Slovenščina 2.0

Let. 9 (2021), št. 2

Univerza v Ljubljani
FILOZOFSKA FAKULTETA

Slovenščina 2.0
Letnik/Volume 9, Številka/Issue 2, 2021

ISSN: 2335-2736

Založila/Published by
Znanstvena založba Filozofske fakultete Univerze v Ljubljani

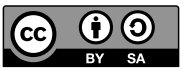
Izdal/Issued by
Center za jezikovne vire in tehnologije Univerze v Ljubljani

Za založbo/For the publisher
Mojca Schlamberger Brezar, dekanja Filozofske fakultete

Publikacija je brezplačna./Publication is free of charge.

Publikacija je dostopna na/Avaliable at: https://revije.ff.uni-lj.si/slovenscina2/index

Revija izhaja s podporo Javne agencije za raziskovalno dejavnost Republike Slovenije./
This journal is published with the support of the Slovenian Research Agency (ARRS).



KAZALO

SOCIOLINGVISTIČNI POSVET: AKTUALNI SOCIOLINGVISTIČNI IZZIVI IN PREDNOSTNE RAZISKOVALNE TEMATIKE

Maja BITENC, Marko STABEJ, Nataša GLIHA KOMAC, Matejka GRGIČ, Monika KALIN GOLOB, Karmen KENDA-JEŽ, Albina NEĆAK LÜK, Sonja NOVAK LUKANOVIČ, Krištof SAVSKI

Bitenc, M., Stabej, M., Gliha Komac, N., Grgič, M., Kalin Golob, M., Kenda-Jež, K., Nećak Lük, A., Novak Lukanovič, S., Savski, K.: Sociolingvistični posvet: aktualni sociolingvistični izzivi in prednostne raziskovalne tematike. Slovenščina 2.0, 9(2): 1–40.

DOI: https://doi.org/10.4312/slo2.0.2021.2.1-40

Posvet o aktualnih sociolingvističnih izzivih in prednostnih raziskovalnih tematikah sta organizirala doc. dr. **Maja Bitenc** in red. prof. dr. **Marko Stabej** z Oddelka za slovenistiko. Potekal je v ponedeljek, 27. 9. 2021, na Filozofski fakulteti Univerze v Ljubljani in s prenosom preko Zooma. V prvem delu so vabljene strokovnjakinje in strokovnjaki predstavili svoje poglede ob izhodiščnih vprašanjih, v drugem je sledila razprava vseh sodelujočih. Objavljeni zapis posnetka so govornice in govorniki uredili po lastni presoji, načeloma s čim manj intervencijami, iz razprave pa so za branje prilagojene in objavljene vsebinsko tehtnejše replike.

Maja Bitenc: Spoštovane in spoštovani, drage in dragi, prisrčno pozdravljeni in pozdravljene na sociolingvističnem posvetu, na dan po evropskem dnevu jezikov, v živo na Filozofski fakulteti in na daljavo na različnih delih Slovenije in sveta!

Ob lastnem iskanju sem se z nekaterimi od vas že mnogokrat pogovarjala o različnih sociolingvističnih izzivih, o tem, kaj in kako je oziroma bi bilo najbolj smiselno raziskovati, večkrat pa sem ob tem pomislila tudi, da bi bilo dobro, lepo in prav, da bi se enkrat srečali skupaj in v širšem krogu razmišljali o prednostnih raziskovalnih tematikah, kot jih doživljamo ob svojem delu. Ob pomladnem izidu monografije *Sociolingvistično iskrenje*, v kateri so predstavljeni

prispevki relevantnejših in tehtnejših raziskav, ki so nastale v procesu formalnega izobraževanja na Filozofski fakulteti, se nama je s soorganizatorjem današnjega srečanja, Markom Stabejem, zdelo, da je to pravi čas, pravi trenutek. Zelo naju veseli in sva hvaležna, da smo danes tukaj in da smo skupaj.

V prvem delu bodo torej na vrsti razmisleki vabljenih strokovnjakinj in strokovnjakov, v drugem delu pa bo čas za razpravo vseh navzočih. Morda se z Markom z namenom piševa Bitenc in Stabej, da imam po abecedi sama lahko tale uvodni nagovor, Marko pa bo, kot zadnji nastopajoči, zaokrožil prvi del.

Izhodiščna vprašanja za današnji posvet so torej:

Kaj so glavni aktualni izzivi za slovensko sociolingvistiko?

Kje in kaj se na področju sociolingvistike že dogaja in kako vrednotite to delo?

Katere nove oziroma dodatne raziskave bi bile najbolj potrebne in v kakšnem okviru bi jih bilo najbolj realno oziroma smiselno načrtovati?

Seveda so dobrodošli tudi drugi relevantni poudarki, morda tudi glede četrtega vprašanja, če bi kdo želel, čeprav je usmerjeno bolj aplikativno: Kaj v sociolingvističnem smislu najbolj potrebuje slovenska jezikovna oziroma govorna skupnost v vsej svoji raznolikosti?

Če mi dovolite še nekaj vsebinskih pomislekov. Z Markom Stabejem sva se v zadnjem letu, ko je zorelo *Sociolingvistično iskrenje*, ki je bilo zasnovano v počastitev 90. obletnice rojstva Brede Pogorelec, lotila tudi pregledovanja in urejanja sociolingvističnih spisov te profesorice, ki naj bi izšli v zbirki njenih zbranih del. Ob branju se pogosto čudim, kako aktualne so njene misli in kako posrečeno zaobjamejo sociolingvistično situacijo in dinamiko. Breda Pogorelec je že sredi 60. let izrazila pobudo za celovitejši opis govorjenega jezika oziroma govorov posameznih mestnih središč, kar se ji je zdelo posebej pomembno zaradi aktualnega uresničevanja norme formalnega jezika v praksi, predvsem v nekaterih medijih, kjer gre za posredno izvajanje jezikovnega načrtovanja, ter v osnovni in deloma srednji šoli. Prav v tistem času je William Labov, pionir sociolingvistike in proučevanja jezikovne variantnosti in jezikovnih sprememb, s svojo raziskavo dokazal, da so vzorci variantnosti, ki vodijo k fonološkim spremembam, v korelaciji s starostjo in družbenim razredom govorcev, obenem pa je obravnaval ozaveščenost njihovega govora.

Takrat so se na primer začele tudi raziskave mestnih govoric na Češkem in Slovaškem in v 90. letih že doživele tudi svojo ponovitev, torej dopolnitev s sodobnim primerjalnim gradivom. Metodologija slovenskih sociolingvističnih raziskav dejanske jezikovne rabe in jezikovnih stališč pa se je v glavnem razvijala na področju slovenščine v stiku z drugimi jeziki.

Večino od vas je profesorica Breda Pogorelec verjetno učila, nekateri od vas pa ste učili mene in sem prav od vas prejela to, kar v strokovnem smislu sem, za kar sem hvaležna. Po dodiplomskem študiju angleščine in slovenščine, kjer sem z zanimanjem spoznavala jezika na različnih ravninah, sem bila obdarjena z možnostjo nadaljnjega študija in raziskovanja v okviru statusa mlade raziskovalke pod mentorstvom Marka Stabeja. Sociolingvistika me je hitro očarala – doživljala sem, da nadgrajuje vse moje dotedanje vedenje in ga povezuje z resničnim življenjem, s tem, kakršen jezik v resnici je in kar za nas, naše odnose in družbo v celovitosti predstavlja. V široki svet sociolingvistike in raziskovanja stališč me je vpeljala profesorica Albina Nećak Lük, prva snemanja in analize sem opravila v okviru spoznavanja novejših metod dialektologije pri Karmen Kenda-Jež. Potem pa statistika, korpusno jezikoslovje in sociologija vsakdanjega življenja in imeli smo solidno podlago za doktorsko raziskavo, v kateri sem proučevala variantnost govorjene slovenščine pri govorcih z Idrijskega, ki delajo ali se šolajo v Ljubljani. Kvantitativno analizo izbranih variabel, predvsem fonoloških, smo prepletali s kvalitativno študijo socialnopsiholoških tem, kot so jezikovna stališča, jezik in identiteta, jezikovna ozaveščenost in izkušnje z jezikovno rabo. Po doktoratu in porodniški sem bila nekaj časa vpeta v sorazmerno sociolingvistična projekta na Oddelku za prevajalstvo pri Darji Fišer o akademski slovenščini in sovražnem govoru.

Možnost za nadaljnje delo na bolj mojem področju se je nato odprla v okviru podoktorskega projekta, kjer podobno zgodbo, kot sem jo pisala z Idrijčani, nadaljujem z Ribničani in Mariborčani. Razveseljujejo me stiki in intervjuji z ljudmi, mučim se z določanjem variabel in seciranjem glasov. Marsikaj je drugače in bolj zahtevno kot v doktorski zgodbi, ker sama ne pripadam skupnosti: težje je dobiti informante, nisem govorka ribniščine oziroma mariborščine, Mariborčanov, ki se vozijo v Ljubljano, je tudi relativno malo. Delo je večkrat precej samotno in bi si želela bolj stalno ekipo, občasno se sprašujem o smiselnosti početja in o nadaljnji poti.

