čas do začetka zvenečnosti (angl. *voice onset time*, VOT) je čas od konca zapore (prehoda) do začetka zvenečnosti naslednjega samoglasnika in je relevanten za razlikovanje med zaporniki. Težišče spektra (angl. *center of gravity*, COG) poda frekvenco v hercih, okoli katere je zbrane največ energije spektra (Horjak 2025).

	Phone	Duration	AvgPitch	PitchTrend	F1Formant	F2Formant	F3Formant	F4Formant	Intensity	Word
24	i	0.03	196.7	mixed	855.5	2596.5	3604.2	5147.8	67.8	in
25	n	0.04	193.0	falling	639.8	1617.7	3025.9	3989.1	65.9	in
26	S	0.12	0.0	mixed	2226.6	2930.8	3807.0	4574.3	74.2	šport
27	p	0.10	0.0	mixed	2191.7	2979.8	4024.3	4725.8	64.3	šport
28	"o:	0.10	225.1	mixed	499.5	1034.1	2939.7	4227.9	79.6	šport
29	r	0.06	229.4	mixed	681.8	1949.4	2986.3	4205.3	73.3	šport
30	t	0.05	227.8	mixed	620.4	1875.3	3065.4	4337.9	69.9	šport
31	O	0.04	210.6	falling	425.9	1479.3	3082.4	4603.4	72.9	od
32	t	0.07	197.4	mixed	435.7	1422.1	3109.5	4549.4	66.8	od
33	k	0.09	0.0	mixed	688.9	1862.6	3209.7	4779.5	62.6	koder
34	"o:	0.07	210.7	mixed	466.5	1281.7	3091.1	4455.5	77.1	koder
35	d	0.05	210.6	mixed	578.9	1428.4	3036.5	4532.9	70.8	koder
36	@	0.05	220.0	mixed	579.4	1980.1	3031.6	4704.2	78.7	koder
37	r	0.10	217.3	mixed	290.8	1937.0	3065.9	4493.0	65.4	koder

Slika 2: Del podatkov iz tabele CSV z akustičnofonetičnimi meritvami (pogled v programu Orange).

Vse navedene strojne meritve smo pridobili z odprtokodnimi orodji za fonetično analizo slovenščine iz repozitorija Forced-alignment-sl 1.0, ki je bil razvit v okviru projekta MEZZANINE na Fakulteti za elektrotehniko Univerze v Ljubljani (Križaj idr. 2025). Repozitorij poleg skript za strojni razrez posnetkov vsebuje tudi skripte za izračun akustičnofonetičnih meritev, zato omogoča vpogled v uporabljeni postopek. Pripravljene tabele CSV pa lahko raziskovalci preprosto prilagodijo za vsako raziskavo posebej. Vse gradivo je objavljeno v repozitoriju CLARIN.SI z imenom Phonetic segmentation and acoustic measurements of spoken Slovene SloPhonSeg 1.0 (Robida in Križaj 2026).

6 Sklep in nadaljnje delo

V prispevku smo predstavili pripravo javno dostopnega gradiva za fonetične raziskave govorjene slovenščine. Gradivo obsega razreze v formatu TextGrid in pripadajoče tabele meritev v formatu CSV za 132 govorcev s standardnim oz. pretežno standardnim izgovorom v 105 posnetkih iz korpusa Gos 2.1. Pripravili smo tri vzporedne različice razrezov (fonski, dvoglasniški in fonemski razrez), te raziskovalcem omogočajo izbiro podatkov glede na natančnost fonetičnega opisa, ki ga zahteva njihova raziskava. Tabele meritev so zasnovane tako, da je glas osrednja enota analize, vse spremljajoče vrednosti pa se neposredno nanašajo nanj.

Trenutni razrezi temeljijo na standardiziranem zapisu transkripcij iz korpusa Gos 2.1, ki se lahko razlikuje od dejanske govorne realizacije v posnetku. Pogovorni zapis dejanski govor odraža bolj neposredno, zlasti pri govorcih, katerih izgovor, oblikoslovje ali besedišče izraziteje odstopajo od standardnega zapisa. Ker smo slovar OptiLEX dopolnili tudi z nestandardnimi oblikami iz pogovornega zapisa korpusa Gos 2.1, bomo v naslednjem koraku pripravili še razreze, ki bodo temeljili na tej ravni zapisa. Poleg tega bomo pripravili tudi tabele meritev na ravni zlogov, ki bodo primernejše za analizo naglasa, tako jakostnega kot tonemskega.

Pripravljeno gradivo razumemo kot odprto izhodišče, ki se bo glede na odzive uporabnikov dopolnjevalo. S tem želimo akustičnofonetične meritve govorjene slovenščine približati tudi jezikoslovcem brez naprednega programerskega znanja in jim omogočiti, da na svoja raziskovalna vprašanja odgovarjajo s pomočjo preglednega in prosto dostopnega gradiva.

Prispevek je nastal v okviru raziskovalnega projekta Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik (MEZZANINE, J7-4642) in raziskovalnega programa Jezikovni viri in tehnologije za slovenski jezik (P6-0411), ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

Literatura

Gonzalez, S., Grama, J. in Travis, C. E. (2020). Comparing the performance of forced aligners used in sociophonetic research. *Linguistics Vanguard*, *6*(1). https://doi.org/10.1515/lingvan-2019-0058

Horjak, L. (2019). Akustična analiza polglasnika: Ob osemdesetletnici izida Orisa slovenskega knjižnega izgovora Franceta Bezlaja. V H. Tivadar (ur.), *Slovenski javni govor in jezikovno-kulturna (samo)zavest* (Obdobja 38, str. 39–47). Znanstvena založba Filozofske fakultete. https://doi.org/10.4312/Obdobja.38.39-47

Horjak, L. (2025). *Fonetično-fonološka načela kodifikacije sodobnega slovenskega govorjenega jezika* [Doktorska disertacija]. Univerza v Ljubljani.

Jurgec, P. (2005). Formant frequencies of standard Slovene vowels. *Govor*, *22*(2), 127–144.

Križaj, J., Dobrišek, S. in Žganec Gros, J. (2025). *Forced-alignment-sl 1.0: Forced alignment of speech for Slovene*. Fakulteta za elektrotehniko, Univerza v Ljubljani. https://github.com/jan3zk/forced_alignment

Križaj, J., Žganec Gros, J. in Dobrišek, S. (2024). Utilizing forced alignment for phonetic analysis of Slovene speech. V Š. Arhar Holdt in T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika: Zbornik konference* (str. 557–566). Inštitut za novejšo zgodovino. https://hdl.handle.net/20.500.12556/RUL-164240

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M. in Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. V *Interspeech 2017*. ISCA. https://doi.org/10.21437/Interspeech.2017-1386

Petek, B., Šuštaršič, R. in Komar, S. (1996). An acoustic analysis of contemporary vowels of the standard Slovenian language. V *Proceedings of fourth international conference on spoken language processing. ICSLP '96* (str. 133–136). IEEE. https://doi.org/10.1109/ICSLP.1996.607055

Pravopis 8.0: *Pravila novega slovenskega pravopisa za javno razpravo*. (b. d.). Založba ZRC, ZRC SAZU. https://pravopis8.fran.si

Robida, N. in Križaj, J. (2026). *Phonetic segmentation and acoustic measurements of spoken Slovene SloPhonSeg 1.0* [Korpus]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/2209

Robida, N., Čibej, J. in Krek, S. (2024). Strojni razrez posnetkov iz korpusa govorjene slovenščine Gos 2.1 in fonetične meritve. V S. Štumberger (ur.), *Predpis in norma v jeziku* (Obdobja 43, str. 267–274). Založba Univerze v Ljubljani. https://doi.org/10.4312/Obdobja.43.267-274

Srebot-Rejec, T. (1988). *Word accent and vowel duration in standard Slovene*. Peter Lang. https://doi.org/10.3726/b12733

Tivadar, H. (2004). Fonetično-fonološke lastnosti samoglasnikov v sodobnem knjižnem jeziku. *Slavistična revija*, *52*(1), 31–48.

Toporišič, J. (2004). *Slovenska slovnica*. Obzorja.

Toporišič, J. idr. (ur.). (2001). *Slovenski pravopis*. Založba ZRC, ZRC SAZU. https://www.fran.si

Verdonik, D., Bizjak, A., Žgank, A., Bernjak, M., Antloga, Š., Majhenič, S., Čakš, P., Pucer, M., Cvetko, M., Zelenik, M., Pavlič, J., Dobrišek, S., Križaj, J., Strle, G., Ivanovska, M., Grm, K., Bajec, M., Lebar Bajec, I., Jelovšek, T., … Bordon, D. (2023a). *ASR database ARTUR 1.0 (audio)*. Slovenian language resource repository CLARIN.SI. https://www.clarin.si/repository/xmlui/handle/11356/1776

Verdonik, D., Dobrovoljc, K., Rupnik, P., Ljubešić, N., Majhenič, S., Čibej, J. in Schmidt, T. (2024). *Training corpus of spoken Slovenian ROG 1.0* [Korpus]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1992

Verdonik, D., Zwitter Vitez, A., Zemljarič Miklavčič, J., Krek, S., Stabej, M., Erjavec, T., Potočnik, T., Sepesy Maučec, M., Majhenič, S., Žgank, A., Bizjak, A., Gril, L., Dobrišek, S., Križaj, J., Bajec, M., Lebar Bajec, I., Jelovšek, T., Trojar, M., Bernjak, M., … Rupnik, P. (2023b). *Spoken corpus Gos 2.1 (transcriptions)* [Korpus]. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1863

Volk, J. (2018). Polglasnik v slovenskem spontanem govoru prebivalcev Obale. *Jezik in slovstvo*, *63*(1), 19–37. https://doi.org/10.4312/jis.63.1.19-37

Zemljak Jontes, M., Kačič, Z., Dobrišek, S., Žganec Gros, J. in Weiss, P. (2002). Računalniški simbolni fonetični zapis slovenskega govora. *Slavistična revija*, *50*(2), 159–169.

Žganec Gros, J., Mirtič, T., Romih, M. in Ahačič, K. (2022). *Slovar izgovarjav OptiLEX*. Založba ZRC, ZRC SAZU. https://doi.org/10.3986/9789610506720

Branje večkodnih hiperbesedil

DOI: 10.4312/Obdobja.45.343-350

Martina Rodela, Pedagoška fakulteta, Univerza na Primorskem, Koper; martina.rodela@upr.si
Sonja Starc, Pedagoška fakulteta, Univerza na Primorskem, Koper; sonja.starc@upr.si

Pismenost v sodobnosti zahteva tudi zmožnost branja večkodnih hiperbesedil. Prispevek obravnava primer branja na vzorcu 59 študentov. Ugotavlja, katere dele vsebine na ekranu dojemajo kot besedilo; kako povezujejo semiotske kode ter kako in koliko jih uporablja hiperpovezave za pridobivanje dodatnih informacij.

večrazsežna pismenost, večkodno besedilo, hiperbesedilo, e-časopis, tehnične omejitve

In the contemporary era, literacy necessitates the ability to read multimodal hypertexts. This article examines a reading case study based on a sample of fifty-nine university students. It investigates which components of screen-based content the participants perceive as text, as well as how and to what extent they use hyperlinks to acquire additional information and make inferences.

multiliteracy, multimodal text, hypertext, e-newspaper, technical constraints

1 Uvod

V zadnjih desetletjih se s hitrim razvojem tehnologije, prevladujočo uporabo večkodnih besedil (iz več semiotskih kodov) in porastom rabe umetne inteligence (UI) način človeškega komuniciranja močno spreminja. In s tem tudi besedilo, saj se to kot semantični pojav oblikuje prav v procesu sporazumevanja (Halliday 2002; Halliday in Matthiessen 2004). Sodobne elektronske tehnologije omogočajo tvorbo pomena iz več različnih semiotskih sistemov tudi v pisnem sporazumevanju. Novosti se pojavljajo predvsem v pomenjenju (tvorjenju pomena), saj je treba besedilni pomen razbrati iz interakcije pomenov vseh uporabljenih semiotskih kodov v nekih okoliščinah (Kress in van Leeuwen 2004), zbranih znotraj nekega okvira; prostorskega pri vidnem kanalu, časovnega pri slušnem.

Omejujemo se na večkodna besedila, posredovana po vidnem kanalu, pri katerih se spremeni bralna pot. Če je v verbalnem besedilu linearna, postane v večkodnem holistična oz. bolj pragmatična, slikovni in kinetični sistemi dopuščajo več možnih vstopnih mest v besedilo. Tako postane tudi pomenjenje bolj dinamično, odprto, odvisno od bralčevih povezav uporabljenih znakov različnih sistemov. Mediji prenosa sporočil, po vidnem ali slušnem kanalu ali obeh hkrati, pomembno vplivajo na odnos med bralcem in besedilom ter posredno tudi na naslovnikovo razbiranje pomena. Bateman (2008: 11) ugotavlja, da lahko bralec v knjigi obrača strani, da se vrne k neki informaciji, zapomni

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

si mesto zapisa (npr. v zgornjem odstavku pred koncem poglavja), v spletnem besedilu ne more tako listati, s kurzorjem se mora pomikati navzdol/navzgor ali klikati na hiperpovezave. Vse te bralcu ponujene možnosti obeh kanalov lahko pomembno vplivajo na razumevanje verbalnega in neverbalnega v različnih žanrih. Se pa spletni medij bolj povezuje z večkodnimi besedili. Iz analize korpusa večkodnih besedil je Bateman (2008: 15–38) izpeljal vzorec, kako se iz družbenih praks oblikujejo večkodni žanri s pomočjo tehnologij, orodij, in ga poimenoval model večkodnih žanrov (GeM model). Besedila so ustvarjena in se prenašajo na določeni fizični podlagi (angl. *canvas*), ta je lahko papir izbranega formata (debeline, vpojnosti itn.) ali ekran (neke velikosti in resolucije) itn. Katerokoli podlago izberemo, vsaka prinaša nekatere omejitve, npr. velika prosojnost papirja zahteva črke določenega sloga in velikosti, da ne presevajo na naslednjo stran. Poleg podlage omejitve pri oblikovanju besedila prinašajo tudi tehnična izvedba (omejitev znakov/besed, barva, velikost grafičnih prvin, časovni roki, cena barvnega tiska, primoranost vključevanja reklam), dostopnost in razumljivost besedila (spletni brskalnik, možnost spremembe oblike pri prenosu dokumenta; (pre)zahtevnost besedila …). Pri tiskanem in elektronskem besedilu prihaja do različne zaznave tudi pri pojmu stran. Z natisnjeno stranjo bralec vzpostavi stik kot s predmetom zaznave na svojem vidnem polju in ga sprejme kot videno dejstvo. Pri digitalnem besedilu postane razumevanje strani odvisno od oblikovanja in interpretacije, kaj stran je. Vsekakor pa je bralna pot odvisna od poudarjenih elementov v besedilu, ki najprej pritegnejo bralčevo pozornost, v smislu predpozornostnega zaznavanja (Bateman 2008: 60).

Pri razbiranju pomena uporabljenih znakov pri vidnem kanalu in pri bralčevem opomenjanju, dojemanju besedilnega pomena, je zelo pomembno njegovo zaporedje zaznavanja in pomenskega povezovanja znakov. Pri branju ima ključno vlogo premikanje očesnega zrkla, ki se spontano odziva na poudarjene, izstopajoče dele besedila (barve, velikost, oblike, premikanje (kinetičnost) itn.). Yarbus (1967, v Starc 2015: 231) dokazuje, da je zaznavanje znaka in njegovega pomena pogojeno s fiziološko lastnostjo očesnega zrkla, in sicer z njegovimi hitrimi gibi od ene fiksacijske točke do druge (ta razdalja je sakada). Kopičenje sakad na določenih delih besedila/slike pomeni, da bralec temu posveti več pozornosti. Omenjena dejstva se močno izražajo prav pri branju večkodnih besedil. Taka so tudi e-besedila, npr. hiperbesedila – spletna besedila s povezavami na druga besedila istega spletnega mesta ali zunaj njega (Djonov 2005, v Zhao 2010). Gledano s tehničnega vidika so ta sestavljena iz blokov besedil in elektronskih povezav med njimi (Ferjan 2017: 15).

Spletna besedila niso nujno le besedila nepretrganega toka, ki zahtevajo linearno branje, temveč kolonije; to so krajša besedila, ki jih tematsko združuje naslov, in jih beremo poljubno, glede na lastno zanimanje (npr. časopis, horoskop, leksikon …). Dopuščajo torej pragmatično branje. Kranjc (2003: 436) ugotavlja, da so spletna besedila praviloma oblikovana z jezikovnimi in nejezikovnimi znaki, da bi se čim bolj približala uporabniku. Tudi Scacco in Muddiman (2015, v Kalin Golob 2025: 565) trdita, da preobilje informacij, dostopnih na različnih nosilcih (računalnik, telefon itn.),

povzroča tekmo za naslovnike. Zato jih morajo naslovi pritegniti, zlasti če so spletni mediji komercialno odvisni od klikov.

Da zahteva sodobna komunikacija tudi drugačno razumevanje pismenosti, je že leta 1996 zapisala skupina jezikoslovcev New London Group (NLG) v svojem manifestu o večrazsežni pismenosti (angl. *multiliteracies*). Ta upošteva dinamičnost družbenih pojavov in z njimi etično komunikacijo, usmerjeno k humanemu sobivanju med ljudmi in med človekom in naravo, kar naj bi pogojevalo potrebo po vpeljevanju učenja drugačnih razsežnosti branja v šole (Law in Matthiessen 2002). Pomembna sestavina večrazsežne pismenosti (Grosman 2011) je večkodna (New London Group 1996; Starc 2011), ki predstavlja zmožnost dinamične rabe semiotskih sistemov v večkodnih besedilih za dosego ustreznega družbenega in kulturnega delovanja posameznika.

Branje postane torej bolj odprto, odvisno od tega, katerim znakom posveti bralec več pozornosti, iz katerih razbira pomen, katere pa celo zanemari.

V nadaljevanju ugotavljamo, kako hiperbesedilo, kot pri pouku še neobravnavano vrsto e-besedila, berejo študenti, ki pa so že pridobili nekaj osnovnega vedenja o večkodnih besedilih.

2 Študija primera

2.1 Namen in raziskovalna vprašanja

Namen študije je ugotoviti, kako študenti 1. in 2. letnika razrednega pouka razbirajo informacije iz večkodnih hiperbesedil, 1) katere znakovne sisteme v večkodnem besedilu razumejo kot nosilce pomena, 2) kateri del ekrana zavzema brano besedilo in kako dolgo je, 3) kako in koliko jih uporablja hiperpovezave za pridobivanje dodatnih, izvenbesedilnih informacij kot potencialnih inferenc. Zanima nas tudi 4) bralna pot. Vzorčno besedilo analize je prispevek V. U., I. Š. in fotografa B. Samca o Delovi osebnosti leta 2025 N. Kovač, objavljen 14. 1. 2026 na Delo.si.[1]

2.2 Metoda

Metoda je kvalitativna analiza podatkov iz vprašalnikov. Študenti 1. letnika (25) so jih izpolnjevali elektronsko v spletni učilnici (na telefonih, tablicah, računalnikih), študenti 2. letnika (34) pa so namesto spletne učilnice večinoma izbrali natisnjeni vprašalnik.

Vprašalnik sestavljata sklopa nalog A in B, reševali so ju zaporedoma. Del A vsebuje pet nalog, povezanih z iskanjem podatkov v danem hiperbesedilu (Tabela 1), del B tri, nanašajoče se na obseg in dolžino strani e-časopisa in bralno pot (Tabela 2).

Za lažje razumevanje rezultatov opisujemo stran na ekranu. Na vrhu ekrana na sredini je zapis Delo, pod njim modra pasica vodoravno od leve do desne strani ekrana združuje vse prikazano na njem v e-časopis. Tekoča stran, ki drsi po nepremičnem ozadju z reklamami na levi in desni, je premična navzdol po sredini ekrana, na njej je časopisno besedilo pomaknjeno v levo, na desni strani so prikazane nepremične

1 https://www.delo.si/delova-osebnost-leta/nocoj-slavnostna-razglasitev-delove-osebnost-leta-2025

povezave na druga besedila. Na računalniku in tablicah je analizirano večkodno hiperbesedilo pomaknjeno v levo, na telefonu na sredino.

Kot vstopna spletna stran e-časopisa so pojmovani vsi podatki, prikazani ob prvem kliku na povezavo e-časopisa.

Vsebinski odgovori študentov so kodirani po kriterijih, iz katerega semiotskega koda je bil razbran zapisani pomen (iz besednega, slikovnega ali kinetičnega) na vstopni spletni strani na ekranu, dolžina strani in podlaga (po Batemanu) branega besedila ter raba ponujene hiperpovezave. Informacije hiperpovezave dobijo v procesu branja vlogo inferenc, ki jih za razumevanje in ustvarjanje koherentnosti besedila kot izvenbesedilne informacije bralec vnese v osnovno besedilo ob nekem leksikalnem signalu (Hoey 2001: 43–49; Shiro 1994: 167–178). V tem primeru je avtor oblikoval besedni signal kot povezavo za pridobitev inferenc.

3 Rezultati in razprava

Nalogo A je reševalo 59 študentov, od teh jih 50 povzema z naslovom napovedano temo (Tabela 1, a) le iz naslova, šest jo razširi z vsebinami iz tekočega besedila, a vse te informacije izhajajo iz besednega (100 %), en odgovor (o večkodni pismenosti in njenem pomenu v družbi) se ne nanaša na vprašanje, dva nista odgovorila. Povzetki (Tabela 1, b) izpostavljajo različne vsebine, največ nagrajenko N. Kovač, Delo kot podeljevalca nagrade. 56 (95 %) študentov podatke v povzetku črpa iz tekočega besedila, pomena besednega, od teh eden tudi iz usmerjevalnih leksemov (podnapisa) pod fotografijo, te pa kot vira vsebine ne vključi nihče. Eden analizira besedilo z orodjem vizualne slovnice. Besedilo je večkodno (»oboje«; Tabela 1, c), jih zapiše 50 (85 %); da je iz jezikovnega sistema, imenovanega tudi tekst, jih zapiše šest, dva ne ločita med semiotskimi kodi, kanali in žanri (*jezikovni*, *vidni*, *zvočni*, *digitalni*; *izsek iz intervjuja in novica*). Podatke o nominirancu G. Božiču (Tabela 1, č), naslove filmov in njihovo žanrsko pripadnost ponuja hiperbesedilna povezava. Da je filmski režiser, zaljubljen v sadovnjak, izvemo iz besednega na vstopni strani, ime je zapisano polkrepko in kot hiperpovezava. Podatek *mojster filma* so študenti prebrali na hiperpovezavi. Ta ponuja naslove filmov in pripadajoče žanre; največkrat je imenovan (Tabela 1, d) film *Zgodbe iz kostanjevih gozdov* (43), šestkrat *Navadna hruška* kot *dokumentarni film*, osemkrat *Zgodbe iz kostanjevih gozdov*, *Navadna hruška*, *Mož, ki ni mogel molčati*. Študenti (83 %) so uporabili hiperpovezavo, a nekateri *Zgodbam iz kostanjevih gozdov* netočno pripisujejo žanra kriminalka, politično-gospodarska drama (Tabela 1, e). Kot avtorja oz. avtorje besedila (besednega) 29 študentov prepozna inicialke V. U., I. Š., nihče kot avtorja ne omeni fotografa, čeprav jih 50 zapiše, da je besedilo sestavljeno iz verbalnega in slikovnega sistema (Tabela 1, f).

Vsebina odgovorov	Vsebina odgovora povzeta …	Št. štud. = 59
a) Tema naslova	**samo po naslovu:** Delova osebnost leta 2025 je Nika Kovač	50
	po celem besedilu: razglasitev Delove osebnosti leta; predstavitev zmagovalke, nominirancev; zgodba o N. Kovač; dosežki posameznika	6
b) Povzetek vsebine	**po besednem** Delova nagrajenka Nika Kovač slavnostna podelitev Delovih nagrad prisotni še drugi nominiranci	55
c) Semiotski siste-m/i/a, ki tvori/jo/ta omenjeno besedilo	**jezikovni**/tekst/verbalni sistem	6
	slikovni/likovni/podobe/slike/vizualni (odebeljene črke, poševne črke, velikost pisave, naslovi)	2
	oboje	50
č) Gregor Božič	**po vstopni strani, besednem**: filmski režiser, mojster filma	52
d) Naslov vsaj enega njegovega dela	**po hiperpovezavi, besednem**: *Zgodbe iz kostanjevih gozdov* idr.	43
e) Žanr	drama: družinsko, politično-gospodarske drame, kriminalka	27
	film: dokumentarno-igrani kratek film	
f) Ime avtorja ali avtorjev besedila	V. U., I. Š.	29
	fotograf Blaž Samec	0

Tabela 1: Rešitve naloge A.

Naloga B preverja, kolikšen del ekrana kot podlage (gl. Tabelo 2) zavzema stran e-časopisa, vtis o dolžini te strani in bralno pot. Nalogo je reševalo 57 študentov. Da se stran (Tabela 2, a) razteza čez ves ekran, odgovarja 40 študentov, 17 pa, da se čez sredinski del ekrana. Nekateri enačijo stran e-besedila in stran članka (*besedilo se mi prikaže čez cel ekran, na robu so predlagani drugi članki*), nekateri pa natančno ločijo med stranjo e-časopisa in članka (*stran časopisa čez cel ekran, vsebina članka pa obsega 3/4*). Za dolžino strani (Tabela 2, b) 15 študentov oceni, da je *precej dolga in zavzema veliko površino, ne prikaže vsega naenkrat, po njej se pomikamo po ekranu navzdol*. Od teh jih šest dolžino strani e-časopisa (Tabela 2, b) oceni s številom odstavkov. Pet jih dolžino e-besedila (11 oz. 7–8 strani) meri z ekranom kot stranjo, to ugotovijo s preštevanjem, kolikokrat se vsebina na ekranu popolnoma zamenja. Tako vpeljejo novo mersko enoto, ki pojem strani še bolj razprši. 13 študentov ugotavlja, da je stran dolga, razlogi za dolžino pa: *velika pisava*, *daljši žanr*, *oglasi*, *dodatne informacije*, *fotografije*. Ostale (24) ugotovitve so razpršene. Izpostavljamo eno, ki kaže na negotovost zaradi novega, očitno še premalo usvojenega pojava: *dolžino težko presojam, ker še nimam usvojenega občutka za ugotavljanje dolžine e-strani.*

Vsebina odgovorov	Odgovori študentov	Št. štud. = 57
a) Velikost strani e-časopisa b) Dolžina strani	**Velikost strani** **čez cel ekran** (z vmesnimi reklamami, z besedilom in drugimi elementi; na telefonu) **le sredinski del ekrana** **Dolžina strani** pribl. 20 odstavkov oz. do 10 vrstic; precej dolga; ni vse prikazano hkrati e-stran je dolga pribl. 8 strani, to sem ugotovila tako, da sem z drsenjem po ekranu navzdol preštela, kolikokrat se je vsebina na ekranu zamenjala precej dolga: zaradi velike pisave, žanra – daljšega besedila, dolge strani zaradi dolge novice zelo dolga zaradi oglasov/dodatnih informacij/fotografij zelo dolga (verjetno zaradi majhnega ekrana)	 40 17 15 5 13
c) Koraki bralne poti: s čim so začeli in kako po vrsti nadaljevali	**Začetek** naslov, podnaslov **Nadaljevanje** **po vrsti** (kot se pojavljata besedno in slikovno) **po vrsti** (kot se pojavljata besedno in slikovno), nato usmerjevalne lekseme pod sliko, prej spuščene; po vrsti, kjer se pojavi slika, naprej sliko in usmerjevalni leksem (podnapis); branje besednega in sprotno ogledovanje fotografij **le besedno**, vse ostalo izpusti ali se kasneje vrne k slikovnemu ali slike le preleti najprej **prebere slikovno**, nato besedno oz. linearno – besedno in slikovno najprej **preleti besedilo**, nato ga začne brati (besedno in slikovno)	 51 40 4 10 3

Tabela 2: Rešitve naloge B.

Na svoji bralni poti 51 študentov najprej prebere naslov, podnaslov. Iz odgovorov prepoznamo štiri različne bralne poti: a) največ jih bere linearno (40), po vrsti, kot se izmenjujeta verbalno in slikovno, štirje od teh odprejo in berejo tudi hiperpovezave, dva zanemarita vse reklame, eden upošteva le tiste, ki ga pozivajo. b) Štirje zapišejo, da berejo le besedno, od teh eden slikovno izpusti, eden se nanj vrne kasneje, dva ga preletita. c) Deset jih bere linearno (besedno in slikovno), od teh eden prebere slikovno, nato besedno. č) Trije besedilo preletijo, nato berejo linearno (besedno in slikovno).

Rezultati kažejo, da študenti še niso usvojili vedenja, kako se pomen vedno tvori v interakciji vseh uporabljenih kodov in kako se vstopa v večkodno besedilo. Večina (89 %) začne brati z naslovom in ohranja linearno branje (70 %) za besedno in slikovno. Spontanemu vodenju zakonitostim večkodnega besedila z vstopanjem vanj ob znaku, ki bralca najbolj pritegne, se prepustijo redki. Slikovnega ne razumejo kot enakovrednega semiotskega sistema jezikovnemu, temveč kot popestritev v besedilu, z vsebino nepovezan pripomoček lažjega branja. To kaže tudi neupoštevanje fotografa kot soavtorja besedila (Tabela 1, f). Vsi odgovori kažejo na razbiranje vsebine le iz jezikovnih znakov. Podobno ugotavlja raziskava o njihovem branju spletnih voznih redov (Rodela 2026).

Nekateri sicer najprej preberejo slikovno (18 %), a je to zavestna odločitev. Razlog je tudi podlaga e-besedila – ekran, ki besedila ne prikaže v celoti, temveč po delih. To omejitev tehnične izvedbe prepoznavajo tudi študenti. Še bolj moteča je primoranost vključevanja reklam, te pri branju največkrat ignorirajo. Nebesednih znakov kot enakovrednih tvorcev vsebine ne upoštevajo, a naučeno odgovorijo (85 %), da je besedilo sestavljeno iz dveh semiotskih kodov. Da študentom raba hiperbesedil ni tuja, kaže dejstvo, da jih je 83 % uporabilo hiperpovezavo, a mnogi niso med seboj povezali pravih podatkov. Pri povezovanju terminologije z novimi tehnologijami je razbrati nelagodje, negotovost pri pojmovanju strani vstopnega e-besedila in strani samostojnega besedila na njej. Zavedati se moramo, da je vstopno e-besedilo organizirano kot besedilo kolonija, sestavlja ga več sporočilno samostojnih enot, ne zahteva linearnega branja. Hkrati sami inovativno določijo še »tretjo« stran – površino ekrana. Študenti opazijo tudi spreminjanje oblike besedila glede na uporabljeno napravo, to največkrat povezujejo z velikostjo strani.

4 Ugotovitve

Čeprav so študenti nekaj osnovnih značilnosti večkodnih besedil (enakovrednost semiotskih znakov pri tvorbi pomena, bralna pot …) že pridobili, še vedno večina spontano bere vse linearno, sporočilo besedila razbira le iz besednega koda. Hkrati pa naučeno opredelijo brano besedilo za večkodno. Izstopa močan vpliv usvojenega linearnega branja besedil. To dejstvo opozarja na pomen osnovno- in srednješolskega usvajanja znanja in bralnih zmožnosti ter na potrebo po učenju večkodne pismenosti postopno, od osnovne šole naprej.

Pri ugotavljanju dolžine elektronskega besedila študenti inovativno vpeljejo lastne merske enote, ki kažejo povezavo s tiskanimi stranmi, število pomikov po podlagi – ekranu navzdol enačijo s številom »strani« e-besedila. Ponujene hiperpovezave je večina uporabila, a mnogi so nekatere podatke med seboj povezali površno in napačno. Kot uporabniki digitalnih tehnologij spletna večkodna besedila berejo z občutkom suverenosti, a jim pri tem umanjka holistično razumevanje besedil.

Literatura

Bateman, J. A. (2008). *Multimodality and genre: A foudation for the systemic analysis of multimodal documents*. Palgrave Macmillan.

