

slovenščina 2.0, letnik 11, številka 2

Slo2.0

2023_2

Univerza v Ljubljani
Filozofska fakulteta



Slovenščina 2.0

Letnik/Volume 11, Številka/Issue 2, 2023

ISSN: 2335-2736

Glavna urednika/Editors-in-Chief

Špela Arhar Holdt, Vojko Gorjanc

Uredniški odbor/Editorial Board

Zoran Bosnić, Simon Dobrišek, Tomaž Erjavec, Ina Ferbežar, Darja Fišer, Polona Gantar, Peter Jurgec, Iztok Kosem, Simon Krek, Nina Ledinek, Nikola Ljubešić, Nataša Logar, Karmen Pižorn, Damjan Popič, Marko Robnik Šikonja, Amanda Saksida, Irena Srdanović, Mojca Šorn, Darinka Verdonik, Špela Vintar

Tehnična urednica/Managing Editor

Eva Pori

Prelom/Layout

Irena Hvala

Založila/Published by

Založba Univerze v Ljubljani

Za založbo/For the Publisher

Gregor Majdič, rektor Univerze v Ljubljani

Izdala/Issued by

Znanstvena založba Filozofske fakultete Univerze v Ljubljani;
Center za jezikovne vire in tehnologije Univerze v Ljubljani

Za izdajatelja/For the Issuer

Mojca Schlamberger Brezar, dekanja Filozofske fakultete

Publikacija je brezplačna./Publication is free of charge.

Publikacija je dostopna na/Avaliable at: <https://journals.uni-lj.si/slovenscina2>

Revija izhaja s podporo Javne agencije za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije. (ARIS)/This journal is published with the support of the Slovenian Research and Innovation Agency (ARIS).



To delo je ponujeno pod licenco Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna licenca (izjema so fotografije). / This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (except photographs).

Kazalo vsebine

PANELNE RAZPRAVE

JAVNI POSVET O ZAHTEVAH PO ZNANJU SLOVENŠČINE..... 1

Marko STABEJ, Boris ČERNILEC, Ina FERBEŽAR, Furkan GUNER, Tina HEFERLE, Ana SAMOBOR, Katarina ŠTRUKELJ

RAZPRAVE

VARIATION OF SPOKEN SLOVENE IN THE CASE OF GEOGRAPHICALLY MOBILE SPEAKERS 29

Maja BITENC

ANNOTATED LEXICON FOR SENTIMENT ANALYSIS IN THE BOSNIAN LANGUAGE 59

Sead JAHIĆ, Jernej VIČIČ

POROČILA

34. EVROPSKA POLETNA ŠOLA LOGIKE, JEZIKA IN INFORMATIKE ESSLI 2023 84

Matej KLEMEN, Aleš ŽAGAR, Timotej KNEZ, Frenk DRAGAR, Marko ROBNIK-ŠIKONJA, Slavko ŽITNIK, Magdalena GAPSA, Mojca BRGLEZ, Katarina BREZOVAR, Špela VINTAR

DHH23 – HELSINKI DIGITAL HUMANITIES HACKATHON 2023 92

David BORDON

Javni posvet o zahtevah po znanju slovenščine

Marko STABEJ, Boris ČERNILEC, Ina FERBEŽAR, Furkan GUNER, Tina HEFERLE, Ana SAMOBOR, Katarina ŠTRUKELJ

Aprila 2023 so se s sprejemom novele Zakona o tujcih¹ v Sloveniji zaostri jezikovni pogoji za priseljene osebe. Zakon za združevanje družine določa znanje slovenščine na vstopni ravni oz. ravni A1, za pridobitev dovoljenja za stalno prebivanje pa znanje na ravni A2, to je na prvi oz. drugi od šestih ravni dokumenta Skupni evropski jezikovni okvir (SEJO)². Svoje znanje slovenščine bodo tako polnoletni državljani in državljanke tretjih držav od novembra 2024 za oba namena dokazovali z obveznim preizkusom iz znanja slovenščine.³ Center za slovenščino kot drugi in tuji jezik Filozofske fakultete Univerze v Ljubljani je v zvezi s tema ukrepoma, še posebej kar zadeva jezikovni pogoj za združevanje družine, večkrat javno izrazil pomisleke, 9. maja 2023 pa je bil na to temo na Univerzi v Ljubljani organiziran javni posvet, na katerem so sodelovali dr. **Boris Černilec**, državni sekretar, Ministrstvo za vzgojo in izobraževanje, doc. dr. **Ina Ferbežar**, Center za slovenščino kot drugi in tuji jezik Filozofske fakultete UL, **Furkan Guner**, Društvo Medkulturni dialog, **Tina Heferle**, državna sekretarka,

1 <http://pisrs.si/Pis.web/pregledPredpisa?id=ZAKO5761>

2 <https://centerslo.si/wp-content/uploads/2015/10/SEJO-komplet-za-splet.pdf>

3 Novembra 2023 je Državni zbor sprejel nove spremembe in dopolnitve Zakona o tujcih. Ta po novem za združevanje družine uvaja obvezni preizkus znanja slovenskega jezika na t. i. preživetveni ravni kot del neformalnih izobraževalnih programov.
<https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2023-01-3326?sop=2023-01-3326>

Stabej, M. et al.: Javni posvet o zahtevah po znanju slovenščine. Slovenščina 2.0, 11(2): 1–28.

3.25 Druga izvedena dela / Other performed work

DOI: <https://doi.org/10.4312/slo2.0.2023.2.1-28>

<https://creativecommons.org/licenses/by-sa/4.0/>



Ministrstvo za notranje zadeve, **Ana Samobor**, Pravna klinika za begunce in tujce pri Pravni fakulteti UL, mag. **Katarina Štrukelj**, Urad za oskrbo in integracijo migrantov. Posvet je moderiral dr. **Marko Stabej**, redni profesor na Oddelku za slovenistiko Univerze v Ljubljani in predstojnik Centra za slovenščino kot drugi in tuji jezik Filozofske fakultete Univerze v Ljubljani.

Zapis je pripravila **Neža Cerinšek**. Pri tem so bile izjave za namene lažjega branja skladenjsko nekoliko prilagojene značilnostim pisnega jezika, vendar po principu minimalne intervencije. Vabljeni govorke in govorci so zapis pregledali in predlagali (nevsebinske) popravke za boljše razumevanje besedila.

Marko Stabej: Dober dan, dober večer na javnem posvetu o zahtevah o znanju slovenščine in dostopu do znanja slovenščine, ki ga prirejamo na Centru za slovenščino kot drugi in tuji jezik Filozofske fakultete Univerze v Ljubljani. Veseli me, da ste prišli v tolikšnem številu. Veselim se že pogovora z gosti, pa tudi vi boste imeli priložnost kaj vprašati in povedati. Najprej pa sem zelo vesel, da med nami lahko pozdravim dekanjo Filozofske fakultete Univerze v Ljubljani, redno profesorico doktorico **Mojca Schlamberger Brezar**. Prosim vas za besedo.

Mojca Schlamberger Brezar: Hvala lepa. Spoštovani udeleženci in udeleženke posveta, spoštovani gostje. Lepo pozdravljeni na javnem posvetu o zahtevah po znanju slovenščine v imenu Filozofske fakultete Univerze v Ljubljani, kjer deluje Center za slovenščino kot drugi in tuji jezik, ki organizira ta posvet. Posebej pozdravljam sodelujoče na posvetu, državno sekretarko predlagateljka zakona Ministrstva za notranje zadeve gospo Tino Heferle, državnega sekretarja Ministrstva za vzgojo in izobraževanje doktorja Borisa Černilca, predstavnico Urada za oskrbo in integracijo migrantov magistrice Katarine Štrukelj, predstavnico Pravne klinike za begunce in tujce pri Pravni fakulteti Univerze v Ljubljani Ano Samobor, predstavnika Društva Medkulturni dialog Furkana Gunerja in predstavnico Centra za slovenščino kot drugi in tuji jezik docentko doktorico Ino Ferbežar ter seveda predstojnika Centra za slovenščino kot drugi in tuji jezik profesorja doktorja Marka Stabeja, ki bo vodil današnjo okroglo mizo. Hvala, da smo se lahko zbrali

tukaj v prostorih zbornične dvorane. V Centru za slovenščino kot drugi in tuji jezik, kjer so zaposleni vodilni strokovnjaki za področje slovenščine kot tujega jezika pri nas, imajo s problematiko posveta bogate izkušnje. Udeležba strokovnjakov z drugih področij pa tudi zbuja zaupanje, da bosta problematika zahtev po znanju slovenščine, kot jih opredeljuje najnovejši Zakon o tujcih, in dostop do znanja slovenščine za državljane in državljanke tretjih držav, predstavljeni z različnih zornih kotov in v širšem družbenem okvirju. Vsem sodelujočim želim uspešno in konstruktivno debato.

Marko Stabej: Hvala lepa gospe dekanji. Med vašim govorom so se spremenile okoliščine in se je napolnila dvorana, kar nas veseli. Za uvod bi prosil, da nam okvir tega, o čemer se pogovarjamo, predstavi kolegica, gospa **Katja Krajnc**. Prosim.

Katja Krajnc: Z novelo Zakona o tujcih, ki je bil po nujnem postopku sprejet aprila letos, so se v Sloveniji zahteve glede znanja slovenščine zaostriale. Za združevanje družine je bilo določeno znanje slovenščine na vstopni, tako imenovani preživetveni ravni oziroma A1, za pridobitev dovoljenja za stalno bivanje pa znanje na osnovni ravni oziroma A2. To sta prva in druga raven na šeststopenjski lestvici dokumenta SEJO in na prosojnicah si lahko ogledate kratek povzetek ravni.

Svoje znanje slovenščine bodo od novembra 2024 dalje polnoletni državljani in državljanke tretjih držav za oba namena dokazovali z obveznim preizkusom iz znanja slovenščine. Za izpolnitev teh pogojev bo na voljo 12 mesecev za združevanje družine oziroma pet let za stalno prebivanje. Določena raven znanja slovenščine je nujna, če si želimo, da se priseljena oseba vključi v skupnost, toda z zahtevo po opravljenem izpitu se Slovenija uvršča med strožje države Evropske unije. Iz utemeljitev in izjav različnih akterjev je mogoče sklepati o njihovem prepričanju, da bo opravljanje izpita rešilo vse težave v zvezi z nemotenim vključevanjem v slovensko družbo. Toda pri tem se morda ne zavedajo dovolj dejstva, da si ljudje jezikovno znanje pridobivajo postopoma in da je integracija dolgotrajen proces. Uspešnost učenja jezika je odvisna od številnih dejavnikov – od posameznikovega prvega jezika ter znanja drugih jezikov in siceršnje izobraženosti,

od nadarjenosti, motiviranosti, priložnosti za uporabo jezika in tako naprej. Zelo pomembno za uspešnost učenja je tudi število ur, ki jih posameznik namenja jeziku, in pa kakovost poučevanja. Poučevanje slovenščine kot nematerne jezika zahteva zelo specifične kompetence, ki jih učitelji pogosto nimajo, delo z ranljivimi skupinami pa je še posebej zahtevno. Zato mora biti ustrezna usposobljenost izvajalcev tečajev slovenščine ena od glavnih priorit. Njihovo financiranje bo namreč le tako prineslo želeni učinek. Po naših podatkih se izpitov na ravni A1, torej vstopni ravni pogosto udeležujejo osebe, ki so slabše izobražene in šibko pismene. Za to skupino je 12 mesecev za doseg ravni A1, kot jo za združevanje družin zahteva zakon, malo. Po novem Zakonu o tujcih bodo tečaji slovenščine za vse zainteresirane brezplačni, opravljanje izpita pa ne. A izkušnje kažejo, da je velika spodbuda za udeležbo na tečajih in sistematično učenje slovenščine prav izpit. A tega si bodo morali tuji državljanke in državljani plačati sami. Zato se v Centru za slovenščino bojimo, da se bodo ljudje udeleževali le izpitov, za tečaje pa se ne bodo odločali. Ker posledično na izpit ne bodo ustrezno pripravljene, lahko pride do goljufanja, ponarejanja listin in podobnega. Razmere bodo namreč težko pregledne, saj je morebitnih kandidat in kandidatov za izpite po nekaterih ocenah blizu deset tisoč. Zakon izvajanje izpitov namreč prepušča vsem izvajalcem javno veljavnega programa Slovenščina kot drugi in tuji jezik. Dogajanje ob sprejemanju Zakona o tujcih je pokazalo, da se odgovorni med sabo premalo usklajujejo. Nas pa, razen izjemoma, postavijo bolj ali manj pred dejstvo samo kot izvajalce, ne pa kot sooblikovalce integracijskih politik, povezanih z znanjem jezika. Ta posvet smo v Centru za slovenščino zato pripravili kot poznavalci področja, zavezani k strokovnemu in etičnemu ravnanju.

Marko Stabej: Hvala lepa. To mizo smo si organizatorji predstavljali nekako kot izmenjavo mnenj. Najprej bo iztočnico dobil vsak od udeleženeh, ki vas je na kratko predstavila že dekanja. Mikrofoni so zadaj, prosimo, da govorite v mikrofone. Potem pa je mišljeno, da se pogovarjamo med sabo in da se kasneje, v drugem krogu v vprašanji na neki točki malce kasneje vključi tudi publika. Predviden čas je do 19.30, če bomo kakšno minuto daljši, nas ne bo nihče vrgel z Univerze.

Zakon o tujcih je torej sprejet, integracijska politika in mesto jezikovnega izobraževanja in znanja v njej pa bo stalnica slovenske družbene in politične realnosti. Kot je bilo rečeno na koncu te predstavitve, je najbrž čas, da začnemo drugače, tesneje sodelovati in se pripraviti na to, da bomo lahko ta proces omogočali in ga naredili humanega in etičnega za vse in da bo učinkovit. Zato kot prvo k besedi vabim gospo Tino Heferle, državno sekretarko Ministrstva za notranje zadeve Republike Slovenije, ki je predlagatelj in nosilec tega zakona. Zakon je, kakršen je, nekaj časa je krožila celo nepreverjena informacija, da pogoja izpita za združevanje družine sploh ne bo, a zdaj je, nekaj je bilo tudi sprememb v zadnjem trenutku ... Kako ste si predstavljali sistem, ki bi te zahteve lahko uresničeval na dober način?

Tina Heferle: Hvala lepa, doktor Stabej, za povabilo na to okroglo mizo. Lep pozdrav vsem v mojem imenu. Kot je že gostitelj povedal, sem državna sekretarka Ministrstva za notranje zadeve RS, ki je nosilec Zakona o tujcih. Za začetek, novela Zakona o tujcih, ki je bila pred kratkim sprejeta po nujnem postopku, je prinašala veliko več vsebin, ki so pomembne za obravnavo tujcev, ki pridejo k nam. Te vsebine seveda zadevajo upravne enote, ki vodijo postopke v zvezi s tujci. Ta novela je bila res nujna, če smo želeli vsaj malo pohitriti upravne postopke in v tej luči narediti sprejem, če se tako izrazim, tujcev, ki pridejo v našo državo, bolj prijazen, bolj human in predvsem hitrejši. Seveda se je pa največja razprava odvijala okrog jezika oziroma pogoja znanja jezika. Na začetku bi izpostavila dve dejstvi, ki smo jih lahko slišali v uvodni predstavitvi, a ne držita povsem točno.

Uvodoma je kolegica [Katja Krajnc] dejala, da so se s to novelo zahteve zaostriale. To v resnici ne drži, ker smo imeli do sprejema te novele v veljavi zakon, ki je določal ostrejši oziroma bolj zahteven pogoj znanja slovenskega jezika. A se ta določba še ni uporabljala, ker je bilo pred dvema letoma določeno prehodno obdobje. Če mi novele, takšne, kot smo jo zdaj sprejeli, ne bi vložili v obravnavo po nujnem postopku, kar bistveno skrajša vse postopke v državnem zboru, bi 27. aprila stopila v veljavo tista ureditev, ki je bila določena izpred dveh let pod taktirko prejšnje vlade, ki je bila za tujce bistveno manj ugodna in bolj zahtevna tudi glede jezika. Kar smo mi naredili, je, da smo

postavili pogoj znanja slovenskega jezika na ravni A1, kot je bilo izpostavljeno, v fazi, ko mora tujec, ki že prebiva pri nas eno leto na podlagi združitve z družino, ta pogoj uresničiti po enem letu. Tukaj je drugo dejstvo, ki je bilo malce napačno predstavljeno. Ne drži, da je zgolj 12 mesecev oziroma da imajo zgolj 12 mesecev časa, da se pripravijo na izpit. Zdaj smo uvedli prehodno obdobje 18 mesecev, v katerem bomo pripravili, kot je bilo rečeno in je izjemno pomembno, strategijo za vključevanje tujcev, ki bo bistveno širša od zgolj znanja slovenskega jezika. Potem pa, kot rečeno, ko bo tujec na podlagi prošnje za združitve z družino prišel v Slovenijo, bo imel časa 12 mesecev. V tem času bo obiskoval tečaje in na koncu – upamo – tudi uspešno opravil ta izpit na vstopni ravni. Kar se tiče koncepta, ki ga imamo v naši državi in je bil primerjan, da je bolj strog kot drugod po Evropi, je odvisno, s katere strani pogledamo. Koncept, ki ga imamo pri nas, je bistveno drugačen, kot ga imajo v Nemčiji, Avstriji ... Tam so že na vstopni ravni, predno tujec pride, postavljeni pogoji, med njimi tudi znanje njihovega jezika. Pri nas se v prvem trenutku, ko tujec pride v našo državo in se združi z družino, torej živi s svojo družino, oziroma za tujce, ki tukaj delajo, ne zahteva ničesar. Šele v časovnem obdobju enega leta se pričakuje, da se tujec integrira v družbo in navadi življenja tukaj, da se privadi, da spozna našo kulturo, način življenja in tako naprej, da potem za boljše vključevanje v našo družbo spozna in se nauči našega jezika. Toliko za uvod. Če na kratko povzamem: glavni namen pogoja, ki smo ga ohranili v zakonu, ampak naredili bolj prijaznega, je ta, da postanemo vključujoča družba, ki bo sprejemala, ki bo dostopna tujim državljanom, ki pridejo v našo državo, da jim damo priložnost. In kako drugače kot s tem, da jim nudimo brezplačne tečaje, da se naučijo našega jezika in s tem tudi lažje vključijo v družbo, v življenje v naši družbi in jim s tem tudi pomagamo. In seveda tudi z drugimi ukrepi, ki bodo sprejeti v okviru nove strategije za vključevanje tujcev.

Marko Stabej: Najlepša hvala. Samo pojasnilo; kolegica je prebrala, kar smo pripravili skupaj in tudi razlika v interpretaciji ni tako zelo velika. Nadaljeval bi z gospo magistrico Katarino Štrukelj, ki je direktorica Urada Vlade Republike Slovenije za oskrbo in integracijo migrantov, ki je tudi po zakonu telo, ki je pristojno za oskrbo in integracijo

migrantov. Kaj vse že imate v načrtu in kakšne so morebitne zagate pri uresničevanju tega načrta?

Katarina Štrukelj: Lepo pozdravljeni. Mi smo izvajalni organ za to zakonsko določbo. Najprej bi predstavila, kakšno je stanje. Pred letom in pol smo prevzeli tečaje začetne integracije priseljencev, ki trenutno potekajo po številnih mestih: Ljubljana, Maribor, Celje, Kranj, Koper, Šoštanj, Nova Gorica, Novo mesto, Krško in drugod. V leta 2021 je bilo vključenih 1545 oseb, v letu 2022 2853 in v letošnjem letu do danes 2184. Število udeležencev se seveda večja in zaradi določenih sprememb se bo še povečevalo. Jezikovne šole niso bile zmožne sprejeti vseh udeležencev, tako da danes obstaja čakalna vrsta. Na to smo se mi odzvali in objavili nov razpis. Razpis smo še razširili po drugih mestih. Razpis se bo v kratkem zaključil in pričakujemo, da se bo število prijaviteljev razširilo in da bomo čakalno vrstno do konca leta tudi zmanjšali. Izziv je, kako zmanjšati to manjšo čakalno vrsto ..., šole poročajo, da je premalo učiteljev, ki izvajajo te tečaje. Mi eno od rešitev vidimo v tem, nismo pa seveda strokovnjaki za to področje, in sicer da bi se znižali kriteriji, kdo lahko tukaj poučuje. Stroka se s tem ne bo strinjala ...

Marko Stabej: Pogovarjamo se.

Katarina Štrukelj: Ja, pogovarjamo se. Jezikovne šole poročajo, da na trgu ni dovolj pedagoških delavcev, ki bi tečaje lahko izvajali. Mi tukaj torej vidimo eno od rešitev. Pa tudi, da se razširijo izvajalci izpitov, kar smo seveda s to novelo tudi naredili. Drugi izziv pa je to, o čemer poročajo osebe, ki se udeležujejo tečajev, in sicer število ur. Ali 60, 80, 300 ur – kolikšno je število ur, ki bi zadostovalo, da bi ljudje tudi na ravni A1 dobili tako znanje, ne samo, da povejo »Jaz sem ...«, ampak da ko grejo na primer k zdravniku, se tam znajo pomenit in seveda v drugih institucijah. Je to 60 ur ali ni? Pričakujem, da se bomo o tem in seveda o diferenciaciji skupin pri pripravi strategije pogovorili s strokovnjaki. Zelo pomembno je predznanje udeleženca. Naše izkušnje in udeleženci govorijo, da je smiselno razmisliti, ali je Začetna integracija priseljencev program – odkar je bil narejen, je že kar nekaj časa – ali je

to program, ki se odziva na potrebe vseh različnih ljudi, ki prihajajo iz različnih držav. Nekdo, ki prihaja iz arabskega sveta, ali nekdo, ki prihaja iz Bosne, imata popolnoma drugačno predznanje. Ali pa nekdo, ki ni pismen in se mora še opismeniti ali ne piše v latinici. To so stvari, za katere pričakujemo, da nam jih bodo povedali strokovnjaki, saj sami nismo strokovnjaki s tega področja, in to bomo potem upoštevali pri pripravi strategije. Hkrati s strategijo bomo pripravili uredbo. Govorili ste o plačljivosti izpitov ... Moram povedati, da trenutno po veljavni uredbi izpiti niso plačljivi. Plačljivost ali neplačljivost se bosta urejali v uredbi, ki jo bo pripravil Urad vzporedno s strategijo.

Marko Stabej: Najlepša hvala. Torej smo že na poti do razrešitve nekaterih vprašanj. Prosil bi kolegico doktorico Ino Ferbežar, ker je nekaj stvari v zvezi s programom Slovenščina kot drugi in tuji jezik in drugimi zadevami. Nekaj besed o tem in o tem, kakšne izkušnje ima Center s populacijo na vstopni ravni, s poučevanjem in merjenjem znanja?

Ina Ferbežar: Hvala lepa. Odprtih je bilo kar nekaj vprašanj, skušala bom problematizirati določene stvari. Omenili smo program Začetna integracija priseljencev, novi Zakon o tujcih pa govori o programu Slovenščina kot drugi in tuji jezik. To sta dva popolnoma različna programa, namenjena za čisto različne stvari. Ali bo uredba še uveljavljala program Začetne integracije priseljencev – tega ne vemo.

Sam program ne govori o tem, kdo so po izvoru ljudje in kako dolgo nekdo rabi, da pride do nekod. Naša želja kot strokovnjakov je, da v jezikovnih programih nikoli ne bi bilo napisano, koliko ur je treba, da oseba doseže neko raven. Zakaj? Zato ker so izhodišča tako zelo različna – jezikovna, izobrazbena, kulturna in tako naprej. In nekateri za dosego določene ravni rabijo 60 ur, za nekatere pa niti 300 ur ni dovolj. Zato si ne želimo programov, ki bi omejevali ure. Naše ministrstvo za izobraževanje zahteva, da so vsi javno uveljavljeni programi ovrednoteni tudi v številu ur in v tem vidimo problem.

V zvezi z vprašanjem o A1 ... Naša izkušnja – že dve leti izvajamo program Slovenščina kot drugi in kot tuji jezik in izvajamo izpite po tem programu, izpiti so na vstopni ravni oziroma ravni A1. Tukaj uresničujemo Zakon o urejanju trga dela in po tem zakonu morajo brezposelne

osebe opravljati izpit, če želijo obdržati status iskalcev zaposlitve na zavodu za zaposlovanje. Velika večina kandidatov na vstopni ravni je na izpitu uspešna – takih je 70 %. Tisti, ki prihajajo iz oddaljenih jezikovnih okolij, pa so manj uspešni. Večina jih opravi 180-urni tečaj. 180-urni tečaj za dosego ravni A1 je za nekatere govorce zelo zelo kratek, sploh za tiste, ki so šibkejši pismeni, ki so slabše izobraženi, ki prihajajo iz oddaljenih okolij, ki morda niso opismenjeni v latinici ali sploh niso opismenjeni in tako naprej. Vse to je potrebno upoštevati pri številu ur in morda določati različno število ur za posamezne skupine ljudi.

Kaj pa A1 sploh pomeni? Omenili ste, da bi se morali znati pogovoriti pri zdravniku. A1 pomeni resnično zelo zelo omejeno zmožnost, omejeno na zelo enostavne, preživetvene, rutinske situacije, ki so ponavljajoče se, na katere človek naleti vsak dan in se nanašajo na čisto konkretne potrebe te ene osebe. Če pričakujemo, da se bodo znali pogovarjati, so pa to popolnoma druge ravni. Tukaj je potrebno gledati sorazmerno ... Omenili ste evropske države – res je, nismo najstrožji, smo pa med najstrožjimi. 7 držav v Evropi je leta 2018 imelo pogoj izpita na A1 pred prihodom v državo, za začasno prebivanje pa 10, 15 maksimalno, niti ne. Pa ne govorim o Evropski uniji, ampak o Evropi, o članicah Sveta Evrope, torej po raziskavah Sveta Evrope. Toliko, mislim, da sem dovolj povedala.

Marko Stabej: Najlepša hvala. Še več točk za nadaljnjo debato se je odprlo, ampak gremo s tem prvim krogom naprej. Gospod Guner, ste predsednik Društva Medkulturni dialog v Ljubljani. Kako gleda na te zahteve in nasploh na jezikovni vidik integracije vaše društvo. Če smem vprašati pa še, kakšne izkušnje imate v zvezi s tem vi osebno kot tuji govorec slovenščine.

Furkan Guner: Najprej hvala za besedo in hvala, ker ste me povabili. Verjetno sem najboljši obraz kot sogovornik in lahko pričakujete, da ne bom govoril tekoče in slovnične slovenščine. Za začetek se zato opravičujem in upam, da bom razumljiv. Kot ste rekli, delam za Društvo Medkulturni dialog že nekaj časa. Za migrante in begunce smo organizirali projekte in aktivnosti v zvezi z integracijo in aktivnim državljanstvom

v Sloveniji. V Sloveniji živim 10 let in lahko rečem, da je slovenščina eden težjih jezikov v Evropi, to že vsi vemo, ampak tudi vem, da učenje ni nemogoče, da se lahko vsi učimo slovenščino, ampak imamo veliko ovir pri učenju. Kot je gospa Katarina že rekla, mislim, da je največji problem število ur. Posebej za ljudi, ki prihajajo iz Turčije, Afganistana, iz Afrike 400 ali 500 ur... Jaz sem jih imel 150, nekaj takega, tudi zavod za zaposlovanje jih da nekaj. Ampak večinoma je med sklopi tečajev veliko odmorov. Tečaji potekajo po 60 ur, torej 60, 60, 60 ... In včasih se vmes čaka tri mesece. Intenzivnost, kvaliteta tečaja in število ur so trije dejavniki, ki so najbolj ključni za učenje jezika. Tako jaz vidim. Mislim tudi, da v Sloveniji manjka praktično usposabljanje za slovenščino. V desetih letih sem videl, da so Slovenci zelo prijazni, aktivni, športniki. Migrantom in beguncem moramo ponuditi tudi aktivnosti, ki so povezane s športom. Ali pa z naravo, sprehode v naravo ... Zdi se mi, da take stvari manjkajo v Sloveniji. Če bi to lahko vključili v integracijske programe za migrante in begunce, mislim, da bi to res pomagalo. Vsi bi se lažje učili in prakticirali slovenščino. Zaenkrat toliko, če imate še kakšno vprašanje ...