Zastavlja se mi vprašanje, ali bi bilo mogoče zasnovati oziroma izvesti celovitejšo raziskavo govorjenega jezika, po možnosti longitudinalno, kako in v kakšnem okviru. Da bi spremljali variantnost in spremembe ter relevantne socialnopsihološke teme na področju različnih narečnih skupin oziroma mestnih središč, tudi še naprej pri mobilnih govorcih – zaradi različne zgodovine, vloge, prestiža in statusa posameznih narečij namreč lahko za različne skupine govorcev pričakujemo različne ugotovitve. Za večjo zanesljivost in veljavnost rezultatov bi bilo zaželeno analizirati reprezentativnejši vzorec govorcev, transkribirati več gradiva in določiti variable na različnih ravninah – ob fonološki torej upoštevati tudi morfološko, sintaktično, leksikalno, pa tudi prozodične lastnosti – tempo, ritem, melodijo govora. Dobrodošle bi bile tudi perceptivne študije, saj sociolingvistični pomen tvorijo vsi udeleženci v pogovoru. Tako bi dobili informacije o poslušalčevem vrednotenju variant oziroma varietet in njihovih govorcev, saj obstaja dilema glede razmerja med poimenovanjem in tem, kar posamezna oznaka predstavlja. Tovrstni podatki bi lahko osvetlili rezultate dosedanjih raziskav o stališčih do posameznih varietet in njihovi rabi. Tudi izkušnje s transkribiranjem kažejo na potrebnost nadaljnjih raziskav percepcije govora oz. posameznih variabel tako pri laičnih osebah kot strokovnjakih, ki zapisujejo besedila za znanstvene namene. Vključiti bi bilo smiselno tudi socialnopsihološke teme o jeziku in identiteti, jezikovnih stališčih, ideologijah, povezavi med osebnostnimi lastnostmi, stališči in jezikovnim prilagajanjem oziroma neprilagajanjem, o stereotipih in predsodkih, izkušnjah z jezikovno rabo in podobno. Posebej aktualna so vprašanja o odnosu do različnih jezikovnih varietet in njihovi rabi v konkretnih okoliščinah in situacijah, v katerih ima govor še posebej pomembno vlogo, recimo v vzgojno-izobraževalnih ustanovah, v podjetjih, v zdravstvu. Kot zanimiva področja, vredna nadaljnje obravnave, se kažejo tudi govor staršev in drugih odraslih z otroki, jezikovna raba v družinah, kjer starša govorita različni narečji oz. jezikovni varieteti, ter v družinah, ki so se preselile v drugo pokrajino, posebej pri otrocih iz takih družin.

Za proučevanje sodobne sociolingvistične realnosti, ki jo zaznamujejo mnoge in hitre spremembe, povezane z naraščajočo mobilnostjo, globalizacijo in novimi tehnologijami, ter za napredek sociolingvistične teorije bi bila potrebna integracija različnih jezikoslovnih, sociooloških in socialnopsiholoških

pristopov, znotraj teh mikro- in makroperspektiv, tudi etnografskih študij in analize diskurza. Vsekakor je dragocena kombinacija kvalitativnega in kvantitativnega pristopa, saj le na podlagi kvalitativnih raziskav in izsledkov lahko smiselno načrtujemo kvantitativne raziskave in smiselno interpretiramo njihove rezultate.

Kaže se, da bi bilo zaradi kompleksnosti in obširnosti področja pri morebitni obširnejši študiji potrebno vključiti strokovnjake z različnih področij, na primer dialektologe, fonetike, psihologe, sociologe, etnologe, računalničarje, strokovnjake za statistiko.

To je torej nekaj izsledkov, ki izhajajo iz mojega dosedanjega dela, sedaj pa predajam besedo vam, ki imate neprimerno več izkušenj in ste vpeti v delo različnih institucij, da problematiko osvetlite vsak s svojega zornega kota.

Vabim kar prvo govornico, docentko doktorico Natašo Gliha Komac, znanstveno sodelavko in namestnico predstojnika na Inštitutu za slovenski jezik Frana Ramovša pri ZRC SAZU.

Nataša Gliha Komac: Lepo pozdravljeni, spoštovane kolegice, kolegi. Najprej organizatorjema zares najlepša hvala za organizacijo tega posveta, ki ga po mojem mnenju slovensko jezikoslovje zagotovo potrebuje. Počaščena sem, da sem del tega omizja, hkrati pa moram priznati, da pogrešam še kakšen glas z Obale, Koroške ali Maribora, čeprav zdaj slišim, da je Maribor delno zajet in da imamo tukaj tudi kolegico s Koroške. Ta misel se mi je porodila, ker je to posvet, namenjen sociolingvistiki, in se mi zdi, da bi bila potem razprava res uravnotežena in bi bil v grobem pokrit celoten slovenski jezikovni prostor. Sicer pa čestitke, saj smo tu res predstavnice in predstavniki različnih pedagoških in raziskovalnih ustanov, raziskovalke in raziskovalci iz Slovenije in zunaj nje, predstavnice in predstavniki različnih starosti, poznavalke in poznavalci najrazličnejših področij.

Tri krovna vprašanja ste nam zastavili za razmislek. Moj je splošen – verjetno boste drugi bolj konkretni, a morda izzove kakšno konkretno, uporabno in iskrivo misel, idejo.

Sociolingvistika je eno tistih področij, ki nam je zaradi svoje prepletenosti z našim vsakdanom tako zelo blizu, da o njej vsi vse vemo in imamo navadno

dokaj izdelana stališča. Morda so prav zato refleksija, strokovna razprava in argumentirano soočenje stališč toliko večji izziv. Jezik in družba sta nujno v navezi, saj je jezik sredstvo sporazumevanja, sredstvo identifikacije, ima svojo simbolno vrednost, svoj status in prestiž, prav prek jezika sleherni od nas sporoča svoje videnje sveta in družbe in jo hkrati soustvarja. In, na kar jezikoslovci vsake toliko radi pozabimo, tudi družba soustvarja in spreminja jezik oziroma vpliva nanj.

Prav zato je jezikovna politika vse prej kot naključna. Organizirane skupnosti navadno vedno premišljeno načrtujejo razmerja med jeziki, poskrbijo za njihove opise ter širjenje jezikovnega znanja. Sociolingvisti – raziskovalci in pedagogi – to stvarnost spremljamo, na dogajanje opozarjamo, razmerja med jeziki ozaveščamo, širimo znanje jezikov ter ga skupaj s stališči prenašamo na nove generacije.

In v tovrstnem celostnem razumevanju vidim glavni izziv slovenske sociolingvistike. V rednem spremljanju in opisovanju ter preverjanju tako imenovane slovenske jezikovne stvarnosti, v vsej jezikovni barvitosti in različnosti jezikovnih rab in praks. Pa naj gre za večplastnost oziroma tako imenovano družbeno zvrstnost slovenskega jezika in njegovo pojavljanje v najrazličnejših besedilnih vrstah ali razmerja med različnimi jeziki, njihovo dostopnost. Po vseh področjih rabe, zlasti javne.

In seveda ob zavedanju in odgovornosti različnih pristojnih služb, zlasti pa raziskovalk in raziskovalcev, pedagoginj in pedagogov, da s svojimi spoznanji, prenašanju teh v javnost, s svojim poučevanjem in usposabljanjem novih kadrov, tudi s svojim lastnim zgledom, rabo jezika, njegovih različic in različnih jezikov, do neke mere usmerjamo in določamo, predvsem pa opozarjamo na potrebo po razvijanju jezikovnih znanj ter zagotavljanju priložnosti za rabo. Kot je pokazala naša raziskava *Jezikovna politika Republike Slovenije in njeni uporabniki* (2017) in kot me v zadnjih dveh letih učijo izkušnje delovanja v različnih strokovnih odborih, zakonske podlage v večini so, in sploh ne slabe, nedomišljene, zalomi se pri uresničevanju določil in zavez. Včasih tudi zaradi njihovega nepoznavanja. Pri vsem tem so zelo pomembni programski dokumenti, za katere ni naključje, da se spreminjajo po določenem časovnem obdobju, in za katere je zelo pomembno, da jih spremljajo empirične raziskave, tako med splošnimi kot med specializiranimi jezikovnimi uporabnicami in

uporabniki. Le tako so potem lahko politične odločitve, razporejanje finančnih sredstev ali celo na primer izobraževalni sistem res v duhu časa. In tu vidim izjemen pomen sociolingvistike kot stroke, ki spremlja tako razmerja in dogajanje v družbi kot tudi razvoj, trende in potrebe znotraj jezika, jezikov in jezikoslovja.

Mislim sicer, da se na področju slovenske sociolingvistike, tudi kar zadeva mednarodno vpetost, veliko dogaja, a morda preveč parcialno, razdrobljeno. Pogosto umanjka kontinuiran, celosten vpogled in pregled, redne in sprotne raziskave. V okviru različnih projektov nastajajo odlični pripomočki, opisi, nastavki teorij, a kaj, ko se včasih vse skupaj, na primer po štirih letih ali s še krajšim projektom, konča. Pa ravno se pridobijo ustrezne izkušnje, znanja, vzpostavijo se mreže in način dela … In nato nekaj časa nič, potem pa nov razpis in zgodba steče od začetka, zopet se vzpostavi sistem itd., umanjkajo pa poglobljene analize. Ali pa drug za drugega preprosto ne vemo, premalo vemo o delu drug drugega in se stvari ponavljajo.

Pogosto se zdi, da ostajamo preveč zaprti vsak v svojem vrtičku, da zlasti pri večjih, bolj dolgoročnih in za širšo jezikovno skupnost pomembnih in odločujočih projektih premalo izmenjujemo in združujemo znanja in izkušnje. Konkurenca je vedno zdrava in kritična mnenja so izjemno dragocena, so pa odločitve in koraki, kjer je treba najti kompromise. In za uspešen vpliv v družbi in politiki, ki bo podpirala tako strateško jezikovno načrtovanje in spremljanje jezikovnih rab, praks, znanj in stališč ter razvoj tudi slovenskega jezikoslovja, je zagotovo zelo pomemben enoten glas stroke. Sicer je veliko enostavneje sredstva preprosto razbiti (in postopoma zmanjšati).