Ferjan, M. (2017). *Poslovno komuniciranje*. Univerza v Mariboru. https://dk.um.si/Dokument.php?id=131694&lang=slv

Grosman, M. (2011). Večrazsežna pismenost – izziv sedanjosti. V M. Cotič, V. Medved-Udovič in S. Starc (ur.), *Razvijanje različnih pismenosti* (str. 19–27). Annales.

Halliday, M. A. K. (2002). Text as semantic choice in social contexts. V J. Webster (ur.), *Linguistic studies of text and discourse* (str. 23–84). Continuum.

Halliday, M. A. K. in Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. Routledge. https://doi.org/10.4324/9780203783771

Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. Routledge.

Kalin Golob, M. (2025). Razvoj naslovja v slovenskih tiskanih in spletnih medijih. *Slavistična revija*, *73*(4), 555–570. http://www.dlib.si/?URN=URN:NBN:SI:doc-RNVCO8UO

Kranjc, S. (2003). Jezik v elektronskih medijih. V A. Vidovič Muha (ur.), *Slovenski knjižni jezik – aktualna vprašanja in zgodovinske izkušnje* (Obdobja 20, str. 435–446). Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete. https://centerslo.si/wp-content/uploads/2015/10/kranjc.pdf

Kress, G. in van Leeuwen, T. (2004). *Reading images: The grammar of visual design*. Routledge.

Law, L. in Matthiessen, C. (2022). *Revisiting Halliday (1990) »New ways of meaning: The challenge to applied linguistics«: What has changed and what still needs to be done?* [Preprint]. ResearchGate. https://www.researchgate.net/publication/362067399_2023_Revisiting_Halliday_1990_New_ways_of_meaning

Michelizza, M. (2015). *Spletna besedila in jezik na spletu: Primer blogov in Wikipedije v slovenščini*. Založba ZRC, ZRC SAZU. https://omp.zrc-sazu.si/zalozba/catalog/view/1310/5524/918-2

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, *66*(1), 60–93. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Rodela, M. (2026). *Krmarjenje po večkodnih hiperbesedilih: Primer branja spletnega mesta javnega železniškega prevoza pri študentih*. Univerza na Primorskem, Pedagoška fakulteta.

Shiro, M. (1994). Inferences in discourse comprehension. V M. Coulthard (ur.), *Advances in written text analysis* (str. 167– 178). Routledge.

Starc, S. (2011). Zmožnost dekodiranja večkodnih besedil kot sestavina besedilne pismenosti. V M. Cotič, V. Medved-Udovič in S. Starc (ur.), *Razvijanje različnih pismenosti* (str. 28–36, 498–499). Annales.

Starc, S. (2015). Decoding multimodal text by untrained readers: A message for mother tongue pedagogy. V S. Starc, J. Carys in A. Maiorani (ur.), *Meaning making in text: Multimodal and multilingual functional perspectives* (str. 223–244). Palgrave Macmillan.

Zhao, S. (2010). Intersemiotic relations as logogenetic patterns: Towards the restoration of the time dimension in hypertext description. V M. Bednarek in J. R. Martin (ur.), *New discourse on language: Functional perspectives on multimodality, identity, and affiliation* (str. 195–218). Continuum. https://www.researchgate.net/publication/285641583_Intersemiotic_relations_as_logogenetic_patterns_Towards_the_restoration_of_the_time_dimension_in_hypertext_description

Vzpon skritih anglizmov v slovenščini

DOI: 10.4312/Obdobja.45.351-358

Eva Sicherl, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; eva.sicherl@ff.uni-lj.si

Ob naraščajočem številu vidnih anglizmov v slovenskih besedilih, ki so že uveljavljeni in/ali leksikalizirani, vedno pogosteje pa rabljeni kot modne in prestižne sposojenke, v sodobni slovenščini lahko opazimo tudi vrsto skritih ali nevidnih anglizmov, katerih zgradba je očitno angleška. To so besede, besedne zveze in fraze, ki se pod angleškim vplivom vključijo v jezikovno rabo. V prispevku predlagamo naslednjo tipologijo nevidnih anglizmov: pomenske sposojenke, besedni, besednozvezni in stavčni kalki, morfo-skladenjski vzorci in besedotvorni vzorci.

skriti anglizmi, pomenske sposojenke, kalki, slovenščina

Along with an ever-growing number of visible Anglicisms in Slovenian texts, which are either established and/or lexicalized, but increasingly used as fashionable or prestige loans, contemporary Slovenian also uses a number of so-called invisible Anglicisms whose structure is clearly influenced by English. These are words, phrases, and phraseological units that have made their way into language use under English influence. This article suggests a possible typology of invisible Anglicisms: semantic loans, calques at the word, phrase, and sentence levels, morphosyntactic patterns, and word-formation patterns.

hidden Anglicisms, semantic loans, calques, Slovenian

1 Uvod

Prispevek prinaša pregled različnih tipov skritih anglizmov, ki se pojavljajo v sodobni slovenščini. Takšne nevidne sposojenke težko izsledimo, ker je njihova oblika pravzaprav slovenska, se pa vedno pogosteje pojavljajo tako v govorjeni kot zapisani slovenščini. V nadaljevanju bomo koncept skritega anglizma najprej definirali, v poglavju 3 pa predstavili možno tipologijo skritih anglizmov, ki se pojavljajo v slovenščini; razdelimo jih lahko v pomenske sposojenke (3.1) in besedne, besednozvezne in stavčne kalke (3.2, 3.3 in 3.4). Razprava se razširi na spremembe v morfo-skladenjskih vzorcih (3.5) in besedotvornih vzorcih (3.6), ki jim botruje vpliv angleščine. V zaključku omenjamo (kvarni) vpliv teh skritih anglizmov, saj slovenska govorjena in pisna besedila zaradi njih izgubljajo svojo idiomatičnost in naravnost, postajajo vedno bolj slogovno revna in uniformirana.

2 Skriti anglizmi

Vidne anglizme, tj. posamezne sposojenke iz angleščine, ki jih slovenščina prilagodi svojemu jezikovnemu sistemu, v našem jeziku opažamo nekje od poznega 18. stoletja

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

dalje, podobno kot v večini evropskih jezikov pa se njihovo število vztrajno povečuje. Povečuje pa se tudi število struktur in oblik, ki so v jezik očitno vnesene pod vplivom angleščine. Gottlieb (2020: 32) in številni drugi avtorji (npr. Fiedler 2017; Pulcini 2023; Witalisz 2025) zato predlagajo razširjeno pojmovanje anglizma, ki naj vsebuje ne le tiste angleške sposojenke, ki oblikujejo novo besedo, morfem ali lokucijo, temveč tudi kakršnokoli pomensko širitev obstoječe besede, morfema ali lokucije v jeziku prejemniku, ki je posledica stika z angleščino, kot tudi morfo-skladenjske inovacije, ki jih lahko pripišemo jezikovnemu stiku z angleščino in so posledica posnemanja angleških struktur v jeziku prejemniku (prim. Picone 1996: 3–4). Tu torej lahko pričakujemo skrite anglizme, ki pa jih le težko odkrivamo: takšne leksikalne enote, morfemi in besedne zveze so se vrinili v jezikovno rabo zaradi intenzivnega stika z anglo-ameriško kulturo in le jezikovno kultivirani uporabniki se zavedajo, da se nekaj čudno sliši, da bi oni to kako drugače povedali ali da generacija njihovih (starih) staršev nikoli ne bi rekla česa takega. Če torej definiramo skriti anglizem v slovenščini kot katerokoli prvino, ki kaže vpliv angleščine oz., še širše, vpliv anglo-ameriške kulture (gl. tudi Gottlieb 2020: 32), potem je očitno, da so skriti anglizmi že prodrli globoko v večino socialnih interakcij; slovenski govorci uporabljajo veliko anglo-ameriških navad/rekel itn., ne da bi se tega sploh zavedali. Zglede takšnih skritih anglizmov, kot jih navajamo v nadaljevanju, lahko zberemo le sporadično, kot se pač pojavljajo v različnih slovenskih besedilih; mnogi med njimi še niso zajeti v slovarjih in drugih jezikovnih priročnikih, sploh kadar presegajo leksikalno raven, njihovo pojavnost pa lahko preverjamo v korpusih in na spletu.

3 Tipologija skritih anglizmov: poskus klasifikacije

3.1 Pomenske sposojenke

Ko jezikovni uporabniki želimo zapolniti leksikalne vrzeli, si seveda lahko pomagamo z direktnim sposojanjem angleških izrazov, ki jih bolj ali manj uspešno prilagodimo slovenskemu jezikovnemu sistemu. Ta prenos pa je lahko tudi bolj zabrisan oz. skrit, kadar pomen neke angleške besede prevzamemo in ga dodamo že obstoječi slovenski besedi. V literaturi se je za tak tip sposojenke uveljavil izraz pomenska sposojenka ali pomenski kalk (prim. Pulcini 2023: 125). Ta pojav je najmanj opazen med vsemi tipi sposojenk, ker besednemu inventarju jezika prejemnika ne doda nove besede, le nov pomen. V slovenščini take dodatne pomene pogosto najdemo pri besedah, ki so izvorno tuje, nova večpomenskost pa je posledica angleškega vpliva. Tak zgled je *ikona*, sposojenka grškega izvora iz 19. stol. in definirana kot ʻsveta podobaʼ v vzhodni cerkvi. Pod vplivom angleščine se je pomen razširil še na dva pomena: ʻsličica na zaslonu (elektronske) napraveʼ in ʻpredstavnik določenega področja popularne kulture, ki je širše znan in visoko cenjenʼ (Slovar slovenskega knjižnega jezika [SSKJ2]). Prvi je zaradi vsesplošne uporabe pametnih telefonov in računalnikov splošno znan, drugi je glede na prvotni pomen v angleščini pomensko nekoliko ožji in rabljen predvsem v novinarskem jeziku in nestrokovnih prevodih.[1] Slednji ima tudi že pridevniško obliko *ikoničen* (ʻki je za kaj

1 Angleški pomeni so preverjeni po HarperCollins (b. d.).

posebej značilen, kaj posebej nazorno prikazuje, ponazarja'), ki jo navaja *Sprotni slovar* (Krvina 2014–), Gigafida 2.2 (Center za jezikovne vire in tehnologije 2025) pa jo zazna od leta 1996 dalje.

Med glagoli je tak primer *nasloviti*, ena modernih inovacij v javnem nastopanju. SSKJ[2] navaja naslednje (splošne/splošnojezikovne) pomene: ʻnapisati naslov prejemnika pošiljke', ʻizražati usmerjenost dejanja', ʻdati naslov', ʻv nagovoru, ogovoru uporabljati naslov'. Vse pogosteje pa je ta glagol uporabljan v pomenu ʻobravnavati', kar je očiten vpliv angleškega *address* (*an issue*). Medtem ko Jezikovna svetovalnica (Dobrovoljc 2018) rabo v tem pomenu odsvetuje in predlaga kar nekaj boljših alternativ (*ukvarjati se z*, *govoriti o*, *osredotočiti se na*; *posvetiti se*, *lotiti se*, *preprečiti*, *odpraviti*), se *nasloviti* že pojavlja tudi v govoru izobražencev.

Primer pridevnika, ki je razširil svoje pomensko polje, je *patetičen*. Kot mnoge sposojenke iz klasičnih jezikov je *patetičen* v slovenščino prišel preko nemščine v 19. stoletju (prim. Snoj 2015), njegov pomen pa je v celoti enak nemškemu *pathetisch*, tj. ʻpretirano navdušen, vznesen', in se navadno veže s samostalniki kot *govor*, *slog*, *vedenje*, *glas* ipd. Tudi angleščina si je sposodila grški pridevnik (iz samostalnika pathos), vendar je pridevnik tam razvil drugačen pomen (ʻbeden') in je pogosto rabljen slabšalno s samostalniki kot *pathetic excuse/sight* (ʻbeden izgovor/prizor'). Očitno med parom *pathetic – patetičen* ni nikakršnega pomenskega prekrivanja, vendar vedno pogosteje v slabih oz. nestrokovnih prevodih (npr. v podnapisih) vidimo vdor angleške rabe (*What pathetic excuses. – Kakšni patetični izgovori.*; namesto: *Kakšni bedni izgovori.*). Očitno je mlada generacija Slovencev take dialoge že tako usvojila, da prvotnega pomena pridevnika *patetičen* (ʻsolzav, pretirano vznesen') sploh ne poznajo več.[2] Slovarji za zdaj navajajo le prvotni pomen, torej je par *pathetic – patetičen* v angleško-slovenskem odnosu lažni prijatelj, medtem ko korpusi že vključujejo zglede rabe v novem pomenu. Stalna raba novega pomena v površnih prevodih (npr. v podnapisih ali sinhroniziranih reklamah) spodbuja novo rabo in jo uveljavlja. Ko takšna substandardna besedila vključujemo v korpuse in jih prevzemajo tudi besedila, narejena z umetno inteligenco, v govorcih otopi občutek za jezik, pridevnik *patetičen* pa bo najbrž postal še eden od večpomenskih pridevnikov z dvema različnima pomenoma.

3.2 Kalkiranje

Tvorjenje kalkov je eden pogostih načinov, kako zapolniti leksikalne praznine. O takšnih »reprodukcijah« je pisal že Weinreich (1953: 50–51). V teh primerih semantiko angleškega modela prenesemo v slovenščino, to pa se zaradi vpliva angleščine dogaja na praktično vseh področjih. Če torej sprejmemo premiso, da je angleški vpliv zadosten parameter za določanje skritih anglizmov, potem so praktično celotne terminologije modernih znanosti (npr. računalništvo, okolje itn.) izvorno angleške. Zaradi

2 To potrjujejo rezultati ankete v sklopu seminarske naloge (pri predmetu Kontrastivna analiza angleškega in slovenskega jezika v študijskem letu 2025–26) študentke N. Gornik kot gradivni vir; od 80 novomeških dijakov, starih med 15 in 18 let, niti eden ni prepoznal ustaljenega pomena pridevnika *patetičen*, 71 dijakov pa je gladko vzpostavilo prevodno ekvivalenco med parom *pathetic – patetičen*.

jezikovnotipoloških razlik med slovenščino in angleščino pa večina kalkov, tvorjenih v slovenščini, spremeni morfo-skladenjsko obliko, saj slovenščina še ostaja zvesta svoji bogati morfologiji. Kalki, v katerih je angleški vzorec dobesedno reproduciran, so navadno kombinacije pridevnika s samostalnikom (npr. *artificial intelligence – umetna inteligenca*; *green transition – zeleni prehod*) ali prislova s pridevnikom (npr. *politically correct – politično korekten*), med skrite anglizme pa lahko štejemo tudi številne zveze in kolokacije, pri katerih ob reprodukciji angleškega vzorca pride do določenih morfoloških sprememb (podobno npr. v italijanščini, gl. Pulcini 2023: 121–122); navadno je samostalnik v vlogi levega prilastka zamenjan s pridevnikom, možni pa so tudi drugi vzorci (npr. *side effect – stranski učinek*; *zero tolerance – ničelna toleranca*; *climate neutral – podnebno nevtralen*). Ponekod angleški samostalnik v vlogi levega prilastka ali prvega dela zloženke ali zveze v slovenščini postane desni prilastek v rodilniku ali tvori predložno zvezo (npr. *brain drain – beg možganov*; *weekend – konec tedna*; *coffee break – odmor za kavo*; *headhunter – lovec na glave*). Imamo pa tudi številne kalke, pri katerih nedvomno najdemo pomenski vpliv angleščine, vendar smo nove tvorjenke ustvarili iz slovenskega jezikovnega materiala z besedotvornimi procesi, npr. z izpeljavo (npr. *influencer – vplivnež*; *whistle-blower – žvižgač*).

3.3 Kalkirane besedne zveze

Kalkiranje se je razširilo od poimenovanja posameznih pojmov vse do oblikovanja določenih besednih zvez in kolokacij, zlasti v pogovornem jeziku. Pri nepazljivem izražanju so iz angleščine dobesedno prevedene celotne besedne zveze (pa tudi večje skladenjske enote), ki se postopoma uveljavljajo v slovenščini. V tem trenutku morda še nekoliko izstopajo kot nenavadne in tuje zveneče, s pogosto rabo pa bodo verjetno postale standardne in nezaznamovane.

Ena takih kolokacij je *zlomiti srce* (angl. *break one's heart*), ki izpodriva ustaljeno in morda nekoliko bolj zborno *streti srce*; kombinacija *zlomiti srce* v korpusu Gigafida 2.2 (Center za jezikovne vire in tehnologije 2025) že prevladuje, njeno rabo pa najverjetneje pospešuje tudi samostalnik *lomilec src* (iz angl. *heartbreaker*). Podobno se uveljavlja zveza *odpiralni čas* (angl. *opening time*), glede na podatke iz Gigafida 2.2 (Center za jezikovne vire in tehnologije 2025) v rabi od leta 1995 dalje; tradicionalno je bil čas odprtosti lokalov in uradov označen z *odprto* ali *delovni čas*. V lokalih, trgovinah in uradih tudi ne uporabljajo več pozdrava *na svidenje*, pač pa vedno pogosteje *lep dan* (iz angleškega *Have a nice day.*).

3.4 Kalki na stavčni ravni

Angleški vpliv je viden tudi na ravni stavkov in povedi; tu so celotni stavki prevedeni dobesedno. Na televiziji lahko vidimo reklamo za piškote Oreo, kjer je uporabljen stavek *smo vsi na isti strani*, ki je dobesedno kalkiran iz angleškega idioma *we are all on the same page*, kar pomeni, da se o nečem vsi strinjamo. O metaforičnem pomenu kalka *smo vsi na isti strani* sicer gledalec lahko le ugiba iz situacije na ekranu, če pa bo stavek slišal neštetokrat, lahko postane čisto sprejemljiv in celo preide v aktivno rabo,

kot se je zgodilo s pridevnikom *patetičen* v novem pomenu. Še ena fraza, ki si utira pot v slovenščino, je *ni zastonj kosila* (angl. *there is no such thing as a free lunch* v več variantah, po ameriškem ekonomistu Miltonu Friedmanu),[3] ki že nadomešča *nič ni zastonj* ali *vse se plača* ali *vse ima svojo ceno* (prim. Meterc 2020–). Tudi Gigafida 2.2 (Center za jezikovne vire in tehnologije 2025) potrjuje povečano rabo te rečenice vse od leta 1991 dalje. Kalkirana je tudi fraza *za vašo informacijo* (angl. *for your information*), ki nadomešča bolj slovenski *če vas (že) zanima*. Mlada generacija Slovencev že pogosto uporablja *Kako je bil tvoj dan?* (kalkirano iz *How was your day?*) namesto *Kako si se imel(a) danes?* (gl. tudi Košmrlj 2024: 120). Še en kalk se je pojavil kot slogan Slovenskega olimpijskega komiteja za Olimpijske igre 2024: *Imamo to, Pariz, prihajamo!* Kalk *imamo to* ali *imaš to* je dobesedni prevod angleškega *we/you've got this* in bi ga seveda lahko nadomestili z *zmoremo* ali *zmoreš*. V novinarskih besedilih tudi vedno pogosteje najdemo frazo *ukrasti šov* (angl. *to steal the show*); verjetno bi bilo primerneje reči *pritegniti vso pozornost*, kadar na neki prireditvi nekdo izstopa.

3.5 Morfo-skladenjski vzorci

Angleški vpliv je jasno razviden tudi v nekaterih morfo-skladenjskih vzorcih, ki so se začeli pojavljati v slovenščini. Tako pridevniki, ki jih prevzemamo iz angleščine (pa tudi nekaterih drugih jezikov, npr. nemščine, francoščine ali italijanščine; prim. npr. Lencek 1982, tudi Sicherl 1999: 59) ostanejo brez obrazil in se ne ujemajo v spolu, sklonu in številu s samostalnikom (npr. *fer tekma*, *pink jopica*). Takšne strukture so odprle pot sodobnim samostalniškim zvezam, kot so *Erasmus študent*, *Hofer cena* in *Subway krompirček*, ki so direkten odsev ustreznih angleških izrazov. Ko namesto *Hoferjeva cena* uporabljamo *Hofer cena*, običajno zvezo pridevnika in samostalnika pravzaprav nadomeščamo s strukturo, ki že postaja nekakšna zloženka, morfološka kompleksnost zveze pa je posledično bistveno zmanjšana.

Drugačen tip angleškega jezikovnega vplivanja lahko opazujemo v kompleksnih levoprilastkovnih vzorcih. Zlasti v neformalnem slogu so takšni kompleksni premodifikatorji že kar pogosti, vpliv angleščine je jasen. Naj to ponazorimo z nekaj zgledi s spleta; tak vzorec se je začel s kodnim preklapljanjem (1), zgleda (2) in (3) pa kažeta, da so takšne kombinacije možne tudi že z izključno slovenskim leksikalnim materialom:

(1) No, to pa je šolski *by the book* primer poročanja o @AstaVrecko, če želite uspeti in dobiti proračunski denar na medijskih razpisih… (Požar 2025)
(2) Ni sicer povsem jasno, kdo je plačal za pijačo in jedačo, a glede na to, da je bil projekt kriminalno preplačan, je sklepati, da gre za »*hvala za davkoplačevalski denar*« potezo s strani izvajalcev. (I. K. 2025)
(3) Iz njega pa se veselo napajajo ministrstva, tudi tisto za »kulturo«, ki ga tako avtoritativno vodi tovarišica »*vsak dan v drugi obleki in plašču*« Asta. (Poniž 2025)

Vpliv angleščine pa lahko zaznamo tudi v spremenjeni vezljivostni shemi določenih slovenskih glagolov. Glagol *zmagati* npr. je zlasti v športnih besedilih pogosto rabljen

3 Prim. npr. https://en.wikipedia.org/wiki/No_such_thing_as_a_free_lunch.

kot navaden prehodni glagol (*zmagati tekmo/prvenstvo* – angl. *win a match/championship*), namesto kot predložni glagol (*zmagati na tekmi/prvenstvu*):

(4) Poleg tega je treba izpostaviti tudi preteklo sezono, zmagati prvenstvo na tujem je velik uspeh. (Mihalič 2023)

3.6 Besedotvorni vzorci

Že samo dejstvo, da so se določeni besedotvorni vzorci, ki jih narekuje angleščina, sploh začeli pojavljati v slovenščini, kaže, kako globoko v sam jezikovni sistem sega vpliv angleščine. Naj to ponazorimo s poimenovanjem *Sirijec*, ki ga glede na dostopne podatke korpus Gigafida 2.2 (Center za jezikovne vire in tehnologije 2025) zazna leta 1991 v članku iz *Mladine*; tudi zgledi, ki jih navaja Nova beseda (Inštitut za slovenski jezik Frana Ramovša ZRC SAZU b. d), so praktično vsi iz časopisja. Raba samostalnika *Sirijec* v besedilih je strmo narasla po letu 2011, kar je sovpadalo z arabsko pomladjo in posledičnimi migracijami Sircev v Evropo. Oblika *Sirijec* očitno posnema angleško poimenovanje *Sirian*, njena uveljavitev v slovenščini pa je plod neznanja ali nerazgledanosti slovenskih novinarjev oz. prevajalcev in njihovega neizdelanega jezikovnega občutka, ki slepo posnema angleško obliko. V slovenščini namreč že dolgo obstaja poimenovanje *Sirec* za prebivalca Sirije (gl. SP 2014 in prejšnje izdaje SP), ki sledi običajnemu besedotvornemu vzorcu (npr. *Slovenija* > *Slovenec* in ne **Slovenijec*; *Britanija* > *Britanec*). Vdor obrazila *-ijec* namesto *-ec* lahko z veliko gotovostjo pripišemo vplivu angleščine in že zgoraj omenjenim površnim prevodom.

Še opaznejši besedotvorni vpliv najdemo pri prekrivankah. Prekrivanje je v slovenščini dokaj nov pojav, saj je slovenski jezikovni sistem besedotvorno bolj nagnjen k izpeljavi ali kateremu drugemu besedotvornemu procesu. Danes pa prekrivanke vedno pogosteje najdemo v govorjeni slovenščini, v oglaševanju in kreativnem pisanju (npr. otroški in mladinski književnosti, na blogih in spletnih forumih; gl. tudi Sicherl 2018). Seveda je sam besedotvorni vzorec prišel iz angleščine in se postopoma udomačil med govorci slovenščine:

(5) angl. *bankster* > slov. *bankster* (banka/bankir + gangster),
(6) angl. *presstitute* > slov. *prestitut*/*prestitutka* (press + prostitut/prostitutka).

Takšne sposojenke iz angleščine s podstavami, ki jih brez posebnih težav prepoznajo in razumejo tudi slovenski govorci, so botrovale nastanku kasnejših slovenskih tvorjenk; vzorec se je prijel in ga danes uporabljajo kreativni govorci s slovenskim besednim materialom:

(7) *Zverjasec* < zverina + merjasec (prevod M. Dekleve za Gruffalo Julie Donaldson),
(8) *Brucšura* < bruc + brošura,
(9) *Mafijanković* < mafija + Janković.

Temu sorodno je tvorjenje z afiksoidi; sami prefiksoidi in sufiksoidi sicer pogosto izvirajo iz klasičnega grško-latinskega leksikalnega bazena, vzorec pa je angleški:

(10) *megastično* < mega- + fantastično,
(11) *blefsikon* < blef- + -sikon,
(12) *čokoholik* < čoko-lada + -holik,
(13) *Vičstock* < Vič + -stock[4] (dijaški festival Gimnazije Vič),
(14) *e-naslov* < e-lektronski + naslov.

Zanimivi so tudi primeri, ki jih lahko uvrstimo med t. i. grafične prekrivanke in jih najdemo zlasti v oglaševanju:

(15) *HOVtel* < hov + hotel (ime hotela za pse),
(16) *HOV du ju du* (ime pasjega salona v Ljubljani, istočasno besedna igra na angleško frazo *how do you do*).

4 Zaključek

Jezikovni stik med angleščino in slovenščino je po letu 1990 postal izjemno intenziven, zlasti zaradi vsesplošne razširitve digitalnih komunikacijskih sredstev. Kot v mnogih drugih jezikovnih skupnostih po svetu tudi mladi Slovenci angleščine ne usvajajo več kot tuji jezik, temveč kot drugi jezik (prim. npr. Gottlieb 2020: 169 za danščino). Ta dvojezičnost mladih slovenskih generacij je povzročila tektonske premike v njihovi vsakdanji jezikovni rabi, zlasti v neformalnem slogu (gl. npr. Košmrlj 2024). V jeziku mladih slovenskih govorcev je kodno preklapljanje med slovenščino in angleščino nekaj običajnega, ob tem pa v jezik vdirajo tudi novi vzorci v obliki skritih anglizmov, ki spreminjajo ali izpodrivajo bolj tradicionalne in izvorne slovenske vzorce. Slovenski bralci, gledalci in poslušalci pa so vsakodnevno izpostavljeni tudi naraščajoči količini besedil, ki so ustvarjena ali prevedena z umetno inteligenco, ob katerih Slovenci enostavno izgubljajo občutek za pravilnost jezika. Posledično se zabrisujejo fine razlike med obema jezikoma, jezikovna produkcija tovrstnih »dvojezičnih« govorcev pa postaja vedno manj idiomatična, vedno bolj poenostavljena in uniformna ter slogovno osiromašena, ne da bi se uporabniki jezika tega sploh zavedali. Rešitev seveda ni le v jezikovni politiki, ki je tako ali tako ne izvajamo, rešitev je v jezikovno osveščenih govorcih, ki se zavedajo, da je jezik del njihove identitete, in ga aktivno gradijo in ohranjajo.

Viri in literatura

Center za jezikovne vire in tehnologije. (2025). *Gigafida 2.2*: *Korpus pisne standardne slovenščine* [Korpus]. Viri CJVT. https://viri.cjvt.si/gigafida/

Dobrovoljc, H. (2018). Glagol »nasloviti« v pomenu ‘obravnavati’. *Jezikovna svetovalnica*. https://svetovalnica.zrc-sazu.si/topic/3175/glagol-nasloviti-v-pomenu-obravnavati

HarperCollins. (b. d.). *Collins Dictionary*. https://www.collinsdictionary.com/

Fiedler, S. (2017). Phraseological borrowing from English into German: Cultural and pragmatic implications. *Journal of Pragmatics*, *113*, 89–102.

Gottlieb, H. (2020). *Echoes of English. Anglicisms in minor speech communities – with special focus on Danish and Afrikaans*. Peter Lang.

4 Tudi v angleščini se *-stock* (iz prvotnega *Woodstock*) občasno pojavi v imenih festivalov.

I. K. (2025, 3. februar). *Tako se je elita mastila na otvoritvi železniške postaje, ki ste jo preplačali za 22 milijonov.* Nova24TV. https://nova24tv.si/takole-se-je-elita-mastila-na-otvoritvi-zelezniske-postaje-ki-ste-jo-preplacali-za-22-milijonov/

Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. (b. d.). *Nova beseda: Besedilni korpus slovenskega jezika* [Korpus]. http://bos.zrc-sazu.si/s_beseda3.html

Keber, J. (2015). *Slovar slovenskih frazemov*. Založba ZRC, ZRC SAZU. www.fran.si

Košmrlj, L. (2024). 'Vsaj en stavek povej normally': Anglicisms and code-switching in Slovene youth slang. *ELOPE, 21*(2), 107–124.

Krvina, D. (2014–). *Sprotni slovar slovenskega jezika*. Založba ZRC, ZRC SAZU. www.fran.si

Lencek, R. L. (1982). *The structure and history of the Slovene language*. Slavica Publishers.

Meterc, M. (2020–). *Slovar pregovorov in sorodnih paremioloških izrazov*. Založba ZRC, ZRC SAZU. www.fran.si

Mihalič, N. (2023, 1. junij). *Lana Ščuka: Pomembno je prepoznati, kdaj si je treba vzeti čas zase*. 24ur.com. https://www.24ur.com/sport/odbojka/lana-scuka_2.html

Picone, M. D. (1996). *Anglicisms, neologisms and dynamic French*. John Benjamins.

Poniž, D. (2025, 7. februar). *Slovenski kulturni praznik 2025: France v gatah in Asta v negližeju*. Plusportal. https://www.plusportal.si/2025/02/07/slovenski-kulturni-praznik-2025-france-v-gatah-in-asta-v-neglizeju/

Požar, B. [@BojanPozar]. (2025, 25. januar). *No, to pa je šolski by the book primer poročanja o @AstaVrecko, če želite uspeti in dobiti proračunski denar na* [Pripeta je slika][Objava]. X. https://x.com/BojanPozar/status/1879568301434794104

Pulcini, V. (2023). *The influence of English on Italian: Lexical and cultural features*. De Gruyter.

Sicherl, E. (1999). *The English element in contemporary standard Slovene: Phonological, morphological and semantic aspects*. Znanstveni inštitut Filozofske fakultete.

Sicherl, E. (2018). English influence on word-formational production in Slovene: The case of lexical blending. *Arbeiten aus Anglistik und Amerikanistik*, *43*(2), 141–158.

Slovar slovenskega knjižnega jezika: Druga, dopolnjena in deloma prenovljena izdaja. (2014). Založba ZRC, ZRC SAZU. www.fran.si

Snoj, M. (2015). *Slovenski etimološki slovar*³. Založba ZRC, ZRC SAZU. www.fran.si

Toporišič, J. (ur.). (2014). *Slovenski pravopis*. Založba ZRC, ZRC SAZU. www.fran.si

Weinreich, U. (1953). *Languages in contact: Findings and problems*. Mouton.

Witalisz, A. (2025). *English loan translations in Polish: Word-formation patterns, lexicalization, idiomaticity and institutionalization*. Peter Lang.

Žele, A. (2011). *Vezljivostni slovar slovenskih glagolov*. Založba ZRC, ZRC SAZU.