Marko Stabej: Najlepša hvala za časovno varčnost. Gospa Ana Samobor, vodite Pravno kliniko za begunce in tujce pri Pravni fakulteti Univerze v Ljubljani. Ne vem, ali sem samo jaz neobveščen, najbrž ne, zato vas prosim, da nam najprej sploh pojasnite, kaj pravna klinika je. Potem pa po svoji presoji, izkušnjah in raziskovanju, kakšen je vaš pogled na tematiko.

Ana Samobor: Najlepša hvala. Hvala za povabilo na današnji pogovor. Najprej nekaj besed o pravni kliniki. To je pravna svetovalnica za begunce in tujce, ki jo že več kot 30 let izvajamo v okviru Katedre za ustavno pravo. Trenutno sem koordinatorka jaz, ki sem asistentka za ustavno pravo na Pravni fakulteti Univerze v Ljubljani. Pred mano je bil to profesor Bardutzky [Samo]. Morda je ime nekoliko zavajajoče, mi ne nudimo direktnih pravnih nasvetov beguncem in tujcem, ampak izobražujemo študente in študentke na področju azilnega prava in sodelujemo z zunanjimi mentorskimi organizacijami, pri katerih študenti, ki so udeleženi, opravljajo 80-urno prakso. Naša osnovna ideja

je izobraziti mlade pravnike na tem področju, kar počnemo skupaj z izbirnim predmetom Pravo migracij in azila, ki ga poučuje profesor Bardutzky. Toliko o pravni kliniki.

Pogled na to tematiko – zdi se mi zelo pomembno, da se poskušamo odmakniti od situacije in pomisliti na to, da ko govorimo o raznih integracijskih pogojih, npr. o jezikovnih integracijskih pogojih, govorimo o pogojih, ki tendirajo v omejitve ali pa kršenje človekovih pravic. Zlasti ko govorimo o izpolnjevanju jezikovnega pogoja za združitev družine in podaljšanje dovoljenja za prebivanje. Tukaj lahko govorimo o hudem posegu v pravico do družinskega življenja. Mogoče še to: če nam nekaj dopušča pravo Evropske unije, ni nujno, da to dejansko uvedemo. Sploh pa ne v obliki pogoja, ki lahko pomeni kršitev človekove pravice. Predstavljam si, če gledamo skozi prizmo sorazmernosti urejanja nekega področja, integracijske strategije, politike, da obstajajo verjetno blažji ukrepi. Oziroma blažji, milejši, bolj spodbudni načini za integracijo tujcev v našo družbo, kot je to trenutno urejeno v Zakonu o tujcih v obliki pogoja za združevanje družine. Da dodam še to prizmo človekovih pravic, zlasti pri pravici do družinskega življenja, ki se mi zdi nekaj, na kar mogoče pozabljamo, pa ne mislim na današnjo razpravo, ampak na splošno. Ko se začnemo podrobno pogovarjati o pogojih, ki jih mora nekdo izpolnjevati, da lahko tukaj prebiva, hitro pozabimo, da je to človek, ki tako kot ostali uživa človekove pravice, in moramo v tem kontekstu to upoštevati, kljub temu da imamo kot družba in država neke interese. Nasproti stoji nek posameznik s človekovimi pravicami, kakršnokoli omejitev človekovih pravic moramo razumeti skozi prizmo sorazmernosti s tem posegom.

Marko Stabej: Najlepša hvala. Nenazadnje, gospod državni sekretar z Ministrstva za vzgojo in izobraževanje, doktor Boris Černilec, s tem zakonom vaše ministrstvo neposredno skoraj nima nič, razen premalo slovenistov in slovenistk ... Seveda ima veliko, ampak ta zahteva je specifična, kot vse bolj spoznavamo tudi v tem okviru. Problem integracije je ena od nalog sodobnega šolskega sistema, ampak ne moremo na široko o tem. Kdo bo pripravil predlog in na kakšen način bomo dosegli neko soglasje, kdo bi lahko izvajal ali pa uresničeval tovrstne pogoje? Ali se ministrstvo s temi vprašanji ukvarja? In ali lahko kakšno rečemo tudi o tihih naslovnih združevanja družin, ki jih ta zakon s to

zahtevano neposredno ne zadeva, to so otroci in mladostniki, tisti, ki so vključeni v slovenski šolski sistem? Vemo, da se precej spreminja in dogaja v pravo smer glede usposobljenosti in pripravljenosti šolskega sistema za integracijo, ampak da je še vedno veliko sivih lis, ko so ljudje sicer tehnično vključeni, ampak ne *de facto* in tako naprej. Vaš komentar, mnenje ... Prosim.

Boris Černilec: To je bil dolg uvod ... Takole mislim: če nam daste delo na šolskem področju, se tega vedno veselimo. Tudi v tem primeru je po moje tako, ampak trenutno imamo težavo v celotnem šolskem prostoru, pa ne samo šolskem, da primanjkuje strokovnih delavcev, ne le učiteljev slovenščine, ampak tudi drugih. Ga ni v Sloveniji, ki bi pristal na to – po duši sem učitelj, bližje mi je matematika kot slovenščina – da bi jaz lahko učil slovenščino na takem tečaju. To ne gre skupaj in ni pričakovati, da bi zelo zelo znižali ta standard. Načeloma je šolski sistem od vrtca do konca srednje šole koherenten in primerljiv, zato ni pričakovati nekih velikih ukrepov. Razen, recimo, ko je bil čas korone, ko so se v družbi že čutile neke posledice zaradi ukrepov, takrat so se sprejele neke interventne stvari, morda za kratko obdobje, da se je rešilo in je veljal blažji ukrep.

Ko sem začel učiti, sem učil v okolju, v katerem je bilo 90 % otrok Neslovencev. V šoli so otroci govorili slovenščino, doma pa materni jezik. To globoko podpiram, to so danes uspešni otroci, ne glede, iz katere republike so bili. Doma sem blizu Kranja, tam je bila industrija zelo močna, delavcev pa ni bilo. Zdi se mi, da se je ta dogodek v 40, 50 letih spet ponovil. Da pa smo vključujoči in smo razvita družba, se vidi v ukrepu, da tujce sprejmemo in jim omogočimo učenje slovenščine. Obenem pa je še nekaj res: dva otroka v istem razredu si nista enaka, popolnoma sta si različna in diferenciacija je edini način, da neko stvar obvladamo, in ne vidim nobene možnosti, da bi se lahko samo iz enega tečaja, 100, 180 ali 500 ur, naučil slovenščine. Če tega ne uporabljáš, je zgodbe konec. Zadnjih 10 let sem delal na Zavodu za gluhe in naglušne Ljubljana. Ko sem tja prišel, nisem vedel ničesar o gluhoti. Moral sem se naučiti vsaj osnov slovenskega znakovnega jezika. Tam nisem bil učitelj, ampak direktor, ampak danes v komunikaciji z gluхими uporabljám slovenski znakovni jezik, oni pa mi pomagajo, ker ga ne znam

dobro. To je ta vključujoča družba, razumevanje družbe do tujcev ali pa do otrok, ki imajo posebne potrebe in so drugačni od ostalih. In na ta način je treba k tej vsebini – to ni problem – k tej vsebini pristopiti pozitivno. Še začeli nismo, pa se že sprašujemo, kaj bo. Poskusiti moramo. Vem pa, da imamo problem, ni učiteljev slovenščine. Imam dve rešitvi, saj sta čisto matematični: učiteljem, ki učijo na tečajih, damo 50 evrov na uro, tujcu, ki gre na izpit, pa rečemo: plačaš ga, če boš izpit opravil, ampak boš dobil denar nazaj, drugače ga boš pa še drugič plačal ... Saj je grdo, a razumete? Ampak to je slovenska oblika.

Marko Stabej: Hvala za rešitve. A kar končamo? [smeh] Najlepša hvala za zelo koherenten odgovor na moje nekoherentno vprašanje. Zanimivo ste opozorili na nekaj, kakorkoli imamo še različne ali pa ne čisto enake poglede na to, kako tehnično izobraziti in testirati znanje, in tudi to, kaj to znanje že pomeni oziroma še ne pomeni, česar pa še nismo rekli in ste vi omenili in je ključno, da izoliran ta proces ne bo poskrbel za integracijo. Moramo se tudi materni govorniki in govornice slovenščine naučiti drugače ravnati v komunikaciji in se marsikdaj prilagoditi na različne načine, če želimo imeti harmonično družbo.

Tina Heferle: Med razpravo sem si zapisovala, kar veliko dobrih iztočnic sem zasledila. Gospod [Guner] je dal dober predlog oziroma poziv, da mora integracija tujcev vključevati tudi športe in tako naprej. Naj vam odgovorim, da ko bomo imeli jutri prvi sestanek delovne skupine za pripravo strategije za vključevanje tujcev in ko se bomo lotili tega projekta, imamo že zdaj v mislih celostno integracijo oziroma vključevanje tujcev. Vključeni bodo vsi resorji, ki pokrivajo določena področja, kar se tiče naše družbe, od športa do kulture, do ministrstva za delo. Kar pomeni, da bomo pripravili strategijo, ki bo tujcu omogočala, da našo družbo spozna tudi iz drugih vidikov, in mu s tem omogočala lažjo integracijo, ne le skozi jezik, ampak skozi spoznavanje kulturnih značilnosti, običajev, značilnosti naše družbe. To zelo pozdravljam, slišala sem vas in to bomo upoštevali.

Drugi poziv je bil s strani predstavnice inštituta [Centra za slovenščino] v zvezi z diferenciacijo oziroma različnim številom ur, potrebnih za učenje jezika glede na sposobnosti kandidatov. Tudi o tem bomo

temeljito razmislili. Naj poudarim, da zakon kot tak, ki je krovni predpis, ne govori o številu ur, ampak bomo to urejali s podzakonskimi akti, npr. z uredbo ali pravilniki, kjer to lahko naredimo bolj življenjsko in bomo temu zagotovo sledili. Seveda s tesnim sodelovanjem kolegov iz šolstva, ki bodo najbolj podkovani za oceno, koliko ur v določenih primerih. Moram pa se odzvati še na eno stvar, kar se tiče razprave kolegice s pravne fakultete v zvezi s človekovimi pravicami in pravico do družinskega življenja. Tudi mi na ministrstvu se zelo zavedamo temeljnih človekovih pravic, spoštovanja in pomembnosti le-teh in med njimi je izjemno pomembna pravica do družinskega življenja. Naj kot pojasnilo povem, da je zakon dovolj skrben – vsaj po naši oceni – ali pa pazljiv predvsem na pravico do družinskega življenja, kajti predvideva določeno izjemo, ki pravi, da v primeru, ko tujec, ki zaprosi za podaljšanje dovoljenja za prebivanje, ne izkaže pogojev, ki so predpisani zakonsko – eden izmed njih je tudi znanje oziroma da se opravi izpit slovenskega jezika – se pretehta, kako trdna je ta družinska vez, kako dolgo je tujec prebival v naši državi in vse ostale značilnosti. Na podlagi tega se potem izda dovoljenje za prebivanje, če se oceni, da je ta družinska vez dovolj trdna za upoštevanje te izjeme. Tako da res verjamem, da so odločevalci na upravnih enotah, ki vodijo te upravne postopke, dovolj ozaveščeni. Na ministrstvu večkrat letno – če ne na mesečni ravni – opravljamo izobraževanja in konzultacije z uslužbenci na upravnih enotah in na to konstantno opozarjamo, naj bodo pozorni na ta vidik zakona. Naj izpostavim tudi vidik zakona, ki posebej obravnava ranljive skupine, kot so na primer otroci, starejši. Zakon izrecno predpisuje, da izpita iz znanja slovenskega jezika ni potrebno opraviti otroku, ki je mlajši od 18 let, ali osebi, ki je starejša od 60 let. Vemo, da obstajajo razlogi za takšno obravnavo. V razpravi smo tudi že slišali, da so nekateri tujci, ki pridejo v Slovenijo, nepismeni, tudi za te obstaja izjema v zakonu. Tudi njim ne bo treba priložiti dokazila o opravljenem izpitu. Pa osebe z okvaro zdravja, ki preprečuje uspešno opravljanje izpita. Novela Zakona o tujcih je na nek način naslovila te najbolj očitne primere, ko tujec potrebuje posebno obravnavo. Seveda pa si ne delam utvar, da je zakon popoln. Med uporabo zakona bo prihajalo do različnih situacij, ki jih bomo morali skupaj z uradniki detektirati na operativni ravni, tudi razčleniti, nasloviti rešitve in morda z novo

novelo situacijo še izboljšati. Na tem področju smo odprti, nimamo zacementiranega predpisa, zakon je živ oziroma pravo je živo in se prilagaja družbenim razmeram tam in takrat, ko je potrebno.

Ana Samobor: Najprej ste omenili individualizacijo. Če nekdo ne bo izpolnjeval pogoja, naj bi v upravnem postopku upoštevali trdnost družinskih vezi, na osnovi katerih bi se izdalo dovoljenje prebivanja, tudi če ne izpolnjuje pogoja. Glede na vsa opozorila – sicer v praksi ne delam na tem področju, ampak poznam opozorila nevladnih organizacij, ki delajo na tem področju – imam občutek, da se ta člen v pravnih postopkih ne uporablja ustrezno. Mislim, da bi ga bilo treba pri uvedbi takega pogoja še izrecno vključiti, torej ustvariti še eno varovalko, ki bi preprečila kršitev človekove pravice. Varovalka bi poskrbela, da ne pride do nepodaljšanja dovoljenja prebivanja in do pretrganja družinskih vezi, če nekdo zaradi posebnih okoliščin ne more izpolniti tega pogoja, kakršnega smo zdaj postavili.

Še ena stvar, na katero sem pomislila že prej, v Centrovi razpravi. Šele zdaj se pogovarjamo, kako se bomo lotili poučevanja, integracije ... In to po tem, ko smo v zakonu določili pogoj za tujce. Zdi se mi, da smo začeli na napačnem koncu. Načeloma s tem ni nič narobe, pravo EU nam ne preprečuje ideje integracije, strategije integracije in postavljanja takih pogojev. Ampak moramo kot država, če gledamo skozi prizmo človekovih pravic, najprej zagotoviti, da bodo tujci lahko izpolnjevali te pogoje. Torej moramo imeti strategijo, preden to postavimo kot pogoj za prebivanje tujcev pri nas ali pa za prebivanje njihovih družinskih članov. Zdi se mi, da smo konceptualno pozni, pa nočem nikomur ničesar direktno očitati, ampak mogoče bi morali ubrati obratni vrstni red.

Tina Heferle: Strinjam se z vami, čisto smo zgrešili časovnico. Ampak naj se malo operem krivde, čeprav jo priznavam. Pogoj znanja slovenskega jezika je bil sprejet pred dvema letoma. V vmesnem času, ko so bili na Ministrstvu za notranje zadeve drugi ljudje, sistema niso pripravili do te točke, da bi se lahko izvajal. Strinjam se. Zdajšnja garnitura na ministrstvu smo takoj, ko se je odprl Zakon o tujcih, to zaznali in naslovili. Zato imamo zdaj 18 mesecev prehodnega obdobja. Strinjam

se z vami, preden zadevo sprejmeš, reguliraš, jo predpišeš, restriktivni ukrep označiš kot obvezen, je treba sistem pripraviti.

Marko Stabej: Res, da je malo pozno, ampak smo tukaj. [smeh]

Ina Ferbežar: Na Centru imamo točno takšne izkušnje tudi z državnimi organi. Zakon o urejanju trga dela je nastal tako, da je Ministrstvo za delo najprej pripravilo projekt skupaj z nami, projekt smo speljali, preizkusili in potem so se izpiti začel izvajati. Torej takšna dobra praksa že obstaja. Bi se pa rada navezala na teh 12 mesecev oziroma, kot ste rekli, je zadeva zamaknjena za 18 mesecev. Ampak še vedno se bodo morali državljani tretjih držav v 12 mesecih naučiti do stopnje A1 za združevanje družine, vam je pa zanimiva ideja, da bi razširili število ur. To bo šlo verjetno skupaj, večje število ur bo zahtevalo tudi časovno razširitev tečajev. V 12 mesecev boste težko strpali na primer 900 ur, to sicer lahko naredite, ampak to pomeni štiri ure učenja vsak sleherni dan. To je en problem. Učinkovit tečaj morajo seveda izvajati usposobljeni učitelji, res usposobljeni. Zato tukaj nivoja ravni usposobljenosti učiteljev ne gre zniževati, tu bi se z vami strinjala. Super je, da ste rekli tudi, da moramo ljudem dati priložnost, da bodo uporabljali slovenščino, kar je naloga nas vseh. Pa še na nekaj bi rada opozorila: če želimo biti vključujoča družba, potem izpit ne more biti pogoj. Izpit ni vključujoč. Izpit vedno izključuje. Izpit disciplinira, nadzira, izključuje. Mi smo pobudniki tega, da bi bil namesto izpita tečaj ali pa da se vse skupaj raztegne, torej da je manevrski prostor ...

Katarina Štrukelj: Ne pravim, da bo predpisanih 400 ur. Izkušnje pri osebah za mednarodno zaščito so, da se lahko 400 ur tečaja opravi v enem letu. Ne govorim, da bomo določili 400 ur, to boste povedali strokovnjaki ...

Ina Ferbežar: Seveda je to mogoče. To sta dva zelo intenzivna tečaja, dva semestra po štiri ure na dan, to je možno. Ampak vam povem primer: smo imeli pilotni tečaj za brezposelne, ampak ljudje niso imeli časa hoditi na tečaj – brezposelne osebe, tako da ... [prekriven govor]

Katarina Štrukelj: To je potrebno prilagoditi, torej tečaje se prilagodi, v nekaterih primerih jih bo potrebno podaljšati. Tečaje se lahko izvaja tudi v soboto, nedeljo ... V novem razpisu so višje urne postavke, naj povem ... Urna postavka se je z 20 povečala na približno 50 evrov ... Z novim razpisom ... Torej to ne bo rešilo vseh težav ... Smo pa objavili nov razpis, da se bodo tečaji izvajali tudi čez vikend, v večernem času in tudi na spletu, kar verjetno ni najboljša rešitev.

Tina Heferle: Še en predlog bo ob pripravi strategije za integracijo, skupaj z ministrstvom za gospodarstvo, da bi tečaj obiskovali in potem izpit opravljali med delovnim časom, torej na primer dve uri na dan določeno obdobje. Tukaj bo potreben dogovor z delodajalci, podpora s strani ministrstva za delo. To je eden od naših predlogov, ki ga bomo dali na mizo in v razpravo in se potrudili, da bi ugledal luč sveta.

Boris Černilec: Naj še enkrat rečem: nam je to izziv na šolskem področju, v celotnem šolskem prostoru pa je seveda izziv kader. V šoli se nam bo zgodilo, da nam bodo generacije drastično upadle. En primer: letos bo v vrtcih 100 oddelkov manj, kar je 200 manj zaposlitev. To so spet ene najmanjših generacij v Sloveniji. Žal pa so tudi tiste generacije, ki so zdaj stare od 22 do 32 let, najmanjše v Sloveniji. Se pravi, nimamo tistih, ki na faksu končujejo, pa da bi bili učitelji, potem so to tako majhne generacije, da je manj družin ... V bistvu smo se torej ujeli v dvojno past in iščemo rešitve. Čez 5 let bo obratno: učiteljev bo preveč in spet bo težko dobiti službo v šolstvu. Podobno se bo zgodilo z dinamiko tujcev. Kot mlajši nisem razumel, zakaj je v Angliji pet milijonov Angležev in 40 milijonov tujcev. Zakaj je v Nemčiji dva milijona tujcev, priseljencev za opravljanje dela. Mislim, da smo v Sloveniji v trenutku, ki je zahteven. Nanj se je treba primerno odzvati, iskati spodbudne rešitve. Sama ideja je, kot je rekla državna sekretarka, zelo spodbudna za ves slovenski prostor.

Marko Stabej: Hvala za pozitivno interpretacijo.

Furkan Guner: A lahko še nekaj dodam? Lahko govorim iz izkušnje, da kot tujec seveda ni zelo enostavno živeti v Sloveniji, najti službo, najti

stanovanje. Kot že veste, imajo tudi Slovenci s tem težave ... Vsi potrebujemo stanovanje in že to traja nekaj, da ljudje najdejo stanovanje, kjer lahko živijo in razmišljajo o prihodnosti. Vsi vemo, da so plače ljudi iz drugih držav, migrantov, beguncev, nizke. In nimajo časa, da bi se učili slovenščino. Težko jim je, veliko delajo in ne grejo na tečaj. In to traja, včasih 20, 30 let, da ljudje govorijo kot jaz ... Tudi jaz sem začel malo kasneje, slovenščino se učim zadnjih štiri, pet let.

Tudi na to moram opozoriti, da ljudje potrebujejo pomoč njihove skupnosti. To počnejo kulturni mediatorji in moram opozoriti, da je v Sloveniji zelo malo kulturnih mediatorjev. V Sloveniji ni veliko skupnosti iz Afrike, iz Turčije, iz Sudana ... Od tam je ljudi zelo malo, ne pridejo živeti v Slovenijo ... Ampak zadnjih šest ali pa sedem let smo tudi videli, da prihajajo tudi od drugod. V Nemčiji, Angliji imajo veliko izkušenj, ker se je to dogajalo od druge svetovne vojne, Slovenija pa ima izkušnje zadnjih deset let, zdaj se soočamo z nečim novim. Tudi danes nisem pričakoval veliko ljudi, ampak mislim, da je udeležba znak za večje potrebe po slovenščini v prihodnosti. Manjkajo medkulturni mediatorji in digitalizacija slovenščine. Več časa preživimo na telefonu ali računalniku, kot pa na primer beremo knjigo. Malo je digitalnih priročnikov za slovenščino; avdio, video, potrebovali bi več financiranja za digitalizacijo. Obstajajo knjige, na primer A1, A2, ampak ljudje se že dolgočasijo ... Zdaj več časa preživljajo na telefonu ... Če jim lahko omogočimo, da se lažje naučijo tega jezika, mislim, da bo to najbolje za prihodnost Slovenije, ne samo za migrante in tujce. Potrebovali bomo več ljudi, ki lahko delajo, živijo, prispevajo v slovensko kulturo.

Marko Stabej: Najlepša hvala. Zdaj pa kar odpiram ...

Andrej Fištravec: Pozdravljeni, prihajam iz Maribora iz jezikovne šole Hiša jezikov, kjer zadnje čase pomagam pri enem zanimivem programu usposabljanja indijskih medicinskih sester za delo. Slovenski sistem jih trenutno ne rabi oziroma so odzivi takšni, kot da jih ne rabi in trenutno se usposabljaajo za delo v Avstriji in Nemčiji. Moja draga prihaja iz Brazilije, zato mi je procedura slovenskega državljanstva preko znanja slovenskega jezika poznana oziroma imam nekaj osebnih izkušenj. Po osnovni stroki sem sociolog in ko gledam situacijo ne le pri nas, ampak

tudi v Evropi, bi rekel, da je zelo pomembno, da bi se vsi, ki pripravljate sistemske intervencije, zavedali, da družbena reprodukcija – ne samo v Sloveniji, ampak v evropskih družbah – ne bo mogoča brez tujcev. Brez tega ne bo mogoče. Tudi danes, ko se pogovarjamo, ne govorimo le o jeziku, ampak o kulturni praksi, o vpetosti jezika v socialni in kulturni prostor. To je kompleksen proces, ki se ne da rešiti z desetimi ali stotimi urami. To se mi zdi izjemno pomembno. Brez dobrih ljudi od drugod, ki doma pustijo družine, otroke, sorodnike, vso svojo tradicijo, kulturo in pridejo v čisto neznan svet, brez njih ta svet ne bo mogel več funkcionirati. Zato morajo biti rešitve ustrezne in ne evropocentrične, slovenijacentrične oziroma egocentrične. Kot razumem, bomo čez 18 mesecev dobili kriterije, ki jih zakon predvideva skozi strategijo, skratka, dolga procedura. Tu se mi zdi nujno, da pri tem sodeluje interdisciplinarni tim strokovnjakov. Hvala za ta posvet, hvala za vašo iniciativo in hvala za normalen pogovor, ki se odvija pred nami.

Bojan Hajdinjak: Predsednik Zveze ljudskih univerz Slovenije in direktor Javnega zavoda Cene Štupar iz Ljubljane. Se mi zdi, da mi vključimo daleč največ tujcev v vse naše izobraževalne programe. Pripravljavcem strategije bi samo položil na srce, da imamo kar nekaj mrež že razporejenih po Sloveniji. Ena taka mreža je mreža ljudskih univerz, ki opravlja tudi javno službo v javnem interesu in bi se lahko tovrstni programi, če lahko uporabim ta izraz, prepredli. Na tak način bi dobili neko stabilnost. Na sami ljudski univerzi imamo javno službo in svetovalce, ki so za to na voljo, ki jih država plačuje. Zakaj ne bi izkoristili to moč. Imamo svetovalna središča, medgeneracijske centre, večgeneracijske centre, družinske centre, različne integracijske programe, izobražujemo tudi medkulturne mediatorje ... Najprej preglejmo, kaj na tem področju v Sloveniji že imamo. Dogaja se, da imamo v azilnem domu mi in filantropija ali pa še kdo drug istočasno pouk, torej morda prihaja do izobilja ponudbe učenja jezika. Morda takšen poziv. Učiteljev dejansko ni. Mi vključimo 1000 oseb na leto v začetno integracijo priseljencev. Učitelji imajo trenutno močno pogajalsko moč. Učitelji v popoldanskem času za 23,25 evra na uro ne želijo več delati. Trenutno je taka situacija. Kazalniki, ki ste nam jih dali z novim razpisom, pa so ogromni. Mi se bomo potrudili po celi Sloveniji te kazalnike doseči.

Ampak prosimo obe ministrstvi, da najdeta nek kompromis vsaj za to prehodno obdobje 18 mesecev. Pri nas se je začela panika, ko se ni vedelo, ali bo ta pogoj sprejet ali ne. Bilo je res noro, prihajale so prošnje, na izpitih je bilo 200, 300 prijavljenih ljudi. To so kapacitete, ki jih zelo težko izpelješ na enem izpitnem roku. Zdaj se je ta zadeva malenkost umirila, ljudje so se začeli odjavljati od izpitov, ampak predvidevam, da bo ta bumerang junija ali pa septembra spet prišel nazaj. Zato prosimo, da poiščete neko rešitev. Po eni strani vemo, da je stroka zastopana maksimalno, ampak hkrati, če želimo te kazalnike doseči ... Nek kompromis je potreben ali pa bo nekaj trpelo: ali kazalniki ali pa bo stroka morala popustiti, da bo te kazalnike dosegla ...