Pa še nekaj me občasno bega. Teoretični modeli – če odmislimo manipuliranje in slabe namene – pojasnjujejo stvarnost oziroma naj bi jo, vsaj po mojem razumevanju, razlagali in – v primeru sociolingvistike – (vsaj posredno) omogočali sporazumevanje in ohranjanje družbene stabilnosti, kohezijo, sobivanje, torej različnim jezikovnim uporabnicam in uporabnikom zagotavljali celostno in učinkovito sporazumevanje in delovanje v družbi ter posledično njeno soustvarjanje. Zato ne smejo biti zgolj sami sebi namen. Zato se zdi pomembno sprotno raziskovanje stvarnosti ter spremljanje dejanskih potreb, rab in stališč jezikovnih uporabnic in uporabnikov. A hkrati je pomembno tudi za zavedanje, da družba in jezik delujeta po nekih ustaljenih mehanizmih, ki so

prenosljivi v času in prostoru. Nista od včeraj ter živita in bosta živela tudi v nekem drugem času in prostoru, brez nas.

Kot slovenistka in raziskovalka Inštituta za slovenski jezik Frana Ramovša, pa tudi kot neodvisna strokovnjakinja v odboru za Evropsko listino za regionalne ali manjšinske jezike pri Svetu Evrope, menim, da Slovenija potrebuje zlasti sistematično in kontinuirano spremljanje (slovenskega) jezika na vseh področjih javne rabe, to je v izobraževanju, sodstvu, javni upravi in javnih servisih, v medijih, na področju kulture, v gospodarskem in družbenem življenju, tudi pri čezmejnih izmenjavah. In sicer tako zapisanega kot govorjenega jezika, pri splošnih in specializiranih jezikovnih uporabnikih, seveda s posebnim poudarkom na spremljanju jezikovnih rab, praks in potreb ranljivih skupin. Pri čemer za dobre in aktualne jezikovne opise slovenskega jezika nujno potrebujemo aktualen jezikovni korpus z raznovrstnimi najnovejšimi besedili, poseben izziv pa je gotovo spremljanje slovenskega govorjenega jezika v vseh različicah.

Nujno je siceršnje sprotno spremljanje znanja in rabe jezikov ter stališč do njih v naši državi nasploh, in sicer zopet tako pri specializiranih kot splošnih uporabnikih ter s poudarkom na ranljivih skupinah in njihovem sporazumevanju. Včasih tudi zgolj in preprosto sledenje in iskanje načinov in poti uresničevanja zakonskih določb in zavez, ki bodo usklajeni s časom in aktualnimi potrebami jezikovnih uporabnic in uporabnikov. Predvsem pa njihovo sprotno preverjanje.

Pomemben je pretok znanj in izkušenj med raziskovalnimi in pedagoškimi ustanovami. Izkušnje učijo, da imamo raziskovalci na raziskovalnih ustanovah morda več priložnosti delati neposredno z gradivom, na terenu, pa naj gre za besedila ali na primer za različne terenske raziskave oziroma podatke s terenskih raziskav. Pedagoški delavci pa se neposredno srečujejo z zadregami študentov, z vprašanji, kako prenesti znanja v prakso ter tako neposredno odkrivajo, kje morda modeli in rešitve, opisi, ki smo jih na primer sprejeli avtorji različnih jezikovnih opisov in priročnikov, ne vzdržijo, niso ustrezni ali pa je potreben ponoven premislek, kako načela pojasniti. Zato mislim, da so izjemno dragocene mešane skupine, kjer sta mogoča prenos in izmenjava mnenj in znanj, predvsem pa se hkrati vzpostavlja medsebojno razumevanje in spoštovanje.

In še čisto za konec: vprašanje smiselnosti in realnosti je vedno aktualno. Jaz bi ga morda zastavila nekoliko drugače: kateri so naši cilji, kam pravzaprav želimo našo »ladjo« pripeljati? Pri tem izhajamo iz realnosti, ki temelji na naših izkušnjah in znanju – vendarle smo se specializirali za to področje, hkrati pa jo nujno spremljamo z raziskovanjem potreb, rab in praks jezikovnih uporabnikov in uporabnic. In podobno je verjetno s smiselnostjo. Saj se verjetno lahko vsi poistovetimo, v kontekstu evropskega dneva jezikov, ki ga obeležujemo z današnjim posvetom, z željo po ohranjanju tako imenovane pregovorne »slovenske« večjezičnosti. Katere nujni del sta tudi znanje in raba slovenskega jezika. Hvala.

Maja Bitenc: Najlepša hvala! K besedi vabim docentko doktorico Matejko Grgič, sodelavko Slovenskega raziskovalnega inštituta SLORI v Trstu in Oddelka za prevajalstvo ljubljanske Filozofske fakultete.

Matejka Grgič: Hvala, lepo pozdravljeni. Moje izhodišče je slovenščina kot manjšinski jezik, predvsem slovenščina v Italiji. Ukvarjam se z jezikovnim stikanjem, se pravi z vplivanjem italijanščine na slovenščino, manj – oziroma skoraj nič – v drugo smer, torej z vplivanjem slovenščine na italijanščino. Ukvarjam se tudi z jezikovno marginalizacijo in z izključevanjem govorcev. Do nedavnega smo o tem, tudi na mednarodni ravni, veliko govorili v odnosu manjšinski–večinski govorci: govorci jezika, ki je na nekem območju večinski, so govorce jezika, ki je na istem območju manjšinski, izključevali iz svojih mrež. Vzrok in hkrati posledica tega je bila, da so govorci manjšinskega jezika zelo slabo obvladali večinski jezik in se zaradi tega tudi sami izključevali iz okolij, kjer se je zahtevala raba tega jezika. Danes opažamo že trend v nasprotni smeri, in to izrazito: nekateri mlajši govorci slovenščine v Italiji zase menijo, da slovenščino govorijo tako slabo, da se potem na primer ne odločajo za študij v Sloveniji – pravijo, da slovensko ne znajo dovolj dobro, da bi prišli študirat na primer računalništvo v Ljubljano, in to po trinajstih letih obiskovanja slovenskih šol. Ali to drži ali ne drži, je vprašanje, ki ga bo treba še raziskati, je pa dejstvo, da je taka percepcija zdaj že zelo močna. Trend se je začel kazati na področju branja: učenci niso več brali knjig v slovenščini, zato so se učiteljice začele spraševati, če so morda slovenske knjige bolj dolgočasne. Nakar so učencem dale Harryja Potterja, torej knjigo, ki je dostopna v obeh jezikih in katere vsebina je v obeh jezikih enaka. Večina otrok je rekla, da lažje

bere v italijanščini – razen besedil, ki jih morajo nujno prebrati v šoli, učenci namreč sicer ne berejo v slovenščini: ne knjig, ne revij, ne objav na družbenih omrežjih. Isto velja za filme in vse druge vsebine, do naročanja hrane na dom. S tem je povezana tudi jezikovna variantnost – kako jo umestimo, kako jo razumemo, kako jo raziskujemo – in transverzalne teme: jezikovno načrtovanje in jezikovna politika, didaktika jezika in priprava jezikovnih virov – katere imamo, katerih še nimamo, katere bi potrebovali.

Zaustavila bi se pri nekaterih bolj aplikativnih vidikih. Na splošno se mi zdi, da raziskovalci veliko delamo, čeprav – tako kot je prej povedala kolegica – na nekih projektih z omejenim časom trajanja in financiranja, kar pomeni, da začenjamo na vsake tri leta stvari postavljati na novo. Včasih so viri financiranja tudi »neraziskovalni« in zahtevajo neke čisto druge cilje, namene in seveda končne rezultate. To pomeni, da je nesistematičnost ena od glavnih težav dela na tem področju. Vendar pa mislim, da je raziskovanje neverjetno dobro glede na dane okoliščine: ogromno raziskav imamo, če upoštevamo vse to, kar je bilo doslej objavljeno, in marsikaj, kar bo še izšlo iz današnjega posveta. Zdi se mi, da je veliko večji problem prenos teh znanj v prakso. Ko govorimo o znanju, ki je zdaj zbrano v tej predavalnici, in znanju kolegov in kolegic, ki nas spremljajo prek Zooma, in potem primerjamo, kaj se zgodi z vsem tem znanjem, opažamo, da je v najboljšem primeru to potem snov za kakšen izpit. Študenti in študentke se nekaj naučijo; eni bolj in drugi manj, eni se bolj napiflajo, drugi to bolj resno vzamejo v roke, a tukaj se zadeva v glavnem konča.

Ko sem pred kratkim pisala strokovni referat, namenjen odločevalcem, mi je Devan Jagodic, direktor SLORI v Trstu, rekel: »Matejka, zaključki niso v redu.« Pa sem rekla: »V katerem smislu niso v redu?« »Ja, v zaključkih moraš po točkah povzeti tisto, kar si prej povedala, ker bodo vsi brali samo zaključke.« Drugič bom torej na enem A4-listu povedala, kar pač moram, ni treba, da pišem članek, dolg 15 strani. Ampak tako pač je. Tu se mi zdi, da obstaja velik problem pri prenosu raziskovalnih rezultatov v prakso.

Za to bi dala še nekaj primerov. Mi nimamo, danes, dvojezičnih virov za jezikovni par italijanščina–slovenščina, slovenščina–italijanščina. Imamo Šlenca, Veliki slovensko-italijanski slovar avtorja Sergija Šlenca, ki je gigantsko delo, sploh če pomislimo, da ga je opravil en sam človek, a danes je slovar z več vidikov neuporaben, tako po naboru gradiva kot po obliki in metodologiji. Nič

drugega nimamo. Opiramo se na EU-vire, ki so večjezični (torej ne specifični za jezikovno kombinacijo), toda to delamo profesionalni prevajalci – težko si predstavljam, da bodo ljudje, ki potrebujejo vsakdanje, hitre rešitve, iskali po Evrotermu, Evrokorpusu in tako naprej. To nekako ne gre. In tukaj nimamo nič: nič se na tem ne dela, nič se o tem ne govori in sploh ni nobenega načrta, da bi se to podprlo na ravni kakšnih bilateralnih zavez ali drugačnih, tudi aplikativnih projektov. Nič. Kot da tega problema ni. Pa je to problem tako za slovensko skupnost v Italiji kot za italijansko skupnosti v Sloveniji in seveda za druge uporabnike in uporabnice.