Vsak začetek ni (enako) težak: korpusni pogled na besedišče pri začetnih govorcih slovenščine kot neprvega jezika

DOI: 10.4312/Obdobja.45.359-370

Mojca Stritar Kučuk, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana;
mojca.stritarkucuk@ff.uni-lj.si

Prispevek s korpusno metodo, na podlagi analize podkorpusov KOST 10k, analizira razlike v produktivni rabi besedišča pri začetnih govorcih slovenščine kot neprvega jezika. Govorci (južno)slovanskih jezikov uporabljajo širši nabor in zahtevnejše leme, vendar pogosteje tvegajo z uporabo neustreznih neslovenskih izrazov. Nasprotno so neslovanski govorci močno omejeni na besedišče na ravni A1. Rezultati potrjujejo ključno vlogo jezikovne sorodnosti pri usvajanju besedišča.

slovenščina kot neprvi jezik, korpusno jezikoslovje, produktivna raba besedišča, jezikovna sorodnost, leksikalna raznolikost

Drawing on an analysis of the KOST 10k sub-corpora, this article employs a corpus-based methodology to examine differences in productive vocabulary use among beginner learners of Slovenian as an L2. The results indicate that speakers of (South) Slavic languages utilize a broader range of more complex lemmas, although they demonstrate a greater tendency to incorporate inappropriate non-Slovenian terms. Conversely, non-Slavic speakers remain heavily restricted to A1-level vocabulary. These findings confirm the pivotal role of linguistic relatedness in the process of vocabulary acquisition.

Slovenian as a nonnative language, corpus linguistics, productive vocabulary, linguistic relatedness, lexical diversity

1 Uvod

Govorci[1] slovanskih jezikov se slovenščino učijo lažje in hitreje kot govorci neslovanskih jezikov – to dejstvo je med učitelji slovenščine kot neprvega jezika utemeljeno na več generacijah izkušenj s poučevanjem. Na Centru za slovenščino kot drugi in tuji jezik Filozofske fakultete Univerze v Ljubljani,[2] ki je osrednja slovenska institucija za poučevanje slovenščine kot neprvega jezika, se denimo pri začetnem učenju slovenščine uporablja različne učbenike: za Neslovane učbenike, ki so uvrščeni na stopnjo A1 po lestvici SEJO (Jerman idr. 2015; Markovič idr. 2013; Petric Lasnik idr. 2018), za Slovane pa na A2 (Jerman idr. 2023).[3] Znanstvena pozornost je bila različnim skupinam

1 V tem prispevku je moški spol uporabljen kot slovnično nevtralen.
2 https://centerslo.si/ (dostop 16. 3. 2026)
3 https://centerslo.si/knjige/ucbeniki-in-prirocniki/#osnovna-stopnja (dostop 20. 7. 2026)

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

uporabnikov slovenščine kot neprvega jezika posvečena predvsem z vidika jezikovnega testiranja (prim. Ferbežar in Eniko 2022) in vpliva jezikovnega prenosa (prim. Balažic Bulc 2004; Nikolovski 2022), o razlikah med usvajanjem slovenščine pri začetnikih Slovanih in Neslovanih pa je bilo izvedenih razmeroma malo raziskav. Klemen (2025) je pri Slovanih potrdil precej višjo stopnjo poznavanja besedišča na receptivni ravni, ta prispevek pa je namenjen produktivni ravni. V njem sem s sodobno, korpusno metodo, s katero sem analizirala besedišče govorcev slovenščine kot neprvega jezika, potrdila, da so med slovansko in neslovansko govorečimi začetniki pri učenju slovenščine bistvene razlike. Analiza se je nanašala na pisno produkcijo odraslih govorcev, katerih prvi jezik ni slovenščina, nastalo v okviru prvih faz učenja slovenščine bodisi kot drugega jezika (torej jezika okolja) bodisi kot tujega jezika.

2 Metodologija

Izhajala sem iz gradiva korpusa slovenščine kot tujega jezika KOST.[4] Ker je prevelik za celostno analizo in ker besedila v njem niso idealno uravnotežena (Stritar Kučuk 2024b: 592), sem naredila izbor predvsem glede na stopnjo zahtevnosti, na kateri so besedila. V korpusu KOST so razdeljena na besedila južnih Slovanov začetnikov (45 % besedil), ostalih začetnikov (9 %), nadaljevalcev (37 % besedil) in izpopolnjevalcev (9 %). Stopnje niso določene po splošneje veljavnih kriterijih, npr. po lestvicah SEJO (prim. Klemen idr. 2026), temveč pragmatično, glede na splošno stopnjo skupine, v kateri se je tvorec besedila organizirano učil slovenščino, ali glede na stopnjo učbenika, ki ga je v času pisanja besedila uporabljal pri učenju. To oceno učitelj na tečaju ali urednik korpusa postavita *ad hoc* in torej ni absolutno zanesljiva. Kot sem že omenila, me v tem prispevku zanimajo besedila južnih Slovanov začetnikov in ostalih začetnikov; med prve se uvrščajo govorci bosanščine, črnogorščine, hrvaščine in srbščine (v nadaljevanju BČHS) ter makedonščine,[5] med druge pa – brez dvoma nekoliko preveč posplošeno – vsi ostali. Pri tem žal nisem mogla upoštevati morebitnega znanja ostalih slovanskih jezikov, ki ga lahko imajo neslovanski govorci in brez dvoma lahko zaznamuje njihovo jezikovno produkcijo v slovenščini.

2.1 Priprava podkorpusov KOST 10k

Analizo sem izvedla na delu korpusa KOST, imenovanem KOST 10k. Sestavljajo ga štirje po velikosti primerljivi podkorpusi (Tabela 1), pri katerih je spremenljivka prvi jezik tvorcev besedil. V podkorpusu KOST 10k NS- so besedila začetnikov Neslovanov, v KOST 10k S- pa začetnikov, govorcev vzhodno- ali zahodnoslovanskih jezikov. Jezikovni sistem slovenščine je slednjim zaradi sorodnosti jezikov razmeroma blizu, na ravni besedišča pa so razlike večje. KOST 10k M- vključuje besedila začetnikov, govorcev makedonščine. Obravnavam jih posebej, saj makedonščina ne pozna sklonov (Arizankovska 2009: 28), zato imajo njeni govorci pri usvajanju slovenščine na oblikoslovni

4 https://viri.cjvt.si/kost/sl/ (dostop 16. 3. 2026)

5 Govorci bolgarščine kot še enega južnoslovanskega jezika v analizo niso vključeni, saj je delež njihovih besedil v korpusu KOST zanemarljiv.

ravni razmeroma več težav od govorcev BČHS (prim. Nikolovski 2022; Stritar Kučuk in Pirih Svetina 2025). Preveriti želim, ali to velja tudi na ravni besedišča. V podkorpusu KOST 10k BČHS pa so besedila začetnikov, govorcev osrednjejužnoslovanskih jezikov. Gre za govorce slovenščini najbolj sorodnih jezikov, zato je pri njih pričakovati najmanj težav in najširši razpon tudi na ravni besedišča (prim. Klinar idr. 2022). Pri izboru sem pazila, da so bila vsa besedila napisana v nadzorovanih pogojih tvorjenja, tako da je mogoče za vsa zanesljivo trditi, da so jih tvorci napisali sami, brez zunanje pomoči (prim. Stritar Kučuk 2024a). Čeprav so vsa besedila spisi oz. eseji, sem skušala vključiti tematsko čim bolj raznolike tekste, tako da bi tudi uporabljeno besedišče lahko pokrivalo različna tematska polja.

Podkorpus	Št. pojavnic	Št. besedil	Prvi jeziki tvorcev
KOST 10k NS-	10.329	110	angleščina, arabščina, argentinska španščina, francoščina, italijanščina, japonščina, kitajščina, madžarščina, nemščina, nizozemščina, romunščina, španščina, turščina
KOST 10k S-	10.102	85	češčina, poljščina, ruščina, slovaščina, ukrajinščina
KOST 10k M-	10.220	78	makedonščina
KOST 10k BČHS	10.291	61	bosanščina, črnogorščina, hrvaščina, srbščina

Tabela 1: Podkorpusi, vključeni v raziskavo.

Analizo za ta prispevek sem izvedla na skupaj 40.942 pojavnicah. Ker je vzorec razmeroma majhen, rezultatov ni mogoče posploševati, temveč jih je treba razumeti kot pilotno študijo za širše zastavljene in bolj utemeljene prihodnje raziskave.

2.2 Priprava seznama lem

Z uporabo funkcije Wordlist v programu Sketch Engine[6] sem za vsak podkorpus ustvarila seznam uporabljenih lem, ga v tabelarni obliki izvozila v program Microsoft Excel in ga ročno prečistila. Izločila sem nerelevantne zadetke, predvsem ločila in osebne podatke, ki so v KOST-u prekriti z enotnimi kodami (Stritar Kučuk 2024c: 103).

2.3 Uvrščanje besed na stopnje po lestvici SEJO

Da bi enotno določila zahtevnost lem, ki so jih tvorci uporabili v besedilih vsakega podkorpusa, sem jih uvrstila na stopnje po mednarodno primerljivi lestvici *Skupnega evropskega jezikovnega okvira* (SEJO). Ta vključuje šest stopenj (Svet Evrope 2011): A1 (vstopna raven), A2 (vmesna raven), B1 (raven sporazumevalnega praga), B2 (višja raven), C1 (raven učinkovitosti) in C2 (raven mojstrstva). Prečiščeni seznam lem iz podkorpusov KOST 10k sem primerjala s seznamom jedrnega besedišča za slovenščino kot J2 (Klemen idr. 2023),[7] ki vključuje 5273 besed, 1214 na stopnji A1,[8] 1451 na stopnji A2 in 2608 na stopnji B1. Žal v njem ni lem na stopnjah, višjih od B1 (Klemen idr.

6 https://app.sketchengine.eu/ (dostop 19. 12. 2025)

7 http://hdl.handle.net/11356/1697 (dostop 19. 12. 2025)

8 Na seznamu jedrnega besedišča sta glede na število učbenikov za slovenščino, v katerih se pojavi določena beseda, ločeni stopnji A1 – jedro in A1 – širše (Klemen idr. 2022: 171). To ločevanje za pričujočo analizo ni pomembno, zato sem kategoriji združila v enotno stopnjo A1.

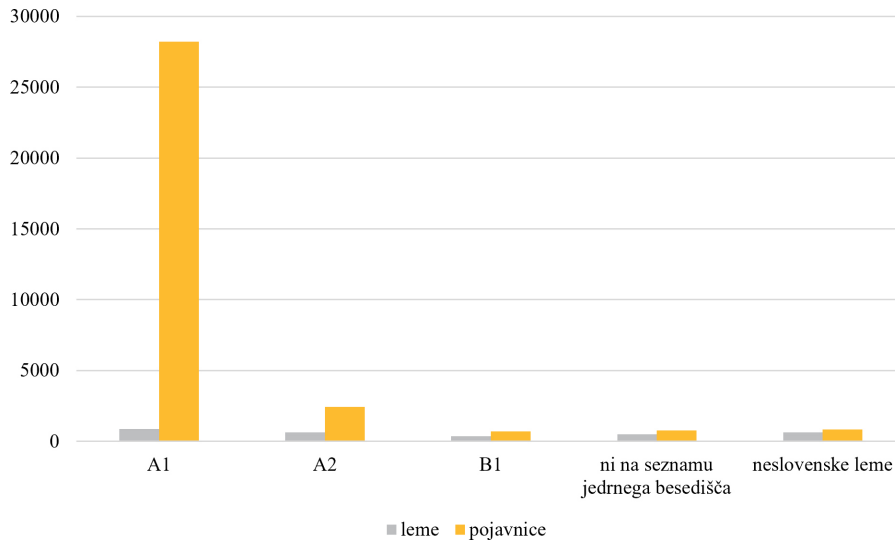
2022: 172). Pri določanju stopenj lem sem uporabila še dva dodatna vira, ki temeljita na konsenzu strokovnjakov, in v njiju poiskala leme, ki jih v seznamu jedrnega besedišča ni. Glavne številke in besede iz besedne družine, vezane na poimenovanje Slovenije (*Slovenija*, *Slovenec*, *Slovenka*; Pirih Svetina 2016: 8), sem v skladu z dokumentom, ki opredeljuje preživetveno raven (Pirih Svetina 2016), umestila na stopnjo A1, nekatere leme iz dokumenta, ki definira sporazumevalni prag za slovenščino (Ferbežar idr. 2004), pa na stopnjo B1.

Pri analizi sem bila pozorna na leme, ki niso vključene v seznam jedrnega besedišča, ter neslovenske besede. Pri teh pojavnicah sem najprej preverila, ali je prišlo do napake pri zapisovanju ali pregibanju leme oz. pri avtomatski lematizaciji, in te primere ročno popravila. Tako denimo avtomatski lematizator oblike *tečeti* ni uvrstil k lemi *teči*. Upoštevati je treba tudi, da je določene neslovenske oblike mogoče interpretirati na različne načine; pretekli deležnik *imal* bi lahko razumeli kot neustrezno obliko glagola *imeti* ali kot neslovensko besedo, torej leksikalno napako. Načeloma sem se pri tem držala napotkov za označevanje napak v korpusu KOST (Stritar Kučuk 2025), vendar se subjektivnega dejavnika pri tem ne da popolnoma izločiti. Umestiti sem morala tudi 91 lem, pri katerih je prišlo do napake zapisa (Stritar Kučuk 2025: 35–36), npr. *dovol* namesto *dovolj*. Ker je v tem prispevku vendarle v ospredju predvsem leksikalna zmožnost tvorcev, sem se odločila, da primere z manjšimi pravopisnimi napakami vseeno uvrstim v analizo (dve tretjini jih sodi na stopnjo A1). Preostale leme, ki niso bile vključene v seznam jedrnega besedišča ali na ostala omenjena seznama, sem razvrstila v dodatne kategorije, podrobneje predstavljene v nadaljevanju prispevka.

3 Rezultati in diskusija

3.1 Skupni seznam uporabljenih lem in pojavnic

Na končnem seznamu lem iz podkorpusov KOST 10k je 3428 lem: 878 na stopnji A1, 628 na stopnji A2, 371 na stopnji B1. Takšno padanje števila lem glede na stopnjo je pričakovano, saj gre za besedila začetnikov. Med lemami, ki jih avtomatski lematizator ni prepoznal, je bilo 486 slovenskih lem, ki niso bile na seznamu jedrnega besedišča med lemami od A1 do B1, 641 pa neslovenskih lem oz. besed iz drugih jezikov. Na podlagi takšnega, skupnega števila lem se zdi, da med stopnjami pravzaprav ni večjih razlik, vendar to seveda ne drži. Graf 1 zato prikazuje tako število uporabljenih lem kot število vseh pojavnic. Iz njega je razvidno, da tvorci na začetnih stopnjah učenja sicer uporabljajo tudi leme, ki presegajo stopnjo A1, vseeno pa izrazito prevladujejo leme na stopnji A1, ki so uporabljene v več kot 85 % pojavnic.



Graf 1: Število uporabljenih lem in pojavnic v vseh podkorpusih KOST 10k glede na stopnjo po lestvici SEJO.

Nekaj kategorij na Grafu 1 ni prikazanih. Med njimi je 60 lem, ki sodijo med citatne besede. Pri tem gre bodisi za besede za predmetnost, ki v slovenščini nima (uveljavljenega) poimenovanja (npr. *carbonara*, *županija*), bodisi za tujejezične izraze, ki jih je tvorec izbral zaradi nezadostne zmožnosti v slovenščini ali zaradi česa drugega. Ker je pri korpusni analizi včasih nemogoče oceniti, posledica česa je raba tujejezičnih izrazov, sem te primere iz analize izključila. Podobno nisem upoštevala 281 lastnih imen, bodisi osebnih bodisi zemljepisnih, ki jih v referenčnih virih nasploh ni (Klemen idr. 2022). Moja izključitev zemljepisnih poimenovanj je sicer nekoliko vprašljiva; že na stopnji A1 (Pirih Svetina 2016) se predvideva, da govorec pozna poimenovanje svoje države v slovenščini. Pri delu s korpusnimi podatki je ugotavljanje, ali je tvorec korpusnega besedila omenil svojo ali katero drugo državo, nemogoče, zato sem tudi zemljepisna imena izključila iz raziskave in se osredotočila le na občnoimenska poimenovanja. Izključila sem tudi osem kratic (npr. *ipd*.) in 25 pogovornih lem. Gre za izraze, ki so pogosti v nestandardni in govorjeni slovenščini (npr. *cajt*, *špilati*). Ker pa niso vključeni v nobenega od referenčnih virov in ker ni jasne opredelitve, na kateri stopnji naj bi govorec obvladal te izraze – ne nazadnje je dokument SEJO namenjen predvsem standardnim jezikom (prim. Svet Evrope 2011: 48) – sem se odločila, da se tudi s temi izrazi v tej analizi ne bom ukvarjala.

Za lažjo predstavo so v Tabeli 2 navedene najpogostejše leme za nekatere izbrane besedne vrste in za vse obravnavane stopnje po SEJO. Pri tem velja izpostaviti samostalnik *človek* in pridevnik oz. celostni zaimek *ves*, ki sodita na stopnjo A1. Zaradi zahtevnejših paradigem ju večina učnih gradiv obravnava na višjih stopnjah učenja, produktivno pa ju uporabljajo že začetniki. Iz tabele je razvidno tudi, kako odvisne so tovrstne raziskave od vsebine obravnavanih besedil. Na stopnji A2 so pogoste leme *plus*,

dejavnost, *dodaten*, kar je brez dvoma posledica tega, da so velik del besedil napisali študenti v modulu Leto plus,[9] ki morajo v okviru učenja slovenščine opravljati tudi dodatne dejavnosti.

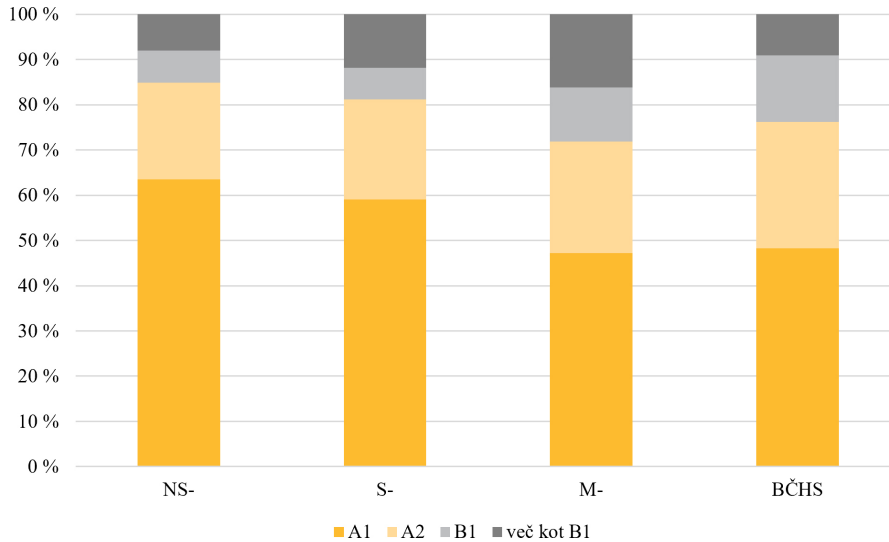
	Glagol	Samostalnik	Pridevnik[10]	Prislov
A1	biti, imeti, iti, živeti, delati	leto, mesto, prijatelj, čas, dan	moj, dober, velik, ves, lep	zelo, veliko, lahko, zdaj, tam
A2	naučiti, upati, uporabljati, zdeti, pogrešati	plus, študij, učenje, dejavnost, predavanje	dodaten, različen, zadovoljen, pomemben, isti	dovolj, največ, najprej, prvič, težko
B1	nahajati, truditi, obstajati, dokončati, tikati	napredek, omrežje, generacija, prednost, cimer	družaben, enostaven, terenski, družben, fizičen	dosti, znova, hkrati, kadarkoli, nikdar
več kot B1	oblikovati, brskati, čuvati, hecati, nadomeščati	fotelj, študiranje, šoljka, tepih, bunda	blond, celoten, prevroč, prihodnji, fonetičen	mnogo, zanimivo, nekje, pravočasno, vsakič

Tabela 2: Največkrat uporabljene leme za vsako stopnjo po lestvici SEJO.

3.2 Pregled rezultatov glede na prvi jezik govorcev

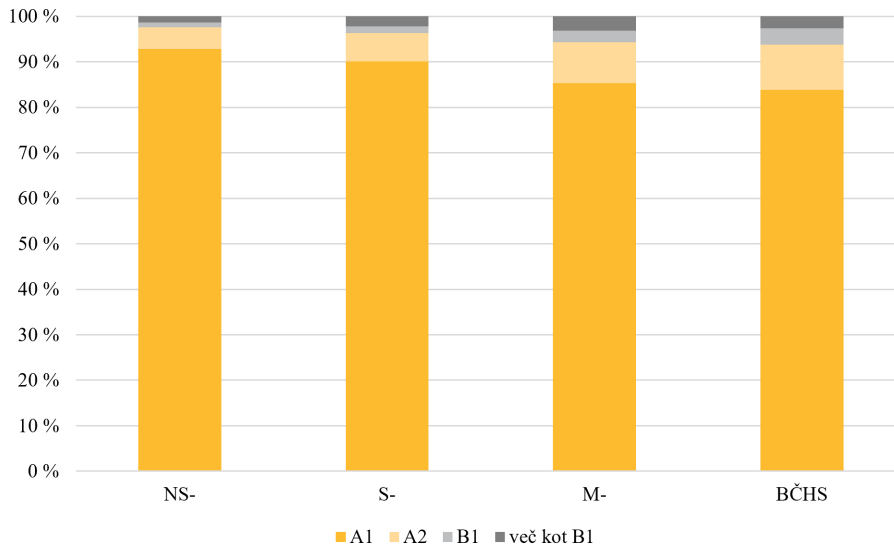
Ko si ogledamo razlike po podkorpusih KOST 10k, postanejo podatki zanimivejši. Graf 2 prikazuje število uporabljenih lem po stopnjah SEJO in glede na prvi jezik tvorcev. V vseh podkorpusih prevladujejo leme na stopnji A1, posebej pa pri neslovanskih začetnikih, ki so na vseh višjih stopnjah uporabili najmanj različnih lem. Pri stopnjah od A2 naprej je opazen velik razkorak med govorci NS- in S- ter med govorci južnoslovanskih jezikov, ki so na višjih stopnjah uporabili bistveno več lem. Zanimivo je še, da je skupina govorcev makedonščine uporabila skoraj dvakrat več lem (196 lem) na stopnjah, višjih od B1, kot govorci BČHS (107).

9 https://www.uni-lj.si/studij/dodiplomski-in-enoviti-magistrski-studij/leto-plus (dostop 17. 3. 2026)
10 Med pridevniki zaradi enake paradigme navajam tudi pridevniške zaimke.



Graf 2: Deleži lem po stopnjah SEJO glede na J1 govorcev.

Graf 3 pa za razliko od Grafa 2 prikazuje deleže pojavnic po stopnjah SEJO za posamezne podkorpuse KOST 10k. Iz njega lahko razberemo še izrazitejšo prevlado pojavnic na stopnji A1, delež pojavnic na stopnjah, višjih od A1, pa je bistveno višji pri govorcih slovenščini bližjih jezikov.



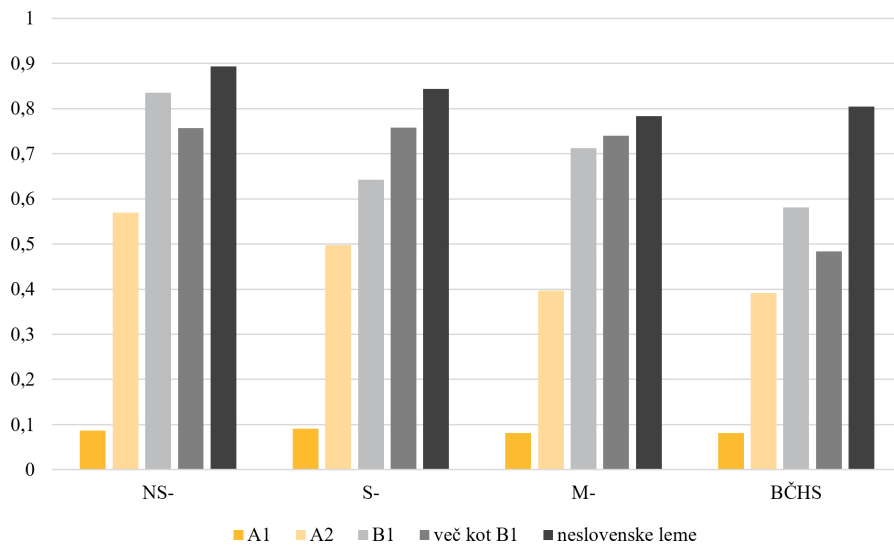
Graf 3: Deleži pojavnic po stopnjah SEJO glede na J1 govorcev.

Za prikaz leksikalne raznolikosti sem uporabila izračun razmerja med številom lem oz. različnic in pojavnic (angl. *type/token ratio*, TRR), ki ga za vsak podkorpus in vsako stopnjo jezikovne zmožnosti prikazujeta Tabela 3 in Graf 4. Višja številka nakazuje

manj ponavljanja oz. bogatejši nabor uporabljenih lem. Upoštevati je treba, da vrednosti ne moremo neposredno primerjati med stopnjami, temveč zgolj med podkorpusi znotraj iste stopnje.

Skupina	A1	A2	B1	Več kot B1	Neslovenske leme	Skupno razmerje TTR
NS-	0,087	0,569	0,835	0,757	0,893	0,142
S-	0,091	0,498	0,643	0,758	0,844	0,161
M-	0,081	0,397	0,712	0,740	0,783	0,163
BČHS	0,081	0,392	0,581	0,484	0,805	0,158

Tabela 3: Razmerje med številom lem in pojavnic po stopnjah SEJO in skupinah J1.

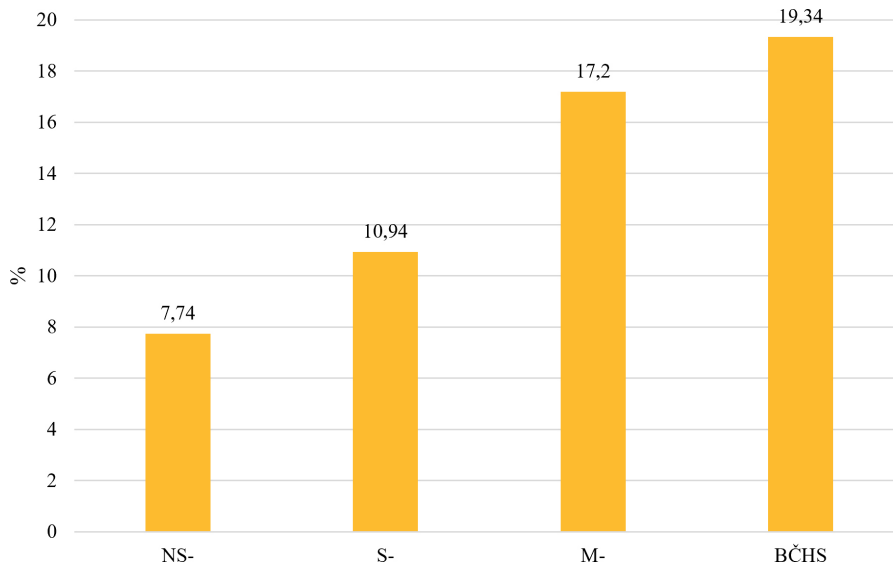


Graf 4: Razmerje med številom lem in pojavnic po stopnjah SEJO in skupinah J1.

Na ravni A1 so si skupine govorcev vseh prvih jezikov precej podobne. Nizko razmerje je statistično logično zaradi velikega vzorca pojavnic in pogostega ponavljanja besed, kot so zaimki, osnovni samostalniki ali glagol *biti*. Na stopnji A2 se razmerje poveča, najbolj pri NS- in S-, a predvsem na račun nizkega števila pojavnic. To velja tudi za višje stopnje, kjer so zanimivejše primerjave med M- in BČHS. Pri govorcih makedonščine je leksikalna raznolikost precej večja. Ekstremno visoka leksikalna raznolikost pri neslovenskih lemah pa potrjuje, da jih govorci uporabljajo skoraj vedno posamično. Med 641 lemami se jih 309, torej slaba polovica, ponovi dva- ali večkrat; najpogosteje, to je 24-krat, *imati* 'imeti' v različnih oblikah. Sledita veznik *i* 'in, tudi' (16 pojavnic) in predlog *u* 'v' (8 pojavnic), ki ga lahko razumemo tudi kot napako črkovanja. Najbolj raznolike neslovenske leme uporabljajo govorci NS-. Izjemno pomenljivo pa je tudi to, da so slovanski govorci neslovenske leme uporabili precej večkrat (od 224- do 235-krat, od 2,6 do 2,9 % vseh njihovih pojavnic) kot neslovanski (159-krat, 1,9 % vseh njihovih pojavnic). Pri rabi besedišča so torej bolj suvereni in bodisi zaradi

pomanjkanja nadzora bodisi zavestno posegajo tudi po besedišču iz svojega prvega jezika, medtem ko se govorci nesorodnih jezikov verjetno bolj zavedajo oddaljenosti svojih jezikov od slovenščine in se bolj držijo besedišča, ki ga obvladajo.

Zanimivo primerjavo da tudi izračun indeksa naprednosti, torej kolikšen delež lem je uporabljen na stopnjah od vključno A2 naprej (Graf 5): skupina govorcev BČHS uporablja 2,5-krat več besedišča na višjih ravneh od skupine neslovanskih govorcev. Pri skupini makedonskih govorcev je ta indeks le malenkost nižji.



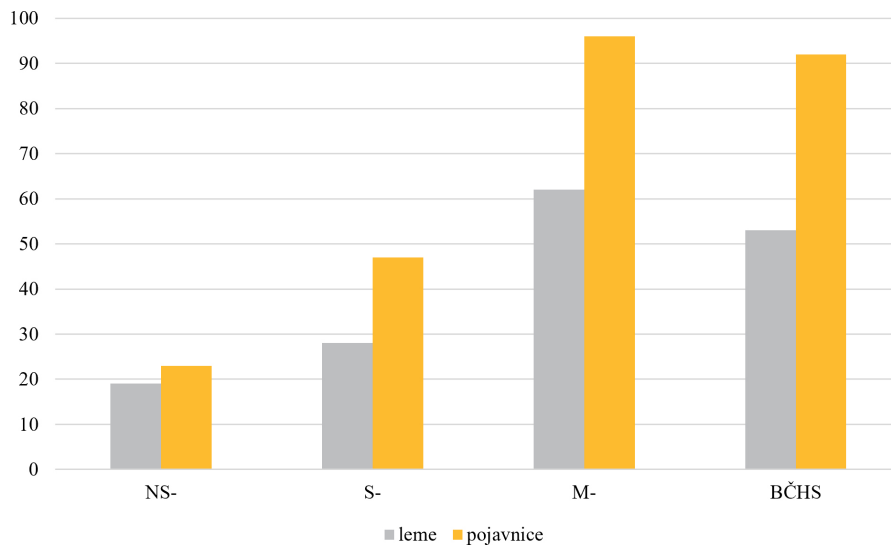
Graf 5: Indeks naprednosti za vsako skupino J1.

Zaključim lahko torej, da so neslovanski govorci (NS-) najbolj odvisni od osnovnega besedišča. Leksikalna zmožnost govorcev vzhodno- in zahodnoslovanskih jezikov (S-) je širša od Neslovanov, a še vedno precej v okviru stopenj A1 in A2. Obe skupini južnoslovanskih govorcev pa sta dovolj podobni, da se ločevanje med njima vsaj na leksikalni ravni zdi nepotrebno. Leksikalno sta naprednejši od drugih dveh skupin, pri čemer ne smemo zanemariti dejstva, da so bile teme besedil, o katerih so pisali govorci M- in BČHS, vendarle večinoma manj osnovne od tem govorcev NS- in S-,[11] kar ima za posledico tudi uporabo zahtevnejšega besedišča. Govorci makedonščine so bili leksikalno nekoliko bolj sofisticirani od BČHS, govorci BČHS pa so bolj produktivno uporabljali besedišče na stopnjah A2 in B1. Ali je ta razlika splošna značilnost govorcev makedonščine ali zgolj naključna značilnost izbranega vzorca, bi bilo treba preveriti s širšo raziskavo.