Iz publike: Dober dan. We are an initiative of asylum seekers and we would like to say a few words about integration. We have come to Slovenia because we were forced to flee from our countries and seek international protection. Firstly, we would like to thank the people of Slovenia for accepting us and considering our difficult situation. We feel welcome here and are doing our best to integrate into society. However, we would like to address an important issue and we hope that we can resolve it in the common interest of all. Most of us have arrived in Slovenia through Croatia and have asked for asylum here. You know the reasons why we didn't stay in Croatia. Over the last years the media, human rights organizations, the courts, and even the European Commission have all made it clear that the Croatian police systematically abuse refugees. Refugees are violently beaten, robbed, tortured, humiliated, and returned to Bosnia without having the chance for a fair asylum hearing. We can confirm these reports are true because we experienced the violence and racism ourselves. We have been shot at, beaten, stripped, humiliated, and denied even the most basic help. Even for those of us who were able to request asylum in Croatia, there was no fair procedure – we were intimidated by the police and then given a paper telling us we must leave the country. We had no choice but to escape to Slovenia, where we are safe from this kind of violence. We would not risk our lives to come here if it was not absolutely necessary. Here we finally found peace and decent treatment, we were able to settle and treat our

trauma, and the trauma of our children. You can imagine our shock when after some months, the Slovenian ministry told us they would send us back to Croatia! How is this possible? People whose basic rights were violated, who were hurt and traumatized, are now being returned to the same conditions – even families with small children, pregnant women, women who are breastfeeding, seriously ill and injured people! We all know that we have no chance for a fair asylum process in Croatia, so what will we do after that? Sending us back is like a death sentence. And imagine how it is for us living under the pressure that any day the police can come to take us and our families away. We have addressed this issue with the ministry many times and provided a lot of evidence. We have invited the ministry representatives to meetings, and also to our public conference. So far, they have never come to listen to us. The only reply we got was a generic statement saying that all of this is legal. And in the public, the ministry continues to deny that anything they are doing is problematic. This very week, just a few days after our public conference, the ministry is arranging deportations of people in the family camp in Logatec. At this time, we were surprised to hear that this very same ministry is preparing a strategy how to attract foreign workers, and also to help foreigners with integration. This makes our situation even more absurd. What about us, who are already here? Who are already studying the Slovenian language? Whose children are already visiting school? Who are already working? Who have made friends and are active in society? Why are we not welcome here? It is hard to believe that these promises about integration are sincere when the ministry is not prepared to accept even the most vulnerable refugees, who are most desperately in need of protection when it doesn't want to listen to us or speak with our organization.

We are here to ask the government representatives to respect their promises about integration. Even if it is completely legal to deport us, it is not legitimate. The fact remains: deportations of asylum seekers to Croatia are inhumane. Croatia does not have a functional asylum system, and the Croatian police are violent towards refugees. Until these conditions permanently and clearly change, all deportations must be called off and asylum requests in Slovenia must be

accepted. Slovenia has all the legal possibilities to do this, it only needs the political will. We are only asking that you respect our basic human rights. If these rights are not respected, we will actively work to prevent any of the refugees in the camps from being deported. Not one more deportation. This is a request sent to the ministry and the government signed by 23 organizations. We ask all of you to support our campaign and help the integration of foreigners to become a sincere and concrete effort, not just another word on paper.

Iz publike: Samo še nekaj bi dodala. Veseli nas, da se danes odvija ta posvet, na katerem se govori o integraciji in vključevanju tujcev v našo družbo. To izjemno pozdravljamo, hkrati pa bi res radi opozorili na stiske ljudi, ki so danes tukaj z nami v dvorani. Gospod iz inštituta Cene Štupar, vsi hodijo k vam na tečaje, tudi drugi, ki ste tukaj s področja integracije in azila, veste, da so to ljudje, ki se vsaj že pol leta učijo slovenščino in ki želijo ostati v Sloveniji. Ne zaradi nerazumnega sistema, kot se mnogokrat očita pribežnikom, ampak dejansko vztrajajo in se borijo za svojo pravico biti tukaj, želijo ostati tukaj, se trudijo, hodijo na tečaje jezika, ne enega, ampak mnogo njih. Vsi že hodijo v službo ali pa iščejo službe in medtem dan za dnem dobivajo odločbo o deportaciji na Hrvaško, kjer so bili tepeni, ponižani, mučeni in tako naprej. Mi se že nekaj časa trudimo skupaj z vsemi, ki so doživeli te grozote na Hrvaškem, ki ne želijo tja in ki so ostali tukaj, ker želijo biti del slovenske družbe, se želijo učiti, delati, sklepati prijateljstva, in vse to so tudi že naredili, želijo ostati tukaj ... Prosim, podprite jih. Če želimo resno debato o integraciji, kar je tema današnjega posveta, mislim, da moramo začeti pri ljudeh, ki želijo ostati tukaj, se želijo integrirati, da jih dejansko sprejmemo in jim dovolimo, da ostanejo tukaj. Naslovila bi predvsem gospo Heferle in gospo Štrukelj, če lahko to temo poneseta naprej pristojnim oziroma to sta tudi vidve, torej da se ta zadeva reši. Kot se ves čas govori, potreba po tujih delavcih je, potreba po spremembi Slovenije v vključujočo družbo je, zdaj imamo to priložnost, da naredimo vsaj prvi simbolni korak in vse te ljudi, ki želijo ostati tukaj, ki se že pol leta trudijo biti del te družbe, sprejmemo in dovolimo, da ostanejo in da jih podpiramo. Hvala. [aplavz]

Goran Lukič: V bistvu ni kaj dodati. Dal bi samo en krajši kontekst k temu, kar se je zgodilo. In sicer bilo je prej rečeno, ja, naredili smo nekaj napak s strani MNZ-ja [Ministrstva za notranje zadeve]. Ampak mislim, da ne gre za napake. Zakaj? Sam sem zelo natančno spremljal spremembe zakonodaje na področju tujcev in ne vem, če je bila napaka v spremembi zakonodaje, da ste sprva dali ven ta pogoj A1, potem ste ga dali pa nazaj noter. Skratka, to ni bila napaka. Zgodilo se je pač to, da se je vmes zamenjala ekipa na MNZ-ju, in potem se je zgodilo, da se je vrnil ta pogoj. Še enkrat poudarjam, to ni bila napaka. Zakaj je prišlo do tega, ne vem. Vem pa, da je pogoj A1 iz predloga Zakona o tujcih šel ven, en teden kasneje pa je spet prišel nazaj noter. V vmesnem času ne vem, kaj se je zgodilo, zakaj se je to zgodilo, na kakšni osnovi ... Zelo mi je žal, da takšnih strokovnih debat, kot smo jo imeli danes, ni bilo predhodno pri pripravi sprememb zakonodaje. Zelo bi si želel, da so takšne strokovne debate del izobraževalnega postopka. Če niso, potem strokovne debate nadomestijo populistične debate. Grozno mi je bilo, ko sem v državnem zboru poslušal namesto strokovnih debat populistične debate. Na osnovi tega je potem marsikakšen poslanec bil v narekovajih od svoje volilne baze pritisnjen, da pritisne kakšen gumb, ki ga v sebi mogoče ni hotel. Grozno je, da to rečem. Seveda imajo poslanci po ustavi pravico in odgovornost, da individualno pritiskajo na gumb, ampak tam mi je delovalo to zelo kolektivno. To je tisto, kar se dogaja. Moram to izpostaviti, da tukaj ni bilo napak. V zakonu je zdaj spet ta A1 in tisti, ki ste bili zraven pri predlogu zakona, veste sami, kaj se je potem dogajalo pri tem, da je na koncu prišel ven ta člen. Vmes so se dogajala pogajanja na nivoju poslancev, poslanskih skupin, vlade, medresorskega usklajevanja, ki so bila skoraj panična. Na dnevni ravni so se spreminjali člani na tem področju. A bo dvoletno prehodno obdobje, enoletno ali pa ga sploh ne bo ... Ali bo s 27. aprilom začel kar avtomatsko veljati ta pogoj ... To se je dogajalo, ker ni bilo strokovne debate. Namesto tega so se usklajevali politiki na nivoju pritiska, populizma s terena, ki pa ni bil teren teh ljudi, ki danes govorijo iz svojih izkušenj, ampak teren tistih, ki imajo svoje politične pritiske, ki prevajajo izkušnje v svoje politične pritiske. In res mi je žal, da je na koncu prišel ven takšen zakon, da se naknadno pogovarjamo, kaj je narobe. In mogoče ena poučna

zadeva iz vsega tega, prošnja, da bi se naučili ... Jutri je dvojna zadeva na MNZ-ju, strategija reševanja tujcev in strategija migracij. Dajmo vsaj pri tem prisluhneti terenu, ki ni populistični, ki ni logistični, ki ni interesni, ampak je strokovni in izkustveni. In dajmo pri pripravi teh dveh strategiji vključiti lokalni teren, dajmo imet strokovne posvete na lokalnem terenu. Čisto konkreten poziv, da ne bo to ljubljancentrično, ministrocetrično, političnocentrično, ampak bo to strokovno in izkustveno. Naj se naredi serija konkretnih lokalnih posvetov, v katere bodo vključene lokalne organizacije, nevladne organizacije, medkulturni mediatorji, ki jih je tako malo, naj imajo besedo pri tem. Ne pa, da je to politično zaprto in potem podvrženo določeni politični odločitvi. Še enkrat, ne ponovite te napake. Moja izkušnja je ta, jaz nočem več poslušati Grimsa.

Marko Stabej: To je zelo specifično. [smeh]

Tina Heferle: Odzvala se bom na razpravo gospoda Lukiča. V eni točki se ne strinjam z vami, in sicer ko pravite, da je šlo za napako, da je ta člen ostal v zakonu. Ne gre za to, da je vmes na našem ministrstvu prišlo do menjave vodstva. Gre za to, da ko smo pripravljali vsebino novele Zakona o tujcih in ko smo zaznali situacijo, ki sem jo že prej omenila, da je prejšnja vlada sprejela člen, ki določa ta pogoj znanja slovenskega jezika, pa hkrati dve leti ni naredila popolnoma ničesar »na terenu«, smo najprej na hitro ocenili, da bo manj škode, če damo člen ven. Ocenili smo, da tega nismo sposobni izpeljati. Potem smo tekom debate, ki ste ji vi rekli panična pogajanja, po mojem pa je bila razprava znotraj koalicijskih partnerjev in poslancev – še preden je prišlo ven dejstvo, da ta člen ostaja, smo mi zelo odgovorno razpravljali – ocenili, da v okviru nekega prehodnega obdobja, v katerem ste vi sami zaznali, da je stvar debate, kakšna časovna dolžina mora bit sprejeta, da bo država pripravljena na novo ureditev, torej ocenili smo, da bomo v okviru določenega časovnega obdobja vzpostavili pogoje, da se brezplačni tečajji slovenskega jezika še vedno izvajajo, in potem z neko motivacijo pripravili tujce, ki obiskujejo te tečaje, da se našega jezika res aktivno učijo, s tem da opravijo na koncu tudi izpit. Še enkrat – ni bila napaka, ampak posledica člena v zakonu, ki je dolgotrajen, nekajmesečni

proces pogovarjanja najprej znotraj ministrstva, potem med resorji in nato časovno sovpadlo z menjavo vodstva. Žal mi je, da se je to zlorabilo na tak način. Odgovorno smo ocenili, da bomo takšno ureditev lahko izpeljali v določenem roku. Se pa strinjam z vami, in tudi jaz sem to ves čas poudarjala pretekla štiri leta, ko sem bila na drugi funkciji, a kljub temu v politiki, da imamo vse premalo strokovnega diskurza. Ne političnega, politika je za to, da postavi neke smernice, če želite, ampak stroka pa mora biti tista, ki postavi črto, po kateri bomo šli. Politika po mojem mnenju je tudi na nekih točkah premalo poslušala stroko. Pri nas se zdaj tega poskušamo v največji meri držati in, kot ste rekli, jutri začnemo dve delovni skupini, kjer bomo pripravljali strategije. Osebno si želim, ker obe delovni skupini vodim, da bomo stroko lahko kar največ vključevali in jo upoštevali, ker le tako bodo rešitve, ki jih bomo dorekli – konec koncev politika sprejme končno odločitev, tako pač je – bodo te odločitve oziroma rešitve legitimne oziroma bodo bolj legitimne, kot če jih zapove zgolj politika. Če imamo podporo v stroki, je potem v družbi tudi drugače in bolje sprejeto.

Marko Stabej: Najlepša hvala. Ima še kdo kakšno mnenje?

Primož Jamšek: Hvala za besedo. Primož Jamšek iz Urada za oskrbo in integracijo migrantov. Mogoče najprej ... Fantje, ki so v zelo veliki stiski zaradi ozračja na Hrvaškem, so zelo pridni učenci slovenskega jezika in vse pohvale okoli tega. Urad za oskrbo in integracijo migrantov deluje na tem področju, ampak nekako jim ne moremo pri tem pomagati. Zavezali smo se, in mogoče bo potrebno predebatirati v strategiji, glede spremembe Zakona o mednarodni zaščiti, da bi tudi učenje slovenščine postalo nekako zakonska pravica tudi prosilcev za mednarodno zaščito, kar zdaj ni. Učenje, ki ga trenutno ponujamo, temelji na dobri volji Urada, pa se nam zdi, da bi se morali ljudje dejansko učiti slovenščino že takoj, ko postanejo prosilci za mednarodno zaščito. Malce smo že omenjali digitalne možnosti učenja, ki se uporabljajo v drugih državah v predintegracijskih paketih. Ko govorimo o integraciji, govorimo o integraciji takrat, ko ljudje pridejo v Slovenijo, in govorimo tudi o predintegraciji, ki pripravi ljudi na okolje, še preden pridejo v Slovenijo, kar se tiče tujcev, migrantov, družinskih članov. Mogoče še, kar je izpostavil

Goran [Lukič]: prvi zakonski predlog ni imel tega, da mora biti sprejeta strategija za integracijo. Osebnostno sem nekaj let delal na tem področju in zelo sem vesel, da bomo končno dobili neko strategijo na tem področju. Tako kot je bila zaveza, da ne bo samo papir, ampak da bo imela konkretne ukrepe in načrte. In tudi finančne posledice, za katere se moramo zavedati, da so zelo pomemben vložek za naprej. Vsak evro, ki bo vložen tukaj, bo malo kasneje nekajkrat povrnjen.

Marko Stabej: Najlepša hvala. Naš čas se počasi izteka. Ima kdo še kakšno mnenje?

Miha Blažič: Hvala lepa za besedo. Najprej, kar se tiče Zakona o tujcih v povezavi z učenjem slovenskega jezika, ali je to bila napaka ali ne ... Mogoče je to diskusija za nek notranji politični krog, kar pa lahko povemo, je, da je pred sprejetjem tega zakona zagovornik načela enakosti jasno povedal, da gre za diskriminatorni člen, enako je povedal varuh človekovih pravic, enako je povedala pravna služba Državnega zbora, enako so povedali v kabinetu predsednice države. Skratka, institucije, ki so pristojne za urejanje teh zadev, in tudi razlog, zakaj je bil prvotno ta člen umaknjen, potem kar naenkrat ni bil več pomemben in dejansko se je vso stroko na področju človekovih pravic in na področju ustavnih pravic tukaj povozilo. O tem moramo biti iskreni. Ali je šlo za napako ali ni šlo za napako, ta mnenja so bila na mizi, bila so jasno povedana na tiskovnih konferencah in predstavljena vsem pristojnim in so bila povožena. Zato smo danes tukaj, kjer smo, in se pogovarjamo, ali bomo lahko zagotovili tečaje ali ne, ali bodo varovalke ali ne bodo varovalke ... Po tem ko sem od blizu videl ta proces, moram iskreno povedati, da ne verjamem tem varovalkam. Ne verjamem tem varovalkam zaradi tega, kar smo danes slišali s strani nosilcev za azil. V četrtek je napovedana deportacija mlajše osebe iz družinskega azilnega doma v Logatcu, njegova sestra bo tu ostala brez njega, imamo družinske člane, nosečnico, osebe z majhnimi otroki, ki obiskujejo šolo in so v deportacijskih postopkih. Zanima me, kje so tukaj varovalke, enostavno jih ni. Ni se izdalo odločbe in potem se zgodi, karkoli se zgodi. Potem se zapleta in se za nazaj ukvarjamo s tem. Vsem tukaj polagam na dušo, da če želijo biti strategije o integraciji karkoli drugega kot piar,

da opravičijo vse te napake in kršitve človekovih pravic za nazaj, če hočejo imeti kakršnokoli kredibilnost, potem je prva stvar, ki se mora zgoditi, varovanje osnovnih in temeljnih pravic najranljivejših, ki so tukaj in se želijo vključiti v družbo. Če tega ne zmoremo za peščico ljudi, ki najbolj iskreno pristopajo k temu procesu, ki so vključeni v družbo in skušajo vse narediti za to, in niti nismo zmožni normalnega dialoga v zvezi s tem, imamo resen problem, kaj pomeni strategija o integraciji in čemu je zares namenjena. Samo to bi vam položil na dušo: za kakršnokoli kredibilnost tega procesa je potrebno najprej poskrbeti za osnovne človekove pravice, in to najranljivejših v družbi. Mislim, da morajo biti spoštovane in da smo k temu zavezani vsi. [aplavz]

Katarina Štrukelj: Gospod Miha, slišimo vas, ampak naša naloga je, sploh Urada, ki bo izvajal to strategijo, da smo prvi, ki se bomo z ministrstvom zavzeli za to, da bo strategija taka, da bo odgovarjala na potrebe na terenu. Kar se tiče Urada, se ne more zgoditi, da nekdo ne bi opravil izpita in bi bil zato izgnan oziroma ne bi več legalno živel v Sloveniji. Jaz in gospa Tina, ki bo vodila to skupino, se bova sigurno zavzeli, da pripravimo tako strategijo, ki bo uporabna v praksi in ne bo samo črka na papirju.

Furkan Guner: Mislim, da sem danes vabljen kot reprezentant tujcev ... Še enkrat moram povedati tole: enkrat smo bili na srečanju in sem vprašal, zakaj ne investiramo ta denar ali pa še več ur ali nekaj takega. Eden izmed reprezentantov zavoda – ni treba imena – je rekel, zakaj investirati v ljudi, ki bodo šli čez eno leto v drugo državo. Ampak tega ne vem, a ne? Tukaj so, skupaj z nami in še ne vemo, če lahko investiramo ali ne ... Ker tukaj res potrebujemo podporo države, zavoda in vseh ... Vem, da se vsi trudite in ste že naredili veliko, za to se vam zahvaljujem, ampak moramo še več investirati v tečaje in izkoristiti vse možnosti za integracijo migrantov in beguncev, ker jih imamo v Sloveniji res zelo malo. In res rabimo vašo podporo. Hvala lepa.

Marko Stabej: Hvala vam. To so kar primerne zaključne besede. Iskrena hvala vsem, ki ste prišli, izrazili svoje mnenje, svoja stališča, odgovarjali na vprašanja in postavljali teze. Če lahko za zaključek rečem: mogoče

zvonimo po toči. Ampak po drugi strani je tak dogodek ... Tudi na Centru, kjer se ukvarjamo s specifičnim, jezikovnim vidikom integracije, za katerega vemo, da ni vse, vemo, da najbrž ni najvažnejši, je pa pomemben, smo tudi na nek način bili frustrirani, da se smemo s tem ukvarjati samo, ko nekdo drug sklene, kaj bo, česar mogoče niti ne razume, vsaj pretežno vsi ne razumejo, o čem sploh govorijo. Puška v koruzi nima kaj početi, zato smo naredili ta posvet, ki bo mogoče samo kaplja v morje, mogoče pa bo začetni košček sestavljanke, ki bo dala lepšo sliko. In to, da ste prišli zelo odzivni, zelo odprti predstavniki različnih organizacij, tudi izvršne oblasti, je en znak, se mi zdi. Se bomo pa morali naučiti dojemati Slovenijo kot odprto državo, kjer ne bomo vsakega tujca gledali, »kdaj boš pa šel nazaj«. To bo trajalo nekaj časa. Celo tisti, ki se ukvarjamo s tem, imamo marsikdaj težave, kaj šele tisti, ki so v kakšnih drugih varnih kotih lastne iluzije. Tega ne moremo prepričati, lahko pa skupaj načrtujemo bolj človeško in vključujočo sceno, če smem reči. Najlepša hvala vsem in upam, da se še kdaj vidimo.

Variation of Spoken Slovene in the Case of Geographically Mobile Speakers

Maja BITENC

Filozofska fakulteta, Univerza v Ljubljani

Based on the pioneering variationist research on Slovene language material, the article explores the authentic language use of speakers of the Cerčno dialect who commute daily or weekly to work or school in Slovenia's capital, Ljubljana. The research corpus comprises excerpts from informants' full-day audio self-recordings, selected to represent variation of location, interlocutors, and conversation topic. Through a quantitative analysis of five phonological variables, the study reveals distinct language strategies with speakers of other varieties, resulting in a classification of one code-switcher, three code-mixers, and one dialect speaker language strategies employed by the informants when communicating with speakers of other varieties, resulting in a classification of code-switchers, code-mixers and dialect speakers. Furthermore, the qualitative data derived from interviews shed light on various sociopsychological aspects of the focal topic, such as the informants' perception of their own speech behaviour, their experiences with language use, and their language attitudes, providing deeper insights into their language variability and highlighting the intricacies of language choice and interpersonal communication.

Keywords: variation of spoken Slovene, Cerčno dialect, speakers' profiles, self-recording, sociolinguistic interviews, language attitudes

Bitenc, M.: Variation of Spoken Slovene in the Case of Geographically Mobile Speakers. Slovenščina 2.0, 11(2): 29–58.

1.01 Izvirni znanstveni članek / Original Scientific Article

DOI: <https://doi.org/10.4312/slo2.0.2023.2.29-58>

<https://creativecommons.org/licenses/by-sa/4.0/>



1 Introduction

The aim of the research presented in this paper is to investigate of the diverse language practice and intralanguage variation of five mobile Slovene speakers of the Cerčno local dialect who commute daily or weekly to work or school in Slovenia's capital, Ljubljana, as well as to examine some the relevant sociopsychological topics.

The study aims to adapt and test the methodological approaches of variationist sociolinguistics (cf. the works of Labov (1972, 2004) and Tagliamonte (2006, 2012) and comparable research in other language communities, as mentioned in section 3) on Slovene language material for the first time, and to define the typology of speech strategies in dialect contacts in the Slovene capital. It seeks to supersede the division and dichotomy of this research area in Slovene linguistics, where the idealized poles of standard language and dialect have been the main focus to date, and to establish a new foundation for researching the field, the need and incentive for which has been expressed previously in the expert literature (e.g., Kenda-Jež, 2004; Stabej, 2010). The topics raised in this context seem to be very relevant to Slovene due to its high level of geographical language diversity, especially in an age of increasingly pronounced geographical mobility.

Studying language variation, particularly in mobile speakers who move between regions and bring different speech varieties and dialectal features into contact, offers valuable insights into the actual language use and the sociolinguistic factors that influence individuals' language behaviour, such as age, gender, education, social networks, personal history, characteristics, and linguistic experiences. Besides, examining how varieties interact and evolve contributes to a better understanding of the processes of language change. Mobility and exposure to different speech varieties also influence individuals' language attitudes and self-perception. By exploring how mobile speakers adapt to different interlocutors and negotiate their linguistic identities in diverse contexts, we gain insights into the complex relationship between language and identity. This understanding can help speakers navigate linguistic landscapes, connect effectively with others, and achieve success in personal and professional domains.

Additionally, such research advances various fields related to language and society, sheds light on issues of social integration and inclusion, power relations between the centre and periphery, and the dynamics between standard and non-standard varieties. Understanding the linguistic experiences and challenges faced by mobile individuals promotes inclusive language policies, educational programmes, and social support systems that foster linguistic diversity, diminish linguistic stereotypes and prejudices, and ensure equal opportunities for all (Labov, 1972; Edwards, 2009; Auer and Schmidt, 2010; Tagliamonte, 2012; Milroy, 2002).

The rest of this paper is structured as follows: Section 2 provides an overview of the sociolinguistic situation regarding the Slovene language and the Idrija region. Section 3 offers a brief review of relevant previous work, with a particular focus on the European and Slovene contexts. The research presentation focuses on two younger informants, both high school students.¹ The methodological framework is presented in Section 4, while Section 5 presents the results and discussion. Subsections 4.1 and 5.1 deal with the quantitative variationist study data, while subsections 4.2 and 5.2 explore the qualitative sociolinguistic interviews. Section 6 concludes the paper with a discussion of the findings and an assessment of their potential implications for future research directions.

2 The sociolinguistic situation in Slovenia and the Idrija region

The sociolinguistic situation regarding Slovene, a South Slavic language now spoken by approximately 2.4 million speakers,² is characterized by several distinctive features. Firstly, there is a significant geographical diversity of the language, encompassing seven dialect groups and over 40 local dialects and subdialects (cf. Škofic et al., 2011, p. 11). Additionally, Slovene has a long history of serving as an L-language in diglossic relations with German, Italian, Hungarian and Serbo-Croatian. Furthermore, the standardization of the written and

1 Two adult female informants are presented in detail in Bitenc and Kenda-Jež (2015).

2 For a general overview of the Slovene sociolinguistic situation, see Greenberg (1997) and Bitenc (2013).

spoken language exhibits an asymmetric nature. While the unification of the written language and the abandonment of regional literary traditions originated from the second half of the 19th century, the development of a unified spoken standard was not as well-established and regional varieties continue to be observed (cf. Pogorelec, 1984; Toporišič, 2000; Stabej, 2010). The Slovene language territory exhibits complex patterns of social and regional stratification of language varieties, with various dialects featuring different levels of prestige and distance from the standard Slovene, as well variations in mutual intelligibility (cf. Bitenc, 2013, pp. 58–61; Bitenc and Kenda-Jež, 2015, pp. 31–32). The classic Slovene theory of language varieties (Toporišič, 2000, pp. 13–27) presents dialect variability in terms of the division between standard (“literary”) and non-standard (“non-literary”) varieties.³ According to critical reflections on the theory, this schematic and rigid interpretation of different language forms and communicative situations can neither explain nor take into account the real sociolinguistic situation in the language community (Stabej, 2010, p. 198). Slovene linguistics, with its prevailing structuralist orientation until the end of the 20th century, has predominantly focused on traditional dialects as well as standard (“literary”) language, since these were supposed to be the only varieties with a systemic character (Smole, 2004, pp. 321–324). The majority of language observation and description has been based on introspection, and in the theory of standard language, the Czech model of Common Czech has been adapted to the Slovene situation (Pogorelec, 1998).

Slovene sociolinguistics has been characterized by special historical circumstances in the Slovene language territory and has gone chiefly in the direction of (standard) language policy and planning, dealing with bi- or multilingualism, language as an indicator of interethnic relations, language ideologies, and so on (Nečak Lük, 1998; Stabej, 2010). The need and incentive for a systematic empirical investigation of spoken Slovene and its variability have been expressed

3 Standard (literary) varieties include formal language and “common” or standard colloquial language; non-standard (non-literary) varieties include rural dialects, regional interdialects, slang, jargon and cant (Toporišič, 2000, pp. 13–27). In the theory, the Czech model of the Prague School with “spisovný jazyk” and “obecná čeština” has been adapted to the Slovene situation (cf. Pogorelec, 1998, p. 59; Wilson, 2010, pp. 34–39).

previously in the expert literature (Pogorelec, 1998; Kenda-Jež, 2004; Stabej, 2010), but it is only recently that concrete suggestions and relevant research have appeared to address this gap.