Drugo tako področje, s katerim se sama sicer ne ukvarjam, ampak ga srečujem in je problematično, je didaktika slovenščine na čezmejnem območju. Kolegica Nataša Pirih Svetina bo veliko več vedela o tem, a načeloma lahko rečem, da model »slovenščina je materinščina« na šolah s slovenskih učim jezikom v Italiji drži toliko kot ta čudovito lepa pravljica o materinščini nasploh. Tam so otroci, ki prihajajo iz zelo različnih družin, in tudi otroci, ki prihajajo iz slovenskih družin, a živijo v povsem italijanskem okolju. Ko gredo naročit sladoled, morajo sladoled naročiti v italijanščini. To pomeni, da se nikoli ne naučijo naročiti sladoleda v slovenščini. Zakaj ne? Zato ker naročanje sladoleda v slovenščini ni del učnih načrtov. Drugi problem je identifikacija z diametralno nasprotno definicijo »slovenščina kot drugi in tuji jezik«. Veliko otrok in mladostnikov obiskuje recimo poletno šolo slovenskega jezika v Ljubljani, ampak vsakič, ko se omenja sodelovanje tudi s Centrom za slovenščino kot drugi in tuji jezik, marsikoga zmoti že ime. Verjetno bo treba tu delati na področju neke specifične didaktike slovenščine kot manjšinskega jezika. Zdaj boste rekli: »Joj, še ena definicija. A bomo zdaj odpirali še en oddelek, še en center?« Ne vem, nimam dokončne rešitve, a dejstvo je, da je ta potreba zdaj že odprla čisto poseben segment dela. Ker so to, kako se ljudje identificirajo in kje se pozicionirajo, elementi, ki jih moramo upoštevati, ko govorimo o prenosu znanj v prakso.

Z didaktiko je povezano tudi pridobivanje novih govork in govorcev. Raziskave so pokazale, da se nobena manjšina ne ohrani, če se zgolj vzdržuje in ščiti. To je slovenska skupnost v Italiji skušala delati 50 let, tudi z zakonodajo, ki zagotavlja varstvo manjšinam in manjšinskim skupnostim. Dokazano je, da takšno varstvo seveda mora obstajati, a je to samo prvi korak; od tod dalje

je potrebnih še veliko drugih korakov in eden od teh je pridobivanje novih govorcev. Pridobivanje novih govorcev je ključno. Na tem področju se dela absolutno premalo in nesistematično. Če vzamemo samo območje Trst–Gorica–Videm, torej obmejno območje, je zadnja raziskava o tečajih slovenščine kot tujega jezika, ki je bila objavljena leta 2015, pokazala, da je na tem območju tečajev veliko, a te tečaje izvajajo razna društva, večkrat tam poučujejo upokojeni učitelji slovenščine, študenti, člani društva ali pa kakšen entuziast. Mislim, da moramo biti, ne samo kot Slovenci, ampak tudi kot strokovnjaki in strokovnjakinje za to področje, malce bolj ambiciozni in moramo postaviti sistem – tako kot je prej rekla kolegica – sistematičnega raziskovanja in sistematičnega prenosa teh znanj v prakso: na primer sistematičnega pridobivanja novih govorcev. Verjamem, da je to veliko pomembnejše od zaščitniškega odnosa, ki ga imamo včasih do slovenščine v manjšinskem položaju. Veliko več bi naredili, če bi strokovno energijo in politični kapital vlagali v kaj drugega kot pa v varstvo manjšine in jezika.

S tem je povezan, nazadnje, odnos tako imenovane matice do manjšinskih skupnosti. Na vseh ravneh se, ko govor nanese na manjšine, veliko govori o ohranjanju. Letno imamo prireditev Dobrodošli doma, na katero so me enkrat tudi povabili, potem pa nič več, ker sem bila mogoče malo preveč kritična. Vedno se tam predstavi kakšna folklorna skupina ali pa kakšna pesnica ali pesnik – čim starejši seveda, iz čim bolj oddaljene vasi, kjer še edino on govori slovensko in piše poezijo v narečju, ki se ga spominja iz čim bolj oddaljenih otroških dni, ko je še z edino babico govoril slovensko. Tudi to je seveda del manjšinske realnosti, vendar pa ni slika celotne skupnosti. Danes zjutraj sem bila na Zoomu z učenci neke šole v Sloveniji, mislim, da so bili iz devetega razreda, in so spraševali, kako je s slovenščino v zamejstvu. »A vas veliko preganjajo?« je bilo eno vprašanje. »A imate še vedno folklorna društva?« Mislim, da je treba tukaj nekaj narediti tudi na podobi, imidžu manjšinskih skupnosti.

Za zaključek bi rekla tole: moje mnenje je, da se, ko gre za manjšinske skupnosti in sploh za ranljive govorce in govorke, zalomi pri prenosu znanj v prakso. V tem smislu, da znanja pridobivamo strokovnjaki in strokovnjakinje po znanstvenih metodah, ki so lahko boljše ali slabše, a imajo neko epistemološko podlago. Potem pa, ko se stvar prenese v prakso, ta kar naenkrat ni več strokovna, ampak postane ali politična ali diletantska – in tu imamo potem

največ težav. Mislim, da moramo na tem še ogromno delati, in mislim, da smo tudi mi, kot strokovna, znanstvena skupnost, poklicani, da zahtevamo neke čim bolj konkretne odgovore in čim bolj konkretne ukrepe. Hvala.

Maja Bitenc: Najlepša hvala. Naslednja na vrsti je redna profesorica doktorica Monika Kalin Golob, predavateljica in dekanja na Fakulteti za družbene vede Univerze v Ljubljani.

Monika Kalin Golob: Dober dan vsem skupaj, lepo vas pozdravljam. Današnji nastop sem res bolj vzela kot posvet, zato nisem pripravila tako lepega besedila kot Nataša, ki je tako lepo nastopila in prebrala, zato bom malce improvizirala in se vnaprej opravičujem, če bo kakšna misel nedodelana. Ko sem razmišljala o tem, kaj danes povedati, sem o tem razmišljala kot jezikoslovka, ki dela zunaj matične fakultete, torej ne na Filozofski fakulteti, ampak na Fakulteti za družbene vede, kjer sem obkrožena z mnogo sociologi, politologi in samimi pametnimi družboslovci. Tako so se mi oblikovale tri točke, ko sem razmišljala o tem, kaj so prioritete, ki jih vidim kot pomembne za sociolingvistično razpravljanje, raziskovanje in poučevanje.

Prva se povezuje s šolo, torej poukom slovenščine, od osnovne šole do fakultete. Pri osnovni in srednji šoli – tudi sama imam trenutno doma maturanta – se mi zdi pomembno to, kar je kolegica Grgič rekla že za drugo didaktiko, torej, kaj danes učiti, kako danes osmisliti slovenščino kot učni predmet in učni jezik. Sama se s sinom velikokrat pogovarjam in me sprašuje: »Pa zakaj se moram to učit?« Nimam vedno dobrega odgovora. Če govorimo o raziskovanju, se mi zdi temeljni projekt o stereotipih o slovenskem jeziku, o tem, kakšen je danes odnos mladih do slovenščine. To je gotovo pomemben projekt, o katerem sva se s kolegom Stabejem nekajkrat tudi že pogovarjala, ampak nisva prišla kaj dlje od pogovora. V osnovni in srednji šoli se mi zdita pomembni predvsem vprašanji, kaj učiti otroka v šoli in kako, kar je seveda povezano ne samo z didaktiko, ampak tudi s tematskim okvirom prenašanja znanja. Na fakultetah se že dolgo ukvarjamo z vprašanji večjezičnosti, z vprašanji angleščine, z vprašanji jezikovne politike na fakultetah, a se pogovor vedno konča v nekem bipolarnem svetu. Svetu, kjer smo absolutno samo za slovenščino, in svetu, kjer absolutno podpiramo angleščino. In vmes iščemo rešitve, ki jih seveda lahko iščemo skupaj in v dialogu, ne pa v večni bipolarni razpravi, ki velikokrat meji na nestrpno razpravo, kot smo videli ob vsakem predlogu zakona, in se bo zdaj

v kratkem ponovila, ko bo na mizi vnovič osmi člen Zakona o visokem šolstvu. To se mi zdi prvi pomembni sklop: šola na vseh stopnjah.

Drugo vprašanje zadeva odnos do jezika in ideologije, kar je tipično sociolingvistična raziskovalna tematika. Tukaj vidim tri podtočke, ki bi jih bilo smiselno raziskovati, predvsem pa osmisliti v praksi. Eno je trenutno zelo aktualna tema, to je sovražnost komuniciranja, ves sovražni govor, žalitve, vse to, kar spremljamo v naši politiki, in gre nato tudi naprej: iz političnega diskurza v širši slovenski prostor. Tukaj je nekaj raziskav, je nekaj idej, več kot to pa najbrž ne. Drugo je politični diskurz sam po sebi, torej vprašanje, kako raziskovati politiko. Imamo nekaj raziskav, kjer govorimo o tem, kako politiki nastopajo v parlamentu, ampak bi bilo to treba najbrž povezati s prejšnjim vprašanjem oz. ravnmi političnega diskurza. En del te sovražne jezikovne politike smo izkusili na lastni koži, ko ste tukaj na Filozofski fakulteti, potem pa še mi na FDV-ju govorili o spolu. Tudi to je tema, ki zelo razburja slovensko javnost, zato se zdi, da ideološke teme v slovenskem prostoru vedno naletijo na dvopolnost: ne najdemo dialoga in se vrtimo okoli dveh zelo nasprotujočih si argumentov, čeprav gre za poskuse ureditve problemov, ki so nujni za civilizirano družbo.