Izpostaviti moram še posebno skupino besed, ki niso umeščene na nobenega od referenčnih seznamov. Gre za mednarodne besede pretežno grškega (npr. *geologija*) ali

11 Podatek o tem, ali so govorci temo lahko izbrali sami s širšega seznama tem ali jim je bila določena, v KOST-u ni na voljo.

latinskega izvora (npr. *semester*), ki so celo v jezikih iz zelo različnih jezikovnih skupin podobne in zato za neprve govorce slovenščine lažje razumljive (prim. Markovič idr. 2009: 6; Uni 2019). Do neke mere je umestitev teh besed na stopnjo, višjo kot B1, za kakršno sem se odločila v tej raziskavi, utemeljena, saj so večinoma povezane z abstraktnimi koncepti. Zdi pa se, da tujim govorcem, ki so te koncepte usvojili v svojem jeziku, prenos v druge jezike ne predstavlja posebnih težav, in z uporabo teh besed lahko hitreje preidejo na višje ravni sporazumevanja. Podobno kot pri ostalih analiziranih stopnjah imajo tudi pri teh besedah govorci slovenščini bližjih jezikov širši razpon (Graf 6) in uporabljajo več različnih lem, te leme pa uporabljajo večkrat. Ali so te razlike zgolj posledica bližine jezikov in s tem povezane jezikovne samozavesti ali drugih dejavnikov, presega namene te raziskave, ki jo omejujejo v korpusu dostopni podatki.



Graf 6: Mednarodne besede; število lem in pojavnic v vseh podkorpusih KOST 10k.

4 Zaključek

V prispevku sem s sodobno, objektivno preverljivo korpusno metodo dokazala, da so pri učenju slovenščine kot neprvega jezika govorci jezikov, ki so slovenščini bližji, v precejšnji prednosti. Omejila sem se na njihovo produktivno uporabo besedišča in ugotovila, da govorci južnoslovanskih jezikov že v začetnih fazah učenja uporabljajo bistveno bolj raznoliko besedišče kot govorci slovenščini manj sorodnih jezikov. Po drugi strani pa to pomeni tudi, da so pri uporabi tega besedišča manj pozorni, kar se na kvantitativni ravni kaže v precej višjem številu neslovenskih lem oz., preprosteje povedano, leksikalnih napak. Ugotovila sem tudi, katero besedišče neprvi govorci slovenščine tvorno uporabljajo že na začetnih stopnjah učenja jezika, zato bi mu kazalo tudi v pedagoškem procesu nameniti več pozornosti (npr. leksema *človek*, *ves*).

Pri sklicevanju na zaključke te analize ne gre pozabiti, da je bila izvedena na razmeroma majhnem vzorcu. Poleg tega temelji na rezultatih avtomatske lematizacije, pri

kateri so mogoče napake. Tovrstna avtomatska obdelava besedil pa tudi ne prepozna, ali so bile leme dejansko uporabljene ustrezno. Pri govorcih sorodnih jezikov, za katere je tipična uporaba lažnih prijateljev, je to lahko resen problem.

Literatura

Arizankovska, L. (2009). Slovenščina kot tuji jezik – težave, s katerimi se srečujejo govorci makedonščine. V M. Stabej (ur.), *Infrastruktura slovenščine in slovenistike* (Obdobja 28, str. 27–32). Znanstvena založba Filozofske fakultete Univerze v Ljubljani. https://centerslo.si/wp-content/uploads/2015/10/28-Arizanovska.pdf

Balažic Bulc, T. (2004). Jezikovni prenos pri učenju sorodnih jezikov: Na primeru slovenščine in srbohrvaščine. *Jezik in slovstvo*, *49*(3–4), 77–89.

Ferbežar, I. in Eniko, M. (2022). »Lah blatschem gotovina?«: Jezikovni profil uporabnika slovenščine na najnižji ravni. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: Slovenščina kot drugi in tuji jezik* (Obdobja 41, str. 99–108). Založba Univerze v Ljubljani. https://centerslo.si/wp-content/uploads/2022/11/Ferbezar-Eniko_Obdobja-41.pdf

Ferbežar, I., Knez, M., Markovič, A., Pirih Svetina, N., Schlamberger Brezar, M., Stabej, M., Tivadar, H. in Zemljarič Miklavčič, J. (2004). *Sporazumevalni prag za slovenščino 2004*. Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani in Ministrstvo RS za šolstvo, znanost in šport.

Jerman, T., Murn, K. in Jugović, T. (2023). *Mozaik slovenščine: Učbenik za učenje slovenščine kot drugega in tujega jezika*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Jerman, T., Pisek, S. in Strajnar, A. (2015). *Slovenščina ekspres 1: Učbenik za začetnike na tečajih slovenščine kot drugega in tujega jezika*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Klemen, M. (2025). Poznavanje pogostih splošnih besed v slovenščini med govorci slovenščine kot drugega in tujega jezika. *Prispevki za novejšo zgodovino*, *65*(3), 243–272. https://doi.org/10.51663/pnz.65.3.10

Klemen, M., Arhar Holdt, Š., Pollak, S., Kosem, I., Huber, D. in Lutar, M. (2022). Korpus učbenikov za učenje slovenščine kot drugega in tujega jezika. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: Slovenščina kot drugi in tuji jezik* (Obdobja 41, str. 165–174). Založba Univerze v Ljubljani. https://doi.org/DOI:10.4312/Obdobja.41.165-174

Klemen, M., Arhar Holdt, Š., Pollak, S., Kosem, I., Pori, E., Gantar, P. in Knez, M. (2023). Building a CEFR-Labeled Core Vocabulary and Developing a Lexical Resource for Slovenian as a Second and Foreign Language. V M. Medved, M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček in S. Krek (ur.), *Electronic lexicography in the 21st century (eLex 2023): Invisible lexicography. Proceedings of the eLex 2023 conference* (str. 664–678). Lexical Computing CZ. https://elex.link/ojs/index.php/elex/article/view/55/39

Klemen, M., Stritar Kučuk, M., Eniko, M. in Huber, D. (2026). SEJO KOST: Razvoj novega jezikovnega vira za slovenščino kot neprvi jezik. V Š. Arhar Holdt, V. Gorjanc in T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika: Zbornik konference* (str. 224–243). Inštitut za novejšo zgodovino. https://doi.org/10.5281/zenodo.22765832

Klinar, M., Pisek, S., Stritar Kučuk, M. in Šter, H. (2022). Poučevanje slovenščine za redno vpisane tuje študente na Univerzi v Ljubljani. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: Slovenščina kot drugi in tuji jezik* (Obdobja 41, str. 185–194). Založba Univerze v Ljubljani. https://ebooks.uni-lj.si/ZalozbaUL/catalog/view/374/701/8143

Markovič, A., Knez, M. in Šoba, N. (2009). *Priročnik za učitelje k učbeniku Slovenska beseda v živo 2*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani. https://centerslo.si/wp-content/uploads/2015/10/SBZ_2-prirocnik-1.pdf

Markovič, A., Stritar Kučuk, M., Jerman, T. in Pisek, S. (2013). *Slovenska beseda v živo 1a: Učbenik za začetni tečaj slovenščine kot drugega in tujega jezika*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Nikolovski, G. (2022). Makedonske »muke« pri učenju slovenščine kot drugega in tujega jezika. *Jezik in slovstvo*, *67*(4), 41–56.

Pirih Svetina, N. (2016). *Preživetvena raven za slovenščino: Za potrebe programa Opismenjevanje v slovenščini za odrasle govorce drugih jezikov*. Univerza v Ljubljani, Filozofska fakulteta, Center za slovenščino kot drugi in tuji jezik. https://centerslo.si/knjige/ucbeniki-in-prirocniki/prirocniki-in-ucno-gradivo/prezivetvena-raven-za-slovenscino/

Petric Lasnik, I., Pirih Svetina, N. in Ponikvar, A. (2018). *A, B, C … gremo: Učbenik za začetnike na kratkih tečajih slovenščine kot drugega ali tujega jezika*. Znanstvena založba Filozofske fakultete.

Stritar Kučuk, M. (2024a). Investigating the usage of machine translation in L2 learning and its impact on writing proficiency. *Lidil*, *70*. https://doi.org/10.4000/12lmd

Stritar Kučuk, M. (2024b). KOST 2.0: Predstavitev korpusa in potek označevanja jezikovnih napak. V Š. Arhar Holdt in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 589–603). Inštitut za novejšo zgodovino. https://doi.org/10.5281/zenodo.13912515

Stritar Kučuk, M. (2024c). Prvi korpus slovenščine kot tujega jezika KOST 1.0. V Š. Arhar Hold in S. Krek (ur.), *Razvoj slovenščine v digitalnem okolju* (str. 93–117). Založba Univerze v Ljubljani. https://doi.org/10.4312/9789612972561

Stritar Kučuk, M. (2025). Označevanje jezikovnih napak v slovenščini kot neprvem jeziku: Splošna načela in praktični preizkus označevanja v korpusu KOST. *Slovenščina 2.0*, *13*(1), 24–57. https://doi.org/10.4312/slo2.0.2025.1.24-57

Stritar Kučuk, M. in Pirih Svetina, N. (2025). Študenti priseljenci in njihove slovenščine: Pisna zmožnost udeležencev Leta plus. *Slovenščina 2.0*, *12*(2), 200–232. https://doi.org/10.4312/slo2.0.2024.2.200-232

Svet Evrope. (2011). *Skupni evropski jezikovni okvir: Učenje, poučevanje, ocenjevanje*. Ministrstvo RS za šolstvo in šport in Urad za razvoj šolstva.

Uni, K. (2019). Benefits of vocabulary of Latin origin for the learners of Swedish and Danish. *The Journal of Social Sciences Research*, *5*(2), 431–435. https://doi.org/10.32861/JSSR.52.431.435

Problematika diskurznega označevalca *ne* pri avtomatski razpoznavi govora in strojnem prevajanju: analiza rabe v korpusu univerzitetnih predavanj

DOI: 10.4312/Obdobja.45.371-378

Tjaša Šoltes, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana;
tjasa.soltes@fri.uni-lj.si

Marko Bajec, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Ljubljana;
marko.bajec@fri.uni-lj.si

Prispevek obravnava problematiko diskurznega označevalca *ne* in njegovo varianto *a ne* pri avtomatski razpoznavi govora in strojnem prevajanju slovenskih univerzitetnih predavanj v angleščino. Analiza petih predavanj, zajetih v sistemu Online Notes, se osredotoča na razlikovanje med rabo *ne* kot nikalnega elementa in kot diskurznega označevalca. Rezultati kažejo, da je v avtomatskih transkriptih diskurzni označevalec večinoma ustrezno zapisan, medtem ko je v strojnih prevodih več napak, saj sistem pogosto napačno interpretira *ne* kot negacijo. Takšne napake lahko povzročijo spremembe pomena izjav in posledično vplivajo na razumevanje prevedene vsebine.

diskurzni označevalec, *ne*, strojno prevajanje, razpoznava govora

This article addresses the discourse marker *ne* in automatic speech recognition and machine translation of Slovenian university lectures into English. The analysis of five lectures included in the Online Notes system focuses on distinguishing between the use of *ne* as a negation element and as a discourse marker. The results show that automatic transcripts generally represent the discourse marker appropriately, whereas more difficulties arise in machine translation, in which the system often incorrectly interprets *ne* as negation. Such errors can alter the meaning of utterances and consequently affect the understanding of the translated content.

discourse marker, *ne*, machine translation, automatic speech recognition

1 Uvod

Diskurzni označevalci predstavljajo pomemben in pogost element govorjenega jezika ter v določenih govornih žanrih tvorijo pomemben delež spontane komunikacije. Schiffrin (1987) jih opredeljuje kot jezikovne enote, ki povezujejo elemente govora. Podobno Fraser (1999) poudarja, da imajo diskurzni označevalci proceduralno funkcijo, ne konceptualne, torej njihova primarna funkcija ni posredovanje propozicijske vsebine, temveč signaliziranje pragmatičnih odnosov med izjavami. Diskurzni označevalci so definirani kot izrazi, ki »ne prispevajo nič ali skoraj nič k vsebini diskurza, pač pa so rabljeni predvsem v pragmatičnih vlogah« (Verdonik idr. 2007b). Ker se tipično pojavljajo

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

v spontanem govoru (Dobrovoljc 2018), imajo pomembno vlogo pri strukturiranju interakcije, organizaciji besedila in izražanju odnosa govorca do sporočanega.

Raziskave kažejo, da je raba diskurznih označevalcev močno odvisna od komunikacijskega konteksta in žanra. Verdonik, Žgank in Pisanski Peterlin (2008) ugotavljajo, da se pogostost in funkcija posameznih označevalcev med različnimi govornimi žanri pomembno razlikujeta. Ker so diskurzni označevalci posebej značilni za spontano komunikacijo, jih govorjeni diskurz vsebuje bistveno več kot pisni jezik. V univerzitetnih predavanjih, ki združujejo elemente pripravljenega in spontanega govora, lahko zato predstavljajo pomemben del govora.

Prav zaradi teh značilnosti diskurzni označevalci predstavljajo poseben izziv za tehnologije avtomatske obdelave govora. Njihova interpretacija je pogosto odvisna od širšega konteksta, intonacije, premorov in položaja v izreki. Heeman, Byron in Allen (1998) opozarjajo, da avtomatsko razlikovanje med različnimi funkcijami enake besedne oblike v govorjenem jeziku predstavlja zahtevno nalogo. Podobno Popescu-Belis in Zufferey (2011) ugotavljata, da pri avtomatskem prepoznavanju diskurznih označevalcev pomembno vlogo igrajo prozodične značilnosti in položaj označevalca v izreki. Ker se takšne informacije pri avtomatski obdelavi pogosto izgubijo ali poenostavijo, sistemi težje prepoznajo pragmatično funkcijo posameznih izrazov.

Tak primer predstavlja slovenska beseda *ne*, ki lahko nastopa bodisi kot nikalni element bodisi kot diskurzni označevalec. V slednji vlogi *ne* nima enoznačne angleške ustreznice, temveč se njegova funkcija izraža z različnimi izrazi, kot so *right*, *y'know* ali *isn't it* (Verdonik idr. 2007b), ali pa se preprosto izpusti. Zaradi odsotnosti enoznačnega prevodnega ekvivalenta in izrazite pragmatične funkcije je takšna raba posebej zahtevna za sisteme strojnega prevajanja.

Od leta 2025 je na Univerzi v Ljubljani na voljo sistem Online Notes, ki omogoča avtomatsko razpoznavo govora slovenskih predavanj in njihovo sprotno prevajanje v angleščino. Sistem zaposlenim omogoča snemanje svojih predavanj ter ustvarjanje avtomatskih transkriptov in prevodov, ki jih lahko študentje spremljajo v realnem času. Primarno je namenjen študentom, katerih prvi jezik ni slovenščina, in študentom s senzornimi oviranostmi, saj lahko s pomočjo aplikacije, ki je dostopna tudi na mobilnih napravah, v živo spremljajo predavanja (Šoltes idr. 2024). Ker so transkripti in prevodi za nekatere uporabnike lahko glavni vir informacij med predavanjem, je natančnost prenesenih informacij ključnega pomena. Pri tem se kot problematična izkazuje raba besede *ne* v funkciji diskurznega označevalca. Prevajalni sistem namreč besedo *ne* v izvirniku pogosto interpretira kot nikalnico, zato jo tudi prevede kot negacijo, kar lahko spremeni pomen celotne izjave in uporabniku posreduje vsebinsko napačno informacijo.

V tem prispevku se osredotočamo na razlikovanje med rabo besede *ne* kot nikalnega elementa in njeno rabo v funkciji diskurznega označevalca v avtomatskih transkriptih in prevodih. Analiza temelji na petih univerzitetnih predavanjih in preučuje pojavnost diskurznega označevalca *ne* in variante *a ne*, njegovo predstavitev v avtomatskih transkriptih ter vpliv njegove interpretacije na kakovost strojnega prevajanja.

2 Gradivo

Analizo smo opravili na petih univerzitetnih predavanjih, ki so bila pilotno izvedena leta 2023 v okviru projekta Online Notes. Zajetih je pet različnih govorcev s predavanj, ki so se izvajala na treh različnih fakultetah Univerze v Ljubljani. V Tabeli 1 je predstavljenih nekaj osnovnih podatkov o vključenih predavanjih. Analiza je bila opravljena na transkriptih, ki smo jih na podlagi posnetkov ponovno ustvarili za namen te raziskave – razpoznava govora je namreč v tem času precej napredovala in v transkriptih opazimo manj napak. V avtomatskih transkriptih smo ročno označili vse pojavitve diskurznega označevalca *ne*, vsak primer smo klasificirali glede na pozicijo v povedi, in sicer na začetno, sredinsko in končno pozicijo ter samostojne primere, v katerih je poved sestavljena le iz diskurznega označevalca *ne* oz. *a ne*.[1]

Avtomatski prevodi, ki smo jih analizirali, so prevodi nepopravljenih avtomatskih transkriptov – ravno takšne namreč na predavanjih in vajah spremljajo študentje, ki se na te prevode zanašajo. Pri podajanju transkriptov in prevodov je pomembno tudi to, da je vsebina na voljo čim hitreje po tem, ko jo predavatelj izreče. Tako jim študentje najlažje sledijo, poleg tega se vsebina izrečenega marsikdaj navezuje tudi na dodatne materiale, kot so prosojnice, formule na tabli ipd. (Šoltes idr. 2024). Za zmanjševanje latence pri prevajanju so zato prevedeni posamezni segmenti (običajno povedi) in ne celotno predhodno besedilo, kar pa vpliva tudi na kvaliteto prevodov. Analiza je torej narejena na avtomatsko prevedenih posameznih segmentih avtomatskega transkripta.

Predavanje	**Fakulteta**	**Trajanje posnetka**	**Število pojavnic v avtomatskem transkriptu**	**Število pojavnic v referenčnem transkriptu**	**Število pojavitev besede *ne***
1	Biotehniška fakulteta UL	01:29:33	11.512	11.533	71
2	Biotehniška fakulteta UL	01:30:53	11.663	11.803	95
3	Fakulteta za matematiko in fiziko UL	00:38:10	5376	5414	21
4	Biotehniška fakulteta UL	01:17:09	9762	7939	140
5	Filozofska fakulteta UL	00:44:08	4735	4655	76

Tabela 1: Predavanja, vključena v analizo.

3 Pregled pojavitev pojavnice *ne*

Iz Tabele 2 je razvidno, da se pogostost diskurznega označevalca *ne* med posameznimi govorci izrazito razlikuje, kar je skladno z ugotovitvami predhodnih raziskav, ki kažejo, da je raba diskurznih označevalcev odvisna od komunikacijskega konteksta, govornega žanra in individualnega govornega sloga (Fuller 2003; Verdonik idr. 2008). Pri univerzitetnih predavanjih lahko na razlike vplivajo tudi različni slogi predavanja in stopnja interaktivnosti (Bellés-Fortuño 2004; Schleef 2008).

V prvih treh predavanjih prevladuje raba *ne* v nikalnem pomenu, medtem ko je v četrtem in petem predavanju število diskurznih označevalcev višje od števila nikalnih

1 Podatki so bili pridobljeni v okviru pilotnega projekta Online Notes in zaradi narave podatkov niso javno dostopni. Anotirani primeri so v anonimizirani obliki na zahtevo na voljo pri avtorjih.

pojavitev. Posebej izstopa četrto predavanje, v katerem se pojavi skupno 99 pojavitev diskurznega označevalca *ne*, od tega 91 *a ne*. Govorci so pretežno uporabljali eno varianto, torej *ne* oz. *a ne*, z redkimi primeri rabe druge variante. Rezultati kažejo, da diskurzni označevalec *ne* predstavlja pomemben element govora tudi v univerzitetnih predavanjih, ki združujejo značilnosti pripravljenega in spontanega govora.

Predavanje	**Število pojavitev v pomenu negacije**	**Število pojavitev v funkciji diskurznega označevalca**	
		ne	***a ne***
1	60	11	0
2	66	27	2
3	16	5	0
4	41	8	91
5	32	44	0

Tabela 2: Pogostost besede *ne* v vlogi nikalnega elementa in diskurznega označevalca.

4 Analiza transkriptov in prevodov

Iz korpusa smo izvlekli vse pojavnice z diskurznim označevalcem *ne* (vključno s pojavnicami *a ne*), ki jih je v petih predavanjih skupno 188. Zanimalo nas je, ali je mogoče iz transkripta razbrati, da gre za diskurzni označevalec. Analiza je pokazala, da niti *ne* niti *a ne* v razpoznavi govora nista problematična in sta v skoraj vseh primerih ustrezno zapisana kot diskurzna označevalca, od preostalega dela povedi ločena z vejicami,[2] dvoumna sta bila namreč samo dva primera, v katerih diskurzni označevalec od preostalega dela povedi ni bil ločen z vejico, npr.: »[…] lahko razločimo rani in kasni les, **a ne** že po sami, tudi pri manjši povečavi vidimo ta svetlejši del, ta pa je temnejši del«. *Ne* v začetni poziciji povedi je mnogokrat brez širšega besedilnega konteksta in poslušanja pripadajočega zvočnega posnetka možno interpretirati na dva načina, torej kot negacijo ali kot diskurzni označevalec,[3] vendar pa to ne vpliva na samo razumevanje vsebine, kot npr. v povedi: »Ne, tu je neko nasprotje.«

Analizirali smo tudi prevode. Te smo ocenjevali zgolj z vidika rabe diskurznega označevalca. Razvrstili smo jih v tri kategorije. Kot ustrezen prevod smo ocenili vsak prevod, v katerem je bil diskurzni označevalec *ne* preveden z ustreznim diskurznim označevalcem v angleščini (npr. *right*, *isn't it*) ali je bil izpuščen. Neustrezne prevode smo uvrstili v dve kategoriji: kritični prevodi in moteči prevodi. Kot kritične prevode ocenjujemo tiste, v katerih je *ne* preveden kot nikalnica in je torej prvotna izjava govorca zanikana, npr. v spodnjem primeru:

transkript: *Skratka, discipliniranje,* ***ne****, ki mu je treba nasprotovati ali mu ubeči, če je le možno.*
prevod: *In short, disciplining,* ***not*** *something to be resisted or escaped from if possible.*

2 »Ne.« kot samostojna izjava je seveda iz te analize izključen – brez konteksta namreč ni mogoče določiti, ali gre za diskurzni označevalec ali za negacijo.

3 V teh primerih smo ustrezno kategorijo izbrali na podlagi širšega konteksta in govora na zvočnem posnetku.

Kot moteče prevode ocenjujemo tiste, v katerih se pojavlja zanikani element, vendar sama izjava ni zanikana, npr.:

transkript: *Če pa je omejeno zaporedje, potem pa ima vsaj eno stekališče,* ***ne****, ker se mora nečemu bližati.*
prevod: *But if it is a bounded sequence, then it has at least one convergence point,* ***no****, because it has to approach something.*

Iz Tabele 3 vidimo, da se diskurzni označevalec *a ne* v korpusu pojavi skupno 93-krat, od tega večinoma (86,0 % primerov) v sredinski poziciji, kjer je delež neustreznih prevodov tudi najvišji. Ustrezna sta namreč le dva primera, kar predstavlja 2,5 % primerov na tej poziciji. V obeh primerih je bil diskurzni označevalec v prevodu izpuščen, npr.:

transkript: […] *da boste dobili malo občutek,* ***a ne****, za za velikosti pri različnih objektivih* […]
prevod: […] *to get a little feel for the sizes with different lenses* […]

Kritičnih prevodov diskurznega označevalca *ne* v sredinski poziciji je bilo 43,8 % (kjer je bil *ne* preveden z *but not*, *but no*, *don't*, *not* in *but it doesn't*), kot npr. v naslednjih dveh povedih:

transkript: *Tako, sestava tal tudi, govorimo o mineralnih rudninskih snoveh,* ***a ne****, koliko je teh snovi dostopnih.*
prevod: *So, the composition of the soil also, we are talking about mineral substances,* ***but not*** *how much of these substances are available.*
in
transkript: […] *jaz probam tudi take prispodobe dati, da razumete,* ***a ne****, to strukturo.*
prevod: […] *I'm also trying to give you such metaphors so that you understand, or* ***don't*** *understand, this structure.*

Kot moteče smo ocenili 53,8 % (kjer je bil *a ne* preveden z *but no*, *but not* ali *no* kot vrinjeni element), npr.:

transkript: *In kako si lahko predstavljate,* ***a ne****, celično zgradbo?*
prevod: *And how can you imagine,* ***but not****, a cellular structure?*

A ne se v funkciji diskurznega označevalca na končni poziciji pojavi v osmih primerih (8,6 %). Od tega sta dva primera prevedena ustrezno, oba s funkcijsko ustreznico, npr.:

transkript: *Ne bo pa čisto tako,* ***a ne****?*
prevod: *But it won't be quite like that,* ***will it****?*

Kot kritičen smo ocenili en primer:

transkript: *Zgleda kot ena taka, kot en trampolinček, če si predstavljate,* ***a ne****.*
prevod: *It looks like one of those, like a trampoline, if you imagine it,* ***but it's not****.*

V dveh tretjinah primerov pa smo prevod ocenili kot moteč (v vseh primerih preveden z *but no*), npr.:

transkript: *Dajte pogledati povečavo, 400-kratno,* ***a ne****.*
prevod: *Let me take a look at the magnification, 400 times,* ***but no****.*

Povedi, ki vsebujejo le samostojni diskurzni označevalec *a ne*, so bile v vseh primerih prevedene s primerno funkcijsko ustreznico *isn't it*.

pozicija	pravilno	moteče	kritično	skupaj
začetna				0
sredinska	2	43	35	80
končna	2	5	1	8
samostojno	5			5
skupaj	9	48	36	93

Tabela 3: Vrednotenje prevodov diskurznega označevalca *a ne* glede na stavčno pozicijo.

Iz Tabele 4 vidimo, da se diskurzni označevalec *ne* v korpusu pojavi skupno 95-krat, s podobnimi deleži v vseh treh stavčnih pozicijah. Rezultati kažejo, da so prevodi, kjer se *ne* v vlogi diskurznega označevalca pojavi na začetku povedi, v angleščino dosledno prevedeni z *no*, kar smo klasificirali kot moteče, npr.:

transkript: ***Ne**, in recimo, če bi bile te dve enaki, bi zgledalo, kakor da sploh nista točki oddaljeni od premice, v bistvu pa sta.*
prevod: ***No**, and let's say if these two were the same, it would look like the points are not far from the line at all, but in fact they are.*

V sredinski stavčni poziciji se diskurzni označevalec *ne* pojavi v slabi tretjini primerov (32,6 %), kjer je največkrat (v 80,6 % primerov) v prevodu izražen z vrinjenim negacijskim elementom *no*, kar smo ocenili kot moteče.

transkript: *»Neskončno jih bo, **ne**, ampak ena tretjina, če rečemo, ne.*
prevod: *There will be an infinite number of them, **no**, but one third, if we say, no.*

Za pet primerov (kar predstavlja 16,1 % primerov) smo prevod ocenili kot kritičen, npr.:

transkript: *Temu rečemo mi režim sušenja, **ne**, kako oster bo režim.*
prevod: *This is what we call a drying regimen, **not** how harsh the regimen will be.*

En primer prevoda smo ocenili kot primeren. V njem je bil diskurzni označevalec v prevodu izpuščen.

transkript: *Mi ne bomo mogli nikoli vedeti, **ne**, ali smo tisti, o katerih se nam zelo živo sanja* […]
prevod: *We will never be able to know whether we are who we are in our vivid dreams* […]

Kot najmanj problematičen se je v prevodu izkazal *ne* na koncu povedi. Večina primerov (86,7 %) je bilo namreč ustrezno prevedenih s funkcijsko ustreznico (v prevodih zasledimo ustreznice: *right*, *is(n't) it*, *doesn't it* in *shall we*), npr.:

transkript: *Ta drugi, **ne**?*
prevod: *This other one, **right**?*
in
transkript: *Zgleda, kot da se manjšajo, **ne**?*
prevod: *It looks like they're getting smaller, **doesn't it**?*

Štiri primere prevodov smo ocenili kot moteče, saj je diskurzni označevalec *ne* preveden z angleškim *no*, npr.:

transkript: *Neskončno jih bo, ne, ampak ena tretjina, če rečemo,* ***ne****.*
prevod: *There will be an infinite number of them, no, but one third, if we say,* ***no****.*

pozicija	pravilno	moteče	kritično	skupaj
začetna		33		33
sredinska	1	25	5	31
končna	26	4		30
samostojno		1		1
skupaj	27	63	5	95

Tabela 4: Vrednotenje prevodov diskurznega označevalca *ne* glede na stavčno pozicijo.

5 Zaključek

Analiza diskurznega označevalca *ne* in variante *a ne* na petih predavanjih, izvedenih v okviru projekta Online Notes, je pokazala, da je raba diskurznega označevalca *(a) ne* pogosta tudi pri govoru na univerzitetnih predavanjih, vendar pa frekvenca rabe precej variira med govorci. Rezultati kažejo, da je iz slovenskega transkripta večinoma mogoče brez težav razbrati, da gre za diskurzni označevalec, medtem ko je to večji izziv za strojno prevajanje govorjenega jezika. Sistemi pogosto ne prepoznajo pragmatične funkcije diskurznega označevalca *ne* in ga interpretirajo kot negacijo, kar lahko popolnoma spremeni pomen izjave govorca, to pa seveda vpliva na razumevanje bralca ter na orodja za obdelavo naravnega jezika, ki sledijo v nadaljnjem procesu obdelave (denimo razumevanje naravnega jezika, povzemanje itn.). Kot ustrezno prevedene smo označili le 19,1 % vseh primerov (od tega 9,7 % vseh primerov diskurznega označevalca *a ne* in 28,4 % primerov diskurznega označevalca *ne*).

Rezultati analize transkripta petih univerzitetnih predavanj so pokazali, da ima na ustreznost prevoda pomemben vpliv tudi pozicija diskurznega označevalca v povedi (na začetku, na sredini, na koncu povedi ali kot samostojna poved). Največ kritičnih napak se pri obeh variantah pojavlja v primerih z diskurznim označevalcem na sredini povedi, kjer ga sistem prevaja kot angleško negacijo. Nasprotno je samostojni *ne* v končnem položaju pogosteje preveden ustrezno, zanimivo pa je, da to ne velja tudi za varianto *a ne*, ki se v končni poziciji večinoma ne prevaja ustrezno.

Čeprav je raziskava opravljena na manjšem obsegu besedil, lahko sklenemo, da bi bilo smiselno v tehnologijah za procesiranje naravnega jezika bolj upoštevati tudi značilnosti govorjenega jezika. Rezultati nakazujejo na to, da bi bilo mogoče kakovost strojnega prevajanja izboljšati z modeli, ki bi poleg leksikalne informacije upoštevali tudi pragmatično funkcijo diskurznih označevalcev, npr. z razvojem postopkov za samodejno razdvoumljanje ali vpeljavo posebnih oznak za diskurzne označevalce v učne podatke. Prihodnje raziskave bi bilo smiselno razširiti na večji korpus govorjenega jezika ter vključiti tudi podrobnejšo analizo prozodičnih značilnosti, ki bi lahko pomembno prispevale k natančnejšemu razdvoumljanju med diskurznimi označevalci in nikalnimi elementi.

Projekt financira ARIS – Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije, št. projekta GC-0002, izbranega v okviru Javnega razpisa za sofinanciranje Gravitacije.

Literatura

Bellés-Fortuño, B. (2004). The influence of lecturing styles in the use of discourse markers within the spoken academic discourse of social sciences. V M. Labarta Postigo (ur.), *Approaches to critical discourse analysis* (str. 1–21). Publicaciones Universidad de Valencia.

Dobrovoljc, K. (2018). Raba tipično govorjenih diskurznih označevalcev na spletu. *Slavistična revija*, *66*(4), 497–513. https://srl.si/ojs/srl/article/view/2018-4-1-6

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*(7), 931–952.

Fuller, J. M. (2003). The influence of speaker roles on discourse marker use. *Journal of Pragmatics*, *35*(1), 23–45.