The Cerkno dialect, spoken in the Idrija region, which is about 60 kilometres to the west of Ljubljana, belongs to the Rovte dialect group, whose structural development differs from the general Slovene tendencies. The dialect is noticeably different from varieties selected as the basis for standard Slovene, and in comparison with the urban dialect of Ljubljana there are noticeable differences in phonological system, morphology and lexicon, the most salient of which is the retention of vocalic quantitative oppositions (cf. Bitenc and Kenda-Jež, 2015, pp. 32–33). The dialect has high status and prestige in the Idrija region and is generally also used in semiformal contexts. This can be attributed to long communication isolation and the period of Slovene-Italian diglossia during the Italian occupation of the region in the interwar period (Kenda-Jež, 2002, p. 67), when the dialect was the only variety of Slovene that was used in the region and therefore developed greater functional flexibility. The city of Idrija is known for having the second largest mercury mine in the world, whose beginnings date back to the late 15th century, and has a remarkable history not only in the economic-technical field, but also in the areas of culture and education⁴ (Kleindienst, 1995, pp. 9–12). It has an important role as a regional centre, and lies at the intersection of the influential regions of two urban centres, Ljubljana and Nova Gorica (Fridl et al., 2001, p. 97), the latter on the border with Italy, which also testifies to its specific sociolinguistic situation. According to demographic data from censuses from the second half of the 20th century, the level of education and the proportion of the population in employment has increased significantly. The size of the working population with a residence in Idrija and a workplace in Ljubljana doubled from 2000 to 2011 (from 237 to 492 inhabitants). In 2002, some 1,114 of 2,682 school and university students residing in Idrija attended educational institutions in municipalities located in other statistical regions. Among the daily

4 The first Slovene theatre was founded in Idrija in 1780 and the first Slovene high school with Slovene as the prevailing language of instruction in 1901.

commuters, the majority commuted to the municipality of Ljubljana⁵ (cf. also Bitenc, 2016b, pp. 164–169).

Recent surveys of the Cerčno dialect for the Slovene and Slavic atlases have enabled a detailed phonological description of the Cerčno dialect (e. g. Rigler, 1981 and Kenda-Jež, 2002). Also relevant for the present research is a discussion of the potential classification of the local dialect of Idrija as a subdialect of the Cerčno dialect (Skalin, 2002).

3 Relevant studies in the Slovene and European context

In Slovene linguistics, there has been a strong focus on the study of traditional dialects, reflecting a dichotomous view on language variability (cf. section 2). Extensive efforts at documenting traditional dialects have resulted in a substantial corpus of dialectal material, compiled through highly structured interviews conducted with non-mobile, older rural informants, and partly presented in dialectological atlases (Škofic, 2011; Škofic, Gostenčnik, Kumin Horvat, Jakop, Kenda-Jež et al., 2016). Dialectological studies of city dialects are rare, however, which can be attributed to the strict theoretical division between rural and city dialects (literally “city speech”) in Slovene linguistics (Kenda-Jež, 2004, p. 270).

The central variant of the capital Ljubljana and its surroundings has been the object of phonological studies conducted on unrepresentative samples of speakers and connected with standardization and spoken standard planning since the second half of the 20th century (cf. Rigler, 1970–1978; Srebot Rejec, 1988; Jurgec, 2011). A million-word corpus of spoken Slovene known as GOS⁶ (Verdonik and Zwitter Vitez, 2011) has been compiled, but in order for its material to be used as a basis for variationist investigation, more data about speakers’ base dialects and more detailed transcription, especially for potential phonological studies, would be necessary.

In two relevant studies on the variation of spoken Slovene, Guzej (1989) investigated the regional variation of migrants, a couple

5 <http://www.stat.si/StatWeb/>

6 www.korpus-gos.net

migrating from Slovenske gorice and Kozjansko to Gorenjska, while Škofic-Guzej (1994) analysed the speech of high school students from Mengeš and Beltinci in Ljubljana. Other studies of the spoken language have focused on different specific research questions, such as discourse markers and self-corrections (Verdonik, 2007), and text types in spontaneous speech (Smolej, 2012).

There have been a number of questionnaire studies on different samples about the use of dialects and attitudes to them on the basis of informants' self-assessment, e.g., Smole (2004), Lundberg (2014) (partly with the methodological approaches of perceptual dialectology), and Zemljak Jontes and Pulko (2019). In a questionnaire survey, attitudes towards different sociolects were investigated (Skubic, 2005) and two experiments with the socio-psychological methodology of the matched-guise technique focused on language attitudes regarding spoken standard Slovene and a distinctive Cerklje local dialect (Bitenc, 2014a and 2014b).

In the European context, linguists from different language communities have explored different aspects of dialect change, in particular dialect convergence or divergence (Auer, Hinskens and Kerswill, 2005). The studies demonstrate the important structural impact of the addition of a standard variety to the language repertoire of large parts of the population as a consequence of modernization and mobility in European societies, the centralization of the state, the spread of mass media, and changes in the education system. However, the outcomes of dialect change in various sociolinguistic situations show significant diversity. They range from situations where traditional dialects are nearly extinct, to those where the influence of the standard variety on the dialect is strong but still permits vital regiolects and base dialects, and to situations where the standard variety plays a relatively minor role in the ongoing dialectal changes (Auer, 1998). European sociolinguists have also dealt with the various social reasons for variety shifts and changes in the course of social and communicative modernization, with a special emphasis on migration (Mattheier, 2000). Within a project that transcends European frames, researchers have focused on the connection between language and space in the investigation of linguistic variation (Auer and Schmidt, 2010).

Especially relevant are sociolinguistic studies in environments that are close to the Slovene sociolinguistic situation in terms of dialect diversity or the structure of language varieties (Germany, the Czech Republic, Norway, Denmark). Notable among these are the investigation of the linguistic behaviour of inhabitants of Erftstadt-Erp, a small town near Cologne (Besch, 1981 and 1983), the analysis of changes in German regiolects and (regional) dialects (Schmidt and Herrgen, 2011), the examination of linguistic accommodation among Moravian university students in Prague (Wilson, 2010), the qualitative exploration of linguistic variation among Norwegians who have migrated to Oslo from rural areas (Mæhlum, 1986), and an investigation of language change in real time among mobile and non-mobile informants from three Danish towns (Monka, 2013).⁷ The insights of these studies have significantly influenced the conceptualization of the research in different ways. They employed a combination of quantitative and qualitative methods to analyse and explain the informants' linguistic behaviour, using variationist analysis and a range of various phonological and morphological variables. Most of them focused on mobile informants, where the potential to observe intermediate linguistic phenomena is bigger, and used an interview-based research design, involving different interlocutors with different language varieties. Schmidt and Herrgen (2011) paid special attention to the dialect-standard continuum and potentially more stable intermediate varieties, while Besch (1981 and 1983) characterized different types of speakers' profiles and speech strategies in interactions with various interlocutors.

4 Methodology

The research is a pilot investigation, focusing on in-depth case studies of five informants: three females (Ula, Ita, and Eva) and two males (Jan and Tim).⁸ Among the participants, three were adults (Ita, Eva,

7 A detailed description and critical evaluation of the methodological approaches for these and other studies can be found in Bitenc (2016b, pp. 123–155 for the Slovene language community and pp. 83–96 for other language communities).

8 To maintain the informants' anonymity, their names are changed in the context of the research and the article.

and Tim), while two were high school students (Ula and Jan). This article specifically focuses on two younger informants to provide a comprehensive illustration of the research process and to highlight the intricate relationship between their life stories and speech profiles.

4.1 Variationist study

4.1.1 *Speech recordings and domains*

With the aim of observing authentic language use and diminishing observer's paradox, the technique of informants' full-day audio self-recording was used in the present study to provide the material for the variationist analysis. For the duration of one entire day, the informants wore a voice recorder (Olympus Digital Voice Recorder WS-210S) in a small bag on their chest. The research corpus is based on a text selection from all of the collected material according to different criteria: especially the interlocutors and the topic of the conversation, but also the time, place and formality of the situation. The individual informants' particularities, as well as the potential for comparing functionally similar speech situations, were taken into account when defining the domains.

The life circumstances and everyday dynamics of the two high school students included in the study were so different that it was impossible to define the same domains for both of them. Ula and Jan both have the domains of local friends and teachers (in public). In addition, Ula has the domains of a roommate and other peers from elsewhere, whereas for Jan, on the other hand, there are the domains of peers at both regular school and at music school. Jan also had some conversations with two teachers during breaks (the teachers' parts of these were kept private) and there was a recording of communal prayer (*Our Father* and *Glory be to the Father*) at church.⁹

9 For the adults, the majority of the speech recordings belong to four main domains: local friends, children, and work colleagues, the latter being divided into the domains of casual topics (holidays, children, housing issues, etc.) and expert topics (professional work).

4.1.2 *Transcriptions and phonological variables*

The recordings were transcribed orthographically, using the 25 letters of the Slovene alphabet and phonetic symbols for specific dialect features (variables).¹⁰

The central part of the investigation is an analysis of five phonological variables (three vowel variables and two consonant variables), chosen on the basis of previous observation (the author is a dialect speaker and used to be a commuter herself) and dialect research in the area (Kenda-Jež, 1999, 2002):

- (1) Variable (ɔ): unstressed central /ɔ/ ↔ local dialect /a/ (the result of “akanje”)¹¹

s[a] sɛ u 'bistu 'ti: [a]tp[ɔ]'wi:dal 'di:lɛzu dɛ jɛ kmɛ'tija [a]s'ta:la ɛ'nɔtna
‘in fact they gave up their share so that the farm remained unified’

- (2) Variable (o): central /o:/ ↔ local dialect /u:/, /uo/

a j 't[o:] zɛ prɛ'wɛf? : 't[u:] 'zɛj dɛ'lujɛ
‘is this too much already?’ : ‘this is now working’

- (3) Variable (e:): central /e:/ ↔ local dialect /i:/, /ie/

u'r[i:]dnu ɔ'ylɛda : 'ɥr[e:]dnɔ ɔ'glɛda
‘worth seeing’ : ‘worth seeing’

- (4) Variable (g): central /g/, [k] ↔ local dialect /ɣ/, [x]

'lɔpɔɔɔ pa na 'muorf 'jet pa'[ɣ]liɛdat : 'ti si 'rɛ:ku dɛ bi 'fɔɔ pɔ'[g]le:dat 'lɔpɔwe
‘you can’t go and see villains’ : ‘you said you would go and see villains’

- (5) Variable (v): central /v/ (before vocals) ↔ local dialect /w/

'tɛ zu'nɔnjɛ s[w]ɛtɔ'[v]a:ɥtse sɔ 'rɛkɫ dɛ nɛ 'bɔjɛ 'li:tas 'nɛf upɔ'ra:blɛl
‘they said they wouldn’t use these outside advisers at all this year’

In addition, attention is paid to certain morphological variables and lexicon. Due to the limited material (the relatively small number of tokens), the results are not subject to statistical analysis. Testing

10 A narrow phonetic transcription was used for the first two, adult female informants. The simplification of transcription conventions enabled a larger database while still preserving an adequate level of transcription detail (cf. Tagliamonte, 2006, p. 54).

11 The development of Protoslovene unstressed and short stressed *o > a.

the reliability of the obtained results with additional data analysis did not reveal any significant changes in distribution patterns.¹²

4.2 Sociolinguistic interviews

Qualitative data were gathered in sociolinguistic interviews (approximately three hours of recordings for Ula and Jan), conducted by the researcher, a dialect speaker herself, with the individual informants. The interviews took place soon after the self-recording. The main topics and open-ended questions were prepared in advance, but the interviewer conducted the interview relatively freely, allowing the interviewees to elaborate and reflect upon issues especially relevant to them, and to influence the course of the conversation. The atmosphere during the interviews was generally relaxed and pleasant, although it did vary to some extent depending on the location (a café with Ula and a school hall with Jan), the timing, and the relationship between the informant and the researcher as an ex-member of the local community (Ula is an acquaintance of the interviewer, whereas she met Jan due to the research). The rich discourse with many personal stories both from the interviewees and the interviewer allowed for in-depth analyses of conversational contributions with reflections and emotions that could have remained unnoticed in a more constrained context.

All of the relevant parts of the interviews were transcribed. In relation to the transcriptions for the variationist study, conventions were further simplified, but the transcriptions still enable observation of variables in the interviews. The selections were sorted according to different topics, and thematic conversational analysis was carried out.¹³

After collecting basic data (informant's family background, personal history, education, social network, personal interests and characteristics), special attention was paid to the informants' perception of their own speech behaviour, their language attitudes, their experience with language use and the connection between language and identity.

12 It is planned to use the results as a basis for a more extensive research project investigating spoken language in a larger geographical area, with adequate exploration techniques (distributional analysis and cross-tabulation).

13 All of the transcriptions as well as the excerpts according to different topics are available on the CD that forms a supplement to Bitenc (2016a).

5 Results and discussion

5.1 Speech patterns

A comparison of informants shows that they differ even in the varieties they use in the local environment, which can be explained by the fact that their parents originate from – and they themselves have resided in – different towns in the Idrija region. Jan grew up in Idrija, acquired the local Idrija dialect in his childhood and retained it after moving to another town in the vicinity (Godovič), which can be attributed to the special status of the Idrija local dialect as a variety of the regional educational, cultural and administrative centre. Ula grew up in Kanomlja with the Kanomlja local dialect. Probably due to her father's Idrija origins and contacts with other people from the area, Ula soon acquired the characteristics of the Idrija local dialect, which were also prevalent in her speech at the time of the recording. For the variationist study, the total length of Ula's recordings is 17 hr 59 min 42 sec, while the total length of Jan's recordings is 11 hr 56 min 22 sec. For Ula, 16 excerpts were selected for analysis with a length of 30 min 33 sec (individual excerpts last from 6 sec to 4 min 36 sec, whereas for Jan, 13 excerpts with a total length of 48 min 52 sec were selected (individual excerpts last from 13 sec to 12 min 48 sec). Ula mostly discussed school issues and plans for seaside holidays with her peers. She also answered her teacher's questions in an oral examination at a history lesson. As a promising musician, Jan mostly conversed about music and concerts with his peers as well as his teachers. At an English lesson he took an active part in a discussion about current issues regarding this school subject.

The proportions of dialect variants for different domains (Table 1) show that Ula uses two distinctively separate codes that she mostly refers to as *idrijščina*¹⁴ (Idrija local dialect) and *ljubljanščina* (Ljubljana urban dialect); her graph (Figure 1, top left) is also distinctively divided in two. Although she claims that she speaks the Idrija local dialect with her roommates, the proportion of dialect variants is lower in this variety than in the variety she uses with local friends

14 The expressions taken and translated from informants' narratives are written in italics.

(87% in comparison with 96%). In her *ljubljanščina* (colloquial speech of Ljubljana) or *pravilna slovenščina* (correct Slovene), a certain share of dialect variants remains (17% with peers from elsewhere, including her boyfriend, and 9% with teachers).

Table 1: Different domains and the distribution of variables for Ula and Jan

Variable	Local friends		Roommate (Ula) / Peers at school (Jan)		Peers from elsewhere (Ula) / Peers at music school (Jan)		Teachers private		Teachers public		Prayer	
	Ula	Jan	Ula	Jan	Ula	Jan	/	Jan	Ula	Jan	/	Jan
(o)	89/15	82/8	23/8	19/9	3/65	14/9	/	5/12	0/16	7/27	/	0/6
	86	91	74	68	4	61	/	29	0	21	/	0
(e)	119/1	129/5	23/1	43/11	21/56	20/12	/	7/9	1/6	15/13	/	0/5
	99	96	86	80	27	63	/	44	14	54	/	0
(o)	61/2	81/1	15/0	29/2	14/35	17/0	/	11/2	0/3	11/9	/	0/2
	97	99	100	94	29	100	/	85	0	55	/	0
(g)	77/1	63/2	9/1	28/1	3/70	13/4	/	9/9	0/7	13/9	/	2/1
	99	97	90	97	4	76	/	50	0	59	/	67
(v)	123/0	124/0	26/1	47/0	15/74	37/3	/	12/1	4/10	16/5	/	1/10
	100	100	96	100	17	93	/	92	29	76	/	9
Average	96	96	87	88	17	79	/	60	9	52	/	15

Note. The first row for each variable: number of dialect/standard variants, the second row: percentages of dialect variants.

Examining the number and average value of dialect variants in different domains for Jan (Table 1), a gradual decrease can be observed from 96% with local friends, through 88% and 79% with peers at school and music school, to 60% and 52% with teachers private and public, and finally 15% in the case of prayer. Jan has a diverse language repertoire with more intermediate speech varieties with different proportions of dialect variants for individual variables, but his speech remains substantially dialectally coloured even in the most formal conversations with his teachers. He is aware of the variability of his speech and accurately states that he is *sometimes really half way* between the dialect and the standard, but would *hardly ever totally switch*. In the comparative graph (Figure 1, bottom left),

Jan's domains of teachers public and private are combined, whereas the domain of prayer is omitted.

Even a cursory glance at the shares of dialect variants reveals a surprisingly accurate match in Ula's and Jan's conversations with local friends (96% for both) and in the column with those peers from elsewhere to whom the informants feel closest (Ula's roommate 87% and Jan's school friends 88%). However, in the others columns – with the share for Ula's peers from elsewhere and Jan's peers at music school, as well as for teachers – the differences between Ula and Jan are considerable and clearly reveal the fact that they use very different language strategies with the majority of speakers of other varieties. Jan's speech remains rather dialectally coloured and his local origin and identity are recognizable in all domains. Even in his communication with peers at music school and with teachers, the proportion of dialect variants is 79% and 60% or 52 %, respectively. Ula, on the other hand, changes her speech to such an extent that only rare dialect traces remain and her speech is, as she herself reports, distinguished as dialect only by the most linguistically sensitive listeners. In her communication with peers from elsewhere and with teachers the share of dialect variants is 17% and 9%, respectively.

In Ljubljana, both Ula and Jan use two different varieties with peers from elsewhere, but the criteria according to which they categorize them are different. For Ula, the criterion is the time when she met certain people: with a roommate from the Gorenjska region, she uses markedly dialectally coloured speech, which she attributes to the fact that they became acquainted just after she came to Ljubljana, when she was not yet used to code-switching.¹⁵ Other peers

15 The concepts of code-switching and code-mixing (and borrowing) are dealt with from different points of view and defined in different ways by different authors; there is no commonality of practice and it seems difficult to distinguish clearly between the terms. They mostly refer to language choice between two or more languages, but also embrace variation within the same language (different language varieties, dialects, styles or registers). Some scholars have reserved code-switching for cases where the two codes maintain their monolingual characteristics and code-mixing for those with some convergence between the two (Gardner-Chloros, 2009, pp. 10–13; Fasold, 1987, pp. 180–183). In the context of this research, we use the terms in congruence with the German sociolinguistic project Erfstadt/Erp (e.g., Besch, 1983; Niebaum and Macha, 2014, p. 198). Code-switching refers to the linguistic and communicative behaviour where varieties (dialect and standard) maintain their relative coherence, whereas in the case of code-mixing, different shares of dialect and standard variants of investigated variables are combined.

from elsewhere include all of the peers not from Idrija whom Ula got to know later on during her school years in Ljubljana, when her speech profile was already different and her *ljubljanščina* was already the natural choice for communication with people from other regions. For Jan, the criterion is the closeness and time spent with his colleagues: his speech is more dialectally coloured with peers at regular school, with whom he spends more time and has a closer relationship, than with peers from music school, whom he knows less personally.

The results of the variationist part of the study show that different types of speakers exist and that we can often speak about continuum-like transitions between the dialect and the standard. The informants use different linguistic strategies: the distribution of dialect variants in different domains shows that Ula uses two distinctive codes with local friends and speakers in Ljubljana, whereas Jan retains a relatively high percentage of dialect variants in most domains and approaches the standard more significantly only when talking to teachers. Using terminology from German sociolinguistics (Erfstadt/Erp Projekt; e.g., Besch, 1983; Niebaum and Macha, 2014, p. 198), we can label the informants as a code-switcher (Ula) and a code-mixer (Jan), although Jan approaches the profile of a dialect speaker with a high percentage of dialect variants in their communication with different collocutors in Ljubljana (Figure 1).¹⁶

16 The adult male informant, Tim, stands as a clear example of a dialect speaker, as evidenced by the high percentage of dialect variants observed in all domains. On the other hand, the two adult female informants, Ita and Eva, can be classified as code mixers. Eva employs the dialect when interacting with local friends, while displaying a gradual yet significant shift towards the standard in the remaining three domains: conversing with children and engaging with work colleagues on both casual and expert topics. Ita's profile aligns more closely with Jan's, with a more pronounced shift towards the standard observed primarily in Ljubljana during interactions with work colleagues on expert subjects.

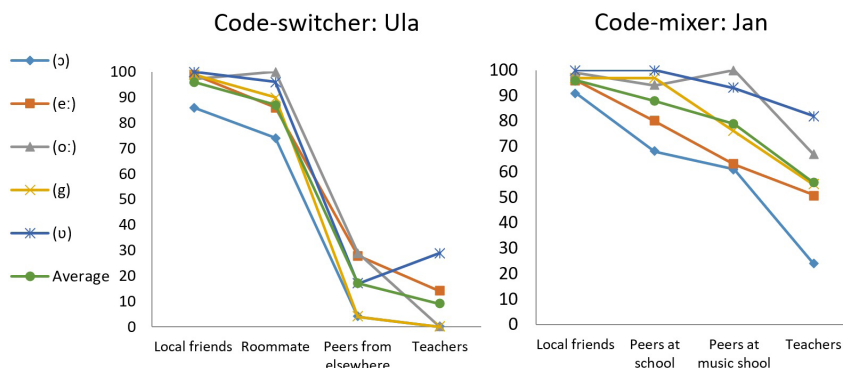


Figure 1: Speakers’ profiles and distribution of variables according to different domains for the younger informants.

Code-mixers’ intermediate forms, in particular, enable the informants to express a kind of double or ambivalent geographical, social and linguistic identity, i.e., social identification with more codes and social groups simultaneously, which can be paralleled with the consideration of all language activity as expressions of “acts of identity” (Le Page and Tabouret-Keller, 1985). Code-mixers’ hybrid repertoire with gradual transitions or relative accommodation can also be interpreted in terms of “strategies of neutrality”. In uncertain situations, speakers tend to use a linguistic variety that is neutral to attributes that they perceive as salient in the situation, attempting to use favourable aspects of potentially salient attributes of each variety to their advantage (Scotton, 1976, p. 940; cf. Mæhlum, 2000, pp. 104–105). The informants’ language situation can be described as uncertain to some extent, and their preferred strategies can be interpreted as neutral.

Due to the small number of informants, it is not possible to determine the relationship between social factors and variability. However, it can be observed that the greater proportion of dialect variants with the male informants is in line with the so-called “sociolinguistic gender pattern” (cf. Fasold, 1996, p. 92), which refers to males speakers’ tendency to use forms that are considered wrong or stigmatized more frequently than female speakers (cf. also Tagliamonte, 2012, pp. 32–33). The difference between younger informants, Ula and Jan, is considerable in Ljubljana – with Ula’s peers from elsewhere and Jan’s

peers at music school, as well as for teachers – Jan’s speech remains rather dialectally coloured, whereas Ula keeps only rare dialect traces in her speech (see Section 6.1 for detailed proportion of dialect variants in different domains).

In addition, there is an obvious use of slang in the speech of both high school students, which is typical for the period of adolescence and serves young people as a marker of ingroup membership and a deviation from the norms of their elders (Chambers, 2003, pp. 186–189). Both Ula and Jan use common slang expressions, some of which derive from English. Examples include *'ful* (slang for totally) and *'tip* (a guy). In addition, Ula very often says *'sori* (sorry), *də 'best* (the best) and *'fɔ:ra* (a joke), as well as using jargon from the school environment, such as *k'la:s* (slang for ‘class’), *pr'foksa* (teacher), *za'liwat* (cut classes), *'mata* (maths), *z'yodla* (history) and *'bajla* (biology). Jan uses *'kul* (cool), *'zakon* (great) and *'badžet* (budget), as well as using expert and jargon expressions in conversations about music, such as *'ja:ki ko'ma:t* (nice song) and *šp'rudlat* (play badly).

Considering the concept of the linguistic market and the connection between language use and education or profession (Sankoff and Laberge, 1978; cf. Ash, 2008, pp. 413–414), the relevance of Ula’s role of “technician of language” in their professional engagement can be pointed out. Her assistant work at expert meetings presented a stimulus for the use of the standard variety. She claims that *at her expert work or at conferences with that high society the dialect wouldn't be acceptable, they would look at you strangely – it's not appropriate, it doesn't fit*, which demonstrates that competence in the standard is associated with social success (cf. Mæhlum, 2010).

5.1.1 *The envelope of variation*

Different variables show different tendencies and the analysis demonstrates that vowel variables change to a greater extent than consonant variables. (ɔ) is changed first and has the lowest percentage of dialect variants for all informants and in most domains,¹⁷ so it may be

17 A common share of dialect variants for all informants was 47% for the variable (ɔ), 68% for (e), 73% for (g), 77% for (o:) and 80% for (v).

possible to speak about a dialect change in progress for this variable. However, it seems to be below the informants' level of consciousness, as it is never mentioned in the interviews. A voiced velar fricative /ɣ/ of variable (g), on the other hand, proves to be a salient feature, a stereotype (cf. Labov, 1972, pp. 314–317). It was most difficult to distinguish between the standard and the dialect variant for the variable (v), especially with noise on the recordings and a fast speech tempo. For the variable (ɔ), there were also cases with intermediate sounds that were difficult to define as either /a/, /o/ or perhaps /ə/. Intermediate or interdialectal phenomena that are part of neither the dialect nor the standard can be interpreted as a result of numerous strategies of code-mixing and code-switching.¹⁸ It seems important to point out that, according to Auer and Hinskens (2005, p. 356):

the driving force behind change in the individual, and also in the community, is not imitation of the language of one's interlocutor but, rather, an attempt to assimilate one's language to the possibly stereotyped characteristics of a group one wants to be part of, or resemble.

In order to properly define the envelope of variation, certain items that behave differently from other members of their class and show no variation were excluded from the analysis (cf. Labov, 2004, p. 7; Tagliamonte, 2012, p. 13), which holds true mostly for vowel variables. An additional questionnaire was completed by the informants and two language experts and dialect speakers in order to define these items.¹⁹ The informants agreed that certain loanwords that pertain to

18 Ula and Jan (also Tim) use a few typical reflexes of the Idrija local dialect, among them the diphthong *uo* as the reflex of *o*, which acquired stress following the regressive accent shift from word-final short syllables onto pretonic short vowels, and short acute *o* in word-final syllables. This sound is, however, interchanged with long and narrow *o*-s and interdialectal short *o*-s, most often in frequently used lexemes, such as *'puo(l)'/po:(l)'/po(l)* (then, later) or *'duol/dɔ:l/dol* (down). Participant observation and some experts' comments show that the decrease in the use of the diphthongs in Idrija local speech occurs not only in the speech of geographically mobile speakers, but also with non-mobile inhabitants, especially among the younger generations, so we might speak of dialect change in progress.