Vidimo torej, da gre vsepovsod za težko iskanje dialoga. In med razmišljanjem, kaj je krivo za to, sem prišla do tretje pomembne točke, in to je položaj slovenistike, položaj jezikoslovja pri nas. Ker tudi v jezikoslovju živimo zelo dvopolni svet, v katerem ne znamo iskati skupnih točk. Zanimivo je, da se to pozna tudi na odnosu družbe do nas: na FDV-ju smo jezikoslovke »lektorice«; nismo raziskovalke, nismo znanstvenice, ampak smo lektorice. Jezikoslovec torej pride prav, ko je treba kaj prebrati, kaj popraviti, kaj več pa tudi ne. Vprašanje spola, vprašanje politike in ideološkega diskurza lahko potem rešujejo sociologi, ne pa jezikoslovci. Tako da se mi zdi, da bi najprej morali opraviti nek metaslovenistični diskurz o tem, kako jezikoslovje danes funkcionira, kaj jezikoslovje je. Je to zgolj neko pomožno lektorsko znanje ali pa zgolj jezikovnosistemska zadevščina, ki potem ostane v ozkohermetičnih člankih in ne pride v prakso. Če bi to znali razrešiti, bi se potem tudi druge polarnosti lahko razreševale v strokovnem, znanstvenem in bolj smiselnem diskurzu, kot se rešujejo danes. Imam izkušnjo, ko smo delali nacionalne programe o jezikovni politiki, kjer se je res izkazalo to, kar je povedala že kolegica Grgič: strokovna ekipa lahko dela

krasne dokumente, a kaj, ko ti dokumenti ne pripeljejo do denarja za postavke in se v petih, desetih letih seveda pišejo novi politični dokumenti, ki pa spet nikoli ne dobijo denarja, potrebnega za takšne projekte. Tako se mi zdi, da dandanes ni veliko optimizma, je pa veliko možnosti, da se s premislekom in temeljnimi projekti stvari tudi spremenijo.

Te tri točke so se mi izoblikovale ob hitrem premisleku: šola, ideološke teme in pa položaj jezikoslovja tukaj v Sloveniji. Zadnja bi potem od spodaj navzgor lahko rešila tudi drugi dve tematiki. Tako na hitro in hvala lepa.

Maja Bitenc: Najlepša hvala. Doktorica Karmen Kenda-Jež, višja znanstvena sodelavka in vodja dialektološke sekcije Inštituta Frana Ramovša pri ZRC SAZU**.**

Karmen Kenda-Jež: Pozdravljeni. Sama se bom osredotočila na eno samo točko, in sicer si to pravico jemljem kot tako rekoč edina priučena sociolingvistka tukaj. Odkar se ukvarjam z dialektologijo, sem – tudi na pobudo jezikoslovne misli Brede Pogorelec – vstopala v sociolingvistiko predvsem skozi iskanje najbolj primerne metodologije in tehnik za raziskovanje govorjenega jezika. O tem se namreč vsaj v našem prostoru, v dialektologiji na začetku mojih terenskih raziskav ni veliko razmišljalo. Zato sem še vedno najbolj povezana z variantnostnim jezikoslovjem in z vsem, kar se je v njem dogajalo od Labova naprej. Pri tem pa me ves čas muči razmerje med dialektologijo in sociolingvistiko v slovenskem prostoru.

Pred leti sem o tem napisala članek, ki je ostal v predalu, in prav vesela sem, da mi danes tega članka ni več treba dokončati, saj je imel naslov *Prevzetnost in pristranost*. Vendar se mi vseeno zdi, da to razmerje še vedno temelji na nekakšnem nesorazmerju, na implicitnih sodbah, ki bi jih za drugo polovico prejšnjega stoletja lahko opredelili kot dve predpostavki. Prva je bila – seveda ne govorim o celotnem jezikoslovju, ampak o tistem, ki je bilo v času mojega študija paradigmatično – da dialektologija ne spada med vede, ki se ukvarjajo s sodobnim jezikom. Druga pa, da empirične sociolingvistične raziskave navadno ne upoštevajo zemljepisnih posebnosti jezikovnega gradiva. Iz njiju sta se izoblikovala dva aksioma, ki še vedno obstajata, morda bolj na nezavedni ravni, čeprav je ravno v tem tisočletju prišlo do velikih premikov. Po eni strani se je povečal obseg empiričnih sociolingvističnih raziskav, se pravi

raziskav jezikovnega gradiva, kar sem kot dialektologinja vedno pogrešala, saj nisem mogla primerjati svojih, zvrstno razmeroma ozkih podatkov, z drugimi. Po drugi strani je nastal težko pričakovani govorni korpus, izoblikovali so se začetki slovenskega variantnostnega jezikoslovja, in sicer na sorazmerno sodobni ravni, saj se je takoj začelo tudi povezovanje kvantitativnih pristopov s kvalitativnimi, kar je rezultat osebne pobude Maje Bitenc. Hkrati pa so bile dopolnjene tudi dialektološke raziskave, bodisi s prvinami zaznavne dialektologije, posameznimi raziskavami, ki so se načrtneje posvečale socialni strukturi govorcev, ali z drugačnimi pristopi k dialektološkim raziskavam mestnih govoric.

Vendar vse to poteka precej bolj počasi, kot sem si predstavljala. Naj naštejem nekaj zgledov za to, da omenjena aksioma, po katerih zemljepisna variantnost ni relevantna za raziskovanje sodobnega govorjenega jezika, socialna variantnost pa ne za raziskovanje slovenskih narečij, še vedno delujeta. Na prelomu tisočletja se je to npr. kazalo v tem, da v zbirki *Najnovejša zgodovina slovanskih jezikov*, ki je izhajala v Opolu na Poljskem v letih 1996–2001, samo v češkem in slovenskem zvezku ni bilo predstavitve zemljepisne razčlenjenosti, kljub temu da je zemljepisna členjenost slovenskega jezika tako rekoč pregovorna. Z dialektološkega vidika na to kaže izločitev področja z neavtohtono poselitvijo oz. t. i. mešanih kočevskih govorov iz dialektoloških raziskav, kot nečesa, s čimer naj se ukvarja sociolingvistika. Anekdotično je to mogoče ponazoriti še z novejšima zgodbama terminološke narave. Andrej Skubic je v svoji monografiji *Obrazi jezika* (2005) angleški izraz *dialect*, ki pravzaprav pomeni jezikovno zvrst, zavestno preimenoval v sociolekt in tako spremenil vrednostno razmerje med socialnim in prostorskim. Zadnji terminološki premik je iz leta 2018: v knjigi Mateja Šeklija *Tipologija lingvogenez slovanskih jezikov* je slovenska dialektologija obravnavana zgolj kot del genetskega jezikoslovja, glavna domena sociolingvističnih raziskav pa je knjižni jezik.

Tako se znajdemo v absurdni situaciji, da je v naših temeljnih narečnih raziskavah narečje pravzaprav obravnavano kot sistem brez govorcev. To je sicer mogoče razumeti kot določen metodološki pristop, ne moremo pa tega razumeti, če celovito opazujemo jezik, ki se govori na nekem določenem področju, ne glede na to, kolikšen jezikovni prostor to opazovanje zajame. S tega vidika je v slovenskem jezikoslovju še vedno vsaj deloma tako, da se sociolingvistika

ne ukvarja z dialektološkimi izsledki, dialektologija pa ne s sociolingvističnimi, pri čemer zadnja pri svojih osnovnih raziskavah ne uporablja niti tistih osnovnih metodologij in tehnik za analizo govorjenega jezika, ki so v sociolingvistiki dobro razvite. In pri tem izgubita obe vedi. Tudi če se raziskovalni paradigmi sploh ne bi spreminjali, ampak bi samo dopolnili svoje raziskave z izsledki in ugotovitvami druge smeri, bi dobili precej bolj izčiščene in veljavne rezultate. Genetskojezikoslovna raziskava je npr. veliko bolj zanesljiva, če terenski raziskovalec zna določiti socialno vrednost sinhrone variabilnosti obravnavanega gradiva, po drugi strani pa je v slovenskem razdrobljenem narečnem in regionalnem prostoru težko zastaviti sinhrono raziskavo brez dobrega poznavanja bodisi diahronih razvojev bodisi tiste osnovne podlage, na katero se naslojijo druge jezikovne zvrsti.

Rada bi poudarila, da prav zaradi tega – te nezmožnosti, da bi pri se pri nas oblikovala sociodialektologija ali da bi se lotili skupinskih interdisciplinarnih raziskav – ostaja popolnoma neraziskano območje med govorjenim knjižnim jezikom in narečji. Pravzaprav redko koga zares zanima, razen mogoče anekdotično, kaj se dogaja v tem vmesnem prostoru, kakšna je njegova dinamika, kakšna je struktura sprememb, kje so žarišča mestnih govoric, kje prihaja do pojavov, ki niso ne knjižni ne narečni. Zelo si želim, da bi se to v prihodnosti spremenilo. Dober vzorec takšnih raziskav ponuja nemško jezikoslovje. Dolgoročni projekt Rede (Regionalni jeziki v Nemčiji)[1] je eden največjih trenutno potekajočih jezikoslovnih projektov, trajal bo okoli 19 let, za njegovo izvedbo pa so prejeli okrog 14 milijonov evrov. Projekt strnjuje vse dosedanje izsledke s področij dialektologije, sociolingvistike in variantnostnega jezikoslovja, z množičenjem, s posebnimi anketami pa intenzivno raziskuje sinhrono regionalno nemščino. Gre za portal, ki na enem mestu združuje vse, kar lahko o jeziku poveš prostorskega.