Heeman, P. A., Byron, D. in Allen, J. (1998). Identifying discourse markers in spoken dialog. V *AAAI Technical Report SS-98-01*. https://cdn.aaai.org/Symposia/Spring/1998/SS-98-01/SS98-01-008.pdf

Majhenič, S., Rojc, M. in Mlakar, I. (2022). Neprototipni diskurzni označevalec zdaj. *Slavia Centralis*, *15*(2), 27–44. https://doi.org/10.18690/scn.15.2.27–44.2022

Majhenič, S. (2024). Communicative intent divergence of discourse markers in simultaneously interpreted speech. V Š. Arhar Holdt in T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika: Zbornik konference* (str. 275–297). Inštitut za novejšo zgodovino.

Popescu-Belis, A. in Zufferey, S. (2011). Automatic identification of discourse markers in dialogues: An in-depth study of *like* and *well*. *Computer Speech and Language*, *25*(3), 499–518. https://doi.org/10.1016/j.csl.2010.12.001

Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press.

Schleef, E. (2008). The »lecturer's OK« revisited: Changing discourse conventions and the influence of academic division. *American Speech Spring*, *83*(1), 62–84.

Šoltes, T., Vasiljević, J. in Bajec, M. (2024). ONLINE NOTES: sistem za razpoznavo govora in strojno prevajanje v realnem času na ravni univerzitetnih predavanj. *Uporabna informatika*, *32*(1), 32–41. https://doi.org/10.31449/upinf.225

Verdonik, D., Žgank, A. in Pisanski Peterlin, A. (2008). The impact of context on discourse marker use in two conversational genres. *Discourse Studies*, *10*(6), 759–775. https://doi.org/10.1177/1461445608096572

Verdonik, D., Rojc, M. in Stabej, M. (2007a). Annotating discourse markers in spontaneous speech corpora on an example for the Slovenian language. *Language Resources and Evaluation*, *41*(2), 147–180.

Verdonik, D., Žgank, A. in Pisanski Peterlin, A. (2007b). Diskurzni označevalci v dveh pogovornih žanrih. *Jezik in slovstvo*, *52*(6), 19–33

Speech Recognition of Standard Czech and Slovenian Using OpenAI Whisper

DOI: 10.4312/Obdobja.45.379-387

Veronika Štěpánová, Ústav pro jazyk český, Akademie věd České republiky, Praga;
stepanova@ujc.cas.cz

Tanja Mirtič, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana;
tanja.mirtic@zrc-sazu.si

Prispevek na podlagi analize prepisov dveh pravljic prinaša primerjavo kakovosti samodejnega razpoznavanja govora za češčino in slovenščino z odprtokodnim razpoznavalnikom OpenAI Whisper. V raziskavi preučujemo zmogljivosti tega globalnega sistema pri obdelavi jezikov z omejenimi učnimi viri, podrobno analiziramo tipologijo napak in ocenjujemo možnosti praktične uporabe razpoznavalnikov v pravorečnih raziskavah sodobnega knjižnega jezika.

razpoznava govora, češčina, slovenščina, govorjeni knjižni jezik, pravorečne raziskave

Based on an analysis of two fairy tale transcriptions, this article presents a comparative quality assessment of automatic speech recognition for Czech and Slovenian using the open-source OpenAI Whisper model. The study investigates the capabilities of this global system in processing languages characterized by limited training resources, provides a detailed analysis of error typology, and evaluates the potential of speech recognizers for practical application in orthoepic research on the contemporary standard language.

automatic speech recognition, Czech, Slovenian, spoken standard language, orthoepic research

1 Introduction

This article presents a comparative quality analysis of Czech and Slovenian automatic speech recognition (ASR) using the OpenAI Whisper model. Its purpose is to determine the current state of speech recognition development for Czech and Slovenian within the open-source OpenAI Whisper model, how this global speech-to-text system copes with medium- and low-resource languages characterized by limited training data, complex morphology, and relatively free word order, and whether speech recognition at this stage of development can already assist in research on contemporary standard Czech and Slovenian. In addition, it investigates potential performance differences between the two languages, driven by the fact that Czech—despite being a globally lower-resource language—benefits from a larger volume of available training data compared to Slovenian.

This practical study evaluates the OpenAI Whisper ASR system—an open-source end-to-end model available free of charge. Whisper demonstrates high transcription

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

accuracy across diverse languages and acoustic environments (Radford et al. 2023) and is widely utilized in ASR research (cf. Graham & Roll 2024). It was also selected because its open-source nature ensures full methodological reproducibility. Furthermore, this study offers a significant contribution by providing an independent evaluation of Whisper's performance, which is particularly crucial because the system's training datasets are not entirely public (Radford et al. 2023). Consequently, this research allows an objective verification of its actual capabilities and limitations.[1]

2 Data preparation

A primary challenge in cross-linguistic comparison is the speaker/recording variable (e.g., speech rate, intensity, and signal quality), which can conflate model performance with external factors. Nonetheless, this study utilizes authentic human speech. To compensate for the absence of perfectly balanced bilingual speakers, we ensured analogous speech situations, identical text genres, and comparable data volumes. Given that Whisper was trained on real-world data (Radford et al. 2023), human speakers provide a more valid test of the system's responsiveness to natural irregularities.

The selected corpora consist of similar narratives: the Slovenian fairy tale "O vetru, ki bi rad pel" (The Wind That Wanted to Sing) and the Czech tale "Jak se tropí kočičiny" (How Mischief Is Made). Both are professionally recorded studio performances by single narrators (Zvone Hribar and Jiří Lábus, respectively), who employ varied rhythm, intonation, and vocal registers to portray multiple characters, creating a polyphonic effect within a monologic delivery. The Slovenian fairy tale, consisting of 8:47 minutes of speech, was authored by Tatjana Kokalj and broadcast on the radio show *Lahko noč, otroci* (Good Night, Children), and the Czech fairy tale, featuring 7:34 minutes of speech, was written by Viktor Měska and aired on the program *Hajaja*, a show named after a beloved radio gnome.

The audio material was sourced directly from the national broadcasting archives of RTV Slovenija and Czech Radio (*Český rozhlas*). These specific programs were selected due to their structural and functional comparability; both are iconic national radio formats designed for the exact same target audience (young children prior to sleep). Furthermore, the speakers in both samples operate within identical communicative settings and speech genres.

3 Analysis

This research presents a comparative analysis of ASR performance for Slovenian and Czech, both of which are classified as low- or mid-resource languages due to their

1 For Slovenian, the primary ASR-specific database is ARTUR, containing 1,067 hours of recordings, 884 of which are manually transcribed (see https://www.clarin.si/repository/xmlui/handle/11356/1772; for Slovenian resource development, see Žgank et al. 2020: 118). For Czech, a key ASR resource is the ParCzech4Speech 1.0 corpus, providing 2,695 hours of aligned speech from 587 speakers. This corpus is based on audio from the Chamber of Deputies of the Parliament of the Czech Republic and their official stenographic transcripts (see https://lindat.mff.cuni.cz/repository/items/8cc95ecb-680f-4454-b041-58a787a5ca5f).

limited annotated speech data compared to English. Joshi et al. (2020: 6284–6285) introduce a taxonomy that categorizes the world's languages into six distinct classes based on the availability of labeled and unlabeled data within the language resource ecosystem. Under this framework, Slovenian is categorized under Class 3 (the "rising stars"), and Czech is positioned higher, in Class 4 (the "underdogs").[2] Furthermore, both languages feature complex morphology and relatively free word order (Žgank et al. 2020: 117).

The recordings were transcribed using MacWhisper (Bruin 2026), a user interface for OpenAI Whisper; specifically, the Large v3 Turbo model. A methodological limitation of this study is the small test corpus, which restricts the generalizability of the findings. The following section presents the transcription analysis for both languages, highlighting key deviations from current orthographic and orthoepic standards.

3.1 Analysis of the Czech transcription

The Czech transcript of the selected recording comprised 855 words. We identified seventy-five instances (8.8%) requiring correction for full compliance with the current Czech orthographic norm. Several errors recurred systematically throughout the transcript (e.g., **bi*[3] instead of *by*, the correct conditional form of the auxiliary verb *být* 'be', occurred five times, whereas the same word was transcribed correctly elsewhere). All deficiencies were categorized into the relevant linguistic domains, which are discussed below in varying degrees of detail.

In eleven instances, word boundaries required correction: five involved a single word written instead of two (e.g., **naopravdové* instead of *na opravdové* 'for real ones', **odmazlení*[4] instead of *od mazlení* 'from cuddling'), and six the reverse (e.g., **kde kdo* instead of *kdekdo* 'just about anyone'). These errors mainly affected less frequent Czech expressions, including neologisms and occasionalisms. Only the segmentation of *veledůležitě* 'with great self-importance' as **vele důležitě* may have been motivated by the presence of two-word stresses: [ˈvɛlɛ ˈduːlɛʒɪcɛ]. Similarly, the merging of the common Czech words *kočičí ne* 'not catlike' into **kočičíne* may have resulted from a phonetic realization with a single stress ([ˈkoʧɪʧiːnɛ]), in which the final syllable is almost elided. In contrast, **krtkané* (instead of *krtka né* 'no mole') and **bané* (instead of *ba ne* 'oh no') were also merged despite lacking any phonetic motivation, producing non-existent forms from otherwise common word combinations. We also identified five punctuation errors, none of which appeared to be influenced by intonation.

In ten instances, the deviation concerned the orthography of initial capital letters. These included inappropriate capitalization of common nouns that are not proper names

2 Croatian and Serbian are also categorized under Class 4. The complete taxonomy is published on the website https://microsoft.github.io/linguisticdiversity/.

3 An asterisk is used to mark non-existent (unattested) or categorically incorrect word forms; phonetic transcriptions are given in IPA.

4 The expression *odmazlení* 'deprivation of cuddling' can in principle be formed in Czech, but, given its somewhat bizarre meaning, it is not used; in actual texts it occurs, for example, as a typo for *omlazení* 'rejuvenation'.

(**Máj* instead of *máj* ‘May’), as well as proper, albeit non-standard, names written in lowercase (**minda* instead of *Minda*, a cat’s name). Because this distinction is not acoustically detectable, it cannot be inferred from the speech signal. In most cases, however, the transcript was correct (e.g., *kočka Guma* ‘the cat Guma’), whereas the repeatedly occurring cat name *Majka* was capitalized inconsistently without any apparent reason.

From the perspective of segmental transcription, several types of correction would be possible. Perhaps the most surprising were incorrect spellings of five common words involving confusion of letters denoting consonants that differ in voicing; for example, **drábky* instead of *drápky* ‘little claws’, **kožih* instead of *kožich* ‘fur’, **ocázku* instead of *ocásku* ‘little tail’, **dlapku* instead of *tlapku* ‘paw’. In the last case, we detected inappropriate (partial) voicing of the initial consonant both auditorily and by spectral analysis in Praat (Boersma & Weenink 2026), indicating a slip of the speaker; in the remaining cases, however, there is no phonetic motivation for such substitutions.

A similar situation applies to most of the eight cases involving incorrect transcription of vowel length. This resulted in non-existent forms such as **kočiči* instead of *kočičí* ‘catlike’ and **antenou* instead of *anténou* ‘with an antenna’, despite the vowels being articulated with the appropriate length. If the cause were sought, for example, in a faster speech rate or occasional vowel reduction, many more instances would have been transcribed similarly rather than only these. Conversely, we also observed the opposite phenomenon: *dokonale* ‘perfectly’ was erroneously “corrected” to *dokonalé* ‘perfect’, and the deliberately non-standard verbal form *neni* ‘is not’ was rendered as *není*.

Another segmental category comprises nineteen cases for which the automatic transcription reflects phonetic rather than orthographic form (e.g., [pʲ̊ɛtsiːɲɛ] as **přecíně* instead of *předsíně* ‘halls’, [paɲː] with vowel reduction as **paň* instead of *paní* ‘lady’). A frequent example is **sme* instead of *jsme* ‘we are’, accurately reflecting the speaker’s pronunciation (without initial [j]), although the identically pronounced [smɛ] is elsewhere transcribed correctly as *jsme*. We also included errors involving *i* and *y*, which are not distinguished in standard Czech pronunciation but remain orthographically contrastive (e.g., **vousi* instead of *vousy* ‘beard’, **odbíjeli* instead of *odbíjely* ‘struck’), requiring morphological and syntactic analysis for automatic correction.[5] Conversely, forms with *y* also occurred where impossible (**zývali* instead of *zívali* ‘yawned’), including a transposition of the graphemes within one word (**vykířem* instead of *vikýřem* ‘dormer window’), itself a common spelling error. Particularly noteworthy is **ebrý* for *prý* ‘allegedly’, caused by preglottalization ([ʔə]; Skarnitzl & Machač 2012), interpreted by the recognizer as [e], together with inappropriate partial voicing of the following consonant. Despite this, the intended word is readily recoverable from context.

5 Similarly, spoken language does not distinguish between the identically pronounced forms [mɲɛ] of the pronoun *já* ‘I’, written as *mě* or *mně* depending on case. This occurred twice in our text; both instances were transcribed as *mě* ‘of me’, although *mně* ‘to me’ was correct.

Finally, several remaining phenomena do not fit the categories above. Errors such as **tichoum se*, **zabručal*, and **křísný* instead of *tichounce* ‘very quietly’, *zabručel* ‘grumbled’, and *přísný* ‘strict’ could be attributed to adverse recognition conditions (e.g., lowered voice, changes in tempo or timbre, or less precise articulation). However, this is contradicted by accurately articulated forms such as *rozvrhem* ‘schedule’, transcribed as **rozvrém*, and *prsknout* ‘sputter’, transcribed as **presknout*. The interjection *hmm* ‘um’ was not reflected in the transcript at all.

3.2 Analysis of the Slovenian transcription

The Slovenian transcript contains 1,062 words, with errors occurring in approximately 101 instances (9.5%). Initial recognition failed during the opening greeting of the program *Lahko noč, otroci* (Goodnight, Children) due to background music, causing the model to generate nonsensical hallucinatory patterns based on learned data (e.g., **Zdravstvo, odroč* instead of *Dober večer, otroci* ‘good evening, children’).

Incorrect word boundary detection was frequent. The model often merged two words into one (e.g., **lepel* instead of *le pel* ‘just sang’, **odgroze* instead of *od groze* ‘out of horror’). Two cases also occur in which a single word is used instead of a prepositional phrase, namely due to completely coalesced pronunciation (e.g., **strahom* instead of *s strahom* ‘with fear’, **vsej* instead of *v vsej* ‘throughout all’). Conversely, single words were sometimes split (e.g., **na okoli* instead of *naokoli* ‘around’), potentially influenced by dual lexical stress in adverbs such as *vsepovsod* ‘everywhere’ (e.g., **vse posod* instead of *vsepovsod*). The letter *v* (phonetically a bilabial [w]) is omitted in the transcript, reflecting the actual pronunciation of the reader, which is a common articulatory simplification in colloquial speech.

A significant group of inconsistencies involved the transcription of schwa [ə]. Because Slovenian lacks a dedicated grapheme for schwa—standardly transcribing it as *e*, the combination of a schwa and the consonant *r* is orthographically rendered as a simple *r*, without a vowel grapheme—Whisper’s performance varied. Although occasionally correct (*temni* ‘dark’, *veter* ‘wind’), the model often omitted it (**vetr*, **vrabc*). For the noun *veter* ‘wind’, the correct spelling appeared fifteen times, and the incorrect form (**vetr*) occurred eight times. The spelling **vrabac* (instead of *vrabec* ‘sparrow’) suggests interference from Croatian or Serbian databases.[6] Furthermore, Whisper incorrectly inserted an *e* before the liquid *r* where the orthography forbids it (**razperl* instead of *razprl* ‘spread open’) and transcribed **vrjeti* for *verjeti* [ʊɛrˈjeːti] ‘believe’, likely because the speaker reduced the first vowel and pronounced it as a schwa [ʊəɾˈjeːti].

Further errors occurred with the final *-l* in the masculine singular active participle, which is pronounced as a bilabial [w]. Whisper transcribed this as *v* (**rjovev* instead of *rjovel* ‘roared’, **pev* instead of *pel* ‘sang’) or substituted Croatian forms (**želio* instead of *želel* ‘wanted’). Voicing assimilation was recorded once (**droptinice* instead of

6 Slovenian *vrabec* ‘sparrow’ and Croatian/Serbian *vrabac* ‘sparrow’ are phonetically similar. The model likely favors the form with higher database representation; Croatian and Serbian have more speakers and more digital training content (cf. Joshi et al. 2020: 6284–6285).

drobtinice ‘breadcrumbs’). In two cases, the affricate *č* replaced the sequence *d* + *š* (**očel*/**učel* instead of *odšel* ‘left’), reflecting the actual merged pronunciation. The model also struggled with the consonant cluster *nj*, reducing it to *n* before vowels and in final positions (**luknica* instead of *luknjica* ‘small hole’, **zan* instead of *zanj* ‘for him’).[7]

In several cases, Whisper also failed to detect word-final segments (e.g., **nikomu* instead of *nikomur* ‘to no one’, **k tlon* instead of *k tlom* ‘to the ground’). In many such instances, the reason can be attributed to the speaker’s imprecise rapid articulation or strong interpretation. The substitution of unstressed *e* with *i* (**ustavljali* instead of *ustavljale* ‘were stopping’, **pritekli* instead of *pritekle* ‘flowed in’) may stem from articulatory imprecision or the higher frequency of the masculine plural form in the training data. Shifts from stressed *e* to *i* were observed in participles (**pil* instead of *pel* ‘sang’, **šumil* instead of *šumel* ‘rustled’). In the substitution of *pil* ‘drank’ for *pel* ‘sang’, which occurs in two instances, a significant semantic shift also takes place. A similar shift is also observed in the pronoun (**mi* ‘to me’ instead of *me* ‘me’). It appears twice, once as a speaker error. In several instances, we noted a change of pretonic *o* to *u*; these examples were either accurately transcribed from the speaker’s actual pronunciation (**ugovoril* instead of *ogovoril* ‘address’) or resulted from model hallucination (**upazi* instead of *opazi* ‘notices’). Conversely, unstressed *u* was sometimes replaced by *o* (**modilo* instead of *mudilo* ‘was in a hurry’). Occasional substitutions occurred (e.g., *t* for *l* in **stišal* for *slišal* ‘heard’, *k* for *h* in **prisluknile* for *prisluhnile* ‘listened’), and some words exhibited cumulative errors (**začil kao* instead of *začivkal* ‘tweeted’, showing the challenge of dual orthographic representation for the bilabial [w] within a single word).

Punctuation errors primarily included missing commas in enumerations and redundant question marks. Inappropriate capital letters also appear (**Miška* instead of *miška* ‘mouse’). Despite Whisper’s high comprehension levels, its Slovenian output remains dependent on articulatory precision and is susceptible to interference from similar languages that have more speakers or larger amounts of training data. A significant limitation is the internal inconsistency of the transcript, in which phonetically identical segments result in varied orthographic outputs.

3.3 Comparative analysis of results

The comparative analysis revealed that, as expected, the recognition accuracy was slightly higher for Czech, reaching 91.2%, compared to 90.5% for Slovenian. In addition to the analysis mentioned above, we also evaluated speech recognition performance using the Word Error Rate (WER), which is defined as follows (Jurafsky & Martin 2026: 21–22):

$$Word\ Error\ Rate = \frac{Insertions + Substitutions + Deletions}{Total\ Words\ in\ Correct\ Transcript} \times 100$$

7 The phonetic realization of Slovenian *nj* remains under-researched. Draft orthographic rules (Pravopis 8.0, Art. 73) state that *nj* is pronounced as a single palatalized or hard consonant before consonants and word finally, but as a two-phoneme sequence before vowels.

We calculated the WER using a Python script, which compares the original text with the transcription in accordance with the standard methodology, identifying incorrectly converted words (substitutions), insertions, and deletions in the transcription.[8] We performed two types of calculations: first, on texts with diacritics removed (which is a more common mechanism for WER evaluation) to focus primarily on word recognition itself; and, second, with diacritics preserved, which evaluates the accuracy of the transcription including diacritical marks. For the Slovenian version, it turned out that there was no difference between the two approaches (WER = 11.2%). In Czech, however, the WER was lower (better) for the version without diacritics (WER = 7.4%; with diacritics preserved WER = 8.3%); this is partly because Whisper sometimes failed to correctly predict vowel length.

In both analyses, a lower error rate was observed for Czech, resulting from the larger volume of training data available for this language (cf. Joshi et al. 2020: 6284–6285).[9]

Table 1 presents a comparative typology of selected errors identified during the recognition of the two texts analyzed; a complete overview of all errors identified is provided in Sections 3.1 and 3.2.

Error type	**Czech (WER = 7.4%)**	**Slovenian (WER = 11.2%)**
Incorrect word boundary	**odmazlení* (*od mazlení* ‘from cuddling’)	**lepel* (*le pel* ‘only sang’)
Transcription reflects phonetic reality (instead of orthography)	**přecíně* (*předsíně* ‘halls’)	**droptinico* (*drobtinico* ‘breadcrumb’)
Impact of sound-to-letter discrepancies (different sounds written identically, or identical sounds written differently)	**kožih* (*kožich* ‘fur’)	transcription of schwa (**vetr* ‘wind’, **vrabc* ‘sparrow’, **razperl* ‘spread open’)
Incorrect capitalization	**Zrzouni* (*zrzouni* ‘redheads’)	**črni peter* (*črni Peter* ‘old maid’)[9]

Table 1: Comparative overview of selected errors.

4 Conclusion

Comparative analysis of OpenAI Whisper—a probabilistic ASR system—reveals high output intelligibility in both Czech and Slovenian, alongside systematic deviations from standard orthography. A common feature is the frequent occurrence of word boundary errors (both merging and incorrect splitting), likely linked to prosodic characteristics such as stress realization and speech fluency. In both languages, the recognizer sometimes prioritizes phonetic reality over orthographic norms, resulting in nonstandard or nonexistent forms.

8 We would like to acknowledge Martin Zedek for developing the custom evaluation Python script used to calculate WER.

9 Performance metrics indicate that Word Error Rate significantly improves as training data volume increases (Radford et al. 2023: 22–23).

10 In one instance, Whisper wrote both components in lowercase, and in another it capitalized the proper noun.

Crucially, most errors could be mitigated by using automatic spell-checkers without manual revision. This suggests that the primary bottleneck is not speech recognition itself, but the mapping between the phonetic and orthographic layers.

Language-specific differences are evident in the nature of these errors. Czech is dominated by phonological neutralizations (specifically the *i*/*y* opposition) and inconsistent capitalization. Whisper also captures fine phonetic details (e.g., vocalic reduction). In Slovenian, the key issue is the absence of a dedicated grapheme for specific phonetic phenomena—primarily schwa [ə]—which leads to high spelling variability. In addition, Slovenian exhibits errors related to the bilabial [w] and interference from closely related languages (such as Croatian and Serbian), which benefit from larger training datasets due to their larger speaker populations.

Both transcripts showed significant internal inconsistency regarding identical phonetic segments. For an end-to-end model such as Whisper, this is expected because the system continuously selects the most probable candidates based on shifting local contexts.

Our research also aimed to determine whether the Whisper model can assist in orthoepic research on contemporary standard Czech and Slovenian. Despite differing error profiles, the fundamental limitations of ASR remain similar in both languages. Practical application necessitates combining ASR with post-processing tools or models that better integrate phonetic and morphosyntactic information.

In summary, AI-based tools can serve as highly effective aids in orthoepic research. However, currently available systems that we have personal experience with still necessitate significant cooperation from an erudite linguist or phonetician.

The preparation of this article was financed within the long-term development program of the Czech Language Institute of the Czech Academy of Sciences (RVO no. 68378092) and the infrastructure program P6-0038, which is financially supported by the Slovenian Research Agency (ARIS).

References

Boersma, P., & Weenink, D. (2026). *Praat: Doing phonetics by computer* (Version 6.4.65) [Computer program]. https://praat.org

Bruin, J. (2026). *MacWhisper* (Version 13.17). https://macwhisper.app

Graham, C., & Roll, N. (2024). Evaluating OpenAI's Whisper ASR: Performance analysis across diverse accents and speaker traits. *Jasa Express Letters*, *4*(2), 025206. https://doi.org/10.1121/10.0024876

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the Association for Computational Linguistics* (pp. 6282–6293). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.560

Jurafsky, D., & Martin, J. H. (2026). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models* (3rd ed.). Online manuscript, released January 6th, 2026. https://web.stanford.edu/~jurafsky/slp3

Pravopis 8.0. Pravila novega slovenskega pravopisa za javno razpravo. Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. Retrieved May 5th, 2026. https://pravopis8.fran.si/

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.), *Proceedings of the 40th International Conference on Machine Learning* (ICML 2023) (pp. 1–27). https://proceedings.mlr.press/v202/radford23a/radford23a.pdf

Skarnitzl, R., & Machač, P. (2012). Míra rušivosti parazitních zvuků v řeči mediálních mluvčích. *Naše řeč*, *95*(1), 3–14. https://asjournals.lib.cas.cz/api/img?uuid=uuid:d0c0b0a4-e687-41a1-afdc-b90e3d921ab9&kramerius_version=k

Žgank, A., Sepesy Maučec, M., & Donaj, G. (2020). Razpoznavalnik tekočega govora UMB Broadcast News 2020: Prehod na arhitekturo z nevronskimi mrežami. In D. Fišer & T. Erjavec (Eds.), *Jezikovne tehnologije in digitalna humanistika* (pp. 117–123). Inštitut za novejšo zgodovino. https://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Proceedings.pdf

Slovenščina gimnazijskih spisov iz časov pouka na daljavo

DOI: 10.4312/Obdobja.45.389-395

Marko Trobevšek, Gimnazija Poljane, Ljubljana; marko.trobevsek@guest.arnes.si

V prispevku skušam prikazati posebnosti dijaške slovenščine, kot jih je razbrati iz spisov iz časa, ko je zaradi bolezni covid-19 pouk potekal na daljavo. Posvečam se odstopanjem od norme oz. rabe, zlasti takšnim, kakršnih učitelji v preteklosti nismo bili vajeni. Primeri so v grobem razvrščeni in komentirani, ob njih pa se odpirajo širša vprašanja v zvezi s presojo inovacij – od iskanja razlogov zanje do domnev o možnosti, da se usidrajo in spremenijo podobo slovenskega jezika.

jezik dijakov, odstopanje, slovenščina, angleščina, vpliv

In this article, I try to show the peculiarities of students' Slovene, as written in essays during the years, when classes were held remotely due to corona virus. I focus on deviations from the norm or usage, especially those that teachers were not used to in the past. The examples are roughly classified and commented on. They raise broader questions regarding the assessment of innovations – from finding reasons for them to assumptions about the possibility of them taking root and changing the image of the Slovene language.

high school student's language, deviation, Slovenian, English, influence

Jeziki, tako kot vlade, po naravi stremijo k izroditvi.
(Samuel Johnson, Uvod k Slovarju angleškega jezika, 1755)

1 Korpus, pristop

Če je v vsakem hudem kaj dobrega, potem je ugodna posledica pouka na daljavo ta, da je slovenščina dijaških spisov iz časov razsajanja bolezni covid-19 ostala bolje dokumentirana kot sicer, ker so jih avtorji in pregledovalci morali pošiljati po elektronski pošti. Anonimno navajanje primerov, nabranih iz domačih in šolskih spisov drugih, tretjih in četrtih letnikov gimnazije v letih 2020 in 2021, verjetno ni etično sporno; mimo tega – kot kažejo tudi občasni javni odmevi na posamezne pojave v sodobni slovenščini – nobenega od prikazanih odstopanj ni mogoče pripisati zgolj posamezniku, kaj šele vrsti šole, ampak so prej odsev splošnega razvoja, na kar sem že opozoril.[1]

1 Ob primerih iz poslanic predsednice države sem pokazal, da se z vidika odstopanj od eksplicitne norme manjšajo razlike med slovenščino v dijaški in reprezentativni rabi (Trobevšek 2025: 382–383). Dodajam soroden primer: slavisti so svoje srečanje in zbornik leta 2023 naslovili *Naslavljanje raznolikosti v jeziku in književnosti*. V SSKJ[2] glagola 'naslavljati' v pomenu 'obravnavati' ni, v svetovalnici ZRC SAZU pa je bila takšna raba odsvetovana (Dobrovoljc 2018).

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

Zanimajo me odstopanja, kakršnih v preteklosti nismo bili vajeni[2] in ki so toliko razširjena, da je med njimi mogoče prepoznati istorodne pojave ter jih obravnavati po skupinah. Primeri zadevajo ne toliko pravopis kot besedišče s frazeologijo, skladnjo in delno oblikoslovje; kajpak pri razvrščanju prihaja do nedoslednosti in prekrivanj. Za vsako skupino takoj za naslovom poglavja (ki je tudi že njegova najsplošnejša razlaga) navedem primere. V oklepajih v prvem poglavju opozarjam na izvor izrazov, torej na tisto, kar je največkrat kalkirano, v naslednjih poglavjih pa navajam nevtralne možnosti ubeseditve. Kot najvidnejšo skupino novih pojavov prepoznavam tiste, do katerih prihaja pod vplivom angleščine.

2 Ko znam angleško

(1) *Tudi ko Kreon pokop Polinejka dovoli,* ***gre*** *Antigona* ***proti njegovi volji****.* (goes against his will)

(2) *Ta naposled vstopi v prostor, kjer ga župnik ogovori in mu reče, da so* ***njegovi trudi*** *zaman, in da naj pusti ljudi na miru.* (his efforts)

(3) *Žena mu je bila nezvesta, najstarejši sin Tone se ga je izogibal in njegov, kot bi rekli prej, tovariš, ga je osramotil, ne samo s tem, da je* ***priznal k poljubu*** *z njegovo ženo* […]. (confessed to kissing)

(4) *Pomembna prelomnica v drami je tudi tam, kjer se Ismena odcepi od njune* ***skupne misije*** […]. (common mission)

(5) *Župnik ima v primerjavi s Kreonom tudi razvidnejše bolj preproste* ***motivacije****.* (motivations)

(6) *Obe z Antigono sta* ***posvečeni*** *njunima družinama* […]. (dedicated)

(7) *Efeška vdova se potem kaj hitro spreobrne in pomaga vojaku, ki je postal že njen ljubimec. Dovoli mu, da na križ* ***manjkajočega*** *trupla obesi njenega moža* […]. (the missing body)

(8) *Na to opozori že* ***originalen*** *naslov Ibsenove drame Hiša lutk* […]. (the original title)

(9) *Lahko bi rekli, da si je nekaj časa pred svetom zatiskal oči, saj so bile* ***rdeče zastave*** *okoli njega že od majhnega.* (red flags)

(10) *Josef je vedel, da s tem obiskom starega prijatelja ogroža vse kar ima, od službe do stanovanja in* ***na koncu dneva*** *celo morda življenja svoje družine.* (at the end of the day)

3 Ko besede nočejo pomeniti tega, kar bi jaz rad/a

(11) *Tebanci bodo na strani Antigone, preprosto bi jo lahko pomilostil in doumel, da je ravnala pravično.* ***Lakomnost*** *ga zaslepi in nerazumno jo obsodi na smrt.* (oblastnost)

(12) *Ponekod v današnjem svetu se vse bolj uveljavljajo pojavljajo voditelji, ki si želijo* ***podrediti*** *oblast le v svoje roke* […]. (dobiti)

(13) *Hana v drami Dogodek v mestu Gogi mora podobno kot Antigona mora tudi sama pokopati svoje* ***okostnjake****.* (mrtve)

2 Zanimajo me bolj odstopanja od implicitne, »notranje jezikovne norme«, kot jo je imenoval Marko Stabej (2008: 89) in ki se delno nanaša na očitne zakonitosti, ki jih ni treba zajeti v pravila. Bolj z vidika slovničnih in pravopisnih napak je o gimnazijskih spisih pisala Irena Krapš Vodopivec (2023).