19 For each lexeme, where there was a dilemma as to whether to regard an individual case as a variant of a variable, the two experts as well as the informant whose utterance was in question stated whether or not they would use the word with the dialect variant, as well. For 40 of the 112 questionable examples, all three agreed that they would not, whereas for the other 72 words their opinions differed. In the latter cases, the answer that occurred twice was taken into account.

expert language do not have dialect variants (e.g., *patɔlɔ'gija* 'pathology' and *ɔd'vɛtniske stɔ'ritvɛ* 'lawyer services') as well as certain proper names (e.g., *'se:ʃa*, the name of the coastal town Seča, or *po'lo:na*, the female personal name Polona). Interesting examples also include lexical items that can be used with the dialect variant when used as common nouns (e.g., *'di:lu* (dialect) vs. *'de:lo* (standard) for 'work'), but would only be used with the standard variant when used as proper nouns (*'de:lo* as the name of the Slovene newspaper *Delo*). Similarly, words such as *'ne:ɣa* 'care' or *'do:m* 'home' have common dialect variants *'ni:ɣa* and *'du:m*, whereas they are never used with the dialect variant when used in a set phrase or in an idiom: *'ne:ɣa o'trɔ:ka* 'child care' as a school subject or *di'ja:ʃki 'do:m* 'boarding school'.

5.2 Participant reflection

5.2.1 *Language attitudes and experiences with language use*

All of the informants assign great significance to the Cerčno dialect for their personal and local identity, but demonstrate different language attitudes towards their own dialects. The two younger informants express mainly positive feelings about it (Jan says it is *great* and *something special*, Ula finds it *really interesting* and *nice*, as it is *so unusual*). In contrast, the adult females, in particular, comment relatively negatively about it (they describe it as *hilly* and *clumsy, not beautiful, hard* and *coarse*). This may be interpreted as an indication of a generational shift towards greater acceptance of and increasing tolerance towards nonstandard regional varieties, as well as the growing orientation towards local and regional identity, also referred to as the "dialect renaissance" (Auer et al., 2005, p. 36).

When people from elsewhere comment on the informants' dialect or dialectally coloured speech, they typically say it is *interesting* (Ula) and *beautiful* (Ula); it *varies the atmosphere* (Jan). They often *smile* (Ula) or *make fun of it in a nice way*; Jan points out that he has *never found out whether they like it or whether it bothers them*. Jan says that he *somehow feels if the dialect doesn't suit somebody* and thinks it *might be intrusive for some people if...[he] spoke entirely in the dialect all*

the time. Ula and Eva have experienced that people in Ljubljana urged dialect speakers from Idrija region *to learn how to speak.*

The informants report a prevailing negative attitude towards the inhabitants of Ljubljana and their speech, which is connected with non-linguistic characteristics in the sense of commodity and superiority. Ula says that the inhabitants of Ljubljana *treat their dialect as holy*, whereas Jan claims that some *simper* and *madly like to use English expressions*, whereas others *speak very beautifully* and their speech *isn't disturbing.*

The informants report that collocutors often do not understand what they say, and that their reaction depends on the individual. For Ula, this was a reason for accommodation: *It really got on my nerves when I was explaining something and then somebody asked: "What did you actually say?"* Jan sometimes asks his interlocutors: *Do you understand? Is it ok?*

Members of the local community often perceive an individual's language accommodation towards a standard variety as a withdrawal from the local community. Ula comments: *I hear it all the time: that I have totally abandoned my dialect and how could I; it is more as a joke, that I have betrayed my homeland!* Jan, who accommodates less, is critical of speakers who accommodate more when talking to speakers of other varieties. As he explains: *I find it a bit stupid that some people want to hide it when they come to Ljubljana. It's really harmful to some people that they start to converse here a bit and then start speaking that way at home, too.* It seems the problem is that in the view of their compatriots speakers with greater accommodation show *some non-affiliation* (Jan), that they deny their origin and local identity and accept another identity instead. The data show that the understanding of the complexity of the linguistic choices depends on personal experience. Jan's family happens *to grimace* when he speaks to his teacher from Ljubljana on the phone; moreover, his father sometimes says: *Don't you just start talking like this sometime!* Jan then defends himself: *Look, don't worry, I've never started speaking like this at home.* In the interview, he comments that *they aren't in this kind of situation every day and it is difficult for them to imagine it.*

5.2.2 Language awareness and metalanguage

The informants mainly claim that their accommodation is automatic. Ula says that she has *never questioned her speech*, and that it is her *automatic reflex* to use *correct Slovene* with people she does not know. Jan claims that in his opinion accommodation works *plainly spontaneously*. He accurately explains these processes: *It is a question of every person making out some kind of system according to which she/he then acts; it is not self-conscious. Then a person sees, now I am in this situation, I will speak like this. It comes with experience. I've been doing this for four years now and a system evolves, a mould according to which a person acts.*

The informants show different levels of language awareness and differ significantly in the accuracy of the description of their own speech behaviour. The terms they use for different language varieties speak for themselves. Jan, for example, uses the term 'literary (colloquial) speech' to refer to one of his varieties, whereas Ula speaks about 'correct Slovene' or *ljubljanščina* (Ljubljana urban dialect) or about speaking *ljubljanško* 'in the Ljubljana way'. Other informants use this expression rarely and only when referring to other people's speech or when expressing attitudes towards this variety. Ula demonstrates some uncertainty regarding these labels: *Also at the faculty I mainly use the correct language or maybe some Ljubljana slang, or I don't know what to call it.* When the adult informants refer to the standard, Ita only uses the expression 'colloquial Slovene', Eva the terms 'beautiful/correct/colloquial Slovene', whereas Tim speaks about 'Slovene' (as opposed to the dialect). The terms show that the informants perceive and evaluate the varieties differently, and highlight that the concepts of a 'beautiful' and 'correct' language still persist within the language community. This reflects the influence of the prescriptive tradition and the standard language ideology in Slovene linguistics, and calls for a better understanding of language variation, appreciating linguistic diversity instead of imposing rigid norms and recognizing the inherent beauty in all forms of language expression.

6 Conclusion

The article presents interlanguage variability in a case study of five commuters from the Idrija region with a distinctive Cerčno dialect to the capital, Ljubljana. The technique of full-day self-recording provided a great deal of authentic material and enabled the selection of various authentic speech situations with the informants' actual acquaintances. Although it is still necessary to identify and account for contextual effects, the observer's paradox is diminished to a minimum: according to the informants, they quickly forgot about the recorder and paid minimal attention to it. However, there are often a lot of useless segments on the recordings, and recording conversations more selectively may therefore be more convenient in future studies.

The variationist analysis of five phonological variables demonstrates that vowel variables change to a greater extent, and that among the variables (ɔ) has the lowest percentage of dialect variants, whereas a voiced velar fricative /ɣ/ of variable (g) proves to be a salient feature, a stereotype. According to different language strategies in contacts with speakers of other varieties, the five subjects represent different speakers' profiles: a code-switcher with two rather distinctive varieties in communication with local friends in Idrija versus speakers in Ljubljana, three code-mixers with different intermediate varieties in different domains, and a dialect speaker with a high percentage of dialect variants in all domains. Intermediate or interdialectal sounds and varieties can be interpreted as a result of numerous strategies of code-mixing and code-switching. They enable the informants to express a kind of double or ambivalent geographical, social and linguistic identity, i.e., social identification with more codes and social groups simultaneously.

The sociopsychological topics dealt with in the in-depth sociolinguistic interviews – such as language attitudes, informants' perception of their own speech behaviour, their experience with language use and the connection between language and identity – additionally illuminate the informants' language variability and the reasons for their different linguistic strategies, as well as testifying to the complexity of language choice and interpersonal communication.

The cases of two high school students, examined in more detail in the central part of the article, show that individual life experiences and everyday dynamics influence language use to a significant extent. With her two distinctly different codes, Ula was defined as a code-switcher, whereas Jan, with his diverse language repertoire with more intermediate varieties, was labelled a code-mixer. Both use common slang expressions and, in contrast to the female adult speakers, express positive attitudes towards their own dialect, which might be interpreted as an indication of a generational shift towards increasing tolerance of nonstandard regional varieties and a growing orientation towards local identity, also referred to as the 'dialect renaissance'. In the future a longitudinal study will be possible, which could reveal potential changes in Ula and Jan's language strategies on their path to adulthood. As for the next research step within the framework of a postdoctoral project, I have decided to focus on adults, since language behaviour and language attitudes (at least in the form of informants' self-evaluation and self-reflection as reported in the interviews) can vary quite significantly through different stages up to adulthood, when language choices are mostly stabilized. I will investigate the language repertoire of speakers from two other dialect regions, Lower Carniola (Dolenjska) and Styria (Štajerska) – specifically, from the town of Ribnica and the city of Maribor – who commute to work in Ljubljana on a daily basis. The findings will enable a comparison of language variation of speakers whose dialects have been involved in the standardization of Slovene in different ways and to a different extent, and of related sociolinguistic issues. Furthermore, they will illuminate the different prestige and status dialects, as well as related stereotypes and prejudices, which can be important for understanding complex interdialect relationships in the Slovene language territory, with its high level of geographical language diversity. The longitudinal plan is to exploit language use and relevant sociolinguistic questions on larger samples, on different dialect areas, and in both urban and rural speech settings, in order to gather relevant data about language variation and potentially more stable interv varieties or regional standards, as well as regional variants of the spoken standard.

The integration of the different research traditions as well as linguistic and sociopsychological methodological approaches that we have pursued seems to provide a solid basis for further investigations, especially in the Slovene context, where there are few comparable studies to refer to, particularly in the present time of new and very complex technological and social issues. The results can contribute to a more objective description of the sociolinguistic picture of the Slovene language community, and to understanding the nature of the language variability of Slovene and its social significance. A more open view of language variability in linguistics, and consequently in the education system, can help to diminish the influence of the reigning standard language ideology and aid speakers of Slovene to achieve more relaxed and efficient communication in different life situations. Since sociolinguistic conditions differ hugely between and within European countries, and since variationist studies in Slavic environments are scarce, the research can help to transcend the prevailing Anglo-American bias in sociolinguistics (Wodak idr. 2011: 1), which is of major importance to sociolinguistic theory itself.

Acknowledgments

The author's postdoctoral research project, titled *Sociolinguistic Variation of Spoken Slovene in the Case of Mobile Speakers: Resources, Methods and Analysis* (project number Z6-9371), is funded by the Slovenian Research Agency, and the author would like to express gratitude for their support. The author is also thankful to her PhD mentors, Dr. Marko Stabej and Dr. Karmen Kenda-Jež, for their valuable assistance in designing the research and providing helpful comments throughout its process. Additionally, the author extends appreciation to Dr. Albina Nečak-Lük and the anonymous reviewers for their insightful feedback on the text. The author would like to acknowledge Dr. Brit Maehlum, Dr. Sali A. Tagliamonte, Dr. Johanna Vaattovaara, and Dr. Peter Trudgill for their valuable advice on methodological issues. Lastly, heartfelt thanks are extended to the informants, without whom the research would not have been possible.

References

- Ash, S. (2008). Social Class. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 402–422). Malden, Oxford: Blackwell Publishing.
- Auer, P. (1998). Dialect Levelling and the Standard Varieties in Europe. *Folia Linguistica*, 32(1–2), 1–9.
- Auer, P., & Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. In P. Auer, F. Hinskens & P. Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 335–357). Cambridge, New York: Cambridge University Press.
- Auer, P., Hinskens, F., & Kerswill, P. (Eds.). (2005). *Dialect Change: Convergence and Divergence in European Languages*. Cambridge, New York: Cambridge University Press.
- Auer, P., & Schmidt, J. E. (Eds.). (2010). *Language and Space: An International Handbook of Linguistic Variation: Theories and Methods*. Berlin: De Gruyter Mouton.
- Besch, W. (Ed.). (1981). *Sprachverhalten in ländlichen Gemeinden: Ansätze zur Theorie und Methode: Forschungsbericht Erp-Projekt 1*. Berlin: E. Schmidt.
- Besch, W. (Ed.). (1983). *Sprachverhalten in ländlichen Gemeinden: Dialekt und Standardsprache in Sprecherurteil: Forschungsbericht Erp-Projekt 2*. Berlin: E. Schmidt.
- Bitenc, M. (2013). Slovene, Between Purism and Plurilingualism. In F. X. Vila (Ed.), *Survival and Development of Language Communities: Prospects and Challenges* (pp. 58–80). Bristol: Multilingual Matters.
- Bitenc, M. (2014a). Stališča gimnazijcev do slovenskih jezikovnih zvrsti: Razi-skava s tehniko prikritih dvojic. *Annales: Anali za istrske in mediteranske študije: Series historia et sociologia*, 24(2), 307–318.
- Bitenc, M. (2014b). Tehnika prikritih dvojic: Primerjava in kritično ovrednotenje dveh poskusov. *Annales: Anali za istrske in mediteranske študije: Series historia et sociologia*, 24(2), 319–330.
- Bitenc, M., & Kenda-Jež, K. (2015). Language variation in Slovene: A case study of two geographically mobile speakers. In E. Torgersen, S. Hårstad, B. Mæhlum & U. Røyneland (Eds.), *Language Variation – European Perspectives V* (pp. 31–42). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Bitenc, M. (2016a). *Variantnost govornje slovenščine pri geografsko mobilnih osebah: Doktorska disertacija*. Ljubljana: Filozofska fakulteta.

- Bitenc, M. (2016b). *Z jezikom na poti med Idrijskim in Ljubljano*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Chambers, J. K. (2003). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*. Malden, Oxford: Blackwell.
- Edwards, J. (2009). *Language and Identity*. New York: Cambridge University Press.
- Fasold, R. (1996). *The Sociolinguistic of Language: Introduction to Sociolinguistics II*. Oxford, Cambridge: Blackwell.
- Fridl, J., Kladnik, D., Orožen Adamič, M., Perko, D., & Zupančič, J. (2001). *Nacionalni atlas Slovenije*. Ljubljana: Rokus.
- Gardner-Chloros, P. (2009). *Code-switching*. Cambridge, New York: Cambridge University Press.
- Greenberg, M. L. (Ed.). (1997). The Sociolinguistics of Slovene. *International Journal of the Sociology of Language*, 124. Berlin, New York: Mouton de Gruyter.
- Guzej, J. (1989). Vpliv migracij na jezik in govor posameznika. *Jezik in slovnstvo*, 35(3), 52–57.
- Jurgec, P. (2011). Slovenščina ima 9 samoglasnikov. *Slavistična revija*, 59(3), 243–268.
- Kenda-Jež, K. (1999). Med narečjem in knjižnim jezikom. In Z. Zorko & M. Koletnik (Eds.), *Logarjev zbornik* (pp. 304–315). Maribor: Slavistično društvo.
- Kenda-Jež, K. (2002). *Cerkljansko narečje: Teoretični model dialektološkega raziskovanja na zgledu besedišča in glasoslovja*. Unpublished doctoral thesis. Ljubljana: Filozofska fakulteta.
- Kenda-Jež, K. (2004). Narečje kot jezikovnozvrstna kategorija v sodobnem jezikoslovju. In E. Kržišnik (Ed.), *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti* (pp. 263–276). Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani.
- Kleindienst, L. (1995). *Bam knapa vzela, bam zméraj veséla: Idrijske anekdote in smešnice*. Ljubljana: Kmečki glas.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, Pennsylvania: University of Pennsylvania Press.
- Labov, W. (2004). Quantitative Analysis of Linguistic Variation. In U. Ammon, N. Dittmar, K. Mattheier & P. Trudgill (Eds.), *Sociolinguistics / Soziolinguistik 1* (pp. 6–21). Berlin, New York: Walter de Gruyter.
- Le Page, R., & Tabouret-Keller, A. (1985). *Acts of identity*. Cambridge: Cambridge University Press.

- Lundberg, G. H. (2014). Dialect Leveling and Local Identity in Slovenia. In J. Witkoś & S. Jaworski (Eds.), *New Insights into Slavic Linguistics* (pp. 241–252). Frankfurt am Main: Peter Lang.
- Mæhlum, B. (1986). *Språklige variasjonsmønstre hos innflyttere i Oslo*. Oslo: Novus.
- Mæhlum, B. (2000). Strategies of Neutrality in the Arctic. In K. Mattheier (Ed.), *Dialect and Migration in a Changing Europe* (pp. 99–108). Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang.
- Mæhlum, B. (2010). Language and social spaces. In P. Auer & J. E. Schmidt (Eds.), *Language and Space: An International Handbook of Linguistic Variation 1: Theories and Methods* (pp. 18–32). Berlin, New York: De Gruyter Mouton.
- Mattheier, K. J. (Ed.). (2000). *Dialect and Migration in a Changing Europe*. Frankfurt am Main, Berlin, Bern, Brussels, New York, Oxford, Wien: Peter Lang.
- Milroy, L. (2002). Introduction: Mobility, contact and language change – Working with contemporary speech communities. *Journal of Sociolinguistics*, 6(1), 3–15.
- Monka, M. (2013). *Sted og sprogforandring. En undersøgelse af sprogforandring i virkelig tid hos mobile og bofaste informanter fra Odder, Vinderup og Tinglev*. Unpublished doctoral thesis. Copenhagen: University of Copenhagen.
- Nečak Lük, A. (1998). Jezik v etničnih študijah: Nekaj raziskovalnih pristopov in izsledkov iz slovenske izkušnje. In I. Štrukelj (Ed.), *Jezik za danes in jutri* (pp. 77–90). Ljubljana: Društvo za uporabno jezikoslovje Slovenije, Inštitut za narodnostna vprašanja.
- Niebaum, H., & Macha, J. (2014). *Einführung in die Dialektologie des Deutschen: 3., überarbeitete und erweiterte Auflage*. Berlin, Boston: Walter de Gruyter GmbH.
- Pogorelec, B. (1984). K vprašanju bilingvizma in diglosije iz slovenske izkušnje. In A. Nečak-Lük & I. Štrukelj (Eds.), *Dvojezičnost - individualne in družbene razsežnosti* (pp. 67–73). Ljubljana: Društvo za uporabno jezikoslovje Slovenije.
- Pogorelec, B. (1998). Jezikovno načrtovanje govornega jezika pri Slovencih: Teorija, praksa in odprti problemi slovenskega zbornega jezika. In I. Štrukelj (Ed.), *Jezik za danes in jutri* (pp. 56–64). Ljubljana: Društvo za uporabno jezikoslovje Slovenije, Inštitut za narodnostna vprašanja.
- Rigler, J. (1970–1978). Akcentske variante I–III. *Slavistična revija*, 18(1–2), 5–15; 19(1), 1–12; 26(4), 365–374.

- Rigler, J. (1981). Cerčno (OLA 6). In P. Ivić (Ed.), *Fonološki opisi srpskohrvatskih/hrvatskosrpskih, slovenačkih i makedonskih govora obuhvaćenih Opšteslovenskim lingvističkim atlasom* (pp. 67–78). Sarajevo: Akademija nauka i umjetnosti Bosne i Hercegovine.
- Sankoff, D., & Laberge, S. (1978). The linguistic market and the statistical explanation of variability. In D. Sankoff (Ed.), *Linguistic Variation: Models and Methods* (pp. 239–250). New York: Academic Press.
- Schmidt, J. E., & Herrgen, J. (Eds.). (2011). *Sprachdynamik: Eine Einführung in die moderne Regionalsprachenforschung*. Berlin: Erich Schmidt Verlag.
- Scotton, C. M. (1976). Strategies of Neutrality: Language Choice in Uncertain Situations. *Language*, 52(4), 919–941.
- Skalin, N. (2002). *Idrijski govor kot podnarečje cerkljanskega narečja?* Unpublished graduation thesis, Filozofska fakulteta, Ljubljana.
- Skubic, A. E. (2005). *Obrazi jezika*. Ljubljana: Študentska založba.
- Smole, V. (2004). Nekaj resnic in zmot o narečjih v Sloveniji danes. In E. Kržišnik (Ed.), *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti*, (pp. 321–330). Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani.
- Smolej, M. (2012). *Besedilne vrste v spontanem govoru*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Srebot Rejec, T. (1988). *Word Accent and Vowel Duration in Standard Slovene: An Acoustic and Linguistic Investigation*. München: Verlag O. Sagner.
- Stabej, M. (2010). *V družbi z jezikom*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Škofic-Guzej, J. (1994). O oblikovanju slovenskega pogovarjalnega jezika. *Slavistična revija*, 42(4), 571–578.
- Škofic, J. (Ed.). 2011. *Slovenski lingvistični atlas 1: Človek (telo, bolezni, družina)*. Ljubljana: Založba ZRC, Znanstvenoraziskovalni center SAZU.
- Škofic, J., Gostenčnik, J., Kumin Horvat, et. al. (2016). *Slovenski lingvistični atlas 2: Kmetija*. Ljubljana: Založba ZRC, Znanstvenoraziskovalni center SAZU.
- Tagliamonte, S. A. (2006). *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Tagliamonte, S. A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Chichester: Wiley-Blackwell.
- Toporišič, J. (2000). *Slovenska slovnica: Četrta, prenovljena in razširjena izdaja*. Maribor: Obzorja.

- Verdonik, D. (2007). *Jezikovni elementi spontanosti v pogovoru: Diskurzni označevalci in popravljanja*. Maribor: Slavistično društvo.
- Verdonik, D., & Zwitter Vitez, A. (2011). *Slovenski govorni korpus GOS*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Wilson, J. (2010). *Moravians in Prague: A Sociolinguistic Study of Dialect Contact in the Czech Republic*. Frankfurt am Main: Peter Lang.
- Wodak, R., Johnstone, B., & Kerswill, P. (Eds.). (2011). *The SAGE Handbook of Sociolinguistics*. London, Thousand Oaks, New Delhi, Singapore: SAGE Publications.
- Zemljak Jontes, M., & Pulko, S. (2019). Odnos do rabe socialnih zvrsti med mladimi. In H. Tivadar (Ed.), *Slovenski javni govor in jezikovno-kulturna (samo)zavest* (pp. 247–254). Ljubljana: Znanstvena založba Filozofske fakultete.

VARIANTNOST GOVORJENE SLOVENŠČINE PRI GEOGRAFSKO MOBILNIH OSEBAH

Članek predstavlja prvo variantnostno študijo na slovenskem gradivu in se osredotoča na avtentično jezikovno rabo govorcev z Idrijskega, ki se šolajo oz. delajo v Ljubljani. Gre za dnevno ali tedensko geografsko mobilne govorce, ki v prvotnem okolju vsaj v neformalnih govornih položajih pretežno govorijo lastno narečje, v osrednjeslovenskem prostoru pa v komunikaciji z govorniki drugih varietet uporabljajo različne strategije govornega obnašanja. V ospredju je predstavitev mlajših informantov, dijakinje in dijaka.

Metodološko se kvantitativna analiza petih fonoloških variabel povezuje s kvalitativnimi podatki iz sociolingvističnih intervjujev. Raziskovalni korpus za variantnostno študijo sestavlja izbor izsekov iz celotnega nabora posnetkov glede na kraj, sogovorce in temo pogovora, analiza se osredotoča na tri samoglasniške in dve soglasniški variabli. Kvalitativni podatki iz sociolingvističnih intervjujev vključujejo osebno zgodovino posameznikov, njihovo samooceno govornega obnašanja, izkušnje z jezikovno rabo, ozaveščenost jezikovnega (ne)prilagajanja, jezikovna stališča ter razumevanje povezave med jezikom in identiteto.

Informanti uporabljajo različne strategije govornega obnašanja, tako da lahko opredelimo različne tipe oz. profile govorcev: kodne preklopljevalce, narečne govorce in kodne mešalce. Na kontinuumu med narečjem in standardnim jezikom se kaže več vmesnih jezikovnih varietet, ki se ne ujemajo z v slovenskem jezikoslovju pogosto pretirano shematizirano dihotomijo med

knjižnim jezikom in tradicionalnimi narečji. Pri posameznih analiziranih variablah je mogoče opaziti različne tendence. Potrjuje se relevantnost različnih družbenih in socialnopsiholoških dejavnikov pri pojasnjevanju posameznikovega govornega obnašanja, kompleksnosti jezikovnih izbir in medosebne komunikacije.

Ključne besede: variantnost govorjene slovenščine, cerkljansko narečje, profili govorcev, samosnemanje, sociolingvistični intervjuji, jezikovna stališča

Annotated Lexicon for Sentiment Analysis in the Bosnian Language

Sead JAHIĆ

Faculty of Mathematics, Natural Science and Information Technologies, University of Primorska

Jernej VIČIČ

Faculty of Mathematics, Natural Science and Information Technologies, University of Primorska; Fran Ramovš Institute of the Slovenian Language

The paper presents the first sentiment-annotated lexicon of the Bosnian language. The annotation process and methodology are presented along with a usability study, which concentrates on language coverage. The composition of the starting base was done by translating the Slovenian annotated lexicon and later manually checking the translations and annotations. The language coverage was observed using two reference corpora. The Bosnian language is still considered a low-resource language. A reference corpus comprised of automatically crawled web pages is available for the Bosnian language, but the authors had a hard time sourcing any corpora with a clear time frame for the text contained therein. A corpus of contemporary texts was constructed by collecting news articles from several Bosnian web portals. Two language coverage methods were used in this experiment. The first used a frequency list of all words extracted from two reference Bosnian language corpora, and the second ignored the frequencies as the main factor in counting. The computed coverage using the first presented method for the first corpus was 19.24%, while the second corpus yielded 28.05%. The second method yielded 2.34% coverage for the first corpus and 6.98% for the second corpus. The results of the study present a language coverage that is comparable to the state of the

Jahić, S. et al.: Annotated Lexicon for Sentiment Analysis in the Bosnian Language. Slovenščina 2.0, 11(2): 59–83.

1.01 Izvirni znanstveni članek / Original Scientific Article

DOI: <https://doi.org/10.4312/slo2.0.2023.2.59-83>

<https://creativecommons.org/licenses/by-sa/4.0/>



art in the field. The usability of the lexicon was already proven in a Twitter-based comparison.

Keywords: Bosnian lexicon, corpus, sentiment analysis, AnAwords, stopwords, log-likelihood, annotation

1 Introduction

Sentiment analysis, also known as opinion mining, is a field of study within the larger discipline of natural language processing (NLP) that aims to determine the sentiment expressed in text, categorizing it as positive, negative, or neutral. The goal of sentiment analysis is to extract meaningful insights from large amounts of unstructured data, such as social media posts (Iglesias and Moreno, 2019) or online product reviews (Wu et al., 2020), in order to understand public opinions and attitudes. In this paper we present the coverage of the first Bosnian sentiment annotated lexicon using two reference corpora.

Although arguably Bosnian is closely related to Serbian and Croatian, there are subtle differences between these three languages that are more evident from the sentiment analysis point of view. The main differences between Bosnian, Serbian, and Croatian lie in the use of vocabulary, grammar, and syntax. Although the three languages share a similar Slavic origin and linguistic heritage, they have evolved differently over time and been influenced by different cultural, historical, and political factors. These differences are particularly pronounced when it comes to sentiment analysis, as the choice of words and the way they are used can significantly impact the sentiment expressed in a text. As such, it is important to consider these differences when developing sentiment analysis tools for the Bosnian language.

The lexicon used in this study has been constructed using two reference corpora and combines NLP and machine learning techniques to assign weighted sentiment scores to the entities within a sentence or phrase. The study covers two approaches to evaluate the performance of the lexicon – the first takes into account the frequencies of the covered and missed words, while the second just counts the words that are covered by the lexicon.

The paper provides a comprehensive overview of the state of the art in NLP and sentiment analysis for the Bosnian language. It explains the methodology used in the process of creating the lexicon, cleaning the corpora, the corpora covered by the lexicon, and annotation. The results of the experiment and the conclusion, along with suggestions for future work, are presented in the last section of the paper.