In za konec še malo samokritike. Tudi mi bi morali drugače pogledati na svoje gradivo. Na podlagi starejših jezikovnih podatkov, tudi z uporabo diahronih podatkov za *Slovenski lingvistični atlas*, in novejših dialektoloških raziskav, ki pri nas v zadnjem času bolj intenzivno potekajo na mikroravni, bi lahko izdelali izhodiščni model za razumevanje regionalne členitve slovenščine, ki bi omogočil izdelavo načrta za nadaljnje raziskave.

1 https://www.regionalsprache.de/

Maja Bitenc: Najlepša hvala. K predstavitvi vabim zaslužno profesorico Albino Nećak Lük, ki je delovala na Oddelku za primerjalno in splošno jezikoslovje Filozofske fakultete Univerze v Ljubljani in je utemeljiteljica študija uporabnega jezikoslovja.

Albina Nećak Lük: Pozdravljeni vsi skupaj. Hvala organizatorju za povabilo. Veseli me, da sem dobila priložnost, da skupaj z vami razmišljam o relevantni sociolingvistični tematiki v slovenskem jezikovnem prostoru. Ko omenjam slovenski jezikovni prostor, mislim na celotni etnolingvistični prostor znotraj Slovenije ter na obeh straneh vzdolž njenih meja na območjih jezikovnega stika s sosedskimi jeziki, pa tudi na slovenske jezikovne skupnosti v različnih predelih sveta.

Marsikje berem, da je sociolingvistika mlada veda. To se kot nekakšno mašilo pojavlja v diplomskih, magistrskih, doktorskih delih, pa tudi v tistih nekaj univerzitetnih programih univerz na Slovenskem, ki sociolingvistiko ponujajo študentom, največkrat kot izbirni predmet. Res je, da začetna sociolingvistična raziskovanja *sensu strictu* postavljamo v 60. leta prejšnjega stoletja in jih povezujemo predvsem z Labovom in z njegovo znamenito sociolingvistično spremenljivko. Vendar naletimo na poskuse postavljanja proste variacije na raziskovalni repertoar jezikoslovcev že mnogo prej. Lahko bi rekli, da so se zametki sociolingvistike rodili, ko je de Saussure, mojster znanstvenega mišljenja, v stremljenju po čim bolj natančni in jasni opredelitvi ali bolje rečeno zamejitvi predmeta svojega raziskovanja, vzpostavil razlikovanje med pojmom jezika in drugimi, z jezikom povezanimi pojmi. Z uvedbo koncepta *parole/govor*, ki de Saussuru pomeni *konkretno rabo jezikovnih znakov v natančno določenem kontekstu*, se začnejo spreminjati pogledi na možnosti raziskovanja pojava variantnosti v jeziku in jezikih. Tudi med h kategorialnosti usmerjenimi jezikoslovci. Kakor vemo, je variantnost, prosta variacija – torej *druga oblika istega s katerekoli ravnine jezikovnega ustroja*, kakor jo je opredelil Toporišič; jaz bi dodala druga oblika istega iz bodisi eno-, dvo- ali večjezičnega govornega repertoarja – pred Labovom veljala za nekakšen tabu, za neoprijemljivo, neopredeljivo in s tem nepreverljivo spremenljivko, za problem, ki ga ni mogoče znanstveno opredeliti in empirično preverjati in preveriti.

Ta kratek uvodni ekskurz v zgodovino razvoja sociolingvističnega raziskovanja omenjam prav na današnjem posvetu, ker je ta namenjen nekakšnemu

pregledu doslejšnjega sociolingvističnega raziskovalnega izplena in urgentnim prihodnjim sociolingvističnim raziskavam. Začetki razvoja sociolingvističnega raziskovanja na Slovenskem se zdijo kot nekakšna replika dogajanja na širši, svetovni sociolingvistični sceni. Razprav o znanstveni relevantnosti sociolingvistike sicer ni bilo veliko, dejstvo pa je, da se stališča, nezaupanje, celo predsodki itd., kakor so jih morali – verjetno povsod – prebroditi začetniki sociolingvističnega raziskovanja v dokazovanju, da je predmet njihovega raziskovanja znanstveno opredeljiv in preverljiv, pri nas zavlečejo kar tja v 90. leta prejšnjega stoletja in čez. S tem nočem reči, da dotlej v Sloveniji ni bilo sociolingvističnih raziskav. Nasprotno, posamezne sociolingvistične raziskave najdemo na Slovenskem že dobro desetletje po uradni uvrstitvi sociolingvistike med medstrokovne jezikoslovne discipline. Vendar je bilo empirično raziskovanje po pravilu uvrščeno v programe dela različnih raziskovalnih inštitutov, medtem ko ga je bilo bolj malo v raziskovalnih programih univerzitetnih in drugih akademskih ustanov, ki se poklicno ukvarjajo z vprašanji jezika, med njimi tudi ali predvsem slovenskega. Tudi od tod morda zgoraj omenjeno mašilo o mladosti sociolingvistike, čeprav tako rekoč pred našimi očmi nastajajo vedno nove interdisciplinarne povezave pri raziskovanju jezika in soodnosnih pojavov in s tem nove medstrokovne, res mlade jezikoslovne discipline. Sociolingvistika danes med njimi dosega skorajda že častitljivo starost. Ob omenjenih zadržkih do sociolingvistike je treba kot svetel zgled izpostaviti zasl. prof. dr. Bredo Pogorelec, ki je s svojo pedagoško in raziskovalno avtoriteto spodbudila sociolingvistično raziskovanje na območjih jezikovnega stika slovenščine z manjšinskimi oz. sosedskimi jeziki. Tudi njenim spodbudam je namreč treba pripisati dejstvo, da imamo danes na Slovenskem najbogatejšo sociolingvistično raziskovalno bero prav na področju raziskovanja jezikovnega stika, medtem ko se v zadnjih desetletjih po zaslugi njenih študentov, danes univerzitetnih učiteljev, tudi sociolingvistična tematika – da ne rečem problematika – slovenskega osredja in slovenskega jezika prebija na raziskovalni repertoar. Toliko o kratkem oziru v preteklost.

Že samo *Sociolingvistično iskrenje*, objavljeno v spomin na delo prof. dr. Brede Pogorelec, ki je vir spodbude za današnji posvet, nudi določen, četudi zožen vpogled v aktualno raziskovalno sociolingvistično bero. Zato puščam ob strani drugo vprašanje organizatorjev današnjega posveta: »Kje in kaj se na področju sociolingvistike že dogaja, kako vrednotite to delo?«. Zavzela pa bi se

za kontinuirano spremljanje in arhiviranje sociolingvistične raziskovalne produkcije. Mogoče bi bilo smiselno za to poskrbeti tudi z uvrščanjem te teme na sezname seminarskih del v nižjih letnikih in bibliografske podatke objavljati v kateri od slovenskih lingvističnih revij. (Morda po vzoru prof. dr. Neldeja, ki je s svojimi sodelavci na ta način oskrbel imeniten vir informacij o evropskih sociolingvističnih študijah in raziskavah v letnih pregledih v mednarodni reviji *Sociolinguistica*.) To bi prispevalo tudi k zmanjševanju večnega problema diseminacije raziskovalnih izsledkov, ki je, kot je opozorila kolegica Grgič, res videti kot gordijski vozel. Ob pisanju različnih strokovnih ocen me dostikrat začudi pomanjkljivo navajanje relevantnih raziskav domačih avtorjev, kar vsiljuje pomisel, da tudi strokovnjaki ne poznamo prav dobro strokovnega izplena svojih kolegov. Potem ni čudno, da naši izsledki, četudi so podani v primerno urejenih poročilih, pač še toliko težje dosežejo ušesa odgovornih za različne družbene odločitve.

Pri razmišljanju o odgovorih na vprašanje »Kaj so glavni aktualni izzivi za slovensko sociolingvistiko?« je treba izhajati iz samega bistva sociolingvistike. Vemo, da sociolingvistika svoj predmet in raziskovalna vprašanja gradi na dejstvu, da se v sporazumevalnem procesu skozi izbiro jezikovnih različic razkrivajo in tudi razkrijejo družbene značilnosti govorca. Na eni strani torej proučuje jezikovne možnosti oz. jezikovna sredstva, ki jih imajo govorci na razpolago pri vzajemni izmenjavi informacij in svojih misli, na drugi pa jo zanima to, kar govorci z izbiro in načinom rabe jezikovnih različic implicitno pripovedujejo o sebi, svoji pripadnosti različnim družbenim skupinam in različnim skupnostim, o svojem dojemanju sogovorcev, njihovih govornih vlog itd. Sociolingvistika je namreč prav to: opazovanje in raziskovanje tega součinkovanja/interakcije oziroma celo sovisnosti med variantnostjo in socialnimi in drugimi pripadnostnimi opredelitvami ter vlogami govorcev. Na eni strani z opazovanjem jezikovne izbire raziskuje implicitna družbena pravila govornega obnašanja, na drugi pa odkriva posledice družbenih razmerij, ki se zrcalijo na spremembah v jezikovnem tkivu.

Seveda lahko slovenski raziskovalci s proučevanjem slovenske sociolingvistične scene, ki je določena v prvi vrsti s slovenskim jezikom, na poseben način prispevajo k razvoju same discipline, torej tudi k novim teoretskim premisam. Gotovo je za vsakega od nas odkrivanje novih smeri razmišljanja o jezikovnih

pojavih, razvijanje novih konceptov in raziskovalnih prijemov velik, pomemben in vedno aktualen izziv. Vsaj za tiste, ki so z dušo zapisani znanosti. Vendar ima sociolingvistično raziskovanje tudi pomembno aplikativno vlogo – konec koncev sociolingvistika sodi v široko področje uporabnega jezikoslovja, kjer je v ospredju dejstvo, da izsledki medstrokovnega jezikoslovnega raziskovanja širijo razumevanje vloge jezika pri delovanju družbe in oskrbujejo s potrebnim znanjem tiste, ki so odgovorni za z jezikom povezane odločitve oz. za odločitve, ki se nanašajo na jezik.