(14) *Antigono **obsodi** za blazno.* (razglasi)
(15) *Obe dejanji se mi zdita smiselni, ker pa obe živita v izrazito patriarhalni družbi, je bilo Antigonino dejanje kaznovano, Hanino pa bi po vsej verjetnosti lahko predstavili kot **poskusni umor*** […]. (poskus umora)
(16) *Kreon je zato moral delovati **nepristojno** in neomajno.* (zunaj svojih pristojnosti/nedostojno?)
(17) *Vsi očetovi grehi so Izidorja **razpolovili** in mu izoblikovali razdvojen, neodločen in šibek značaj*. (razdvojili)
(18) *Takrat je bil Khallan že bolj star in odločil se je, da bo **uvažal** vino iz Vipave.* (vozil)
(19) *Na koncu pa so le **razčistili** Agatino ime.* (oprali)

4 Ko sem utrujen/a in nezbran/a, ko se mi mudi

(20) *Zato se z **Antigonovim** dejanjem lahko le delno poistovetim.* (Antigoninim)
(21) ***Linda** bi lahko pustila, da Krogstad od Torvalda zahteva pismo.* (Lindejeva)
(22) ***Vzroki** za njihovo odločitev **so** verjetno občutek slabe vesti in želja po izkazovanju spoštovanja mrtvim.* (Vzroka, sta)
(23) *Vsak **poslednji** dan, ko sestri iščeta brata* […]. (naslednji)
(24) *Ljudstvo je strah, saj bi Antigonina resnica pomenila **vladarjevo zmotljivost in šibkost, ki ga** v času po vojni zelo potrebujejo.* (zmotljivost in šibkost vladarja, ki ga …)
(25) *Zaradi težnje po tem da **vzdrži** svojo avtoriteto in **pohlepom** začne iskati nesmrtnost.* (ohrani, pohlepa)
(26) […] *ampak je s pozicije moči samo upošteval svoje mnenje, ki je **mejilo** na njegovi oblasti in moči.* (temeljilo)
(27) *Prav tako **jima je podobna samotnost** v dejanjih.* (jima je skupna samotnost)
(28) *A v vsakem primeru njeno globoko prepričanje kaže na pomembnost **o dostojnem in spoštljivem ravnanju človeških ostankov**.* (dostojnega in spoštljivega ravnanja s človeškimi ostanki)
(29) *Zaradi Izidorjevega posebnega odnosa z **mati**, se mi zdi, da je eden izmed dogodkov, ki ga je najbolj zaznamoval, smrt **mati**.* (z materjo, smrt matere)

5 Ko sem ustvarjalen/ustvarjalna

(30) *Hana je v svoji zamrznjenosti ter **bojaznosti** nezmožna premagati svojo preteklost* […]. (bojazljivosti)
(31) *Pokop brata ima zanjo **globjejši** pomen.* (globlji)
(32) *Vedela je za posledice, ki **se ji bodo storile**, če nadaljuje z dejanjem.* (jo bodo doletele)
(33) *V tem lahko **potegnem rdečo črto in jo na drugem koncu povežem** z glavnim likom Porkasveta – Vojcem.* (prepoznam podobnost)
(34) *Zaradi nepoznavanja bolezni sama ne bi **pripomorala*** k reševanju problema. (pripomogla)
(35) *Rastignac skozi celotno knjigo poskuša **splezati po lestvi** v pariške visoke **družbe**.* (se skuša povzpeti po družbeni lestvici, v pariško visoko družbo)
(36) *Slednji vzpostavi skozi knjigo še posebno močno vez **z soimenovancem** knjige* […].

(z naslovnim junakom (očetom Goriotem))

(37) ***Temu je seveda res**, vendar tisti, ki pa to povejo* […]. (To je seveda res)

(38) *Za Izidorja bi bilo pričakovati »običajno«, da bo bi verjel v »svojega človeka« ter **v svoje šeste čute**, toda tukaj je ponovno imel njegov ponos enega od glavnih vplivov* […]. (v svoj šesti čut)

(39) *Torej se zelo dobro vidi da noče, da ga Bog in **Bogoslužne osebe** vidijo s nekom, ki ni katolik.* (duhovniki)

6 Ko povem kar po domače

(40) ***Kreonu gredo vse dlake pokonci*** […]. (Kreon se zgrozi)

(41) *Prav je tudi, da goji nekaj ljubosumja do sestre, saj ji je ta **speljala** idejo* […]. (prevzela, ukradla zamisel)

(42) *Tudi, ko se Ismena odtrga od nje, **jo to ne briga**, dejanje nadaljuje do konca.* (ji za to ni mar)

(43) ***Probal** je priti v guinessovo knjigo rekordov* […]. (Poskusil)

(44) *Bil je **zatežen** in nič mu ni bilo prav.* (zlovoljen)

(45) ***Primo** ima tako, prav tako kot Tantal, ob svoji žeji pred seboj vodo, ki je izven njegovega dosega.* (Primo Levi)

(46) *Zaradi nesrečne ljubezni **nastradata** ženski.* (sta prizadeti)

(47) *Še posebej pa nisem morala **prebavljati** Izidorjeve osebnosti,* šel mi je na živce […]. (prenašati, spravljal me je ob živce)

(48) *Takih odnosov si, **služba ali ne, šef ali ne**, po mojem mnenju pač ne bi smeli nikjer privoščiti.* (v službi ali zunaj nje, s šefom ali s kom drugim)

(49) *Nihče ji ni **rabil pomagati**, da je Polinejka pokopala* […]. (Nikomur ji ni bilo treba pomagati)

7 Prava mala jezikovna katastrofa?[3]

Zglede zadnje skupine, torej preveč pogovorne rešitve, je najlažje presoditi. Jezikoslovje zadnjih desetletij bi jih obravnavalo kot posledico liberalne presoje *primernosti*, kot izraz pomanjkljivega jezikovnega čuta oz. zavedanja o potrebi po tem, da se mora beseda prilagoditi okoliščinam. Vse kaže, da je mladina največ v stiku s tistim (obsežnim) delom javno dostopne slovenščine, ki je v odnosu do pravil bolj sproščena, in da ob tem ločnica med *javnim* in *pravilnim* oz. *ustreznim* postaja še manj razvidna.[4] Opraviti imamo torej s pojavom, ki spremlja demokratizacijo knjižnega jezika.[5]

3 Apokaliptični naslov povzemam po Janezu Gradišniku, ki je leta 1993, po osamosvojitvi, menil: »Ne da bi javnost opazila, kaj se dogaja, se nam je v nekaj kratkih letih v naš jezik in v naše življenje zasejala takšna množica angleških (in še nekaj malega drugih) besed, da ne pretiravamo dosti, če rečemo, da gre za pravo malo jezikovno katastrofo« (Gradišnik 1993: 10).

4 Problem je v intervjuju za *Delo* pred leti osvetlil Kozma Ahačič: »Pojav interneta je povzročil, da lahko javno v knjižnem jeziku piše mnogo več ljudi kot kadar koli prej. Zato ne moremo reči, da je knjižni jezik slabši kot kdaj v preteklosti. Kakovostnih besedil v kakovostnem knjižnem jeziku je gotovo več kot pred desetletji, so tudi neprimerno bolj raznovrstna. Je pa res, da so skrita v množici drugih besedil, saj danes lahko javno objavlja svoja besedila prav vsak« (Roglič 2022).

5 Opozorimo še na pomen družbenega konteksta oz. kulturnega izročila: različna okolja so različno

Če nadaljujem proti začetku, lahko združim naslednje tri skupine: dijaki občasno uporabljajo besede v napačnem pomenu, nevede izumljajo nove, prenavljajo fraze, iščejo lastne slovnične rešitve. Včasih se pri tem, zlasti ko ponesrečeno povežejo besede v stavke in stavke v povedi, približajo meji razumljivosti. Poglejmo še enkrat zgled (25). Da razumemo, kaj je mišljeno z *vzdrži*, zahteva nekaj pozornosti, težje pa je prepoznati, da bi *pohlep* moral biti v rodilniku, sploh ker manjka vejica. Iz izkušenj vem, da dijaki včasih končnic ne uporabijo smiselno, ampak se ravnajo po neki bližnji besedi v besedilu; menim, da je avtor/ica prepisal/a *m* iz besede *tem*, ki je bližje od predloga, po katerem bi se moral ravnati sklon. Pojav ni niti pogost niti neopazen.

Ugotovitve, da nove tehnologije ne vplivajo ugodno na zbranost, zmožnost branja in razmišljanja, verjetno ni več treba dokazovati;[6] isto velja za povezanost vsega trojega z obvladovanjem pisanja. Šolniki kajpak razmišljamo o tem, kaj lahko storimo za pišočo mladino; ker potrebuje več stika z dobro slovenščino, bi bila ena od možnosti ta, da jo skušamo navdušiti za branje dobrih knjig. Tu vidim pot – nekaj bi lahko dosegli, če bi domače branje obogatili z izborom iz sodobne literature, slovenske in prevedene.

Na vprašanje iz naslova poglavja odgovarjam z *ne*, kar bom v naslednjem poglavju skušal utemeljiti v zvezi z vplivom angleščine, na tem mestu pa bi rad opozoril na nek teoretični vidik obravnavanih odstopanj. Guy Deutscher (2006: 63–73) je, ko je razpravljal o jezikovni evoluciji, povzel jezikoslovna spoznanja o globljih gibalih sprememb v jeziku, in sicer v treh principih, ki veljajo mimo medjezikovnega vplivanja oz. nad njim – *ekonomičnost*, *izraznost* in *analogija*. Ob primerih iz spisov moramo ne glede na to, koliko so nam smešni ali zoprni, ugotoviti, da je del njih tako ali drugače v skladu z normalnim jezikovnim razvojem.[7] Vsaka poenostavitev (kot »z mati«) izraža težnjo po ekonomičnosti; kdor je zapisal: »V tem lahko potegnem rdečo črto in jo na drugem koncu povežem z glavnim likom Porkasveta – Vojcem,« je nedvomno nekaj izrazil; kdor je pridevnik stopnjeval v »globjejši«, je izjemo poskusil ukrotiti s pravilom.

8 Angleščina

Da nove tehnologije povečujejo vpliv angleščine na slovenščino, je toliko očitno, da pretres primerov prve skupine, anglicizmov, rajši začenjam s temo, ki sem jo načel v prejšnjem poglavju, z branjem knjig. Že pred desetletji se je dogajalo, da me je kak/a dijak/inja prosil/a, ali bi knjigo za domače branje angleško pišočega avtorja smel/a brati v izvirniku. Načeloma nisem nasprotoval, vendar se je pri preizkusu zmeraj izkazalo, da ima prosilec oz. prosilka v angleščini daleč prešibek besednjak in da pomene besed ves

tolerantna do jezikovne sproščenosti. Ko je tiskovni predstavnik Bele hiše označil izstrelitev satelita kot *strašno kul* (angl. *awesomely cool*), to ni bilo moteče, pri Nemcih ali Francozih pa si je kaj takšnega »v višjih socialnih registrih« težje predstavljati (povzemam in navajam po Fischer 1999: 177). Dodajmo: pri Slovencih tudi.

6 Nekaj zadevnih ugotovitev povzema razprava Zakaj je pomembno branje na višji ravni v reviji *Knjižnica* (Schüller-Zwierlein 2024).

7 Ko govorimo o evoluciji, moramo vsaj teoretično upoštevati možnost, da se jezikovna napaka deduje kot mutacija in se prime celo tedaj, ko ni koristna. Kot meni Steven Mithen (2024: 54–55), to ustreza genetskemu zdrsu. O njem vemo, da je lahko koristen, nevtralen ali škodljiv.

čas po malem ugiba iz konteksta, in nisem dovolil. Pri tem je ostalo do danes – še vedno ne dovolim, toda delno iz drugih razlogov. Dijaki sicer vse bolje znajo angleško, a, kot kažejo primeri iz spisov, vse bolj angleško tudi razmišljajo.[8] (Vem pa, da je branje knjig v angleščini med dijaki razširjeno in da ga doživljajo drugače kot branje leposlovja za šolo – z več užitka.) Kako nevarno je to, presojamo različno.

Jeziki so od nekdaj v stiku in vplivajo drug na drugega, se oplajajo, in dvojezičnost posameznika je že dolgo marsikje običajno stanje. Tisti jeziki, ki so bili v zgodovini v kateremkoli smislu močnejši (uradni, z več govorci, prestižni, prevladujoči, predpisani, državni, privilegirani, uglednejši, sveti …), so vedno bolj vplivali na druge. Z globalizacijo je sicer nastopilo novo stanje, ko je *lingua franca* prvič svetovni pojav, in posledice niso zares predvidljive,[9] vendar v slovenski zgodovini obstaja vzporednica, namreč z nemščino in s strahovi, ki jih je vzbujala v 19. stoletju.

Za Levstika so leta 1858 napake pisanja pomenile tisto, kar danes nastaja pod angleškim vplivom, v njegovem času pa je pod nemškim. Posvetil se ni toliko prevzetim besedam kot prevzetim vzorcem, besedotvornim in skladenjskim. Odsvetoval je besedi *drevored* in *kolodvor* (»res okrogla, sama na sebi lepa beseda, pa vender napak zložena«), ravno tako *predstaviti*, *predložiti*, *predlog* in *predpis* (»čisto napak«). Priporočil je besedi *volkosnedež* in *letorasl*, namesto *parobrod* pa predlagal *parogonja* (ker imajo Srbi besedo *vetrogonja* (mlin)). Obregnil se je ob Pohlina (»Ko bi ta knjiga ne imela spredaj letnic, gotovo bi mislil, da je pater Marko njeni oče; le on je znal barbariti …«). Kritiziral je še ne deset let mrtvega Prešerna (»ima obilo tacih napak«); za zgled je postavljal kmečki jezik, a ne nekritično (»Naslednji stavek gotovo ni za nič, akoravno ga je govorila baba, domá iz kraja, kjer je slovenščina dobro čista …«). In, končno, težko bi rekli, da ni bil črnogled (»In tacih napák ni kraja ne konca, kakor bi ne imeli razun slovnic užé dovolj druzih pisarjev, ki nam jezik pridijo«).

Napake slovenskega pisanja (1858) so poglobljena in tehtna jezikoslovna študija, ki izhaja iz drugih okoliščin, a izkazuje podobne strahove, kot jih marsikomu vzbuja današnje stanje. V tistem, v čemer se je Levstik motil ali pretiraval, so njegovi pogledi v nekem smislu najbolj aktualni, tolažilni, ker sejejo dvom o tem, koliko je v jezik smiselno in sploh mogoče posegati na tak način ter koliko ga prevzemanje ogroža.[10] Kdor

8 To ni tako enostavno. Prevzemanje včasih dejansko izkazuje razmišljanje v dveh jezikih, denimo v slengu, ki je prostor radožive ustvarjalnosti. Zabaven primer je izjava, ki sem jo iz mladih ust slišal nedavno, »Vodnjak, vodnjak, vodnjak« in se poigrava z angleškim »Well, well, well«. Tak humor kaže na zavest o jeziku, ki ni samoumevna. To, da govorec izraz utegne poznati iz priljubljene igre *Hogwarts Legacy*, ni nič hudega.

9 Kot je zapisal Nicholas Ostler (2010: 5): »V takšnih časih razumljivo obstajajo dvomi o tem, da bi nas preučevanje zgodovine moglo voditi na poti v prihodnost.« Isti avtor po drugi strani ne dvomi, da bo angleščina v dogledni prihodnosti nehala biti svetovni jezik, do česar bo pripeljalo novo geopolitično razmerje moči ob pomoči novih tehnologij: »Vendar je lingua franca, tudi najbolj univerzalna, breme. Prenašamo ga le tedaj, ko ga upravičuje neko višje dobro. Če tega ne bo, ga bomo ob prvi priložnosti odložili. Nekega dne bomo tudi angleščino, poslednjo linguo franco, ki služi večjezičnemu svetu, odložili. Potem bo vsakdo govoril ali pisal v jeziku po lastni izbiri, in svet bo razumel« (Ostler 2010: 286).

10 Ogroženost vodi v smrt. Tovrstni strahovi so sicer iracionalni, a ne tako redki, zato ne bo odveč povedati, da slovenščini statistično kaže dobro. Po razvrstitvi, ki jo povzema Harald Haarmann

zna nemško, more – če hoče – hitro uvideti, koliko so slovenska frazeologija ter besedotvorni vzorci, in še kaj, prepojeni z nemškimi vplivi; kdor zna angleško, lahko opazi nekaj podobnega, eden in drugi pa lahko pojav razložita čisto različno, bodisi kot znamenje ogroženosti bodisi kot strategijo preživetja. Kam se nagibam avtor prispevka, je, menim, dovolj jasno.

Viri in literatura

Ahačič, K. (2020). *Kozmologija*. Založba Goga.

Dobrovoljc, H. (2018). Glagol »nasloviti« v pomenu ‘obravnavati’. *Jezikovna svetovalnica*. https://svetovalnica.zrc-sazu.si/topic/3175/glagol-nasloviti-v-pomenu-obravnavati

Deutscher, G. (2006). *The unfolding of language: The evolution of mankind's greatest invention*. Arrow.

Fischer, S. R. (1999). *A history of language*. Reaktion Books.

Gradišnik, J. (1993). *Slovensko ali angleško?: Priročnik za dobro slovenščino*. Mohorjeva družba.

Haarmann, H. (2010). *Weltgeschichte der Sprachen: Von der Frühzeit des Menschen bis zur Gegenwart*. C. H. Beck.

Krapš Vodopivec, I. (2019). Napake dijaškega pisanja: Vrednotenje slovnično-pravopisnih napak v šolskih esejih. V H. Tivadar (ur.), *Slovenski javni govor in jezikovno-kulturna (samo)zavest* (Obdobja 38, str. 513–521). Znanstvena založba Filozofske fakultete Univerze v Ljubljani. https://doi.org/10.4312/Obdobja.38.513-521

Levstik, F. (1858). Napake slovenskega pisanja. *Kmetijske in rokodelske novice*, *16*(1). https://sl.wikisource.org/wiki/Napake_slovenskega_pisanja

Mithen, S. (2024). *The language puzzle: How we talked our way out of the Stone Age*. Profile Books.

Ostler, N. (2010). *The last lingua franca: English until the return of Babel*. Penguin.

Roglič, T. (2022, 9. september). Obljubljajo kvaliteto, pa nimajo za vejico. *Delo*. https://www.delo.si/kultura/razno/obljubljajo-kvaliteto-pa-nimajo-za-vejico

Schüller-Zwierlein, A., Mangen, A., Kovač, M. in van der Weel, A. (2024). Zakaj je pomembno branje na višji ravni. *Knjižnica*, *68*(1), 43–73. https://doi.org/10.55741/knj.68.1.3

Stabej, M. (2008). Kako pa govoriš!? (Spis o normi in normalnem). V P. Vitez (ur.), *Spisi o govoru* (str. 87–100). Znanstvenoraziskovalni inštitut Filozofske fakultete.

Trobevšek, M. (2024). Al prav se reče kákšen ali kàkšen: O gimnazijskem pravorečju. V S. Štumberger (ur.), *Predpis in norma v jeziku* (Obdobja 43, str. 381–387). Založba Univerze v Ljubljani. https://doi.org/10.4312/Obdobja.43.381-387

(2010: 342), sodi v skupino »milijonskih« jezikov, se pravi tistih, ki jih govori vsaj milijon ljudi, in njihovi govorci predstavljajo 90 % svetovnega prebivalstva. Leta 2010 sta kontinuirano naraščali tako število teh jezikov kot število njihovih govorcev. Glede prevzemanja je zgovoren podatek, da ima angleščina, jezik, ki naj bi ogrožal največ drugih, približno pol besednjaka prevzetega, največ iz normanske francoščine in latinščine, ki nista germanski.

Linguists' Attitudes toward the Use of AI in Language Consulting

DOI: 10.4312/Obdobja.45.397-406

Urška Vranjek Ošlak, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana;
urska.vranjek@zrc-sazu.si

Hana Mžourková, Ústav pro jazyk český, Akademie věd České republiky, Praga;
mzourkova@ujc.cas.cz

Nedavni razvoj na področju umetne inteligence je v aktivnosti jezikovnega svetovanja uvedel novo vrsto orodij. Članek predstavlja raziskavo stališč jezikoslovcev do uporabe umetne inteligence pri jezikovnosvetovalnem delu. Med zaposlenimi v jezikovnih svetovalnicah v več evropskih državah je bila izvedena anketna raziskava. Njen cilj je prispevati k bolj poglobljenemu razumevanju vloge, ki bi jo lahko umetna inteligenca imela pri nadaljnjem razvoju dejavnosti jezikovnega svetovanja.

uporabno jezikoslovje, jezikovno svetovanje, umetna inteligenca, jezikovna kultura, anketna raziskava

Recent advancements in artificial intelligence have introduced a new category of tools in language consulting. This study investigates linguists' attitudes toward integrating AI into language consulting practice. A questionnaire was distributed to staff members of language consulting services across several European countries. The research seeks to provide a nuanced understanding of the potential role of AI in the future development of language consulting activities.

applied linguistics, language consulting, artificial intelligence, language culture, questionnaire

1 Introduction

Language consulting is offered as a free public service in several European countries through national research institutions. In Slovenia, the Language Consulting Service is part of the Fran Ramovš Institute of the Slovenian Language, and in the Czech Republic the Language Consulting Center operates within the Institute of the Czech Language.[1] Other language consulting centers (hereafter LCCs) affiliated with academic or university institutions are present in Bulgaria, Croatia, Poland, Slovakia, the Baltic and Nordic countries, and other regions. For an overview of LCCs in the Slavic countries, see Štěpánová (2022) and, for institutional language consulting in Europe, see Ludányi (2026).

1 For a more detailed account of the activities of the Slovenian and Czech language consulting centers, see Mžourková (2024).

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

LCCs address a broad spectrum of linguistic problems. They frequently function as specialized research centers, with staff gathering information from language users regarding their needs, particularly concerning language handbooks, their content, and form, as well as attitudes toward language norms and linguistic authority, which the LCC embodies for its users.[2] LCCs play an important role in science communication and the dissemination of scientific knowledge. They serve as intermediaries between experts and the general public, facilitate the practical application of research findings, and contribute to advancing education in society. In addition, LCCs mirror broader social changes, both in the nature of contemporary linguistic problems and in the methods and tools employed to formulate and address language queries.

During the twentieth century, queries were typically addressed through traditional mail or telephone, with printed books serving as primary resources for preparing responses. Since the 1990s, advancements in information technology have transitioned language consulting into the digital environment of the internet. Staff at LCCs now use electronic databases, dictionary portals, and expanding language corpora. The rapid development of artificial intelligence, language models, and chat-based systems currently marks another significant milestone in language consulting. Artificial intelligence is exerting a broad influence on society, affecting both general language users and LCC staff. In current language consulting practice, users increasingly present LCC staff with answers to language queries, often entirely incorrect, generated by AI systems.[3] Furthermore, some LCC staff are actively engaging with AI by evaluating the quality of its responses to language-related queries.

Although artificial intelligence tools have been utilized and examined within linguistics for some time (see Jakubíček & Rundell 2023; and an overview and comparative analysis in de Schryver 2023), to our knowledge, only one study in language consulting (i.e., Vranjek Ošlak 2025) has addressed both the potential applications and current limitations of AI in this context. The scarcity of relevant data motivates the present investigation into the use (or non-use) of AI in language consulting and language consultants' attitudes toward these technologies.

Our research, the course, methodology, and results of which are presented below, seeks to improve this situation and to contribute to a better understanding of AI within European language consulting. Specifically, our aim was to determine whether and how the staff of European LCCs use artificial intelligence tools and what attitudes they adopt toward them.

2 Several publications provide information about users' language problems as well as about the users themselves, such as Dobrovoljc et al. (2022) and Mžourková and Beneš (2024).

3 The term *AI* is used throughout this article to refer to generative artificial intelligence, including large language models and AI tools that make use of such models.

2 Methodology

2.1 Questionnaire structure

The primary research method was an online survey using a questionnaire. This section details the questionnaire's design, structure, and administration, which aimed to capture respondents' perspectives and practices regarding the use of AI in language consulting. The questionnaire had four main sections: the first collected demographic information, the second addressed the respondent's language consulting context, the third examined the use of AI, and the fourth focused on attitudes toward AI use and related practices. For participants that indicated non-use, the questionnaire ended after the third section because subsequent questions were relevant only to users of such tools. This structure prevented respondents from encountering irrelevant items.

The survey was developed using 1KA,[4] an open-source web application for designing, distributing, and analyzing online questionnaires, developed and maintained by the Academic and Research Network of Slovenia (ARNES).

2.2 Timeline of the survey

Given the questionnaire's focus, only staff members of European LCCs were targeted as respondents. Accordingly, the questionnaire was not distributed via websites or social media; instead, invitations to participate were sent by email to specific preselected institutions. Language consulting centers from fifteen European countries were contacted.[5] To ensure data comparability, the survey was limited to staff employed at research and government institutions, and universities. Recipients were also asked to assist with further distribution of the questionnaire (to employees of other LCC institutions within their personal or professional contacts), which resulted in the inclusion of data from additional countries.

Before primary data collection, a preparatory phase was conducted, during which two colleagues, also staff members of our respective LCCs, provided detailed feedback on the initial questionnaire version. Their comments were incorporated into the final version. The main data collection period was from March 2nd to 25th, 2026.

3 Survey results

3.1 Respondent demographics

Fifty respondents employed at LCCs across Europe completed the questionnaire in full or in part.[6] The sample comprises participants from a diverse range of countries, encompassing both Slavic and non-Slavic linguistic contexts and various institutional traditions. Although the distribution across individual countries is uneven, the data offer

4 Available at https://1ka.arnes.si/.

5 Bulgaria, Croatia, the Czech Republic, Denmark, Estonia, Finland, Hungary, Latvia, Lithuania, Norway, Poland, Serbia, Slovakia, Slovenia, and Sweden.

6 Not all respondents answered every question; some questions were answered by fewer respondents, primarily because the questionnaire terminated after the third section for a subset of participants.

a sufficiently broad empirical foundation for a general overview of current practices. The number of respondents from each country is as follows: Bulgaria (eleven), Croatia (one), the Czech Republic (eleven), Estonia (three), Finland (two), Hungary (four), Italy (one), Latvia (two), Norway (three), Slovakia (two), Slovenia (nine), and Sweden (one).

Regarding institutional affiliation, given the previously described limitations of institution types, most respondents (thirty-six) are employed at LCCs established by national language institutes. Fewer respondents are affiliated with government agencies (seven) or LCCs established by universities or academies (six). One respondent is employed at an LCC established by a private research institute.

The age distribution of respondents reflects a heterogeneous professional community, with ages ranging from twenty-five to sixty-four (Table 1). This diversity aligns with the reported range of professional experience in language consulting, which includes both early-career consultants and highly experienced practitioners (Table 2). The presence of these groups indicates both continuity and gradual generational renewal within the field.

Age	n	Percentage
Under 25	0	0
25–34	15	30
35–44	14	28
45–54	10	20
55–64	8	16
65 and over	3	6
Total	50	100

Table 1: Respondents' age structure.

Years	n	Percentage
0–4	11	22.4
5–9	9	18.4
10–19	15	30.6
20 and over	14	28.6
Total	49	100

Table 2: Respondents' language consulting experience.

Consultation formats vary, with respondents indicating engagement across multiple modes. Written responses, especially those delivered via email or online platforms, are most prevalent, accounting for approximately two-thirds of responses based on a multiple-choice question. Although online and social media chats are not very common, their frequency appears to be increasing. Paper letters are now extremely rare. Telephone consultations continue to be used but are less central, comprising close to 30% of responses,[7] and face-to-face interactions are infrequent. This distribution illustrates the

7 These data align with the high respondent turnout at the Czech Language Consulting Center, where inquiries are primarily handled by telephone.

ongoing transition of language consulting into digital environments; see Mžourková and Dvořáková (in press).

In terms of specialization, most respondents indicate involvement in multiple linguistic areas rather than focusing on narrowly defined subfields. Core areas include orthography (spelling, punctuation, and capitalization), grammar, lexicology, and stylistics, frequently addressed in combination. These findings support the generalist profile commonly attributed to language consulting, in which consultants are expected to address a broad spectrum of user queries.

Workload indicators reveal substantial variation among respondents (Table 3). The number of responses prepared per month ranges from low volumes to several dozen cases, with rare instances of two hundred or more queries. These differences are primarily influenced by institutional size and consultation format.

Responses per month	n	Percentage
1–4	10	20.8
5–19	6	12.5
20–49	15	31.3
50–100	9	18.8
Over 100 (specify)	8	16.7
Total	48	100

Table 3: Language consultants' workload.

Open-ended responses that elaborate on the workload question provide further context for interpreting these patterns. These responses indicate considerable variation in both the volume and organization of consulting work. A persistent challenge is quantification because consultations are frequently collaborative, not always systematically documented, and vary in complexity. Several respondents note that they primarily address specialized or complex queries rather than routine or general consultation tasks.

3.2 Information regarding the use of AI tools

The survey included questions designed to obtain both general and specific information about the language consultants. Respondents were asked about their home institution, their individual practices in language consulting, and their attitudes toward the use of AI in these activities. When asked about the use of AI tools in language consulting work, twenty respondents reported using such tools, and twenty-eight indicated that they do not. The respondents that indicated that they do not use AI tools were asked to explain their decision by selecting one or more of four preformulated responses (see Table 4) and/or by providing justification in the comment section of the "Other" option.

Reason	Frequency	Percentage
I do not need AI tools for my work.	8	21.0
In my opinion, AI tools are not reliable.	13	34.2
AI tools do not provide sufficiently detailed or expert answers.	12	31.6
I do not wish to state the reason.	0	0
Other (specify):	5	13.2
Total	38	100

Table 4: Reasons for not using AI tools in language consulting.

Table 4 indicates that the most common reasons for not using AI tools are concerns about their reliability (thirteen responses) and the perception that they do not provide sufficiently detailed or expert answers (twelve responses). Five respondents cited additional reasons: some reiterated the preformulated answers ("AI tools are not reliable" or "other tools were sufficient"), others highlighted the need for collective consensus within a specific LCC ("we just haven't discussed it as an option yet"), and one respondent referenced dissatisfaction with the tone and style of AI-generated content ("[AI has an] annoying tone and style").

The survey results further demonstrate that, among respondents that use AI tools, ChatGPT is the most frequently used tool (seventeen responses), followed by Gemini (ten responses). Other tools were mentioned infrequently, with MS Copilot and Claude each receiving two responses, and Perplexity and Google AI each receiving one response.

Regarding training in working with AI, eleven respondents indicated they had not had any such training, whereas eight reported having done so. This group was asked to provide further specification. Those that had received training specified participation in workshops and webinars, often organized by employers, as well as independent information-seeking motivated by personal interest. Several respondents noted that these training activities were not specifically focused on applying AI in language consulting.

Additional survey questions addressed the use of AI tools within language consulting practice. Nineteen respondents provided information on the frequency of AI tools usage in their language consulting work (in 2025). The most common frequency was several times a year but less often than once a month (eight respondents), followed by more than once a week but less often than once a day (five respondents). No respondents reported using AI tools more than once a day. According to Table 5, language consultants most frequently use AI tools for translation from foreign languages (twelve responses), also quite often for information retrieval (eleven responses), and brainstorming (ten responses). Two respondents selected the "Other" category to specify additional uses: one for formulating assertive yet polite responses, and another for structuring written responses.

Use	Frequency	Percentage
Searching for information	11	17.2
Verifying retrieved information	2	3.1
Summarizing information or references	9	14.1
Translating from a foreign language	12	18.7
Translating into a foreign language	6	9.4
Drafting written responses	6	9.4
Checking formulations (phrasing) or tone	6	9.4
Brainstorming	10	15.6
Other (specify):	2	3.1
Total	64	100

Table 5: Reported uses of AI tools in language consulting work (multiple-choice question).

Respondents were asked to evaluate the strengths and weaknesses of AI tools across specific linguistic domains, using a Likert scale with the range *very weak – weak – neither weak nor strong – strong – very strong*, as well as the options *not applicable / insufficient experience to evaluate*.