In summary, the development of the Bosnian sentiment annotated lexicon is a step towards better understanding and analysing public opinion expressed in the Bosnian language. The results of the study suggest that the lexicon has good coverage, and the methodology used in the construction of the lexicon can serve as a reference for future work in this field.

2 State of the art

There has been quite extensive research in the area of sentiment analysis, and many types of models and algorithms have been proposed depending on the final goal of the analysis of the interpretation of user feedback and queries, such as fine-grained sentiment analysis (based on polarity precision) (Chen et al., 2020), emotion detection, aspect-based sentiment analysis (Suciati and Budi, 2020), and multilingual sentiment analysis (Kia et al., 2016). All those algorithms and models can be divided into one of three basic classes: rule-based systems (relying on long-used linguistic methods, rules, and annotated linguistic materials such as annotated lexicons), automatic (corpus-based) systems, and hybrid systems that combine properties from both previous types. In the latter, hybrid systems use machine learning techniques together with NLP techniques developed in computational linguistics, such as stemming, tokenization, part-of-speech tagging, parsing, and lexicons.

Lexicons have been widely used for sentiment analysis. One of the first-known, human-annotated lexicons for sentiment analysis is the General Inquirer lexicon (Hartman et al., 1967), which contains 11,788 English words (2,291 labeled as negative and 1,915 as positive, with the rest, labeled as objective).

Sentiment lexicons exist for most Slavic languages, including Bulgarian (Kapukaranov and Nakov, 2015), Croatian (Glavaš et al.,

2012), Czech (Veselovská, 2013), Macedonian (Jovanoski et al., 2015), Polish (Wawer, 2012), Slovak (Okruhlica, 2013), Slovenian (Kadunc, 2016) and Bosnian (Jahić and Vičič, 2023b), with the last of these containing 1,219 entries labeled as positive and 3,935 as negative.

Important questions for natural language researchers, general linguists, and even teachers and students are how much text coverage can be achieved with a certain number of words from the lexicon in a given language, since the number of terms in the lexicon is smaller by a few magnitudes than the number of terms in the corpus.

Studies of vocabulary coverage have been carried out for many languages, such as German (Jones, 2006), where a study based on the BYU/Leipzig Corpus of Contemporary German has shown that a basic vocabulary of 3,000 high – frequency words can account for between 75% and 90% of the words in the text. Moreover, with Spanish (Davies, 2005) it is claimed that it is enough to know 4,000 words to cover or recognize more than 90% of the words in native texts. Moreno-Ortiz and Pérez-Hernández (2018) presented Lingmotif-lex, a wide-coverage, domain-neutral lexicon for sentiment analysis in English, and stated that it achieves significantly better performance than the other lexicons for English, with coverage of up to 75% and 84% (F1-score) for two datasets.

In a study aimed at developing resources for sentiment analysis in Slovene, Bučar, Žnidaršič and Povh (2018) collected more than 250,000 news items from five Slovenian online media sources as the basis for their resources, which corpora, annotations, and a lexicon. To evaluate the quality of the annotation process, they used five different measures of correlation. The results showed good internal consistency across all levels of granularity, although the values decreased slightly when applied to the smaller units of text.

Corpus-based and lexicon-based model methods have been increasingly used to compare language coverage, and the comparison of hundreds of thousands or even millions of words/lemmas from a corpus with a few thousand words/lemmas from a lexicon is one of the main types of corpus comparison.

3 Construction of the lexicon

The Bosnian sentiment-annotated lexicon is presented and analysed in this paper. For this purpose, our data consists of the “core” lexicon (Jahić and Vičić, 2023b), a list of stopwords, and a list of AnAwords (Affirmative and Non-affirmative words, such as “ekstremno” (“vrlo”) – extremely, “jedva” – barely) (Jahić and Vičić, 2023a), as clarified in Figure 1.

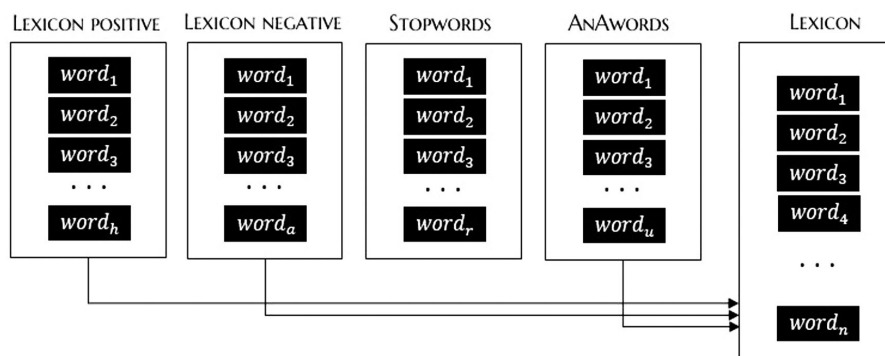


Figure 1: Construction of the lexicon.

The lexicon creation process is comprised of taking entries (word forms) from the Slovene sentiment lexicon KSS 1.1 (Kadunc, 2016) and translating them into Bosnian. We also allow some variance of the same lemma as part of the lexicon. The process of creating the Bosnian translation was undertaken in a dual-phase approach. In the initial phase, the transformation of the Slovenian lexicon into Bosnian took place through well-defined steps. Initially, the Slovenian lexicon underwent translation into English through the utilization of the translators from Google and Microsoft. Subsequently, this intermediary English version was subjected to translation into the Bosnian language, which is visually depicted in Figure 2. Moreover, the first phase involved these steps:

- Translation using Microsoft Translator for the Slovene sentiment lexicon KSS 1.1.
- Translation using Google Translator for the same lexicon.
- Manual comparison and merging of the two lists, removing duplicate entries.

- Manual cross-checking to ensure that words had matching or similar meanings.
- The result was the creation of the Bosnian_MG_Translated Lexicon.

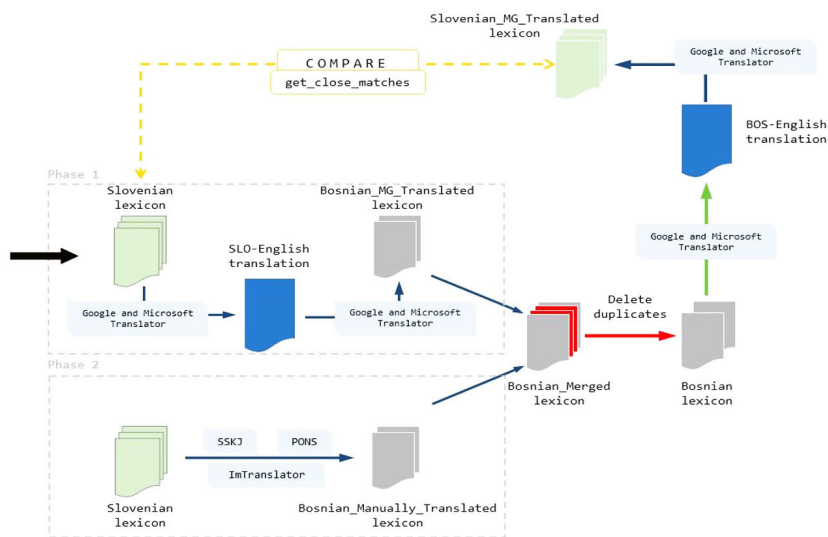


Figure 2: The lexicon creation process.

The last phase witnessed the creation of the lexicon in a two-fold manner. Firstly, word forms from the Slovenian lexicon were manually translated into Bosnian through manual means. This comprehensive process encompassed a verification of each term using various tools, including Pons,¹ Google Translate, ImTranslator,² and the Dictionary of Slovenian Literary Language (SSKJ – Slovar slovenskega knjižnega jezika³). The results of this process yielded the “Bosnian_Manually_Translated” lexicon.

These two lexicons (“Bosnian_MG_Translated” lexicon and “Bosnian_Manually_Translated” lexicon) were subsequently united and merged into a cohesive entity, referred to as the “Bosnian_Merged” lexicon. The refinement process further entailed the removal of duplicate entries, resulting in the initial iteration of the Bosnian sentiment

1 <https://sl.pons.com/>

2 <https://imtranslator.net/>

3 <https://fran.si/>

lexicon. To ensure the accuracy and robustness of the lexicon, a back-translation procedure was executed. This involved translating the newly constructed Bosnian lexicon back into the Slovenian language, as depicted in Figure 2.

The goal of the back-translation procedure was to retranslate the obtained Bosnian lexicon into a Slovenian lexicon and then compare this translated lexicon (in Slovenian) with the initial KSS 1.1 lexicon. What we found during the back-translation process is that many words were translated into a form that is not present in KSS 1.1, while the infinitive form of those words is indeed available in KSS 1.1.

To circumvent this challenge in the evaluation phase and also in the process of using the lexicon in the sentiment analysis process, we used the *'get_close_matches'* function (part of the *difflib* module in *Python*). By using this function we effectively pinpointed the closest approximations to the target string from a pool of candidate strings. This process substantially improved the coverage and reliability of our lexicon, amplifying the precision of our sentiment analysis efforts.

The method works by comparing the target string with each candidate string, using a defined similarity ratio, and then returning the matches with the highest similarity ratio. The number of matches returned and the similarity ratio threshold can be controlled through the *n* and *cutoff* parameters, respectively. The order of close-matched strings is based on the similarity score, so the most similar string comes first in the list.

This function accepts four parameters:

- *word*: This is the string for which we need the close matches.
- *possibilities*: This is usually a list of string values with which the word is matched.
- *n*: This is an optional parameter with a default value of 3. It specifies the maximum number of close matches required.
- *cutoff*: This is also an optional parameter with a default value of 0.6. It specifies that the close matches should have a score greater than the cutoff.

In our case, we pick the first element from the close-matched strings list (with the highest similarity score). More cutoff values were considered, and the best confidence-accuracy score was reached with a cutoff of 81%.

Table 1: Comparing the Slovenian lexicon before and after the translation process

	Slovenian lexicon (lemmas) (Kadunc, 2016)	cutoff	Slovenian MG_Translated lexicon		Comparing accuracy
			translated words	matched words	
Positive	1911	80%	1758	1829	-
		81%	1781	1686	88.23%
		82.5%	1790	1627	85.14%
		85%	1806	1550	81.11%
		90%	1838	1369	71.64%
		100%	1858	1235	64.63%
Negative	5125	80%	4572	4999	-
		81%	4654	4604	89.83%
		82.5%	4690	4432	86.48%
		85%	4739	4125	80.49%
		90%	4846	3514	68.57%
		100%	4898	3067	59.84%

The accuracy score was counted by comparing the primary lexicon of the Slovenian language (Kadunc, 2016) and the back-translated lexicon of the Slovenian language.

$$\text{Comparing_accuracy} = \frac{\text{Number of matched words (positive/negative)}}{\text{Number of all words in the lexicon (positive/negative)}}$$

The equation used is as follows:

The Bosnian sentiment lexicon consists of 3,935 negative words (Lexicon negative), and 1219 positive words (Lexicon positive). Besides that, we also added a list of 394 Bosnian stopwords (such as: “gosp.” (“Mr.”), “je” (“is”), “juli” (“July”), and so on), and list of AnAwords. Stopwords usually refer to the most common words in a language, and there is no single universal list of these. The first

Bosnian sentiment lexicon was tested by using it to label tweets written in the Bosnian language (Jahić and Vičić, 2023a, 2023c).

4 Methodology and work

The core emphasis of this paper is on assessing the coverage achieved by the lexicon, rather than on its creation, although a comprehensive account of this is also presented. More specifically, the focus of is on evaluating how many lemmas the lexicon covers in bsWaC and bsNews, as detailed below.

The language coverage of the lexicon was evaluated through two different corpora:

- The Bosnian web corpus bsWaC 1.1 (Ljubešić and Klubička, 2014). The bsWaC 1.1 corpus was part of a collection of corpora, named the {bs, hr, sr}WaC – Web corpora of Bosnian, Croatian, and Serbian languages. The number of seed URLs (crawled web pages) was 8,388 for bsWaC, 11,427 for srWaC, and 14,396 for hrWaC. The bsWaC corpus consists of more than 285 million tokens (286,865,790, to be precise) written in Bosnian. The corpus is also morphosyntactically annotated and lemmatized. At the time of writing, this corpus was the *de facto* reference corpus for the Bosnian language.
- The Bosnian news corpus 2021 bsNews 1.0 (Vičić, 2021), which is a collection of web news articles crawled at the start of 2021. The corpus contains a balanced set of at most 2,000 of the most recent news articles from each identified web news portal in Bosnia and Herzegovina. The list of portals is maintained by Press Council in Bosnia and Herzegovina.⁴ The corpus contains news articles from 46 portals. This corpus was used as a contemporary and balanced source. The sentence tokens are morpho-syntactically annotated with MULTEXT-East morpho-syntactic annotations for Croatian, Version 6⁵. The corpus was morpho-syntactically annotated and lemmatized with ToTaLe (Erjavec et al., 2015). It consists of more than 36 million tokens in the Bosnian language.

4 Vijeće za štampu u Bosni i Hercegovini: <https://www.vzs.ba/index.php/vijece-za-stampu/internet-portali-u-bosni-i-hercegovini>.

5 <http://nl.ijs.si/ME/V6/>

Two different approaches are applied:

- First, all lemmas with their frequencies were considered,
- Second, the frequencies for lemmas were ignored.

A list of lemmas with frequencies was extracted from each corpus and cut off at five occurrences to avoid clutter.

The list of lemmas extracted from the first corpus (Ljubešić and Klubička, 2014) consisted of 348,988 different lemmas with frequency. The lemmas are ordered in increasing order by frequency, where the lowest value is five (cutoff) (“batkovi - drumsticks” ...) and the highest value is 16,652,066 for the lemma “biti - to be”.

The list of lemmas extracted from the second corpus (BsNews 1.0 corpus (Vičič, 2021)) consisted of 101,771 lemmas ordered in decreasing order, the most frequent lemma again being “biti – to be” with the frequency of 2,350,487, and with the lowest frequency of five for lemmas such as “polegnuti – lay down”.

Not all lemmas can be included in the analysis. Symbols, equation marks, and numbers, even if part of the corpus, cannot be part of the lexicon, especially a sentiment annotated lexicon.

The following items were thus removed from both corpora in the cleaning process: emoticons, punctuation like quotes, exclamation marks, etc., numbers, and hyperlinks.

4.1 The first approach: lemmas with their frequency were included in the analysis (all appearances of lemmas were used for each corpus)

In the first approach used in our analysis, we considered lemmas along with their frequency as the basis for our investigation. This means that we included all instances of lemmas found in each corpus for our analysis.

Figure 3 shows the procedure for checking the existence of given words from the lexicon in the corpus.

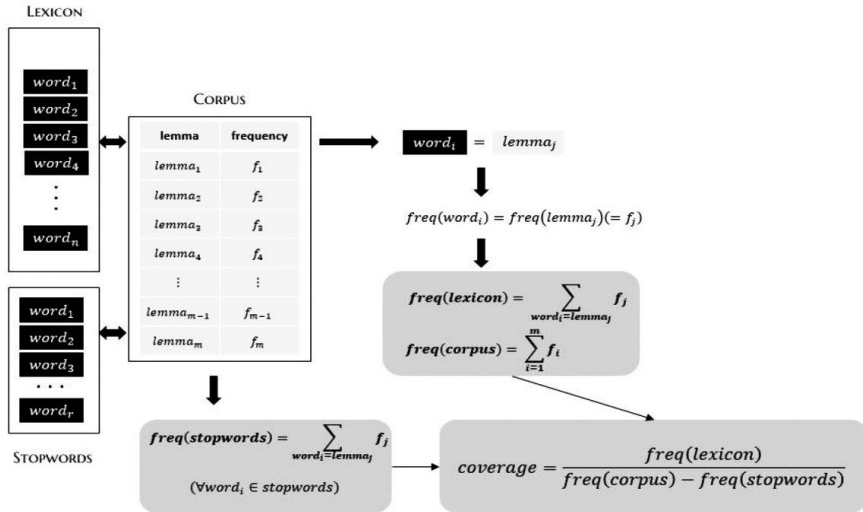


Figure 3: Process of matching words from the corpus with words from the lexicon.

If this statement is true and the word exists in the corpus, the value of $freq(\text{lexicon})$ is accumulated for the value of the frequency of each word, otherwise, the value 0 is added to $freq(\text{lexicon})$.

The sum of all word's frequencies from the corpus is given as $freq(\text{corpus})$ and $freq(\text{stopwords})$ presents the sum of all frequencies of the stopwords that appears in the corpus.

The coverage is counted as:

$$coverage = \frac{freq(\text{lexicon})}{freq(\text{corpus}) - freq(\text{stopwords})} \quad (1)$$

where all stopwords were excluded from the corpus.

4.2 The second approach: using the accuracy of words without the influence of frequency

Given that the sentiment value is not at the forefront at this stage of research (we are looking for language coverage), lists of 1,219 positive and 3,935 negative words were united in a unique lexicon.

In addition to lexicons, two other groups of words – stopwords (394 of them) and AnAwords (Affirmative and Non-affirmative words) – play a significant role in this process. Jahić and Vičič (2023a) pointed out that stopwords usually refer to the most common words in a language and that there is no single universal list of these.

However, 139 words from the AnAwords list were created by Jahić and Vičič (2023a), and it has been proven (Osmankadić, 2003) that most of these are intensifiers.

The consideration of words from the AnAwords list significantly impacted the corpus coverage, as elucidated in the second stage of the second approach. These words were evaluated in a manner similar to stopwords, given their absence of inherent sentiment value. Consequently, they were excised from the corpora, in line with the objective of eliminating non-sentiment-bearing terms.

Taking this into account, the AnAwords were also subject to examination, considering their lack of any discernible sentiment value. Consequently, they were treated analogously to stopwords, leading to their exclusion from the analysis

The process of annotating the lexicon went through several stages, and they were all based on the following equation:

$$\frac{FOUND}{NOT_FOUND}$$

(2)

where FOUND presented the list of all words in the corpus that were matched with words from the lexicon, NOT_FOUND opposite.

In more details, these stages are as follows:

- Simple coverage of the corpus by a lexicon was shown in the first stage. The stopwords were part of the corpus at this stage.
- In the first stage, stopwords were integrated into the corpus, contributing to the text's initial structure. However, as the coverage process unfolded in the second stage, the corpus coverage was achieved without the incorporation of stopwords, in addition to the exclusion of AnAwords words. The rationale behind these

decisions stems from the fact that the number of stopwords and AnAwords is almost negligible in comparison to the total number of elements in the corpus. As such, a substantial variance in coverage during this stage was not anticipated.

- Guided by the results of research conducted for the corpus-based lexical analysis of subject-specific university textbooks in English (Hajiyeva, 2015), in the third stage coverage was observed by the frequency distribution of words.
- In the fourth stage, the question arises as to whether it is possible to group similar words (such as “anđeo” and “anđel” (angel)) and view them as a single word. As Davies (2005) stated, one of the solutions to this problem is grouping words according to word families. Given this possibility of grouping, matching functions were applied between corpus words and lexicon words.
- In the fifth stage the log-likelihood was computed for each word in the lexicon. Following Rayson and Garside (2000), the word frequency list is then sorted by the resulting log-likelihood values. This gives the effect of placing the largest log-likelihood value at the top of the list representing the word that has the most significant relative frequency difference between the two corpora. This method enables a comparison of the most indicative (or characteristic) words in both corpora.

5 Results

This section showcases the results of the two approaches described earlier. We started by cleaning the corpora, which led to the inclusion of 263,969 words from the first corpus and 84,859 words from the second in our subsequent analysis (see Table 2).

Table 2: *Number of lemmas left after pre-processing the corpora*

	CORPUS1	CORPUS2
The overall number of lemmas	348,988	101,771
Cleared lemmas	263,969	84,859
Percent (%)	75.64	83.38

In the first approach (influence of frequency was considered), $\text{freq}(\text{corpus})$, the sum of stopwords frequencies $\text{freq}(\text{Stopwords})$ and the overall sum of all frequencies of the words from the lexicon ($\text{freq}(\text{lexicon})$) were computed.

By using equation (1) coverage of the corpus1 is 19.24%, and coverage of the corpus2 is 28.05% (see Table 3).

Table 3: Coverage of the corpora's lemmas with words from sentiment lexicon

	Freq(corpus)	Freq(lexicon)	Freq(stopwords)	Coverage
CORPUS1	187,957,442	28,174,959	41,542,468	19.24%
CORPUS2	3,0168,771	6,371,417	7,456,808	28.05%

The second approach (when the influence of frequency is ignored) was to compute the overall coverage of the corpora without using word frequencies. The motivation behind this approach was to count how many different lemmas from the corpus are already present in the sentiment lexicon. There are several stages in this approach.

- *First stage:* In this first stage, 1.523% coverage of the first corpus and 4.098% for the second corpus was achieved.

Table 4: Coverage of corpora's lemmas with words from the sentiment lexicon (without any additional changes being made)

	CORPUS1	CORPUS2
FOUND	3,959	3,341
NOT_FOUND	260,010	81,518
Coverage (%)	1.523	4.098

Table 4 presents lemmas that were matched with words from the lexicon (FOUND) and that were absent from the lexicon (NOT_FOUND).

Maximum coverage of corpora is possible if all words from the lexicon are included in the corpora. This means that the maximum coverage for the first corpus is 1.99% and for the second corpus is 6.47%. In contrast, the coverage of the lexicon by the corpora is 76.81% and 64.82%. This means that of 5,154 words from the lexicon, 3,959 were presented in corpus1, indicating 76.81% use

of the lexicon, and 3,341 were presented in corpus2, indicating 64.82% use.

- The *second stage* increases the number of words in FOUND since all words that are stopwords or AnAwords have been detected in the corpus. In this case, coverage of corpora is increased to 1.7% and 4.62% for corpus1 and corpus2, respectively.

Table 5: Coverage of corpora's lemmas with word from the sentiment lexicon

	CORPUS1	CORPUS2
FOUND	4,406	3,747
NOT_FOUND	259,533	81,112
Coverage (%)	1.7	4.62

- The *third stage* is distributing words by frequency and counting the number of lemmas that were or were not covered by words from the lexicon.

From a total of 5,154 words from the lexicon, 3,257 (63.19%) were included in the 50,000 most frequent lemmas from corpus1 (see Figure 4 (left)). Meanwhile, of the 15,000 most frequent lemmas from corpus2, 3,071 were in the lexicon. Since the overall number of words from the lexicon is 5,154, this means that gave 59.58% of all words from the lexicon are found in the 15,000 most frequent lemmas from corpus2 (see Figure 4 (right)).

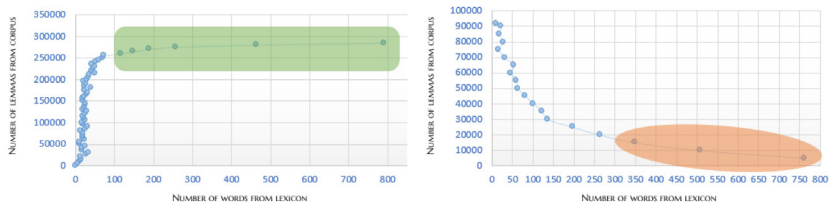


Figure 4: Annotated lexicon by distributed lemmas from corpus1 (left) and corpus2 (right).

- *Fourth stage:* In the fourth stage the lexicon annotation was increased to more than 2.2% for the first corpus. Even though it looks like this contradicts the claim that the maximum coverage for the first corpus is about 1.54%, it does not. The reason for this is

because the `get_close_matches` function was applied with several cutoffs and `n=1` (one possibility).

`get_close_matches(word, possibilities[, n][, cutoff])`

The function works in such a way that all almost similar words (82.5% and 85% matching in this case for the first and second corpora, respectively) are considered as one word. For example, *anđel* (Engl. angel), *anđelko* (Engl. little angel), and *anđela* (“I saw an **angel**”), all three words were replaced with the word *anđeo*. We have found that for a cutoff lower than of the 82.5% matching function returns words that are not matched or related to the root word.

The impact of `get_close_matches` is presented in Figure 5 on a small part of corpus2 with the matching word “anđeo” (angel).

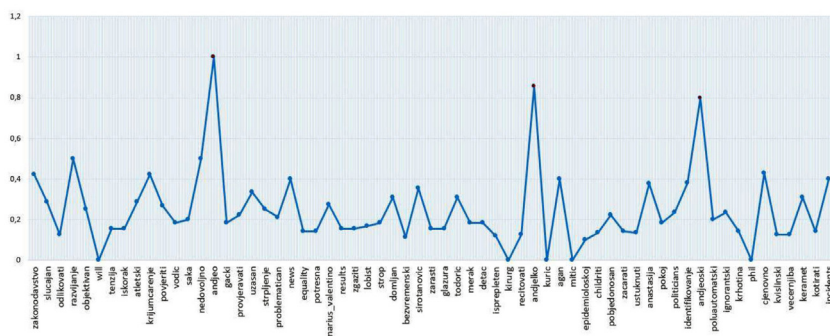


Figure 5: Implementation of `get_close_matches` on corpus1 of the word “anđeo” (an angel).

In the figure shown above there are some words whose matching factor with the word “anđeo” was greater than 0.85. Those words (*anđeo*, *anđelko*, *anđeoski*) were replaced with the word “anđeo”. Having that in mind, the number of words (for corpus1) in the corpora decreased (see Table 6). If an 85% cutoff is applied, the overall number of words in corpus1 is reduced to 242,615, and the annotation increases to 2.22%. This means that out of 242,615 words in corpus1, 5,280 were found in the lexicon, stopwords, and AnAwords groups. The same applies to corpus2, where an annotation of 4,706 words from the lexicon in corpus2

was detected. Moreover, when a cutoff of 82.5% is applied, we get an annotation of 2.34% and 6.98% for the first and second corpora, respectively.

Table 6: Coverage of corpora's lemmas with words from the sentiment lexicon

Cutoff	82.5%		85%	
Corpus:	CORPUS1	CORPUS2	CORPUS1	CORPUS2
No. of lemmas	233,157	74,134	242,615	77,437
FOUND	5,327	4,838	5,280	4,706
NOT FOUND	227,830	69,296	224,939	72,731
Coverage (%)	2.34	6.98	2.22	6.47

- *Fifth stage:* In the fifth stage the log-likelihood was computed for each of the 5,000 most frequent lemmas from both corpora (10,000 overall) and only those that were common for both corpora were counted (4,207 in total).

In this way, the opportunity was given to compare the frequencies of word form occurrences in two texts (for this purpose two corpora) and obtain a statistical measure of the significance of the differences.

To compute the log-likelihood, a *two-by-two contingency table* (see Table 7) of frequencies for each word was constructed.

Table 7: Contingency table for word frequencies

	CORPUS1	CORPUS2	TOTAL
Freq. of word	a	b	a+b
Freq. of other words	c-a	d-b	c+d-a-b
TOTAL	c	d	c+d

In Table 7, *c* and *d* present the number of words in the corpora. In this case, $c = 183m481,818$ and $d = 28m690m802$. *c* and *d* were obtained by summing all the frequencies of all 4,207 words.

Following Rayson and Garside (2000), the equations:

$$E_1 = \frac{c \cdot (a + b)}{(c + d)}, E_2 = \frac{c \cdot (a + b)}{(c + d)}$$

were used to calculate the expected values, and:

$$LL(\text{word}) = 2 \cdot \left(a \cdot \ln \frac{a}{E_1} + b \cdot \ln \frac{b}{E_2} \right),$$

to calculate the log-likelihood for each word.