To spoznanje me vodi do poskusa odgovora na tretje, s prvim tesno povezano vprašanje: »Katere nove oz. dodatne raziskave bi bile najbolj potrebne?«. Repertoar sociolingvističnih tem je obsežen in pester in sivih lis je veliko. In vedno je treba upoštevati avtonomnost raziskovalca. Tega se zavedam in to spoštujem. Prepričana pa sem, da je poleg bolj ali manj solističnih sociolingvističnih raziskav čas za celovito raziskavo jezikovnega stanja v slovenskem jezikovnem prostoru, takšno, ki bi osvetlila smeri jezikovne spremembe in dinamiko družbeno-jezikovnih razmerij. Mislim na kompleksno raziskavo, oprto na teorijo variantnosti, teorijo etnolingvistične vitalnosti, teorijo komunikacijskega prilagajanja, in na še kakšno od relevantnih, tudi novejših teorij … Mislim na pravo, pristno longitudinalno raziskavo, ki bi zajemala podatke na zajetnih, dobro strukturiranih vzorcih istih govorcev z ustreznimi dopolnitvami, v recimo 3- do 5-letnih panelnih rezih (primer japonske raziskave od sredine 50. let prejšnjega stoletja), s sproti posodobljenim instrumentarijem, ki ga pripravijo strokovnjaki računalniške sociolingvistike (no, ta veda je res mlada), ter z natančno načrtovanimi analizami. Mislim na raziskavo, ki bi prinesla verodostojne posplošitve. To bi morala biti nacionalna raziskava, za katero bi se zavzela država in jo dolgoročno financirala zunaj običajnih ARRS razpisov in bdela nad njenim kontinuiranim dolgoročnim izvajanjem.

Naj na hitro omenim samo najbolj očiten aplikativni izplen takšne raziskave:

1. ogromen jezikovni korpus, uporaben za najrazličnejše nadaljnje jezikoslovne raziskave na različnih strokovnih in medstrokovnih področjih;

2. verodostojni podatki za konsistentno, kontinuirano, dolgoročno načrtovanje jezikovno političnih ukrepov;

3. verodostojni podatki za nadaljnje jezikovno načrtovanje;
4. verodostojni podatki za preverjanje ustreznosti jezikovno političnih ukrepov
5. itd., itd.

In nenazadnje, takšna raziskava bi odgovorila na zaskrbljenost tistih, ki slovenščini napovedujejo črno prihodnost. Upravičenost te zaskrbljenosti bi bodisi potrdila bodisi ovrgla. Eksaktno.

Seveda bi bilo mogoče takšno raziskavo izvajati samo v tvornem timskem sodelovanju kompetentnih raziskovalcev, v ciljnem sodelovanju, da ne rečem v športnem duhu, ne glede na ustanovo, kjer so trenutno zaposleni. Smo pač mali narod z malim jezikom z malo strokovnimi močmi, pa z enako velikimi ali še večjimi potrebami kakor večji narodi in jeziki. Zato je za velike projekte treba združiti moči.

Pozdravljam ta mini sociolingvistični posvet v upanju in z željo, da je to začetek dolgega sodelovanja, če že ne začetek čudovitega prijateljstva, pri raziskovanju jezikovne spremembe ter prepoznavanju zunajjezikovnih, družbenih impulzov, ki do nje pripeljejo.

Hvala za vašo pozornost!

Maja Bitenc: Najlepša hvala. Zdaj pa vabim redno profesorico doktorico Sonjo Novak Lukanovič, predavateljico na Oddelku za primerjalno in splošno jezikoslovje ljubljanske Filozofske fakultete in direktorico Inštituta za narodnostna vprašanja.

Sonja Novak Lukanovič: Spoštovane kolegice in spoštovana kolega. Najprej bi se rada zahvalila organizatorjem, da so se odločili in sklicali takšno srečanje, ki ga osebno ne dojemam kot predavanje o pomenu sociolingvistike oziroma predstavitev sociolingvističnih tem, ki jih vsi dobro poznamo, ampak ga dojemam kot srečanje kolegov oziroma srečanje raziskovalcev, kjer se bomo sestali in začeli z diskusijo, pogovorom o izzivih v sociolingvistiki. Čeprav sem profesorica, torej delam s študenti, pa sem po duši raziskovalka. In tukaj bi se najprej zahvalila svoji mentorici in sodelavki, ki me je vpeljala v čar in svet raziskovanja, profesorici Albini Nećak Lük, s katero sem

vstopila v raziskovalno področje uporabnega jezikoslovja. Skratka, že pred mnogimi, skoraj 30 leti, smo začeli z raziskavami o etnolingvistični vitalnosti, o problematiki jezika v stiku, o stališčih do jezika. Seveda ne samo v smislu teoretičnega razpravljanja, ampak predvsem z empiričnim pristopom in mislim, da je taka empirija, torej terensko delo, tisto, kar mora biti srčika raziskovanja in mora vedno ostati najpomembnejše v našem sociolingvističnem raziskovanju.

Seveda empirično raziskovanje ni lahko. To zahteva zelo veliko teoretičnega znanja, znanja iz raziskovane tematike, metodološkega znanja, statistike in nenazadnje tudi znanja o okolju in področju, v katerem raziskava poteka. Zelo lahko je narediti vprašalnik, zelo lahko je napisati teoretično delo, zelo težko pa je priti na teren in tam ugotoviti, kaj se dogaja. Zakaj? Predvsem zaradi jezikov v stiku. Vsi vemo, da družbeni, zgodovinski, ekonomski in politični faktorji ustvarjajo ter oblikujejo jezikovno stična območja, kjer pa ima vsak jezik svoj status in položaj. Izbira jezika je vedno odločitev posameznika, toda v takem okolju je ta odločitev družbeno oziroma politično determinirana in odraža klimo, v kateri posameznik lahko živi v svojem jeziku oziroma ima pravico živeti v svojem jeziku. To sta izhodišči, ki nam zelo natančno pokažeta, kaj se dogaja z jezikom oziroma s posameznikom, ki ta jezik govori. Stična območja bi morali raziskovati, saj na njih živijo in govorijo predstavniki manjšine, večine in tudi drugi. In razlika med jezikom večine in manjšine se v različnih okoljih seveda ne odraža samo v številčnosti, ampak predvsem v obsegu pravic, privilegijev in pa tudi moči. Prav ta odnos med jezikom in močjo je zlasti prisoten v jezikovno stičnih območjih. In ko govorimo o našem področju, mnogokrat pozabljamo, da jezik ni samo sredstvo komunikacije: jezik ima veliko simboliko, ima pa tudi moč. Moč, ki jo mi, sociolingvisti, premalo poudarjamo in izpostavljamo v družbi. Družba se ne zaveda, kakšno vlogo in kakšen pomen ima jezik. In tu se dotaknemo znanosti: ali ni sociolingvistika del znanosti? Čemu je namenjena znanost? Kaj so rezultati našega znanstvenega dela? Rezultati našega dela niso samo objave člankov, predavanja študentom … rezultate našega dela je mogoče aplicirati tudi na širšo družbo. In nenazadnje bi nas morala politika poslušati in bi morala biti pripravljena izvesti spremembe na področju jezikovne politike, zlasti spremembe, ki se nanašajo na vlogo, položaj in učenje jezika.

Težko mi je, ko vam govorim, da se družboslovci in humanisti, h katerim sodimo jezikoslovci, pogosto čutimo nekako odrinjeni, ko se v družbi obravnavajo dosežki in pomen znanosti. Pri tem se vedno spomnim misli iz knjige *Koristnost nekoristnega*, da sta družboslovje in humanistika, kamor sodi tudi jezik, kot plodovna tekočina, v kateri se razvijajo ideje demokracije, enakopravnosti, razvoja. Zato bi morali družbi predstaviti naš pogled na vlogo in moč jezika. Jezik nima samo simbolnega pomena komunikacije, ampak odraža tudi odnose med skupinami, med večino, manjšino in drugimi. In vse to smo mi raziskovali že mnogo let, prav na projektu o medetničnih odnosih v slovenskem etničnem prostoru, ki je bil tako v metodološkem kot v empiričnem pogledu vzorčen – vse je potekalo tako, kot je napisano v teoretičnih delih. In prav v tem projektu se je tudi pokazalo, kako jezik odraža odnose med posameznimi skupinami, predvsem, ali te skupine živijo ena z drugo ali ena mimo druge. Tudi ko obravnavamo manjšinsko tematiko – sem pač raziskovalka na Inštitutu za narodnostna vprašanja, ki se ukvarja z manjšinskimi zadevami, z etnično raznolikostjo, tako v Sloveniji kot tudi v zamejstvu – vedno stremimo k raziskovanju prostora oz. celotnega regionalnega področja. Nikoli se ne usmerjamo samo v raziskovanje manjšinske problematike, ampak širše. Kajti če želimo, da se jezik ohrani, je pomembno, da ga na nekem določenem nivoju sprejema tudi večina. In v tem kontekstu pridemo do raziskovanja stališč do jezika. Seveda se pri obravnavanju jezika v najširšem smislu dotaknemo tudi koncepta enakopravnosti, neenakopravnosti, nediskriminacije, diskriminacije, kajti te tematike so vedno zelo močno povezane z jezikom. Seveda se vsi zavedamo, da so vsa ta okolja ne samo narodnostno mešana okolja, ampak tudi obmejna področja, kjer jeziki predstavljajo tudi neke vrste mejo med kulturami. Tudi o tem še ni bilo dovolj raziskanega.