Domain	Very weak	Weak	Neither weak nor strong	Strong	Very strong	Not applicable / Insufficient experience to evaluate	Total
Phonology and pronunciation	2	4	1	1	0	9	17
	11.8%	23.5%	5.9%	5.9%	0%	52.9%	100%
Orthography (spelling)	0	6	3	2	1	6	18
	0%	33.3%	16.7%	11.1%	5.6%	33.3%	100%
Orthography (punctuation)	0	5	3	2	0	7	17
	0%	29.4%	17.6%	11.8%	0%	41.2%	100%
Orthography (capitalization)	0	4	4	1	0	8	17
	0%	23.5%	23.5%	5.9%	0%	47.1%	100%
Lexicology, terminology, semantics	0	2	5	6	0	5	18
	0%	11.1%	27.8%	33.3%	0%	27.8%	100%
Morphology	1	2	6	0	0	7	16
	6.3%	12.5%	37.5%	0%	0%	43.7%	100%
Syntax	1	3	5	0	0	7	16
	6.3%	18.8%	31.3%	0%	0%	43.8%	100%
Pragmatics	0	2	3	2	0	10	17
	0%	11.8%	17.6%	11.8%	0%	58.8%	100%
Stylistics	0	2	2	4	1	7	16
	0%	12.5%	12.5%	25.0%	6.3%	43.7%	100%
Etymology	0	2	3	3	0	8	16
	0%	12.5%	18.8%	18.8%	0%	50.0%	100%
Other (specify)	0	0	1	0	0	5	6
	0%	0%	16.7%	0%	0%	83.3%	100%

Table 6: Strengths and weaknesses of AI tools in specific linguistic domains.

Due to space constraints, Table 6 only summarizes the collected data. The category *not applicable* was selected most frequently overall. Within specific linguistic domains, the highest number of strong ratings was recorded for *lexicology, terminology, and semantics* (six responses), *stylistics* (four responses), and *etymology* (three responses). In contrast, the domains most frequently rated as weak were *orthography (spelling)* (six responses), *orthography (use of punctuation)* (five responses), and *phonology and pronunciation* (four responses).

The discussion section below examines the usefulness of AI tools regarding time efficiency and activity types. Responses to open-ended questions offer valuable insights into language consultants' attitudes toward this technology.

4 Attitudes behind the use or non-use of AI

The discussion examines the reported use or non-use of AI in language consulting. Respondents that do not use AI were asked to specify the requirements that would prompt them to consider adopting such a tool, whereas those that do use AI identified its main advantages in consulting-related activities. Both groups also described their principal reservations regarding the use of AI-based tools in language consulting. Despite starting from opposing positions, the respondents reveal a notable degree of overlap in the attitudes expressed.

4.1 Non-users of AI tools

Respondents that do not use AI in language consulting primarily express concerns regarding reliability, source validity, and domain-specific competence. AI outputs are often characterized as inaccurate and misleading or based on nonexistent sources, with particular emphasis on hallucinations and insufficient training for less-resourced languages such as Bulgarian, Czech, or Latvian. A recurring issue is AI systems' inability to address context-sensitive queries, which are typical of language consulting and require interpretation based on detailed situational and linguistic information.

Several respondents question the compatibility of AI with the field's epistemic standards. Language consulting is described as an activity grounded in verifiable sources, expert judgement, and responsibility towards users. AI is perceived as lacking transparency in both data and reasoning processes. Some responses further extend this critique to ethical and professional concerns, including the potential to diminish consultants' expertise.

The requirements identified for the acceptable use of AI include training on curated linguistic resources, explicit citation of sources, and adaptation to specific languages and terminology. For example, one respondent emphasized the importance of institutionally developed and local language models based on reliable resources such as dictionaries and grammars. This approach was argued to reduce the influence of low-quality online content in training data and to improve response reliability by limiting hallucinated or unsupported outputs.

Non-users of AI appear to be considering ways to address current AI shortcomings for use in language consulting through improvements to AI itself. However, respondents emphasize that final responsibility should remain with the human expert. Only a few respondents categorically reject the applicability of AI, arguing that it cannot meet the standards required for professional language consulting.

4.2 Users of AI tools

Respondents that report using AI tools demonstrate a nuanced yet cautious perspective. Their accounts indicate that AI is integrated into practice in a limited and controlled manner, primarily for peripheral or supportive tasks rather than addressing substantive linguistic problems. The most frequently cited advantages pertain to efficiency, such as saving time, rapid scanning of large volumes of information, and assistance with repetitive tasks. Specifically, AI is employed for drafting and reformulating responses, adjusting tone and style, summarizing information, and supporting translation or communication in foreign languages.

However, these advantages are consistently accompanied by significant reservations that closely mirror those expressed by non-users. Users emphasize that AI cannot be relied upon for accurate linguistic analysis, particularly in complex or language-specific contexts, and that hallucinated references and incorrect explanations remain persistent issues. Consequently, AI-generated outputs are subject to systematic verification and revision (post-editing), and their use is often deliberately limited.

In summary, the attitudes of both groups largely converge, with the primary distinction being the degree of acceptance of AI's shortcomings. The findings suggest that, similar to corpus tools, AI tools are likely to become part of the language consulting toolkit in the future. However, it is not anticipated that AI will replace the expert staff of the LCCs.

5 Conclusion

The findings demonstrate that AI currently serves as a supplementary tool in language consulting practice rather than a primary source of authoritative answers. Both users and non-users of AI tools agree that linguistic expertise and responsibility to the users of language consulting services remain essential. Differences are primarily observed in the degree of trust and the scope of acceptable use, especially when comparing routine tasks to expert judgement.

The survey indicates that language consultants adopt a cautious and critical approach to AI. They either refrain from using it because of concerns about its current limitations or use it selectively for clearly defined tasks. Intermediate forms of AI use therefore predominate. Respondents that use AI report restricted, task-specific applications, accompanied by systematic verification and clear awareness of their limitations. Consequently, AI is regarded as a supportive tool within a human-led process, rather than as an autonomous authority. These findings provide valuable insights for the future use of AI in language consulting. The results confirm the assumption that, for AI to

become a more trustworthy and usable tool for language consultants, it is necessary to develop specialized AI models that operate in accordance with the standards of language consulting[8] and are trained on verified data for specific languages. Further research work will extend this study within a broader European framework and examine institutional practices in greater detail. Achieving success in this research task requires international collaboration, which will continue to be pursued.

The preparation of this article was financed through the P6-0038 program, funded by the Slovenian Research and Innovation Agency (ARIS), and within the long-term development program of the Czech Language Institute of the Czech Academy of Sciences (RVO no. 68378092). We would like to thank Klára Dvořáková and Manca Černivec for providing detailed feedback on the initial questionnaire version. The research data supporting the findings of this article are available at https://doi.org/10.5281/zenodo.21345017.

References

de Schryver, G.-M. (2023). Generative AI and lexicography: The current state of the art using ChatGPT. *International Journal of Lexicography*, *36*(4), 355–387.

Dobrovoljc, H., Gliha Komac, N., Michelizza, M., Vranjek Ošlak, U., Lengar Verovnik, T., & Weiss, P. (2020). *Kje pa vas jezik žuli? Prva pomoč iz Jezikovne svetovalnice*. Založba ZRC, ZRC SAZU.

Jakubíček, M., & Rundell, M. (2023). The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? In M. Medveď, M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček, & S. Krek (Eds.), *Electronic lexicography in the 21st century (eLex 2023): Invisible lexicography* (pp. 518–533). Lexical Computing CZ s.r.o.

Ludányi, Z. (2026). Directions of research on institutional language consulting: A comprehensive overview. *Slovo a slovesnost*, *87*(1), 51–78.

Mžourková, H. (2024). Czech and Slovenian language consulting: What do they tell us about language and language users? In S. Štumberger (Ed.), *Predpis in norma v jeziku* (Obdobja 43, pp. 225–231). Založba Univerze v Ljubljani.

Mžourková, H., & Beneš, M. (Eds.). (2024). *Kdo se zeptá, ten se dozví*: *Více než 250 odpovědí jazykové poradny*. Nakladatelství Lidové noviny.

Mžourková, H., & Dvořáková, K. (in press). Výsledky průzkumu uživatelů české jazykové poradny v evropském kontextu. *Slavia*.

Štěpánová, V. (2022). Moderní jazykové poradenství ve slovanském světě, zejména v Česku: zpřístupňování dat prostřednictvím webové databáze. In J. Chomko & A. Romanik (Eds.), *Horyzonty slawistyki w XXI wieku. Stan obecny i perspektywy rozwoju* (pp. 179–194). Uniwersytet w Białymstoku.

Vranjek Ošlak, U. (2025). How effective is AI as a language consultant? In I. Kosem, M. Jakubíček, M. Medveď, K. Zgaga, Š. Arhar Holdt, T. Munda, & A. Salgado (Eds.), *Electronic lexicography in the 21st century (eLex 2025): Intelligent lexicography* (pp. 499–512). Lexical Computing CZ s.r.o.

8 Available at https://ujc.cas.cz/wp-content/uploads/2024/09/Standard-of-language-consulting_FV.pdf.

Priložnosti in izzivi za uporabnike brajeve pisave v dobi sodobne tehnologije

DOI: 10.4312/Obdobja.45.407-414

Nina Zavašnik Hofer, Ljubljana in Gradec; n.zavasnik@gmail.com

Prispevek povzema razvoj brajice v slovenskem jezikovnem prostoru in osvetljuje vlogo knjižnice za slepe in slabovidne. Osredinja se na izsledke empirične raziskave o rabi brajice in bralnika zaslona v Sloveniji. Ugotovitve kažejo, da zaradi ekonomičnosti prevladuje raba zvočne tehnike (bralnik zaslona in zvočni zapisi). Izpostavljena je problematičnost rabe vključujočega jezika in digitalne piktografije.

brajica, sodobne tehnologije, bralnik zaslona, brajeva vrstica, sekundarna nepismenost

This article outlines the development of braille within the Slovenian linguistic context and highlights the role of the library for the blind and visually impaired. It focuses on the results of an empirical study on the use of braille and screen readers in Slovenia. The results indicate that, due to cost-effectiveness, the use of audio technology (screen readers and audio books) prevails. The problematic nature of using inclusive language and digital pictography is also pointed out.

braille, modern technologies, screen reader, braille display, secondary illiteracy

1 Brajica skozi čas

Mednarodni svet za izobraževanje ljudi z okvaro vida (ICEVI) in Svetovna zveza slepih (WBU) sta leta 2026 začela z globalno kampanjo za opismenjevanje v brajici, ki je temelj pismenosti za ljudi brez vida, kajti gre za edini način, ki jim omogoča videti besedo (Bavčar 2009). Po številnih poskusih iskanja ustrezne pisave za ljudi brez vida je Charles Barbier de la Serre leta 1815 razvil sonografijo – prvi taktilni pisalni sistem z dvignjenimi pikami. Leta 1824 ga je Louis Braille izboljšal tako, da je število pik zmanjšal z dvanajst na šest in prešel s fonetičnega na ortografski zapis, kar je omogočilo upoštevanje pravopisnih pravil namesto zgolj izgovorjave (Brešar 2025: 86). Do leta 1920 je brajica postala splošno sprejet svetovni standard za ljudi brez vida. Kljub temu še danes nastajajo nove različice (Brešar 2025: 88).

Slovenska brajica je verjetno nastala kmalu po letu 1900, saj je iz šolske kronike Odilien-Blindeninstituta v Gradcu in iz časopisa *Slovenec* z dne 19. 7. 1919 razvidno, da je s. Frančiška Apolonija Bezlaj slovenske učence poučevala v materinščini, pri čemer je uporabljala slovenske brajeve čitanke. Kasneje naj bi s. Frančiška Glaser znake za šumnike pridobila iz češke brajeve abecede (Brešar 2025: 100).

27. 2. 1920 je bila na sestanku v Ljubljani poenotena brajica v Kraljevini Jugoslaviji. Izbrana je bila hrvaška brajeva različica, ki je bila najbolj dodelana in je bilo v njej

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

napisane največ literature. Raba velike začetnice je bila po mnenju takratne stroke odveč.[1] Nastala je t. i. jugoslovanska brajica, ki je združila cirilico in latinico in je vsebovala 37 znakov. Sklenili so še, da se povsem odpravi raba Kleinove igličaste pisave (Brešar 2025: 61).[2] Iz članka, ki ga je 25. 1. 1945 objavil časopis *Slovenec*, je razvidno, da so to pisavo v slovenskem zavodu za slepe še vedno poučevali. Verjetno so z njenim poučevanjem prenehali po letu 1950 (Brešar 2025: 62).

Brajico uporabljajo v 142 državah in v 133 jezikih. Najdemo jo celo v esperantu, ki je bil nekoč mednarodni jezik slepih. Leta 1951 je Unesco poenotil brajico po vsem svetu. Takrat je bila spremenjena tudi jugoslovanska brajica, prav tako pa so svojo brajevo pisavo dobili tudi Makedonci. Po mednarodni konvenciji se brajica vedno piše od leve proti desni, tudi za jezike, ki se običajno pišejo drugače. Lahko rečemo, da je brajeva pisava resnično globalni jezik dotika (Brešar 2025). Slovenska brajica je zadnjo posodobitev doživela leta 2016, pred tem pa leta 1974. 27. 8. 2025 je bila vpisana v Register slovenske nesnovne kulturne dediščine (Ministrstvo za kulturo 2025).

Z začetkom digitalizacije je v sedemdesetih letih prejšnjega stoletja nastala osemtočkovna brajeva pisava, t. i. elektronska ali računalniška brajica, saj klasična šesttočkovna brajeva pisava ni več zadostovala za prikaz nabora znakov takrat modernih računalnikov. Povečanje brajeve celice za dve piki je omogočilo prikaz velike začetnice, številk in nekaterih najpogosteje uporabljenih znakov z enim brajevim znakom. Pomemben mejnik v Evropi je bil razvoj standarda Eurobraille, ki je bil določen leta 1986 na konferenci v Frankfurtu na Majni z namenom, da se mednarodno poenoti prikaz znakov v računalniški brajici (Epheser 1986). O tem, kdaj je v Slovenijo prispela prva brajeva vrstica in kdaj je bila kreirana prva kodna tabela[3] za slovenski jezik, ni podatkov.

1.1 Brajev kratkopis

Brajev kratkopis je standardiziran zapis v brajici, pri katerem se z enim brajevim znakom zapiše skupina črk ali se z nekaj znaki zapiše beseda, npr. *nm* za *namreč*, in se uporablja za hitrejše pisanje in branje (Gregorc idr. 2016: 5). Prvi brajev kratkopis se je pojavil leta 1871 v Angliji. Z njegovo uporabo so uspeli zmanjšati obseg knjig za 40 %. Po njegovemu zgledu je nastal kratkopis za francoščino, kasneje pa še za druge jezike. Za naše geografsko področje je zanimiv razvoj nemškega kratkopisa (Brešar 2025: 122). Jezikoslovci pravijo, da je morda vzrok počasnega uvajanja oz. razvijanja kratkopisa v slovansko govorečih državah v večji kompleksnosti jezikov. V slovanskih jezikih so slovnične oblike kompleksnejše in z manj ponavljanja (Brešar 2025: 121). Leta 1955 sta učitelja France Rožanec in Ciril Musar izdelala prvi slovenski brajev kratkopis, ki je

1 Znak za veliko začetnico se je v slovenski brajici začel uvajati šele v sedemdesetih letih 20. stoletja (Brešar 2025: 99).

2 J. W. Klein je leta 1807 razvil Kleinov iglopis, ki je omogočal tisk velikih tiskanih črk na papirju. V rabi so bili trije slogi reliefnega tiska: tisk z iglo, tisk s črtami in tisk z luknjicami. Kot večina njegovih predhodnikov in sodobnikov je bil tudi Klein prepričan, da je slepe treba naučiti pisave videčih (Brešar 2025: 58).

3 Brajeva kodna tabela je nabor pravil, ki določajo, kako se znaki iz običajnega tiska pretvorijo v šest- ali osemtočkovno brajico in obratno.

obseg knjig zmanjšal za 28 %. Minka Skaberne[4] je takoj dala v kratkopis prepisati nekaj mladinskih del (Brešar 2025: 123). »Učenje kratkopisa in tiskanje knjig v kratkopisu sta pri slovenskih slepih naletela na velik odpor [...] Učenje, branje in pisanje v kratkopisu je namreč zahtevnejše od pisanja brez okrajšav. Zaradi revolta slepih so se v kratkopis nehale prilagajati knjige« (Brešar 2025: 124).

1.2 Knjižnica slepih in slabovidnih Minke Skaberne

Prvo pobudo o prepisovanju knjig v brajico za slovenske oslepele vojake je dala Franica Vrhunc (s. Klara), ki je v Gradcu živeče Slovenke pisno pozvala k prepisovanju. Minka Skaberne, ki je delovala v društvu Dobrodelnost, se je v Gradcu pri slovenskih učiteljicah seznanila z njihovimi idejami in izkušnjami glede organizacije dela z ljudmi brez vida in prepisovanja knjig v brajico, kar je prenesla na področje tedanje Slovenije (Brešar 2025). 1. 3. 1918 je imela v prostorih društva Dobrodelnost predavanje o potrebah ljudi brez vida. Knjižnica slepih in slabovidnih Minke Skaberne (KSS) omenjeni datum beleži kot svoj ustanovni dan. Skaberne je organizirala prepisovalke knjig v brajico za potrebe oslepelih vojakov in v prvem letu je več kot 100 prepisovalk prepisalo 60 del slovenskih pisateljev v obsegu 150 brajevih zvezkov s po 130 stranmi. S tem je bil položen temelj za slovensko knjižnico slepih in slabovidnih.

Prepisovanje je med drugo svetovno vojno zamrlo in se nadaljevalo po njej. Način pretvarjanja v brajico se je spreminjal: sprva sta se uporabljala brajeva tablica in šilo, zatem mehanski in nato električni brajev pisalni stroj, danes pa se z računalniško strojno in programsko opremo prilagajajo besedila hkrati v brajico in elektronski zapis. Od leta 1994 se v KSS brajeve knjige tiskajo obojestransko in v več izvodih. Poškodovane knjige je danes mogoče hitro in učinkovito popraviti (Pavlin idr. 2020: 142).

Od leta 1958 so sestavni del KSS tudi zvočne knjige, ki so bile sprva posnete na magnetofonske trakove, od leta 1989 na kasete, od leta 2006 pa se snemajo v stisnjenem digitalnem zapisu MP3. KSS je edina ustanova v Sloveniji, ki za ljudi z okvaro vida prilagaja leposlovno in strokovno gradivo v njim dostopnih tehnikah (brajev, zvočni, računalniški zapis) in ga tudi izposoja. Redkokatera založba je izdala posamezne knjige v zvočnem zapisu na kasetah ali zgoščenkah, a to še zdaleč ne more zadostiti vsem potrebam ljudi z okvaro vida po branju, saj berejo veliko več od slovenskega povprečja (Pavlin idr. 2020: 143).

Do 26. 3. 2026 je knjižnični fond KSS obsegal 7806 naslovov leposlovnih del in 1198 učbeniških in strokovnih del v MP3, od tega 410 v formatu daisy. V brajico je prepisanih 1841 leposlovnih ter 246 učbeniških in strokovnih del, elektronska knjižnica pa obsega 1378 naslovov leposlovnih in 294 učbeniških in strokovnih del. Najstarejše v brajico prepisano delo iz leta 1918 hrani NUK. V brajico se običajno ne pretvarja obsežnih del (več kot 400 strani). Posebnosti v besedilu, zlasti v učbenikih in strokovnih besedilih, se zapišejo v opombe, mednje pa ne sodijo informacije o oblikovanju besedila (ležeči tisk, velike tiskane črke). Včasih se brajeve knjige vračajo poškodovane oz.

4 O njenem življenju in delu gl. Poljak Istenič in Testen (2018).

umazane od hrane ali pijače, v obdobjih, ki prinašajo večjo verjetnost suhe kože na dlaneh, npr. pozimi, pa tudi od krvi.[5]

1.3 Avtorske pravice

Minka Skaberne je avtorske pravice uredila z vzorčnim pismom, ki ga je naslovila na tedanje slovenske avtorje in jih prosila za dovoljenje za prepisovanje njihovih del v brajico. Kot razberemo iz odgovora Ivana Cankarja, je bila zanj posebna čast, da bodo njegova dela dostopna tudi oslepelim vojakom: »Čestitam vam, velecenjena gospodična, da ste si bili dali tako upravičeno in plemenito nalogo! Ljudje vam je ne bodo plačali, ali plačal vam jo bosta prav gotovo vaše srce in Bog« (Cankar 1918).

Po več kot stotih letih od začetkov prepisovanja knjig v brajico se ljudje z okvaro vida v Sloveniji kot tudi drugod po Evropi kljub *Marakeški pogodbi o olajšanem dostopu do objavljenih del za slepe in slabovidne osebe ter osebe z drugimi motnjami branja* (World Intellectual Property Organization 2013), ki je največkrat ratificiran dokument v zgodovini človeštva, še vedno srečujejo z nedostopnostjo tiskanih del, ki jih potrebujejo za izobraževanje, kaj šele tistih, ki bi jih radi prebrali zgolj zato, ker jih zanimajo. Ljudje z okvaro vida so še naprej prevečkrat obravnavani kot kriminalci, ki se jim očita možnost zlorabe elektronskih del, zaradi česar jim je onemogočen dostop do potrebnih gradiv, ob tem pa še nihče ni pomislil, da nedostopnost gradiv pomeni prepoved do izobraževanja in informacij, kar je kršenje temeljnih človekovih pravic.

1.4 Brajica in sodobne tehnologije

Osnovni element brajice je brajeva celica, prostor za šest pik pri šesttočkovni brajici, kar omogoča oblikovanje 64 različnih brajevih znakov. Z razvojem sodobnih tehnologij in pojavom prvih brajevih vrstic[6] sta bili za potrebe računalniške brajice dodani še sedma in osma pika, kar je povečalo število možnih znakov na 255 (Gregorc idr. 2016). Raba prilagojene računalniške opreme ljudem brez vida zdaj omogoča in od njih zahteva poznavanje oblike zapisa, kot je v črnem tisku (Gregorc idr. 2016: 4). Poleg tega so sodobne tehnologije prinesle nove znake (npr. @ ali €), ki jim je bilo treba določiti brajev znak.

Brajeva vrstica se običajno uporablja v kombinaciji z bralnikom zaslona.[7] Najpogostejša sta Jaws in NVDA. Ker je vrstica drag pripomoček, si ga veliko ljudi brez vida po svetu ne more privoščiti in se zanašajo le na bralnik zaslona. Raba bralnikov zaslona prevladuje tudi zaradi hitrejšega branja in pri rabi mobilnih naprav.

5 Vzrok za krvave madeže je suha koža na dlaneh, zaradi česar so blazinice prstov občutljivejše za ureznine, ki lahko nastanejo tudi pri branju s prsti in jih bralci velikokrat ne opazijo. Po pričevanju ljudi brez vida iz Severne Makedonije naj bi jih v osnovni šoli za slepo in slabovidno mladino v Skopju učitelji vse do devetdesetih let opozarjali, da si morajo pred branjem brajevih knjig umiti roke z milom in jih namazati s kremo, da bi preprečili morebitne krvavitve, saj so bile brajeve knjige redke, drage in dragocene.

6 Brajeva vrstica je zunanja strojna naprava, ki omogoča tipno branje besedila z računalniškega zaslona ali mobilnih naprav (Le in El-Glaly 2025).

7 Bralnik zaslona je programska oprema, ki vsebino na zaslonu uporabniku posreduje v govorni obliki po sistemu TTS (text to speech) ali preko vrstice (Web Accessibility Initiative 2022).

Zmotno mislimo, da se ob izgubi vida ostala čutila izostrijo in da je vsakdo, ki ne vidi, tudi uporabnik brajice. V resnici gre za proces nevroplastičnosti,[8] ki ga lahko ovirajo različni dejavniki, prav tako pa se nekateri prav zaradi bralnika zaslona ne odločijo za učenje brajice.

2 Vloga brajice danes

Za objektivno opredelitev brajice v sodobnem času v Sloveniji je bil med njenimi uporabniki izpeljan vprašalnik, ki je bil med 1. 3. in 20. 4. 2026 dostopen na spletni platformi www.1ka.si. Izpolnilo ga je 51 oseb, od tega 53,1 % moških in 46,9 % žensk. Vprašalnik so večinsko izpolnili respondenti stari nad 40 let (25,5 % med 50 in 60 let ter 23,5 % med 40 in 50 let), zaradi česar rezultati ne morejo veljati za celotno populacijo ljudi brez vida. Opaziti je še naslednje zanimivosti: pri anketiranih, starih med 30 in 40 let, se pojavi nekaj odstotkov tistih (prva generacija), ki so se brajice naučili v času šolanja v redni šoli. Mlajše so generacije, višji je njihov odstotek. V starostni skupini 40 do 50 let in 50 do 60 let so respondenti, ki so kasneje izgubili vid in od katerih se jih 9,3 % ni naučilo brajice, ker je zahvaljujoč bralnikom zaslona ne potrebujejo. Kar polovica sodelujočih ne vidi od rojstva, 16 % jih je pred izgubo vida videlo slabo, ostali pa so vid izgubili v različnih življenjskih obdobjih.

34,1 % sodelujočih ocenjuje, da brajico berejo zelo hitro (več kot 60 besed na minuto), 25 % jih po samooceni prebere več kot 50 besed na minuto, 15,9 % se jih samoocenjuje za počasne bralce in zanimiv podatek je, da jih 9,1 % bere zelo počasi in med vrsticami potrebujejo prazno vrstico, t. i. presledek. 63,9 % uporabnikov brajice meni, da je brajica na papirju na otip prijetnejša od računalniške (na vrstici).

40,9 % sodelujočih pozna pisavo videčih (tiskane in pisane črke), predvsem je visok odstotek pri tistih med 30. in 60. letom. Gre predvsem za respondente, ki so vid izgubili kasneje in pisavo videčih poznajo že od prej. 54,8 % tistih, ki ne vidijo od rojstva, pozna le velike tiskane črke. 13,6 % se jih zna le podpisati in 4,5 % si pisave videčih ne zna predstavljati.

Pri izobraževanju brajico uporablja 79,1 % kandidatov. Prevladuje kombinacija brajevih in zvočnih knjig (27 %). Po podatkih KSS se učbeniki v brajici tiskajo za šolajoče se v prvi triadi, kasneje pa uporabljajo učbenike v elektronski obliki.

8 Proces nevroplastičnosti je sposobnost možganov, da se reorganizirajo in ustvarijo nove nevronske povezave, kar je bistveno pri prilagajanju na izgubo vida. Pri branju brajice pride do t. i. navzkrižne nevroplastičnosti. Možgani vidnega korteksa (dela možganov, ki predeluje vizualne informacije) ne pustijo neaktivnega, temveč ga reorganizirajo za predelavo tipnih informacij. Pri branju besedila v brajici vidni korteks prevzame vlogo prepoznavanja omenjenih vzorcev kot jezikovnih znakov. Če se ljudje brez vida ne morejo naučiti brati brajice, to z vidika nevroplastičnosti pomeni, da se taktilni in vidni korteks nista strukturno povezala za opravljanje te naloge. Brez intenzivnega učenja in ponavljanja se sinaptične povezave med tema dvema področjema ne utrdijo. Nevroplastičnost je najmočnejša v zgodnjem otroštvu. Če pride do izgube vida kasneje v življenju in se oseba ne more naučiti brajice, gre pogosto za posledico dejstva, da plastičnost vidnega korteksa z leti upada, zaradi česar je reorganizacija za predelavo kompleksnih tipnih informacij precej težja (Sadato idr. 1996).

66,7 % ljudi brez vida pri izobraževanju uporablja brajevo vrstico in 53,8 % bralnik zaslona. Zanimivo je, da jih je kar 82,1 % kot pripomoček pri izobraževanju navedlo brajev pisalni stroj, kar se nanaša na dejstvo, da so vprašalnik izpolnili pretežno respondenti nad 40 let in v času njihovega izobraževanja še ni bilo sodobnih tehnologij oz. se je njihova raba šele začela.

Pri delu brajico uporablja 87,8 % sodelujočih. Prevladuje raba bralnika zaslona (70,3 %), sledita pa mu brajeva vrstica in brajev pisalni stroj (67,6 %). Kar 40 % sodelujočih bere vsak dan. 63,2 % jih posluša zvočne knjige, po 18,4 % pa jih bere elektronske knjige – po podatkih KSS posegajo po knjigah v brajici predvsem otroci in starejši.

Sodelujoči brajico uporabljajo pri branju elektronskih besedil (59 %), e-sporočil (56,4 %), izpolnjevanju e-obrazcev (51,3 %) in branju leposlovja (48,7 %), strokovne literature in spletnih vsebin (46,2 %). Navajajo težave, ki se pretežno nanašajo na bralnik zaslona in le majhen odstotek na nedostopnost spletišč. Večina težav izvira iz neznanja o rabi bralnika zaslona, zato so izobraževanja o rabi sodobnih tehnologij za ljudi brez vida nujna. Samostojna raba sodobnih tehnologij ljudem brez vida prinaša neodvisnost in konkurenčnost. Ob tem ne smemo pozabiti na dostopnost vsebin in na dejstvo, da mora nekdo brez vida imeti več znanja od videčega kolega, da pri rabi tehnologij lahko opravi določen korak.

Pri ljudeh brez vida je primarno orodje za delo z računalnikom bralnik zaslona (70,6 %), ki je hitrejši in dostopnejši, vendar zgolj slušno sprejemanje informacij vodi v sekundarno nepismenost, saj uporabnik ne zazna vseh ločil, zapisa imen ali tujk. Treba je omeniti, da številni uporabniki bralnik zaslona poslušajo pri dokaj visoki hitrosti, ki je za videče nerazumljiva in je namenjena hitremu pridobivanju informacij.

82,5 % jih meni, da je brajica v vsakdanjem življenju premalo zastopana. Navajajo potrebo po brajevih napisih na embalaži prehranskih, kozmetičnih in higienskih izdelkov, v javnih ustanovah, prevoznih sredstvih, na postajah in domofonih, gospodinjskih aparatih, v brajici pogrešajo navodila za uporabo zdravil, dokumente, obrazce in račune (slednje omogočajo le nekatere države kot sta npr. Francija in Velika Britanija). Brajica ni zgolj bralni medij, temveč omogoča večjo samostojnost, orientacijo in mobilnost.

2.1 Zapisovanje enakosti spolov v jeziku in težave uporabnikov brajice pri branju

Za izražanje enakosti spolov v jeziku se najpogosteje uporabljajo podčrtaj, dvopičje ali zvezdica. Kot navaja Nemška zveza slepih in slabovidnih (Deutscher Blinden- und Sehbehindertenverband [DBSV] 2024), ti znaki znotraj besed za ljudi z okvaro vida niso le estetska motnja, temveč tehnična ovira, saj ljudem z okvaro vida, ki berejo (povečan) črni tisk, predstavljajo oviro pri tekočem branju, bralnik zaslona pa jih pogosto interpretira tako, da prekine razumevanje besedila. V slovenščini se priporoča raba podčrtaja, toda Nemška zveza slepih in slabovidnih vse tri oblike (podčrtaj, dvopičje in zvezdico) označuje za neuporabne in poziva skupine, ki čutijo, da njihov spol ni zajet v binarni obliki, naj se dogovorijo glede poenotenega zapisovanja in predlagajo nekaj rešitev, izmed katerih bi izbrali tisto, ki ne povzroča težav pri branju ljudem brez vida in ljudem

s težavami pri branju (DBSV 2024). Za zdaj Nemška zveza slepih in slabovidnih predlaga zvezdico kot najdostopnejšo obliko, vendar opozarja, da tudi ta ni primerna in da je nujna nova rešitev. Tudi v Sloveniji je treba najti boljšo rešitev, vendar se omenjene težave sploh ne upošteva. Velja razmisliti, ali je jezik morda zrel za nevtralne končnice, ki bi zajele vse spole.

Slovenski uporabniki brajice so v vprašalniku s 73,5 % odgovorili, da jih zapisovanje enakosti spolov v jeziku ne moti, vendar je pogosto sledil komentar, da gre bolj za njihov odnos do družbe in kako nikogar ne želijo diskriminirati, ali pa, da naj oblik s podčrtaji ali poševnicami v besedilu ne bo preveč, saj jih 25 % zaradi njihove pogostnosti izgubi rdečo nit, enak odstotek pa v bralniku izključi funkcijo izgovarjanja nekaterih ločil, zaradi česar imajo težave drugod (npr. zaradi neprebrane zvezdice ne vedo, katero polje je pri izpolnjevanju obvezno).