Using these equations a word frequency list (LL_list) was created, and the words were sorted from the smallest to largest values, where the largest value represents the word that has the most significant relative frequency difference between the two corpora.

As such, the most characteristic words of one corpus, as compared to the other corpus, were listed at the bottom of the given list, while words with almost the same relative frequency were listed at the top.

To evaluate the result and identify the N number of words that have similar interpreted values, we needed another method. As stated by Kilgarriff in *Comparing Corpora* (Kilgarriff, 2001), the simplest method that could be used for this is applying Sketch Engine. For each word the quotient of frequency was computed, and if the value of the quotient is 1 it indicates that its frequency is identical in both corpora. The higher the score in the Sketch Engine frequency word list (SE_list), the greater the difference between corpora. However, it should be noted that the given score can only be used for comparing differences, and it does not give clues as to what exactly is different between the corpora.

Because of this, the identification of words with similar interpreted values was done. This means that the percentage of coverage of N's highest keyness score words with words from the lexicon for both lists LL_list and SE_list was computed.

The 500 most relevant words for both lists were identified. These words distinguish one corpus from the other, and also present the strengths of one corpus over the other. First the log-likelihood was computed and the LL_list was created. A list of 500 words with the biggest frequency differences between the two corpora was created. Then the two corpora were compared

by using Sketch Engine and the SE_list was created. As for the LL_list, the 500 words with the biggest frequency differences between the two corpora were identified.

The aim was not to compare corpora but to check the coverage of the most relevant words – those that distinguish the two corpora from each other – with words from the lexicon.

Using these comparison methods produced a matching factor of 55.2% between LL_list and SE_list of the 500 most relevant words.

Table 8: Coverage of 500 most relevant words from the lexicon group and distribution of words from the lexicon, stopwords, and AnAwords lists (LSAnA group)

	LL_list				SE_list			
FOUND words	156				136			
Coverage (%)	31.2%				27.2%			
By words from	Lexicon		Stopwords	AnAwords	Lexicon		Stopwords	AnAwords
	pos	neg			pos	neg		
	30	50	62	14	44	42	35	15

As can be seen in Table 8, from the 500 words there were 156 from the LSAnA group that matches them (31.2% for the LL_list) and 136 words from the LSAnA, representing 16.8% of coverage from the SE_list.

Words from the lexicon had 51.28% coverage $((30+50)/156)$ and annotation of the lexicon in those 500 words from LL_list had 16% coverage $(80/500)$. For the SE_list this annotation had about 27.2% coverage, and the overall impact of lexicon words with regard to all the words covered by the SE_list was about 63.24% $((44+42)/136)$.

Even though the third and fifth stages present insights into the annotation of the most frequent words, for overall annotation the most important stages were the first, second and fourth ones, since they produce the overall coverage of the corpora by lexicon (see Table 9).

Table 9: Annotation of corpora

Approach:	Coverage of corpora	
	CORPUS1 (bsWaC)	CORPUS2 (bsNews)
First	by using the accuracy of words with the influence of frequency	
	19.24%	28.05%
Second	by using the accuracy of words without the influence of frequency	
	First	4.098%
	Second	1.7%
	Fourth	2.22%-2.34%
		6.47%-6.98%

6 Conclusion

Although Bosnian is arguably closely related to Serbian and Croatian, there are subtle differences between these three languages that are more evident from the sentiment analysis point of view. This paper presents the annotation of the first Bosnian sentiment lexicon that has been proven on a sentiment basis in earlier work. The lexicon includes about 5,500 words (1,219 positive, 3,935 negative, 394 stopwords, and 139 AnA words) and covers more than 19% of the words in the first observed corpus (corpus1) (Ljubešić & Klubička, 2014), and more than 28% of words in the second corpus, BsNews 1.0 corpus (corpus2), (Vičić, 2021). If the emphasis is on coverage of different words from the corpus by the lexicon, then coverage is 1.7% for corpus1 and 4.62% for corpus2. This coverage will increase by applying some matching functions between the corpora's words and lexicon's words (as described in the fourth stage of the second approach in Section 4). In that case, the coverage rises to 2.34% for corpus1 and 6.98% for corpus2. From 85.07% to 93.67% of words from the lexicon were found in corpus1 (between 360 and 849 words from the lexicon were not found) and between 82.75% and 92.84% in corpus2, which means that between 407 and 981 words from the lexicon were not found in corpus2.

The results show that about a quarter of the words from the corpora have their sentiment value annotated in the lexicon, which greatly helps in the sentiment annotation of the sentences (tweets or regular text).

Stopwords and AnAwords were also included in the analysis, which leads to the possibility that the LSAnA group becomes a representative group for sentiment words, stopwords, and intensifiers (all written in Bosnian).

The language coverage of the lexicon is comparable with the current state of the art, and the values can be compared (Moreno-Ortiz & Pérez-Hernández, 2018).

During the process of creating our lexicon, we were aware that there would be deviations during the translation. The Slovene sentiment lexicon KSS 1.1 also includes multi-part words, which are words composed of multiple individual words joined by “_”, such as “dobro_sprejet”, “dobro_upravljan”, “dobro_voden”, “dobro_vzgojen”, “energetsko_varčen”, “funkcijsko_bogat”, and similar terms. Most of these types of words do not have an equivalent in the Bosnian language. However, during the manual review of our lexicon we noticed that some of these words could be included, such as “prekomerna_teža” (Bosnian: predebelo, English: too fat) or “srce_parajoč” (Bosnian: srceparajuće, English: heartbreaking). Despite this, these words did not make it into the primary version of our lexicon.

Table 10: Comparison of lexicon terms

Language	Positive		Negative	
	core terms	terms with “_”	core terms	terms with “_”
Slovenian	1,911	61	5,152	276
Bosnian	1,126	-	3,868	-

According to Table 10, it is evident that the Bosnian lexicon can be updated by finding appropriate translations for multi-part words from KSS 1.1. This update would have an immediate impact on the lexicon’s annotation, as incorporating more terms would allow for more comprehensive annotation. Furthermore, during our analysis we discovered that among these multi-part words some contained elements from the AnAwords list, which we treated separately. Examples of such cases include “hudo_bolan” (Bosnian: veoma bolan, English: very painful), “zelo_poceni” (Bosnian: veoma jeftin, English: very cheap), “povsem_prava” (Bosnian: potpuno pravo (tačno), English: completely right), and others.

Additionally, we found that there were entire expressions in the Slovenian lexicon that were not included in the Bosnian lexicon. Some examples of these are “nič_hudega_sluteč” (Bosnian: ne slutiti ništa loše, English: unaware of any harm), “obesiti_na_klina” (Bosnian: objesiti o klin, English: hang on a nail), and “veliko_hrupa_za_nič” (Bosnian: mnogo buke oko ničega, English: much ado about nothing), “zvit_kot_lisica” (Bosnian: lukav kao lisica, English: sly as a fox), among others.

The focus in future work will be on developing and improving the LSA_{nA} group. All members of the group should be extended, which means that we expect to have more items/words labelled as positive or negative in our “core” lexicon, as well as extending lists of stop-words and AnAwords. To increase coverage, we will try to create a lexicon with all possible words, and in doing so we will contain all the grammatical rules found in the Bosnian language itself (declination, conjugation, change of words by gender, number, and so on). Although the process of annotation, as well as improvement of the first Bosnian lexicon (Jahić and Vičić, 2023b), is still in development, the results shown here are comparable with those reported for other related languages, and also for language families, as shown in Davies (2005) and Bučar, Žnidaršič and Povh (2018).

Acknowledgments

The authors gratefully acknowledge the European Commission for funding the InnoRenew CoE project (Grant Agreement #739574) under the Horizon 2020 Widespread-Teaming programme and the Republic of Slovenia (investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund).

References

- Bučar, J., Žnidaršič, M., & Povh, J. (2018). Annotated news corpora and a lexicon for sentiment analysis in slovene. *Language Resources and Evaluation*, 52, 895– 919. doi:10.1007/s10579-018-9413-3
- Chen, C., Hu, X., Zhang, H., & Shou, Z. (2020). Fine grained sentiment analysis based on Bert. *Journal of Physics: Conference Series*, 1651.

- Davies, M. (2005). Vocabulary range and text coverage. insights from the forthcoming routledge frequency dictionary of spanish. *Selected Proceedings of the 7th Hispanic Linguistics Symposium* (pp. 106–115).
- Erjavec, T., Ignat, C., Pouliquen, B., & Steinberger, R. (2015). Massive multi lingual corpus compilation: Acquis communautaire and totale. *Archives of Control Sciences* 15.
- Glavaš, G., Šnajder, J., & Bašić, B. D. (2012). Semi-supervised acquisition of croatian sentiment. *Proceedings of the International Conference on Text, Speech and Dialogue, 7499* (pp. 166–173). Brno, Czech Republic. doi:10.1007/978- 3- 642- 32790- 2_20
- Hajiyeva, K. (2015). *A corpus-based lexical analysis of subject-specific university textbooks for english majors, 2*, 136–144. doi:https://doi.org/10.1016/j.amper.2015.10.001
- Hartman, J. J., Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1967). The General Inquirer: A Computer Approach to Content Analysis. *American Sociological Review*, 4. doi:10.2307/1161774
- Iglesias, C., & Moreno, A. (2019). Sentiment Analysis for Social Media. *Sentiment Analysis for Social Media*, 1–4. Retrieved from https://www.mdpi.com/journal/applsci/special
- Jahić, S., & Vičić, J. (2021). Determining sentiment of tweets using first Bosnian lexicon and (AnA)-affirmative and non-affirmative words. *Advanced technologies, systems, and applications V*, 142, 361–373. doi:https://doi.org/10.1007/978-3-030-54765-3_25
- Jahić, S., & Vičić, J. (2023a). *Lists of stopwords and AnAwords of Bosnian language (1.00) [Data set]*. doi:10.5281/zenodo.8021150
- Jahić, S., & Vičić, J. (2023b). Sentiment polarity lexicon of Bosnian language. 361–373. Univerza na Primorskem; CERN. Retrieved from https://zenodo.org/record/7520809#.Y8-4L3bMLi0
- Jahić, S., & Vičić, J. (2023c). Impact of Negation and AnA-Words on Overall Sentiment Value of the Text Written in the Bosnian Language. *Applied Science*, 13, 7760. doi:10.3390/app13137760
- Jones, R. L. (2006). An analysis of lexical text coverage in contemporary German. In *Brill, Language and Computers* (pp. 115–120). Leiden, The Netherlands: Brill. doi:https://doi.org/10.1163/9789401202213_010.
- Jovanoski, D., Pachovski, V., & Nakov, P. (2015). Sentiment analysis in Twitter for Macedonian. *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 249–257). Hissar, Bulgaria: INCO-MA Ltd. Shoumen. Retrieved from https://aclanthology.org/R15-1034

- Kadunc, K. (2016). *Določanje sentimenta slovenskim spletnim komentarjem s pomočjo strojnega*. Ljubljana: Fakulteta za računalništvo in informatiko Univerze v Ljubljani. Retrieved from <https://repozitorij.uni-lj.si/IzpisGradiva.php?lang=eng&id=91182>
- Kapukaranov, B., & Nakov, P. (2015). Fine-grained sentiment analysis for movie reviews in Bulgarian. *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 266–274). Hissar, Bulgaria: INCOMA Ltd. Shoumen. Retrieved from <https://aclanthology.org/R15-1036>
- Kia, D., Soujanya, P., Amir, H., Erik, C., Ahmad, H. Y., Alexander, G., & Qiang, Z. (2016). Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. *Springer Link – Cognitive Computation*, 8, 757–771. doi:10.1007/s12559-016-9415-7
- Kilgarriff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics*, 6(1), 97–133. doi:<https://doi.org/10.1075/ijcl.6.1.05kil>
- Ljubešić, N., & Klubička, F. (2014). bs,hr,srWaC - web corpora of Bosnian, Croatian and Serbian. *Proceedings of the 9th Web as Corpus Workshop (WaC-9)* (pp. 29–35). Gothenburg, Sweden: Association for Computational Linguistics. doi:10.3115/v1/W14-0405
- Moreno-Ortiz, A., & Pérez-Hernández, C. (2018). Lingmotif-lex: a wide-coverage, state-of-the-art lexicon for sentiment analysis. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 2653–2659). Miyazaki, Japan: European Language Resources Association (ELRA). Retrieved from <https://aclanthology.org/L18-1420>
- Okruhlica, A. (2013). *Slovak sentiment lexicon induction in absence of labeled data*, Master's Thesis. Comenius University Bratislava.
- Osmankadić, M. (2003). A Contribution to the Classification of Intensifiers in English and Bosnian. 50–62.
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. *Proceedings of the Workshop on Comparing Corpora WCC'00*. 9 (pp. 1–6). USA: Association for Computational Linguistics. doi:10.3115/117729.117730
- Suciati, A., & Budi, I. (2020). Aspect-Based Sentiment Analysis and Emotion. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 11(9), 179–186.
- Veselovská, K. (2013). Czech subjectivity lexicon : A lexical resource for czech polarity classification. *Proceedings of the 7th international conference Slovko* (pp. 279–284). Bratislava.

- Vičić, J. (2021). Bosnian news corpus 2021. Retrieved from <http://hdl.handle.net/11356/1406>
- Wawer, A. (2012). Extracting emotive patterns for languages with rich morphology. *International Journal of Computational Linguistics and Applications*, 11–24.
- Wu, F., Shi, Z., Dong, Z., Pand, C., & Zhang, B. (2020). Sentiment Analysis of Online Product Reviews Based On SenBERT-CNN. *International Conference on Machine Learning and Cybernetics (ICMLC)* (pp. 229–234). Adelaide, Australia: IEEE. doi:10.1109/ICMLC51923.2020.9469551

Razpoloženjsko označeni leksikon v bosanskem jeziku

Prispevek predstavlja prvi razpoloženjsko označeni leksikon bosanskega jezika. Postopek in metodologija označevanja sta predstavljena skupaj s študijo uporabnosti, ki se osredotoča na jezikovno pokritost. Sestava izhodišča je bila izvedena s prevajanjem slovenskega označenega leksikona in kasnejšim ročnim preverjanjem prevodov in oznak. Jezikovna pokritost je bila preverjana z uporabo dveh referenčnih korpusov. Bosanski jezik še vedno velja za jezik z malo jezikovnimi viri. Za bosanski jezik je na voljo referenčni korpus, ki ga sestavljajo samodejno preiskane spletne strani, vendar so avtorji ugotavljamo, da korpus z jasnim časovnim okvirom vsebnega besedila ni dosegljiv. Z zbiranjem novic z več bosanskih spletnih portalov je bil sestavljen korpus sodobnih besedil. V raziskavi sta bili uporabljeni dve metodi jezikovnega pokrivanja. Pri prvi je bil uporabljen frekvenčni seznam vseh besed, ekstrahiranih iz dveh referenčnih korpusov bosanskega jezika, druga metoda pa je prezrla frekvence kot glavni dejavnik pri štetju. Izračunana pokritost po prvi predstavljeni metodi za prvi korpus je bila 19,24 %, drugi korpus pa 28,05 %. Druga metoda daje 2,34 % pokritost za prvi korpus in 6,98 % za drugi korpus. Rezultati študije predstavljajo jezikovno pokritost, ki je primerljiva s znanimi metodami na tem področju. Uporabnost leksikona je bila dokazana že s primerjavo na Twitterju.

Ključne besede: Bosanski leksikon, korpus, analiza sentimenta, potrtilne in nepotrtilne besede (PnPbesede), ustavne besede, logaritemska verjetnost, označevanje

34. evropska poletna šola logike, jezika in informatike ESSLLI 2023

Matej KLEMEN,¹ Aleš ŽAGAR,¹ Timotej KNEZ,¹ Frenk DRAGAR,³
Marko ROBNIK-ŠIKONJA,¹ Slavko ŽITNIK,¹ Magdalena GAPSA,²
Mojca BRGLEZ,² Katarina BREZOVAR,² Špela VINTAR²

¹ Fakulteta za računalništvo in informatiko, Univerza v Ljubljani

² Filozofska fakulteta, Univerza v Ljubljani

³ Fakulteta za naravoslovje, Univerza v Utrechtu

1 Uvod

Med 31. julijem in 11. avgustom 2023 smo na Fakulteti za računalništvo in informatiko v Ljubljani organizirali poletno šolo *European Summer School of Logic, Language and Information* (ESSLLI). Dvotedenska poletna šola s tečaji, delavnicami in večernimi predavanji omogoča interdisciplinarno povezovanje udeležencev s področij logike, jezika in informatike. Za organizacijo 34. izvedbe poletne šole smo skrbeli sedemčlanski organizacijski odbor, dva prostovoljca in t. i. zaupna oseba¹ pod vodstvom Slavka Žitnika, za sestavo programa pa programski odbor, v katerem sta bila Juha Kontinen in Kaja Dobrovoltjc. Poletna šola je bila skoraj v celoti izvedena v živo.² Privabila je 318 udeležencev iz 37 držav, med njimi največ študentov in drugega akademskega osebja, šole pa se je udeležilo tudi nekaj predstavnikov iz zasebnega sektorja. Predavalo je skupno 93 predavateljev in vabljenih govorcev iz 23 držav.

1 Organizacijsko ekipo sestavljamo vsi soavtorji tega poročila.

2 Izjemoma smo po predhodnem dogovoru oddaljeno predstavljanje omogočili vabljenim govorcem in avtorjem prispevkov na delavnicah in študentski sekciji ter dvema predavateljema, ki se zaradi višje sile poletne šole nista mogla udeležiti v živo.

Klemen, M. et al.: 34. evropska poletna šola logike, jezika in informatike ESSLLI 2023. *Slovenščina* 2.0, 11(2): 84–91.

1.19 Recenzija, prikaz knjige, kritika / Review, book review, critique

DOI: <https://doi.org/10.4312/slo2.0.2023.2.84-91>

<https://creativecommons.org/licenses/by-sa/4.0/>



Poleg glavnega programa je poletna šola vključevala tudi nekaj bolj sproščenih dogodkov, katerih cilj je bil spodbuditi mreženje med udeleženci: uvodno pogostitev, večerjo za predavatelje, študentsko zabavo za udeležence, izlete po Sloveniji in tradicionalno nogometno tekmo, v kateri s(m)o sodelovali tako udeleženci kot organizatorji poletne šole. V nadaljevanju se bomo osredotočili na opis vsebinskega programa poletne šole.

2 Program

Program poletne šole je bil sestavljen iz tečajev, delavnic, študentske sekcije in (tudi širši javnosti odprtih) večernih predavanj.

2.1 Tečaji

V dveh tednih poletne šole je bilo na voljo 49 različnih tečajev, vsebinsko razdeljenih na tri interdisciplinarna področja: jezik in logika (16 tečajev), jezik in računalništvo (16 tečajev) ter logika in računalništvo (17 tečajev). Polovica tečajev je potekala v prvem tednu, polovica pa v drugem tednu poletne šole, pri čemer je večina izvajalcev organizirala en tečaj, pet pa jih je organiziralo dva tečaja ali tečaj in delavnico. V nadaljevanju na kratko opišemo vsebinska področja poletne šole, naslove vseh tečajev pa podamo tudi v Prilogi 1.

Prvi sklop tečajev povezuje področji jezika in logike. Za ta sklop je značilno, da jezikovne prvine obravnava simbolno preko logičnih pravil in matematičnih izrazov ne glede na konkretno vsebino izrekov. Na tečajih so bile obravnavane teme analize jezikovnih struktur (npr. tečaj *Current topics in the semantics and pragmatics of plural expressions* in tečaj *Semantic properties and combinatorial restrictions of attitude predicates*) in razčlenjevanje pomena fraz (tečaj *Decomposing the meaning of modals*). Nekateri tečaji so se osredotočili tudi na logične operacije (tečaj *Logical Operations Among Conditionals as Conditional Random Quantities*).

Področje jezik in računalništvo vsebuje tečaje, ki se vsebinsko nahajajo na spektru med jezikoslovjem in informatiko. Konkretnije ti tečaji ponavadi obravnavajo raziskovanje aktualnih jezikoslovnih tem s pomočjo računskih metod. Na letošnjih tečajih so bile na primer obravnavane teme strojnega prevajanja (tečaj *Neural Machine Translation*),

učenja jezikovnih modelov neposredno iz govornih posnetkov (tečaj *Deep Language Learning: Modeling language from raw speech*) in detekcija semantičnih premikov (tečaj *Computational approaches to semantic change detection*). Nekateri tečaji so se osredotočili tudi na proces pred uporabo računskih metod – na primer problem subjektivnosti označevanja in različnega razumevanja preučevanih pojavov (tečaja *Limitations in NLP: Disagreements, Misunderstandings, and other Challenges in Data Perspectivism in Computational Linguistics*). Izjema je bil tečaj *Computational Game Theory in Julia*, ki se je oddaljil od osrednje teme področja in je obravnaval koncepte teorije iger in njihovo implementacijo v programskem jeziku Julia.

Področje logika in računalništvo je obravnavalo abstraktnejše teme v povezavi z računalništvom in izračunljivostjo. Tečaji so obravnavali teme, kot so dokazovanje (tečaji *Proofs and explanations, An Introduction to Proof Equivalence in Proof theory of arithmetic*), epistemologijo (tečaja *Tools for Formal Epistemology: Doxastic Logic in Probability and Default Logic, Quantification in Dynamic Epistemic Logic*) in lambda račun (tečaja *Time and Space for the lambda Calculus in The lambda-calculus: from simple types to non-idempotent intersection types*). Pozornost je bila posvečena tudi specifičnim aplikacijam, tečaj *Logics for Safe AI* je na primer predstavil uporabo logike za ustvarjanje dokazljivo pravih sistemov umetne inteligence.

Ker je udeležencem na poletni šoli omogočena prosta izbira in menjava tečajev, o uradnih statistikah udeležbe po posamičnih aktivnostih ni mogoče poročati. Vsi tečaji so bili sicer zmerno obiskani, zaradi velikega števila udeležencev na nekaterih pa smo med poletno šolo te morali premakniti v večje predavalnice. Če sklepamo po teh podatkih, je bilo največ zanimanja za tečaje s področja jezik in računalništvo, zlasti za tiste, ki so obravnavali teme, povezane z nevronskimi mrežami in globokim učenjem.

2.2 Delavnice

Vzporedno s tečaji poletne šole so potekale še štiri fokusne delavnice, po dve vsak teden. Njihov program je v skupnem seštevku sestavljalo 12 vabljenih predavanj in 31 predstavitev sprejetih prispevkov.

V prvem tednu sta potekali delavnici *Procedural and computational models of semantic and pragmatic processes* v organizaciji Sonie Ramotowske in Fabiana Schlotterbecka ter delavnica *Internal and external pressures shaping language*, ki jo je organizirala Milica Denić. Prva je naslavljala združevanje statističnih pristopov in modelov kognitivnega procesiranja za odkrivanje in razlaganje mišljenega pomena izrekov, druga pa se je osredotočala na notranje (npr. kognitivne) in zunanje (npr. zgodovinske in geografske) dejavnike, ki vplivajo na oblikovanje in raznolikost jezikov.

V drugem tednu je potekala delavnica *First-order Modal and Temporal Logics: state of the art and perspectives*, ki sta jo vodila Valentin Goranko in Dmitry Shkatov, ter dvodnevna delavnica *Modalities in substructural logics: Applications at the interfaces of logic, language and computation*, ki sta jo organizirala Michael Moortgat in Mehrnoosh Sadrzadeh. Na prvi delavnici so se posvečali združevanju filozofskih in računalniških pristopov, ki za presojanje o nujnosti, možnosti in časovnosti uporabljajo logiko prvega reda, ter snovali nadaljnje korake za razvoj področja. Na drugi, nekoliko krajši delavnici so naslavljali modalnost s pristopi neklasične logike in predstavljali primere tovrstnih študij na preseku logike, jezika in računalništva.

2.3 Študentska sekcija

Vsak dan je bila ena ura programa poletne šole namenjena študentski sekciji, ki sta jo vodila Omri Doron in Janek Guerrini, na njej pa je 24 študentov predstavljalo svoje raziskovalno delo. Osem dni je bilo namenjenih petnajstim študentskim predstavitvam, dva dneva pa predstavitvam osmih posterjev.

Obiskovalci so slišali o raznolikih temah na preseku logike, jezikoslovja in računalništva, na primer o analizi metafor (*Towards a Definition of Novel Metaphor for Automatic Metaphor Detection*), odvisnosti logiki (*Implementation of Elements of Dependence*) in nevronske modelih (*Neural Models for Syllogistic Logic*). Govora je bilo tudi o vse večji uspešnosti velikih jezikovnih modelov in morebitnih posledicah za raziskave doktorskih študentov (*Troubling times for PhD research on text categorization? ChatGPT for Automatic Genre Identification*).

Zadnji dan sta vodji podelila nagrado za izstopajoče prispevke študentov. Nagrado za najboljšo predstavitev je prejela Jia Ren za prispevek *Weak Reciprocity without the Cumulative Operator*, nagrado za najboljši poster pa Valeria Gradimondo za prispevek *Negation in Italian until-clauses: effective or expletive?*.

2.4 Večerna predavanja

Štiri večere so na poletni šoli potekala večerna predavanja, kjer so strokovnjaki s področja logike, jezika in informatike predavali o aktualnih temah s področij, na katerih delujejo.

Marko Robnik-Šikonja je v prvem večernem predavanju z naslovom *Large Language Models for Cross-lingual Transfer* predstavil razvoj velikih jezikovnih modelov vse od začetnih transformerskih modelov do trenutnih ogromnih modelov, kot je na primer ChatGPT³. Osredotočil se je na problem neenakomerne dosegljivosti virov v različnih jezikih in predstavil medjezikovni prenos kot možno rešitev.

O jezikovnih modelih je predavala tudi Malvina Nissim v drugem večernem predavanju *Language Technology <preposition> Society*. Osredotočila se je na vpliv, ki ga imajo jezikovni modeli na družbo, in vpliv družbe na jezikovne modele. Veliko pozornosti je namenila neželjenim pristranskostim modelov ter vprašljivi uspešnosti tehnik njihovega odstranjevanja.

Tretje večerno predavanje *The Cost of the Lambda Calculus and the Semantics of Sharing* se je vsebinsko oddaljilo od prejšnjih jezikovno usmerjenih predavanj. V prvem delu je Beniamino Accattoli predstavil področje modelov z obvladljivo kompleksnostjo za lambda račun, v drugem pa razširitve lambda računa z deljenjem prvega reda.

V četrtem večernem predavanju *The Good, the Bad, and the Ugly of Language Technology Infrastructure*, ki je bil posvečen spominu na preminulega profesorja Richarda Oehrla, je Darja Fišer predstavila infrastrukturo za trajnostno shranjevanje in deljenje virov CLARIN, ki sledi načelom FAIR (*findability, accessibility, interoperability, reusability*). Poleg same infrastrukture je predstavila tudi nekaj raziskav citiranja in dostopnosti raziskovalnih podatkov, poudarila pomembnost deljenja virov in podala smernice, kako skrbno deliti svoje podatke.

3 <https://chat.openai.com/> (8. 9. 2023)

3 Sklep

Po 12 letih smo poletno šolo ESSLLI ponovno organizirali v Ljubljani. Dvotedenska poletna šola je z raznoliko ponudbo predavanj, delavnic, večernih predavanj in družabnih dogodkov privabila mednarodno publiko in poskrbela za ustvarjanje novih in krepitev obstoječih povezav ter deljenje idej na preseku področij logike, jezika in informatike. Naslednje leto se poletna šola seli v Belgijo, 35. izvedba poletne šole bo namreč potekala v Leuvnu.

3.1 Priloga A: Seznam tečajev poletne šole

Seznam tečajev prvega tedna poletne šole je prikazan v Tabeli 1, tečajev drugega tedna poletne šole pa v Tabeli 2.

Tabela 1: Tečajji prvega tedna poletne šole

Jezik in logika		Jezik in računalništvo		Jezik in logika	
Trivalent and Dynamic Theories of Conditionals	Semantic properties and combinatorial restrictions of attitude predicates	Neural Machine Translation	Deep Language Learning: Modeling language from raw speech	Elements of Kleene Algebra	The Safety Fragment of Temporal Logics of Infinite Sequences
<i>Lorenzo Rossi, Paolo Santorio</i>	<i>Deniz Özyıldız, Ciyang Qing</i>	<i>Antonio Toral, Arianna Bisazza</i>	<i>Gašper Beguš</i>	<i>Tobias Kappé</i>	<i>Luca Geatti, Angelo Montanari</i>
Formal Theories of Properties (Salvatore Florio, Carlo Nicolai)	Logical Operations Among Conditionals as Conditional Random Quantities	Tree-Adjoining Grammars: Theory and implementation	Data Perspectivism in Computational Linguistics	Proofs and explanations	Linear arithmetic theories: algorithms and applications
<i>Salvatore Florio, Carlo Nicolai</i>	<i>Giuseppe Sanfilippo</i>	<i>Kata Balogh, Simon Petijtjean</i>	<i>Enrica Troiano, Valerio Basile</i>	<i>Francesca Poggiolesi</i>	<i>Alessio Mansutti, Christoph Haase</i>

Jezik in logika		Jezik in računalništvo		Jezik in logika		
Let's talk about Frodo: An Introduction to the Semantics of Fiction (<i>Merel Semeijn, Louis Rouillé</i>)	Current topics in the semantics and pragmatics of plural expressions	Natural Language Syntax and Statistical Semantics with Modal Lambek Calculus	Limitations in NLP: Disagreements, Misunderstandings, and other Challenges	Tools for Formal Epistemology: Doxastic Logic, Probability and Default Logic	First-order Modal and Temporal Logics: Philosophical and Computational Aspects	
<i>Merel Semeijn, Louis Rouillé</i>	<i>Keny Chatain, Benjamin Spector</i>	<i>Mehrnoosh Sadrzadeh, Gijs Wijnholds</i>	<i>Michael Roth</i>	<i>Aleks Knoks, Eric Pacuit</i>	<i>Valentin Goranko, Dmitry Shkatov</i>	
Propositional Quantifiers (<i>Peter Fritz</i>)	The Semantics of Reciprocity	Computational Game Theory in Julia	Formal Techniques for Neural-symbolic Modeling	Possibility Semantics	Logics for Safe AI	Logics with Probability Operators and Quantifiers
<i>Peter Fritz</i>	<i>Yoad Winter</i>	<i>Eric Pacuit</i>	<i>Kyle Richardson, Vivek Srikumar</i>	<i>Wesley Holliday</i>	<i>Brian Logan</i>	<i>Nebojša Ikodi-nović, Dragan Doder</i>

Tabela 2: Tečaji drugega tedna poletne šole

Jezik in logika		Jezik in računanje		Jezik in logika	
Common ground*	Probability logic, language, and cognition	Treebanking: methodology, tools and applications	Formal Language Theory and Neural Networks	Quantification in Dynamic Epistemic Logic	Time and Space for the lambda Calculus
<i>Bart Geurts</i>	<i>Niki Pfeifer</i>	<i>Bruno Guillaume, Kim Gerdes</i>	<i>Ryan Cotterell</i>	<i>Rustam Galmullin, Louwe B. Kuijer</i>	<i>Beniamino Accattoli</i>
Communitarian Semantics	Monotonicity: Grammar, Processing, and Neural Reflections	Introduction to Linguistic Data Science	Computational approaches to semantic change detection	An Introduction to Proof Equivalence	Logic, Data Examples, and Learning
<i>Elin McCready, Grégoire Winterstein</i>	<i>Luka Crnic, Yosef Grodzinsky</i>	<i>John P. McCrae</i>	<i>Lidia Pivovarova, Andrey Kutuzov</i>	<i>Matteo Acclavio, Paolo Pistone</i>	<i>Balder ten Cate, Carsten Lutz</i>

Jezik in logika		Jezik in računanje		Jezik in logika	
Decomposing the meaning of modals	The semantics of visual communication. Theoretical approaches to visual meaning aspects in co-speech gestures and sign language	Natural language syntax: parsing and complexity	A Uniform Meaning Representation for NLP Systems	The lambda-calculus: from simple types to non-idempotent intersection types	Proof theory of arithmetic
<i>Annemarie van Dooren and Anouk Dieuleveut</i>	<i>Cornelia Ebert and Markus Steinbach</i>	<i>Timothee Bernard and Pascal Amsili</i>	<i>Martha Palmer and James Pustejovsky</i>	<i>Giulio Guerrieri</i>	<i>Anupam Dwas</i>
Advanced Topics in Formal Semantics Based on Modern Type Theories	Explaining anaphoric accessibility: navigating non-veridical environments in dynamic semantics	Computational Creativity	The probabilistic Language of Thought	Conditional logics of preference: how to make the best choice	Logics of dependence and independence
<i>Zhaohui Luo</i>	<i>Patrick Elliott, Lisa Hofmann</i>	<i>Tim Van de Cruys</i>	<i>Fausto Carracci, Michael Franke</i>	<i>Louwe B. Kuijer</i>	<i>Fan Yang</i>

Opomba. Tečaj, označen z *, je bil odpovedan.

DHH23 – Helsinki Digital Humanities Hackathon 2023

David BORDON

Univerza v Ljubljani, Filozofska fakulteta

Maja 2023 je v Helsinkih potekal humanistični *hackathon*.

Ob omembi besede *hackathon* se marsikomu pred očmi izriše slabo prezračevan kletni prostor, kamor se čez konec tedna – namesto da bi preživeli nekaj časa na svežem zraku, občasno oplazili kakšno zelenico in nadomestili kronično pomanjkanje vitamina D – zapre nekaj programerjev ali razvijalcev videoiger in hudo neprespani po četrtem obroku, ki ga predstavlja pica, kot vse obroke prej, skuša ugotoviti, zakaj je VS Code tako zelo poln klicajev. Klasičen primer, ko si preobremenjeni ljudje vzamejo odmor od »cruncha«¹ z *drugačnim* »crunchem«.

V Helsinkih je potekal, kot omenjeno, humanistični hackathon, kar je nekoliko drugačna zadeva. Klet je bila prostornejša, svetlejša in dobro prezračevana, izbor hrane je bil bogat, neprespanost je bila posledica posameznikove neodgovornosti, delo je potekalo ob rednem delovnem času, konec tedna pa je bil popolnoma prost. Ukvarjali smo se s tem, da smo na nekoliko bolj igriv način izvedli znanstveno raziskavo, ki je povezovala področji humanistike in računalništva, končni rezultat pa je bil konferenčni poster, ki smo ga predstavili na zaključnem dogodku.

1 Obdobje zelo intenzivnega dela, ki se nikoli ne konča.

Bordon, D.: DHH23 – Helsinki Digital Humanities Hackathon 2023. Slovenščina 2.0, 11(2): 92–101.

1.19 Recenzija, prikaz knjige, kritika / Review, book review, critique

DOI: <https://doi.org/10.4312/slo2.0.2023.2.92-101>

<https://creativecommons.org/licenses/by-sa/4.0/>



Po tem splošnem orisu se lahko posvetimo temu, kaj Helsinki Digital Humanities Hackathon sploh je. Dogodek ima že močno tradicijo – prvič so ga priredili leta 2015, od tedaj (z izjemo 2020) poteka letno pod okriljem Univerze v Helsinkih in v praksi demonstrira njihov pristop k poučevanju digitalne humanistike, kjer je poudarek na posameznikih in njihovih veščinah, obenem pa gre za poligon, kjer lahko ti posamezniki razvijejo skupni jezik in plodno sodelujejo na skupnem projektu. Hackathon soorganizira Univerza Aalto, dodatno podporo pa nudita konzorcija CLARIN in DARIAH, ki sta denimo letos krila stroške potovanja in namestitve za 20 udeležencev.

Aktivnosti potekajo v »kleti« centra Oppimiskeskus Minerva (Učni center Minerva), v kateri je poleg skupne dvorane z delovnimi mizami na voljo še pet dobro opremljenih zastekljenih »učilnic«, kjer lahko skupine delajo na svojih projektih. Zgoraj, v pritličju, je menza, s katero upravlja lokalna študentska organizacija – iz pogovora z enim od koordinatorjev naše ekipe sem izvedel (upam, da nisem česa razumel narobe), da se študentska organizacija helsinške univerze zelo dobro hrani z oddajanjem zemljišč ob glavni železniški postaji, kjer bi (preden je zrasla postaja) univerza morala imeti svoj kampus, sedaj pa so tam trgovine in poslovalnice multinacionalnk, ki plačujejo neskromno najemnino. Dobiček gre tudi v sistem subvencionirane prehrane za študente, koncept, ki si ga na domači grudi, kjer imajo akterji tipa ŠOU povsem drugačne apetite, zelo težko predstavljamo.

Subvencije za prehrano so veljale tudi za vse udeležence dogodka, poleg tega pa so nam bile na voljo neomejene količine kave ali čaja – še več – vzporedno je potekalo tudi tekmovanje, kdo bo konzumiral največ omenjenih stimulantov, na koncu pa razglasitev obeh zmagovalcev, nevrotično tresočega Italijana in temperamentnega Grka, ki sta na glavo spila okrog 5 skodelic dnevno.

Hackathon je potekal 10 dni, od tega sta bila dva dneva, sobota in nedelja, popolnoma prosta. Nekateri smo ta čas izkoristili za dnevno potepanje po Helsinkih in obisk estonskega Tallina, ki je zgolj prijetno uro plovbe stran. Vsekakor, če je možnost, priporočam, da si bodoči udeleženci po dogodku vzamejo še kak dodaten dan dopusta, saj je mesto čudovito in kulturna ponudba izvrstna.

Namen dogodka je spodbujanje interdisciplinarnega sodelovanja med humanisti in računalničarji, o katerem je eden izmed organizatorjev, Eetu Mäkelä, govoril na vabljenem predavanju konference JTDH 2022². Obenem je dogodek leta 2021 zaradi oteženih pogojev potovanja potekal na spletu, organizatorji pa so s tem pridobili ogromno povratnih informacij, kako dogodek optimalno zastaviti v bodoče. O tem je na TwinTalks delavnici na konferenci DH23 Graz govoril še en organizator, Mikko Tolonen, čigar prispevek si lahko ogledate na YouTube kanalu CLARIN ERIC³.

Formalno se je hackathona udeležilo približno 60 oseb, ki so bile razdeljene v štiri različne skupine. Vsaka skupina je dobila podatkovno zbirko, na podlagi katere so osnovali znanstveno raziskavo o določeni temi. Te so bile sledeče:

- Metapodatki pisem in razumevanje družbe
- Interakcijska dinamika spletnega diskurza
- Zgodnje novoveške znanstvene publikacije
- Politična polarizacija v parlamentu

Slovenijo smo, poleg pisca tega prispevka, predstavljali še Katja Meden (INZ), Vid Klopčič (FRI), ukrajinska sodelavka na INZ Anna Kryvenko in v vlogi enega izmed vodij naše tematske skupine makedonski kolega Bojan Evkoski, takrat še raziskovalec na INZ – vsi smo bili del parlamentarne skupine, ki je raziskavo oblikovala okrog podatkovne zbirke ParlaMint.

Pred samim dogodkom smo imeli dve uvodni srečanja na spletu, kjer smo se spoznali med seboj, s podatkovno zbirko, postavili okvirna raziskovalna vprašanja in pretehtali metodološke pristope. V vmesnem času je računalniški del skupine že izvajal predprocesiranje podatkov, humanisti pa so zbirali gradivo za potencialni teoretski okvir.

Dogodek je bil izvrstno organiziran. Prvi dan so nas odgovorni seznanili s pogoji dela, roki za predstavitev napredka in vmesnih rezultatov, stranskimi projekti, družabnimi dogodki in večkrat poudarili, na koga se lahko obrnemo, če se čutimo v stiski zaradi kakršnih koli razlogov. Za tem smo pričeli z delom, ki je potekalo neprekinjeno vse

2 <https://www.sdjt.si/wp/dogodki/konference/jtdh-2022/zbornik/#vabljena>

3 https://youtu.be/kTjahw2q_5g?si=SmbgPEA0-z4w58Xe&t=9223

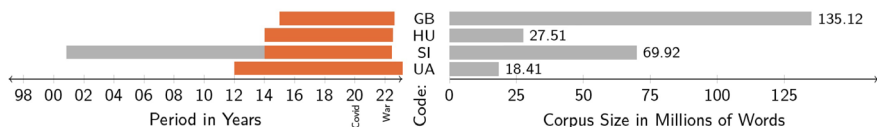
do predstavitve posterjev. Vzporedno smo imeli še nekaj »stranskih projektov«, kjer smo na nekoliko bolj eksperimentalen in umetniški način s popolnoma prostimi rokami uporabljali generativno umetno inteligenco za različne namene (recimo, multimedialno prevajanje poezije). Poleg dela v skupinah je bilo nekaj segmentov vezanih na sodelovanje med skupinami, denimo, letos smo vzpostavili kanal na družbenih omrežjih, s katerim upravljajo sodelujoči in ga bo v bodoče možno predajati naslednjim generacijam.

Sodelovanje s člani drugih skupin je bilo omogočeno v skupnem centralnem prostoru, kjer so bili delovni otoki. Najbolj plodne debate so bile na sporedu po koncu delovnega dneva, ko smo se zbrali tam, kjer je *žejni učenjak* doma – prostor, kjer smo tudi poglobili obstoječe vezi in sklenili nove.

Če se sedaj posvetim dejanskemu delu parlamentarne skupine – ukvarjali smo se z vprašanjem polarizacije v parlamentih. Naša raziskava se je osredotočila na odgovore na ključna vprašanja, povezana s polarizacijo, kot so:

- 1) Kako je mogoče meriti polarizacijo z računalniškimi metodami?
- 2) Kako se polarizacija kaže skozi čas?
- 3) Kako določene teme polarizirajo različne parlamente?

Za merjenje polarizacije, opazovanje njenega izražanja skozi čas in analizo, kako določene teme polarizirajo parlamente, smo uporabili računalniške metode in podatkovno zbirko ParlaMint 3.0 (različica beta), ki jo je predhodno izdal CLARIN ERIC. Gre za primerljiv večjezični korpus parlamentarnih razprav. Izbrali smo štiri države: **Združeno kraljestvo, Madžarsko, Ukrajino in Slovenijo** v spodaj navedenih obdobjih (Slika 1).



Slika 1: Levo: časovni obseg korpusa besedil, ki smo jih vključili v raziskavo (oranžno). Desno: velikost korpusov v milijonih besed. Vertikalno v sredini: država.

Naša metodologija je bila razdeljena na tri glavne dele:

- 1) Tematske podkorpuse
- 2) Numerične reprezentacije govorov
- 3) Merjenje polarizacije

Za izbiro specifičnih tem iz parlamentarnih govorov smo uporabili multidisciplinarni pristop, pri čemer smo izločili ključne besede za cepitev korpusov ter uporabili vektorsko vložitev ključnih besed po metodi LDA. Določili smo tudi pomembne leme in besedne zveze za vsako temo, pri čemer smo upoštevali generičnost in metaforično rabo. Prispevke in njihov vpliv na polarizacijo smo analizirali s filtriranjem govorov na podlagi teh ključnih besed ter jih ovrednotili s pomočjo »matrike zmede« (confusion matrix) in števila govorov (Tabela 1).

Tabela 1: Število govorov za določeno podtemo, pridobljenih s ključnimi besedami, ki so bile vključene v analizo

Tema	SI	GB	HU	UA
EU	13943 (11 %)	38938 (10 %)	5933 (14 %)	5863 (8,4 %)
vojna	8686 (7,0 %)	15039 (4.0 %)	2141 (5.0 %)	7101 (10 %)
zdravstvo	13802 (11 %)	45022 (12 %)	4583 (11 %)	4546 (6.5 %)

Opomba. V oklepaju: procent pojavnosti v celotnem korpusu.

Naš cilj je bil ustvariti primerljive reprezentacije večjezičnih govorov. Uporabili smo model SBERT za vektorske vložitve govorov, kar nam je omogočilo generiranje visokodimenzionalnih vektorjev za optimizirane semantične primerjave. Poleg tega smo z ekstrakcijo sentimenta s pomočjo regresije na podlagi modela RoBERTa določili vrednosti sentimenta za vsak govor.



Slika 2: Ocena sentimenta za štirimesečja. Oranžna – pozitiven sentiment; modra – negativen sentiment.

Za vrednotenje polarizacije smo uporabili štiri tehnike, in sicer vložitve razlik med nasprotnimi strankarskimi skupinami, analizo sentimenta in primerjavo nasprotnih poudarkov na podtemah BERTopic. Najprej smo preučili razlike v vložitvah govora (vektorjih) med nasprotnimi strankarskimi skupinami. To nam je pomagalo razumeti, kako različne stranke izražajo svoja stališča in obseg polarizacije med njimi. Izvedli smo tudi analizo sentimenta, da bi ugotovili splošno razpoloženje političnih strank. S primerjavo nasprotujočih si stališč o določenih podtemah smo dobili vpogled v njihovo polarizacijsko naravo. Nazadnje smo izvedli diahrono analizo, da bi preučili, kako se je polarizacija spreminjala skozi čas, in sicer z analizo sentimenta in vložitev.

Naše delo je privedlo do več pomembnih ugotovitev:

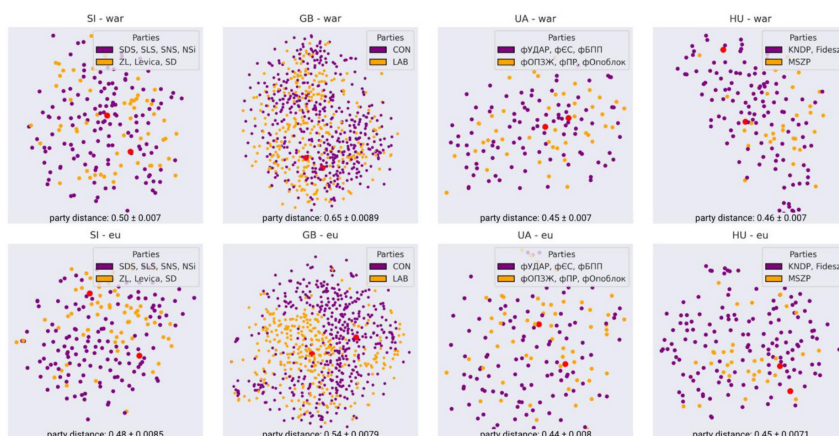
- **Grozdjenje tem:** Z uporabo tehnike BERTopic smo iz krovnih tem uspešno izluščili podteme. To nam je omogočilo, da smo se poglobili v govore in primerjali usmeritve zgodovinsko nasprotujočih si strank v posameznih državah. S prepoznavanjem in združevanjem teh podtem smo celovito razumeli vprašanja, ki prispevajo k polarizaciji parlamentarnih razprav (Tabela 2).

Tabela 2: Izbrane podteme BERTopic, ki poudarjajo razkol med nasprotujočimi si političnimi skupinami

Great Britain			
Theme	Topic	Focus % (CON / LAB)	Sentiment (CON / LAB)
EU	Brexit Referendum	11.1 / 17.6	0.10 / -0.08
War	Ukraine-Russia War	13.4 / 10.6	-0.08 / -0.30
Healthcare	Covid	30.1 / 22.5	0.17 / -0.35
Ukraine			
		Focus% (Pro-UA* / Pro-RU*)	Sentiment (Pro-UA* / Pro-RU*)
EU	Language Policy	10.30 / 47.50	-0.34 / -0.55
War	Legislations in War	9.60 / 22.40	-0.41 / -0.58
Healthcare	Organ Transplantation	16.80 / 1.10	0.03 / -0.88

		Hungary	
		Focus % (Fidesz-KDNP / MSZP)	Sentiment (Fidesz-KDNP / MSZP)
EU	Corruption Charges	16.70 / 30.9	-0.38 / -0.48
War	Constitution Defense	13.4 / 29.5	-0.32 / -0.52
Healthcare	Covid	24.3 / 9.5	-0.13 / -0.52
		Slovenia	
		Focus % (SDS / SD)	Sentiment (SDS / SD)
EU	Tax Coffers	6.90 / 10.8	-0.04 / 0.02
War	Veteran Pensions	2.2 / 5.0	-0.30 / -0.32
Healthcare	Healthcare System	14.2 / 8.0	-0.26 / -0.48

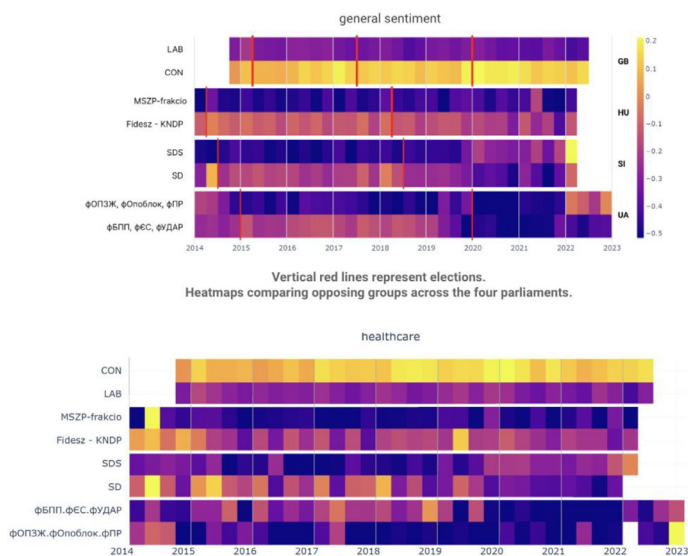
- Osi polarizacije:** Z zmanjšanjem dimenzionalnosti in izrisom t-SNE smo vizualizirali odnose med govorniki iz različnih strank znotraj vsake države in posamezne teme. Ta tehnika vizualizacije je razkrila jasne vzorce in delitve, zlasti v Veliki Britaniji, ko smo vektorje z visoko dimenzionalnostjo pretvorili v dvodimenzionalne vložitve; prepoznali smo osi polarizacije, ki kažejo, v kolikšni meri so bile stranke pri določenih vprašanih razdeljene. Poleg tega je analiza sentimenta razkrila polarizacijo med prevladujočimi strankami, pri čemer je bil pozitiven sentiment večinoma opažen v koalicijah, negativen pa v opozicijskih skupinah. To je pokazalo velike razlike v pogledih in stališčih med političnimi skupinami, ki smo jih vključili v raziskavo (Slika 3).





Slika 3: Osi polarizacije.

- Sentiment političnih strank:** V naši raziskavi smo poleg ugotavljanja polarizacije na širši ravni raziskali tudi polarizacijo čustev pri posameznih temah. Ugotovili smo, da so različna vprašanja, vključno z zdravstvom, vojno in Evropsko unijo, pokazala pomembno polarizacijo sentimenta med političnimi strankami. S to ugotovitvijo smo potrdili kontroverznost tem ter različna stališča in mnenja, ki jih imajo različne stranke. Razumevanje čustev, ki se povezujejo s temi temami, je ključno za razumevanje globine polarizacije in njenega vpliva na parlamentarni diskurz (Slika 4).⁴



Slika 4: Analiza sentimenta. Primerjava dveh nasprotujočih si strank iz vključenih parlamentov.

- 4 Omejitev odgovornosti in obvestilo o avtorstvu: Vsebinsko tehničnega dela na straneh 96–99 je kolektivno spisala skupina za potencialno objavljane na blogih ali v poročilih, vodilno vlogo pri pisanju je imela kolegica Nikoleta Jablonczay. Avtor prispevka sem izvirno besedilo prevedel v slovenščino in ga rahlo prestrukturiral.

V sklopu raziskave smo naleteli na nekatere izzive, ki so nam ponudili priložnosti za nadaljnje raziskovanje:

- **Omejitve podatkov:** Različni obsegi podkorporusov so omejili možnost posploševanja modelov sentimenta in ovirali ponovljivost tematskega modeliranja.
- **Metodološke razlike:** Čeprav so pristopi, ki temeljijo na LLM (velikih jezikovnih modelih), ponudili dragocen vpogled, so razlike v razlagi tem predstavljale izziv za kvalitativno analizo.
- **Kvalitativna analiza:** Krepitev okvira kvalitativne analize po opravljeni raziskavi bi povečala zanesljivost naših metod.

Možnosti za nadaljnje raziskave vključujejo razvoj konkretne ontologije polarizacije za pomoč pri izbiri in prepoznavanju lastnosti, standardizacijo računalniških metodologij ter vključitev razumljivih metod umetne inteligence in »bele škatle« za boljšo interpretacijo, večjo transparentnost in naknadno analizo.

Raziskavo smo predstavili v obliki konferenčnega posterja⁵, na spletu pa so na voljo tudi posterji ostalih treh skupin^{6,7,8}.

Za zaključek pa nekaj splošnih nasvetov za bodoče udeležence – Skandinavija je, za slovenski standard, draga, predvsem kar se tiče nočitev. Priporočljiv je skupen najem nastanitve, kar so spodbujali tudi organizatorji s predčasno vzpostavitvijo temu namenjenega kanala na projektnem Slacku. Stroški prehranjevanja niso visoki, predvsem zaradi odlične menze v objektu Minerva, nestandardno dragi so samo potencialni *promilni priboljški*, za katere je v povprečju treba odšteti dvokratnik slovenske cene, je pa stranski učinek intelektualnega doprinosa ob spremljajočem druženju neprecenljiv. Helsinki so razmeroma majhna prestolnica, ki je obvladljiva peš, javni promet je izvrsten, na voljo so tramvaji in podzemna železnica, vozovnice pa je ob vstopu na postajo podzemne ali na tramvaj moč kupiti prek aplikacije HSL. Ob odsotnosti direktnih povezav je Helsinško letališče zelo dobro povezano z evropskimi vozlišči, karte pa ne bi smele biti predrage, če se potovanje načrtuje dovolj zgodaj.

5 <https://www.helsinki.fi/assets/drupal/2023-06/dhh23-parliament-poster.pdf>

6 <https://www.helsinki.fi/assets/drupal/2023-06/dhh23-disc-poster-comp1.pdf>

7 <https://www.helsinki.fi/assets/drupal/2023-06/dhh23-letters-poster.pdf>

8 <https://www.helsinki.fi/assets/drupal/2023-06/dhh23-earlymodern-poster.pdf>

In še čisto za konec – Helsinki Digital Humanities Hackathon je izvrsten dogodek, kjer posameznik spozna, da je delo v skupini, ki jo sestavljajo osebe iz drugačnih akademskih ozadij, ki imajo povsem drugačne pristope in tradicije, vznemirljivo. Obenem je ta pogled onstran svojega vrtilčka v zgodnji fazi raziskovalnega udejstvovanja izjemnega pomena, saj pri posamezniku pusti ihtečo željo po novih oblikah sodelovanja, in ne zapiranja v varnost lastnega balončka. Vsem, ki imate možnost, da se dogodka udeležite, svetujem, da spremljate relevantne spletne strani⁹ in v prihajajočih letih izkusite, kaj vse vam lahko da.

⁹ <https://www.helsinki.fi/en/digital-humanities>