2.2 Raba digitalnih piktogramov

Sodobni način komunikacije prinaša več vizualizacije, med drugim tudi rabo digitalnih piktogramov, ki ljudem brez vida niso nedostopni, vendar je njihova interpretacija odvisna od tehničnih opisov v ozadju, saj so emodžiji v standardu Unicode definirani z edinstvenimi kodami, ki jih bralniki zaslona povežejo z besedilnimi opisi iz zbirke CLDR. Videči emodžije dojemajo intuitivno in hitro, ljudje brez vida pa jih procesirajo kot besedilo. Od vsebine besedila je odvisno, kako (in ali) bo nekdo brez vida razumel pomen emodžija (Unicode Consortium 2023). Iz rezultatov vprašalnika je razvidno, da ima 69 % sodelujočih težave s prepoznavanjem pomena emodžijev, 26 % ne razume pomena nekaterih emodžijev, 21,7 % pa jih ima težave, ker njihov bralnik emodžijev ne prepozna. Kljub temu jih 48,5 % emodžije uporablja, in sicer jih 31,2 % uporablja tiste, za katere poznajo pomen oz. jih uporabljajo ljudje, ki vidijo.

3 Sklep

Brajica ostaja temelj opismenjevanja ljudi brez vida, saj omogoča neposreden stik z besedo. V Sloveniji se uporablja od začetka 20. stoletja in je pomemben člen pri izobraževanju, delu in povsod, kjer je potrebno natančno delo z besedilom. Sodobna tehnologija je prinesla pripomočke, kot so brajeve vrstice in bralniki zaslona. Slednje ljudje brez vida zaradi hitrosti in večje dostopnosti pogosteje uporabljajo, vendar zgolj poslušanje vodi v sekundarno nepismenost. Uporabniki brajice opozarjajo na pomanjkanje brajevih napisov v vsakdanjem življenju in na težave pri rabi oblik za enakost spolov v jeziku in tudi pri rabi digitalnih piktogramov.

Prispevek je nastal v sodelovanju z Zvezo društev slepih in slabovidnih Slovenije ter Knjižnico slepih in slabovidnih Minke Skaberne. Hvala Diani Žalik in Luku Pavlinu za aktualne podatke o knjižničnem fondu.

Viri in literatura

Bavčar, E. (2009). Le regard d'aveugle entre le mythe, la métaphore et le réel. V M. Chottin (ur.), *L'aveugele et le philosophe: Ou Comment la cécité donne à penser*. Éditions de la Sorbonne.

Brešar, D. (2025). *Dotik besed*: *Zgodovina opismenjevanja slepih in slabovidnih*. Združenje prijateljev slepih.

Cankar, I. (1918, 7. junij). *Pismo Minki Skaberne* [Pismo]. Zapuščina Ivana Cankarja, Ms 513, Rokopisna zbirka, Narodna in univerzitetna knjižnica v Ljubljani.

Deutscher Blinden- und Sehbehindertenverband. (b. d.). *Gendern*. https://www.dbsv.org/gendern.html

Epheser, H. (1986). *Die Euro-Braille-Schrift: Entstehung und Struktur einer universellen Computer-Punktschrift*. Deutscher Blinden- und Sehbehindertenverband.

Gregorc, J., Jenčič, M., Jeraša, M., Repe Kocman, M., Murn, T., Pavlin, B. in Wraber, T. (2016). *Standard slovenske brajice: Posodobitev slovenske 6-točkovne brajice*. Komisija za slovensko brajico. https://www.zveza-slepih.si/wp-content/uploads/2016/06/Standard-slovenske-brajice-3.0-PDF.pdf

Le, H. in El-Glaly, Y. N. (2025). *Advancements in refreshable Braille display technology*. Chapman University Digital Commons. https://digitalcommons.chapman.edu/engineering_articles/232/

Ministrstvo za kulturo. (2025, 2. september). *Slovenska brajica vpisana v register slovenske nesnovne kulturne dediščine*. https://www.gov.si/novice/2025-09-02-slovenska-brajica-vpisana-v-register-slovenske-nesnovne-kulturne-dediscine/

Pavlin, L., Padežnik, S. in Vertačnik, V. (ur.). (2020). *Od točkopisa do eBralca*: *100 let organiziranega delovanja slovenskih slepih in slabovidnih*. Zveza društev slepih in slabovidnih Slovenije.

Poljak Istenič, S. in Testen, P. (ur.). (2018). *Minka Skaberne* (*1882–1965): Pobudnica in ustanoviteljica prve slovenske knjižnice za slepe*. Znanstvenoraziskovalni center SAZU.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G. in Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*(6574), 526–528. https://doi.org/10.1038/380526a0

Skaberne, M. (1918, 29. maj). *Pismo Ivanu Cankarju* [Pismo]. Zapuščina Ivana Cankarja, Ms 1152, B/27, Rokopisna zbirka, Narodna in univerzitetna knjižnica v Ljubljani.

Unicode Consortium. (2023). *Emoji list, v15.1*. https://unicode.org/emoji/charts/emoji-list.html

Web Accessibility Initiative. (2022, 31. marec). *Screen readers*. World Wide Web Consortium. https://www.w3.org/WAI/perspective-videos/screen-readers/

World Intellectual Property Organization. (2013). *Marrakesh treaty to facilitate access to published works for persons who are blind, visually impaired, or otherwise print disabled*. WIPO. https://www.wipo.int/wipolex/en/treaties/textdetails/13169

Digitalna opremljenost za opolnomočenje uporabnikov in uporabnic slovenščine v Italiji: primer *Primorskega dnevnika*

DOI: 10.4312/Obdobja.45.415-422

Zarja Zver, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; Slovenski raziskovalni inštitut, Trst; zarja.zver@ff.uni-lj.si

Prispevek predstavlja magistrsko raziskavo o razširjenosti jezikovnih virov in tehnologij med novinarji in uredniki *Primorskega dnevnika*. Hipoteza, da je uporaba jezikovnih orodij za slovenščino v manjšinskem položaju omejena, se umešča v družbeni kontekst razmisleka o tem, kako digitalno okolje oblikovati na način, da bo opolnomočilo manjšinsko skupnost in blažilo jezikovne neenakosti, ki nastajajo v odnosu do večinske in do širše referenčne skupnosti. Rezultati kažejo, da je uporaba orodij tesno povezana z uporabniškim zaupanjem, kar zahteva participativni pristop in razpršeno diseminacijo, ki ju je treba pospremiti z ažurno jezikovno politiko.

manjšinski jeziki, slovenščina v Italiji, *Primorski dnevnik*, jezikovna orodja, participativni pristop

This article presents research for a master's thesis on the use of language resources and technologies among journalists and editors of the newspaper *Primorski dnevnik*. It hypothesizes that the use of language tools tailored to Slovenian as a minority language remains limited, situating this issue within a social discussion on how the digital landscape can be shaped to empower the minority community and mitigate linguistic inequalities vis-à-vis the majority and the wider reference community. The results show that the use of tools is closely linked to user trust, underscoring the need for a participatory approach, widespread dissemination, and an up-to-date language policy.

minority languages, Slovenian in Italy, *Primorski dnevnik*, language tools, participatory approach

1 Uvod

Februarja 2025 je potekala druga mednarodna konferenca o jezikovnih tehnologijah za vse, katere osrednja tema je opolnomočenje skupnosti prek razvoja jezikovnih tehnologij (LT4All 2025 b. d.). Čeprav sta po poročilu Unesca z jezikovnimi orodji opremljena le 2 % od več kot 7000 jezikov (Mariani, v tisku), mednje spada tudi slovenščina. Potreba po oblikovanju temeljnih virov in tehnologij je bila prepoznana z Resolucijo o nacionalnem programu za jezikovno politiko 2014–2018 in Akcijskim načrtom za jezikovno opremljenost (Ledinek 2017), uresničuje pa se s projekti, kot je Razvoj slovenščine v digitalnem okolju (RSDO), in delovanjem Centra za jezikovne vire in

Stritar Kučuk, M. (ur.). (2026). *Slovenščina v dobi pametne tehnologije* (Obdobja 45).
Založba Univerze v Ljubljani.
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-933-1, 978-961-297-943-8 (PDF)

CC BY SA

tehnologije (CJVT). Vendar pa se k razvoju jezikovnih orodij na nacionalni ravni pristopa z vidika osrednjega standarda, o čemer priča zadnji program za jezikovno politiko, ki v poglavju o jezikovni opremljenosti (Resolucija o nacionalnem programu za jezikovno politiko 2021–2025: 51–70) Slovencev[1] zunaj meja Republike Slovenije (RS) ne naslavlja z nobenim ciljem (Grgič in Zver 2024: 223). Ta vrzel odpira pomembno vprašanje, saj ključ ni v razvoju jezikovnih tehnologij kot takšnih, temveč v oblikovanju rešitev, prilagojenih posameznim skupnostim (LT4All 2025 b. d.). Da je to posebej pomembno za ohranjanje in promocijo manjšinskih in regionalnih jezikov (Gerken 2022: 22), potrjuje primer slovenske skupnosti v Italiji, ki v luči negativnih vplivov na sporazumevalne prakse in identitetne izbire ter pomanjkanja jezikovnopolitičnih mehanizmov razvija lastne pristope.

2 Izhodišča

Sociolingvistični položaj slovenščine v Italiji je rezultat pozne vzpostavitve ustreznega pravnega varstva, saj je skupnost splošno zaščito prejela šele leta 2001, ter odsotnosti jezikovne politike, ki jo obe državi prepuščata manjšini (Brezigar idr. 2022: 10–11). V odsotnosti teh struktur je slovenska skupnost v Italiji razvila lastne institucije in zaprto družbeno mrežo, kar je srednjeročno zaustavilo izgubo jezika, a spodbudilo specifične sporazumevalne prakse in omejilo stik z drugimi različicami slovenskega jezikovnega kontinuuma (Grgič in Guardiancich 2025). V skupnosti smo tako priča procesom opuščanja jezika, med drugim jezikovnemu separatizmu, pri katerem se lokalne variante nekega jezika zaradi jezikovnega stikanja in/ali marginalizacije od referenčne skupnosti vse bolj oddaljujejo od širšega jezikovnega kontinuuma (Grgič in Popič 2023: 152).

Primer slovenske skupnosti v Italiji potrjuje, da je dejansko rabo jezika mogoče doseči šele ob stalni izpostavljenosti jeziku v vseh kontekstih sporazumevanja (Brezigar idr. 2022: 14). Ker se družbena in jezikovna krajina v 21. stoletju selita v virtualni prostor, postajata digitalna prisotnost in jezikovnotehnološka opremljenost vse pomembnejša dejavnika ohranjanja jezika (Grgič in Zver 2024: 216).

V prispevku se termin *jezikovna orodja* uporablja kot krovni izraz za *jezikovne vire* in *jezikovne tehnologije*. Jezikovni viri so zbirke jezikovnih podatkov v strojno berljivi obliki (npr. korpusi, leksikoni, terminološke baze) in tvorijo podlago za razvoj, nadgradnjo ali evalvacijo jezikovnih tehnologij, tj. sistemov za obdelavo naravnega jezika in govora (ELRA Language Resources Association b. d.), ki vse pogosteje temeljijo tudi na umetni inteligenci (UI) (npr. črkovalniki, strojni prevajalniki, klepetalni roboti). Pomen digitalnih pripomočkov se kaže na različnih področjih, od komunikacije in dostopa do storitev (Popič idr. 2024: 170) do usvajanja jezika in razvoja jezikovnih kompetenc (Grgič in Zver 2024: 222–223). V spremenjenih geopolitičnih in družbeno-psiholoških okoliščinah, v katerih je večinsko prebivalstvo v Italiji slovenščini bolj

1 Zagovarjamo vključujočo rabo jezika, vendar z namenom strnitve besedila uporabljamo moško obliko za označevanje katerega koli spola. To velja za vse vrste poimenovanj, ki se nanašajo na osebe.

naklonjeno (Jagodic 2024: 34), lahko dovolj zanesljiva orodja odpravijo razloge za opuščanje manjšinskega jezika v stikih z večinsko skupnostjo, s krepitvijo sporazumevalnih zmožnosti med pripadniki manjšine pa pripomorejo k preseganju procesov jezikovnega separatizma. Na ta način lahko jezikovna orodja prispevajo k enakopravnejšemu statusu slovenščine tako v italijanskem kot v širšem slovenskem jezikovnem prostoru (Zver 2026: 38).

Ker pa orodja za večinske jezike pogosto ne ustrezajo potrebam manjšinskih skupnosti, so ključne prilagojene rešitve, ki upoštevajo sociolingvistične značilnosti uporabnikov (Popič idr. 2024: 169). Čezmejne skupnosti imajo prednost, saj lahko prevzemajo in prilagajajo orodja referenčne države (Gerken 2022: 19), vendar je uspešnost prenosa odvisna od institucionalne podpore, virov in motivacije skupnosti (Gaspari idr. 2023: 51–52). V primeru slovenske skupnosti v Italiji je prav motivacija bistveni dejavnik: opažamo, da skupnost kljub odsotnosti sistemske podpore krepi jezikovno opremljenost. Osrednjo vlogo pri tem ima Slovenski raziskovalni inštitut (SLORI), ki s podporo Centralnega urada za slovenski jezik Dežele Furlanije - Julijske krajine nadgrajuje obstoječe in razvija lastne specializirane digitalne rešitve, usmerjene v podporo slovenščini v manjšinskem položaju.

Med glavnimi dosežki je portal Jezik na klik (JNK), zbirališče virov in tehnologij, ki izhajajo iz potreb na lokalni ravni. Med njimi izstopa Zvezek standardiziranih terminov (Zvezek) z izrazi z upravno-pravnega področja in sistemov javnih storitev. Pomembno vlogo ima tudi korpus lektoriranih slovenskih besedil z območja italijansko-slovenskega jezikovnega stika STIKit, ki ponuja vpogled v specifične jezikovne prakse na območju stika z italijanščino in podlago za razvoj naprednejših jezikovnih tehnologij (Zver 2026: 44–45). V tem kontekstu posebno mesto zavzema najnovejša različica jezikovnega asistenta Loris 1.1, ki uporabnike slovenščine v Italiji ob vnosu besedila v besedilno polje opozori na zaznamovane jezikovne elemente (npr. paronime, kalke) ali druge neskladnosti, ki odstopajo od osrednjega standarda, in ponudi nezaznamovane elemente (Popič idr. 2024: 168). Hkrati Loris 1.1 z integracijo splošnih orodij za slovenski jezik (Sopomenk 2.2, Vejic 1.0 CJVT) in strojno naučenih modelov (EssayHelper in SloNSpell) odraža povezanost z referenčno skupnostjo in nakazuje potencial nadaljnjega razvoja prilagojenih UI-rešitev (Popič idr. 2024).

Pomemben del razvoja jezikovnih orodij sta ocena njihove uporabnosti in identifikacija možnosti za nadgradnje. Uporaba orodij za slovenščino v Italiji se je doslej spremljala kvantitativno, z beleženjem uporabniških iskanj (Popič idr. 2024: 168), objava različice Loris 1.1 pa je pokazala potrebo po poglobljeni kvalitativni evalvaciji, ki bi razvijalcem ponudila temeljit uvid v potrebe in pričakovanja uporabnikov. Ker se uporaba teh orodij umešča v širši ekosistem jezikovnih virov in tehnologij, so bila v opisani raziskavi upoštevana tudi orodja za splošne uporabnike slovenskega jezika (priročniki Inštituta za slovenski jezik, orodja CJVT in drugih slovenskih ponudnikov ali pobud, npr. Amebisa, RSDO) ter UI- in druga orodja, ki omogočajo sporazumevanje v slovenščini (npr. Glosbe, Google Prevajalnik, ChatGPT).

3 Metodologija

Raziskava je bila izvedena v okviru magistrskega dela aprila 2025. Za opazovano skupino so bili izbrani novinarji in uredniki *Primorskega dnevnika* (PD), ki jih lahko opredelimo kot aktivne uporabnike slovenščine, saj jezik uporabljajo vsakodnevno za produkcijo novih vsebin, hkrati pa imajo kot sodelavci osrednjega manjšinskega dnevnika pomembno jezikovnoozaveščevalno vlogo (Zver 2025: 35–36). Temeljni cilj raziskave je bil ovrednotiti uporabnost orodij za slovenščino v Italiji in seznanjenost skupnosti z njimi ter predlagati izboljšanja njihove tehnološke ali vsebinske zasnove in promocije. Pri tem se je preverjala hipoteza, da je uporaba specializiranih orodij za slovenščino v manjšinskem položaju med novinarji in uredniki PD redka. Obenem je raziskavo vodilo vprašanje, kako in v kolikšni meri lahko jezikovna orodja pripomorejo k opolnomočenju aktivnih uporabnikov slovenščine v Italiji, konkretno sodelavcev PD, ter kako in v kolikšni meri blažijo jezikovne neenakosti v odnosu do italijanskega jezika v Italiji in do slovenskega jezika v Sloveniji (Zver 2025: 34).

Raziskovalno vprašanje in hipoteza sta se preverjala z metodo delno strukturiranih intervjujev. Sodelovalo je 10 sodelavcev tržaške in 1 sodelavec goriške redakcije PD, od tega 8 žensk in 3 moški (n = 11) (Zver 2025: 36). Pogovori, ki so se snemali, so trajali v povprečju 45 minut. Potekali so na sedežu PD v Trstu, le srečanje z osebo z uredništva v Gorici je potekalo preko videokonferenčne platforme Zoom. Udeleženci so pred začetkom intervjuja podpisali obveščeno soglasje, pripravljeno v skladu s smernicami Komisije za etiko Univerze v Ljubljani (Zver 2025: 37).

Za karseda podroben vpogled v digitalni položaj slovenščine v Italiji, okoliščine uporabe jezikovnih orodij in možnosti, ki jih ponujajo za jezikovno opolnomočenje, so se vprašanja nanašala na pet glavnih tem: jezikovne izkušnje in percepcije udeležencev, njihovo vlogo na PD ter vlogo PD za skupnost; proces tvorjenja (novinarskih) besedil v slovenskem jeziku, vključno z najpogostejšimi jezikovnimi težavami; strategije reševanja jezikovnih dilem s poudarkom na primerjavi različnih virov pomoči (npr. človeka, tradicionalnih tiskanih in sodobnih digitalnih pripomočkov) glede na naravo problema oz. okoliščine; jezikovna orodja s poudarkom na motivih za uporabo posameznih orodij, uporabniški izkušnji in potrebi po dodatnem usposabljanju in/ali promociji ter predlogih za nadaljnji razvoj specializiranih tehnoloških rešitev; potencial digitalizacije ter jezikovnih virov in tehnologij za opolnomočenje posameznika in manjšinske skupnosti kot celote (Zver 2025: 37).

Pridobljene zvočne datoteke so bile prepisane s pomočjo programa za prepoznavanje govora Transkriptor, pri čemer so bila z namenom varstva osebnih podatkov izpuščena vsa mašila, ponavljanja in druge govorne posebnosti. Zaradi majhnosti vzorca oz. skupnosti, ki ji udeleženci pripadajo, je bila posebna pozornost namenjena kontekstualni psevdoanonimizaciji: vse informacije, ki bi lahko posredno razkrivale identiteto udeležencev, so bile nadomeščene z nevtralnimi oznakami (npr. [institucija], [prijatelj]) (Zver 2025: 37–38). Na ta način obdelano raziskovalno gradivo je bilo analizirano po temah, za katere je na splošno značilno, da se tesno prepletajo.

4 Rezultati

Rezultati raziskave kažejo, da jezikovna orodja krepijo jezikovno občutljivost in kompetence novinarjev in urednikov PD. Respondenti v uredništvu opažajo porast jezikovnih dilem, kar pripisujejo večji ozaveščenosti, ki je nastopila z večjo razpoložljivostjo digitalnih virov in tehnologij. Kot poudarjajo, so »ti spletni pripomočki […] vsekakor izboljšali situacijo«, saj so *ad hoc* rešitve nadomestili s preverjanjem rabe in iskanjem ustreznih izrazov (Zver 2025: 44). To v opazovanem kontekstu potrjuje tezo o opolnomočenju uporabnikov z jezikovnimi orodji.

Med digitalnimi pripomočki za slovenščino so v ospredju splošni jezikovni viri, kot sta portal Fran in korpus Gigafida, v katerem sodelavci PD lastno rabo primerjajo z rabo v širši jezikovni skupnosti. Tovrstna uporaba priča o pragmatičnem in sociolingvističnem potencialu referenčnega korpusa, ki uporabnikom omogoča umeščanje lastne rabe v širši jezikovni prostor in presojanje ustreznosti posameznega jezikovnega koda (Zver 2025: 61). Splošni jezikovni viri ozaveščajo o raznolikosti slovenskega jezikovnega kontinuuma, s čimer prispevajo k premagovanju procesov jezikovnega separatizma in h krepitvi položaja slovenščine v Italiji v odnosu do osrednjega standarda. Pomembno komplementarno vlogo imajo specializirani viri, zlasti Zvezek, ki novinarje in urednike PD seznanja s pojavi jezikovnega stikanja in jih podpira pri poimenovanju za italijansko družbo značilnih pojavov in upravno-pravnih struktur. Domnevamo lahko, da Zvezek tudi na ravni skupnosti prispeva k rabi slovenščine v javnih in institucionalnih kontekstih ter k zmanjševanju jezikovne neenakosti v odnosu do italijanščine, vendar respondenti opozarjajo na omejeno razširjenost zunaj jezikovno bolj ozaveščenih skupin. Po pričevanjih udeležencev javni uslužbenci še vedno posegajo po neustreznih dobesednih prevodih (npr. *dnevni red* namesto *predlog skupnega stališča* za *ordine del giorno*), kar je najverjetneje povezano s pomanjkljivo diseminacijo in posledičnim nepoznavanjem standardiziranih rešitev (Zver 2025: 92–93). To kaže, da potreba po promociji jezikovnih orodij presega nivo posameznih institucij ali poklicnih skupin in zahteva širši pristop: udeleženci izpostavljajo izobraževalno okolje, v katerem bi z vključevanjem digitalnih orodij in usposabljanjem učiteljev lahko vplivali na jezikovne prakse mladih in dolgoročno okrepili kompetence celotne skupnosti (Zver 2025: 84). Poleg tega mladi kot najvidnejši akterji digitalne komunikacije odločilno sooblikujejo spletno prisotnost skupnosti in s tem količino razpoložljivih digitalnih podatkov, ki neposredno pogojujejo njen jezikovnotehnološki potencial (Gaspari idr. 2023: 52).

V nasprotju z jezikovnimi viri je uporaba jezikovnih tehnologij med novinarji in uredniki PD manj razširjena. Med specializiranimi tehnologijami je bil v središču zanimanja Loris 1.1, po katerem respondenti posežejo »morda enkrat na mesec« ali pa jim sploh »ne pride na misel« (Zver 2025: 66). Redkost uporabe je posledica slabe uporabniške izkušnje, pri nekaterih še s starejšo različico orodja (1.0), nezadostne promocije in neskladja z jezikovno-tehnološkimi potrebami (Zver 2025: 67–69). Udeleženci namreč izpostavljajo potrebo po podpori na skladenjsko-strukturni ravni, zlasti pri besednem redu, ki je v manjšinskem okolju »gotovo zelo pod vplivom italijanščine« (Zver 2025: 64), hkrati pa zanj ne obstaja še nobeno orodje, niti za splošne uporabnike slovenskega

jezika. To odpira vprašanje nadaljnjega razvoja tehnologij, ki bi temeljile na naprednih jezikovnih modelih in specializiranih učnih množicah, kot je STIKit (Zver 2025: 94). Če torej pojem jezikovna orodja razumemo v širšem smislu tako jezikovnih virov kot tehnologij, rezultati raziskave delno ovržejo hipotezo o redki uporabi orodij za slovenščino v Italiji: medtem ko so prilagojeni viri med sodelavci PD nepogrešljivi, prilagojene tehnologije ostajajo na obrobju (Zver 2025: 93).

Izven zastavljenih hipotez je raziskava osvetlila še en pomemben vidik: razmerje med percipirano jezikovno zmožnostjo in uporabo UI, predvsem strojnih prevajalnikov. Rezultati namreč potrjujejo izsledke preteklih raziskav, da so povprečni uporabniki bolj odprti za uporabo strojnih prevajalnikov kot jezikovni strokovnjaki (Bowker 2021). V kontekstu PD se to kaže tako, da respondenti z nižjo samooceno jezikovnih zmožnosti strojne prevajalnike uporabljajo pogosto, medtem ko jih tisti z več formalnega jezikovnega znanja in izkušenj uporabljajo selektivno in na splošno v zvezi z UI izražajo večji »strah« (Zver 2025: 96). Čeprav bi lahko pričakovali, da bodo spretnejši uporabniki UI uporabljali pogosteje že zato, ker so bolj ozaveščeni o njenih (z)možnostih in jo znajo uporabljati ciljno, denimo z namenom obogatitve jezika (Zver 2025: 77), rezultati kažejo na kompleksnejše razmerje med jezikovno kompetenco, zaupanjem in uporabo najnaprednejših tehnologij (Zver 2025: 96).

Prakse in prepričanja respondentov kažejo, da ima pri uporabi jezikovnih orodij bistveno vlogo uporabniško zaupanje, ki vključuje več dimenzij (npr. zanesljivost, ustreznost, uporabnost), v splošnem pa ga lahko opredelimo kot prepričanje uporabnikov, da bo orodje nalogo opravilo skladno z njihovimi pričakovanji (Yan idr. 2013: 639). Takšno zaupanje v občutljivih družbenih kontekstih ni samoumevno. Kot potrjuje pričujoča raziskava, je ključen razvoj prilagojenih rešitev v sodelovanju z manjšinskimi skupnostmi: participativni pristop, ki vključuje končne uporabnike, omogoča oblikovanje konteksta in izzivom ustreznih orodij, s čimer se poveča njihova sprejetost in zmanjšajo tveganja za pomanjkljivosti (Denecke idr. 2025: 360), kakršne so se pokazale pri orodju Loris 1.1.

5 Sklep

Raziskava o razširjenosti jezikovnih virov in tehnologij med novinarji in uredniki PD je pokazala, da tako splošna kot specializirana orodja krepijo jezikovno kompetenco uporabnikov slovenščine v Italiji, vendar njihova uporaba ni odvisna zgolj od razpoložljivosti, temveč predvsem od ustreznosti rešitev in zaupanja v tehnologijo. To potrjuje izhodiščno tezo, da je treba za doseganje enakopravnosti slovenščine kot manjšinskega jezika – tako v odnosu do večinske italijanščine kot tudi do slovenščine v Republiki Sloveniji – razvijati kontekstu prilagojene digitalne rešitve, ki nastajajo z vključevanjem skupnosti. V tem okviru se odpira vprašanje prihodnjega razvoja jezikovne politike, zlasti, ali bo novi nacionalni program za jezikovno politiko 2026–2030 presegel dosedanjo osredotočenost na osrednji standard in sistematično naslovil tudi jezikovnotehnološke potrebe Slovencev zunaj RS.

Viri in literatura

Bowker, L. (2021). Can machine translation meet the needs of official language minority communities in Canada? A recipient evaluation. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, *8*, 123–155. https://doi.org/10.52034/lanstts.v8i.248

Brezigar, S., Grgič, M. in Jagodic, D. (2022). Koncept deželne jezikovne politike za slovenščino: Teoretska izhodišča, cilji, področja ukrepanja in institucionalni okvir. V D. Jagodic (ur.), *Tretja deželna konferenca o varstvu slovenske jezikovne manjšine* (str. 133–158). Slovenski raziskovalni inštitut. https://www.slori.org/publikacije/tretja-dezelna-konferenca-o-varstvu-slovenske-jezikovne-manjsine/

Denecke, K., Rivera-Romero, O., Giunti, G., Van Holten, K. in Gabarron, E. (2025). Key components of participatory design workshops for digital health solutions: Nominal group technique and feasibility study. *Journal of Healthcare Informatics Research*, *9*, 359–379. https://doi.org/10.1007/s41666-025-00199-4

ELRA Language Resources Association. (b. d.). *What is a Language Resource?*. Pridobljeno 15. julija 2026. https://www.elra.info/about/what-is-a-language-resource/

Gaspari, F., Grützner-Zahn, A., Rehm, G., Gallagher, O., Giagkou, M., Piperidis, S. in Way, A. (2023). Digital language equality: Definition, metric, dashboard. V G. Rehm in A. Way (ur.), *European Language Technologies* (str. 39–73). Springer. https://doi.org/10.1007/978-3-031-28819-7_3

Gerken, M. (2022). *Facilitating the implementation of the European charter for regional or minority languages through artificial intelligence*. Council of Europe.

Grgič, M. in Guardiancich, I. (2025). Minority-language socioeconomic policies and their impact on speakers' choices, habits and attitudes: The Slovenian community in Italy. *European Journal of Language Policy*, *17*(1), 27–44. https://doi.org/10.3828/ejlp.2025.3

Grgič, M. in Zver, Z. (2024). Načrtovanje rabe jezika za opolnomočenje manjšinskih skupnosti v umetnointeligenčni dobi. V M. Grgič (ur.), *Slovenščina v Italiji* (str. 213–231). Založba Univerze v Ljubljani. https://doi.org/10.4312/9789612974725

Jagodic, D. (2024). Demografska podoba slovenske skupnosti v Italiji. V M. Grgič (ur.), *Slovenščina v Italiji* (str. 31–55). Založba Univerze v Ljubljani. https://doi.org/10.4312/9789612974725

Ledinek, N. (2017). Opremljenost jezikovne skupnosti v RS z jezikovnimi viri in jezikovnimi tehnologijami kot jezikovnopolitično vprašanje: Potrebe in pričakovanja jezikovnih uporabnikov. V N. Gliha Komac in M. Jesenšek (ur.), *Jezikovna politika Republike Slovenije: Zbornik povzetkov*. Univerza v Mariboru.

LT4All 2025. (b. d.). *Overview*. Pridobljeno 20. februarja 2026. https://www.lt4all2025.eu/overview/

Mariani, J. J. (v tisku). Language technology for all: A challenge. V *UNESCO Report on languages*. https://hal.science/hal-04415222v1

Popič, D. in Grgič, M. (2023). Procesi jezikovnega separatizma pri čezmejnih jezikovnih manjšinah: Prevzemanje, prilagajanje in prevajanje covidne terminologije med Slovenci in Slovenkami v Italiji. *Annales. Series historia et sociologia*, *33*(1), 151–166. https://doi.org/10.19233/ASHS.2023.07

Popič, D., Bordon, D., Franza, J. in Bezgovšek, L. (2024). Jezikovna orodja za slovenščino v Italiji: Jezikovni asistent Loris 1.1. V M. Grgič (ur.), *Slovenščina v Italiji* (str. 167–194). Založba Univerze v Ljubljani. https://doi.org/10.4312/9789612974725

Resolucija o nacionalnem programu za jezikovno politiko 2021–2025. (2021). Pravno-informacijski sistem Republike Slovenije. https://pisrs.si/pregledPredpisa?id=RESO123

Yan, Z., Dong, Y., Niemi, V. in Yu, G. (2013). Exploring trust of mobile applications based on user behaviors: An empirical study. *Journal of Applied Social Psychology*, *43*(3), 638–659. https://doi.org/10.1111/j.1559-1816.2013.01044.x

Zver, Z. (2025). *Razširjenost jezikovnih virov in tehnologij med aktivnimi uporabniki in uporabnicami slovenskega jezika v Italiji: Primer Primorskega dnevnika* [Magistrsko delo]. Univerza v Ljubljani. https://repozitorij.uni-lj.si/IzpisGradiva.php?id=170946

Zver, Z. (2026). Jezikovna politika in jezikovno načrtovanje za slovenščino: Vloga digitalne opremljenosti za jezikovno enakopravnost slovenščine v Italiji. V V. Mikolič, D. Košir in M. Rolih (ur.), *Jeziki sožitja, jeziki konflikta* (str. 37–47). Annales ZRS.
https://doi.org/10.35469/978-961-7276-12-1