Ampak vsa ta področja, ki so danes jezikovno stična, mejna področja, niso neki zaprti otoki, ampak so polotoki, ki se jih dotikajo tudi valovi migracijskih tokov in mobilnost prebivalstva, kar seveda vpliva na spremenjeno strukturo prebivalstva, na vzorec življenja in na drugačne oziroma različne jezikovne vzorce, v katerih se spreminja tudi mesto in vloga manjšinskega jezika. Tradicionalno jezikovno-kulturno mešano območje pod takimi pogoji seveda izgublja, spreminja svojo identiteto, ampak izgubljata in spreminjata jo tudi manjšina in avtohtono večinsko prebivalstvo. Skratka, vsi smo podvrženi

migracijskim tokovom in spremembam. Tudi to je nekaj, kar na nek način še vedno kliče po nadaljnjem, poglobljenem raziskovanju.

Na inštitutu smo skupaj s profesorico Nećak raziskovali tudi že, kaj je dvojezičnost, kaj je večjezičnost. To se pravi, ali večjezičnost danes obstaja na enak način, kot je obstajala pred mnogimi leti. Ali lahko govorimo o medsebojnem vplivu, interakciji večjezičnosti na eni strani in dinamiki transnacionalizma, globalizma, evropeizma na drugi strani? Ena pomembnih posledic te dinamike je seveda tudi velika sprememba funkcij in položajev različnih jezikov v primerjavi enega z drugim oziroma jezikov v stiku, kar se odraža tako na nacionalni, regionalni kot na globalni ravni in zaznamuje nacionalne jezikovne politike in jezikovne prakse v takih okoljih.

Seveda se raziskovalno na inštitutu ukvarjamo tudi z našimi narodnostno mešanimi območji, v katerih se izvaja institucionalna dvojezičnost. Ugotavljamo uspešnost take institucionalne dvojezičnosti: kot je že povedal nekdo pred mano, imamo izredno izdelan, tako rekoč edinstven pravni sistem, v realnem življenju, v praksi, pa je mnogokrat drugače. In kaj so vzroki za to? Ali so instrumenti jezikovne politike ustrezni ali neustrezni? Pri tem se dotaknemo tudi ekonomskega vidika: država zagotavlja dodatek za dvojezičnost, finančno jo nagrajuje že od poznih 50. let prejšnjega stoletja. Preseneča nas, da so se takrat, v času socializma, odločili finančno nagraditi nekoga, ki govori dva jezika, se pravi lahko uporablja še italijanščino ali pa madžarščino. Zakaj je bilo to uvedeno, je seveda predmet širšega raziskovanja. Ampak vse to raziskujemo in ugotavljamo povezave z jezikovno politiko. Z raziskovanjem etnolingvistične vitalnosti, jezikovnega prilagajanja, družbenih procesov jezikovnega stika pridemo tudi do ekonomskega vidika jezika, do katerega so nas na inštitutu privedle predhodne raziskave. Nismo ga začeli raziskovati, ker bi v neki teoriji prebrali, da je ekonomski vidik pomemben: naše predhodne raziskave, raziskave v 90. letih prejšnjega stoletja, so nam nakazale močno povezavo ekonomskega vidika z rabo jezika. Mobilnost, medgeneracijski stiki, mešani zakoni – vse to so tematike, ki bi jih lahko še poglobljeno raziskovali.

Druga tematika, ki jo v raziskovanju moram omeniti, so Romi in romski jezik. Trenutno na inštitutu poteka pilotni projekt poučevanja romščine, kjer s pomočjo dodatnega pouka romščine za romske otroke spremljamo napredek njihovega opolnomočenja in posledično boljše jezikovne kompetence, tako v

romščini kot v slovenščini. Z enakimi cilji in namenom smo se lotili večletnega eksperimentalnega projekta uvajanja novih večjezičnih pedagoških pristopov s poudarkom na romščini. Večjezičnost je torej nova realnost in skupni imenovalec vseh raziskovalnih udejstvovanj, ki bi jih bilo treba zastaviti čim širše in multidisciplinarno. Skratka, tudi to je eden od sociolingvističnih izzivov.

Tematik za raziskovanje je tako zelo veliko. Kar bi želela poudariti in kar osebno čutim, da nam primanjkuje, je prav povezanost nas vseh. Kajti projekte izvajamo nekako parcialno in koristno bi bilo, da bi skupaj zastavili longitudinalne projekte, kajti prav longitudinalni projekti so tisti, ki odpirajo nova vprašanja in dajejo nove odgovore. Kako prepričati državo, družbo, da je jezik pomemben? Politika je za nas to, kar je za nek tehnološki inštitut gospodarstvo. Tehnološka znanost tudi ustvarja inovacije, nove produkte, ki pa ne pridejo vsi vedno v tovarno, da bodo izdelani. Tako mora tudi družba ustvarjati nove poglede, modele, opozarjati na probleme, kajti jezik je pomemben indikator identitete, ki lahko vodi v diskriminacijo, nespoštovanje, to pa je lahko vzrok nestabilnih razmer, lahko vodi v konfliktne situacije, toda o tem nikoli ne govorimo. Ni raziskav, ki bi ugotavljale, ali je jezikovni stik v Sloveniji jezikovni konflikt. Želim si, da bi nekako prišli do politike, ki nas bo poslušala in ki nam bo omogočala longitudinalno raziskovanje, katerega namen ni, da o rezultatih predavamo študentom ali objavimo članek, ampak predvsem to, da se izboljša družba v vseh pogledih. Hvala organizatorju.

Maja Bitenc: Najlepša hvala. Naslednji govornik je doktor Krištof Savski, ki je diplomiral na tukajšnjem Oddelku za prevajalstvo, magistriral in doktoriral v Veliki Britaniji pri znameniti profesorici Ruth Wodak, sedaj pa poučuje na Tajskem, na Prince of Songkla University in je z nami na daljavo.

Krištof Savski: Vsem lep pozdrav in hvala lepa za povabilo. Kot je povedala Maja, sem podiplomski študij opravil v Veliki Britaniji, po sicer zelo lepem študiju na Filofaksu, na Oddelku za prevajalstvo, zdaj pa že peto leto delam na Tajskem. Ravno te dni mineva 10 let, odkar sem se prvič odpravil iz Slovenije, tako da bom danes razmišljal o slovenski sociolingvistiki bolj od zunaj. V doktoratu sem se sicer ukvarjal s slovensko jezikovno politiko, od takrat naprej pa večinoma pišem bolj o sociolingvističnih vprašanjih angleščine in drugih jezikov v jugovzhodni Aziji, predvsem seveda na Tajskem. Torej bom morda predstavil nekoliko drugačen pogled, a upam, da bo zato toliko bolj

zanimiv. Res pa hvala za organizacijo tega zelo zanimivega in zelo dobrodošlega dogodka.

Za začetek bi se vprašal, kaj je sociolingvistika. Seveda bi lahko preprosto rekli, da je sociolingvistka veda o razmerjih med jezikom in družbo, vendar pa smo danes videli, da je to vseeno treba veliko natančneje definirati. Iz vseh prispevkov je jasno videti, da je to zelo empirična veda: če po objavah, knjigah in konferencah ocenimo, kaj sociolingvistika je, vidimo, da je to večinoma veda, usmerjena k empiričnemu preučevanju realnosti jezikovne rabe. Sociolingvistika ima danes izjemno robusten nabor kvantitativnih in kvalitativnih raziskovalnih metod, od jezikoslovnih v ožjem smislu, torej fonetike, fonologije, sintakse in drugih, do veliko bolj sociooloških in antropoloških metodologij, kot recimo analiza diskurza in etnografija, ki sta moji specialiteti – jaz sem, za razliko od večine drugih govorcev tukaj, stoodstotno kvalitativni raziskovalec. Kot je ravno pred menoj povedala Sonja Novak Lukanovič, je sociolingvistika tudi kritična veda, torej takšna, ki ne razmišlja samo o jeziku in se samo posredno dotika razmerij moči, ampak ta razmerja tudi neposredno preiskuje in o jeziku razmišlja v luči njegove vloge v razmerjih neenake moči v družbi.

Najbolj bi se osredotočil na tretjo karakteristiko današnje sociolingvistike, namreč njeno usmeritev k dinamiki v smislu, da se vedno bolj zaveda, da so statične definicije jezika in jezikov ter družbe in družb problematične. S tem mislim na strukturalistično pojmovanje jezika kot sistema s pravili in sredstvi, ki je bilo brez dvoma dominantno v 20. stoletju, tudi v sociolingvistiki, in je še danes izjemno pomembno. Variantostna sociolingvistika, ki jo je definiral William Labov in se sedaj razvija v svoj tretji oziroma morda celo četrti val, je še vedno vezana na to sistemsko pojmovanje jezika, s tem da poskuša ta sistem statistično navezovati na družbene spremenljivke, kot so prostor, spol, starost, socioekonomski razred in druge. Danes je ta sociolingvistika brez dvoma še naprej aktualna, a se hkrati pojavlja vedno več vprašanj o smiselnosti takšnega dela, zlasti spričo študij interakcije v diskurzu, ki kažejo, da se jezikovna sredstva opomenjajo kot del družbenih dejanj, torej da so umeščena v kontekst rabe; o tem na primer govori Ron Scollon, zdaj že preminuli ameriški sociolingvist. V bolj diskurzivno usmerjeni sociolingvistiki se problematizira rigidna navezava določene jezikovne varietete na neko vrsto družbene identitete: če si od tukaj, govoriš tako; če si ženska, govoriš tako;

če si heteroseksualec, govoriš tako. Naj poudarim, da se tako razmišljanje ne kaže vedno kot popolnoma napačno, pač pa se poskuša bolje upoštevati kompleksno realnost vseh družb.

To vprašanje razmerja med jezikovnim sredstvom in družbenim pomenom (indeksikalnost) ima že dolgo zgodovino. Vsestranski ruski literarni teoretik Mihail Mihajlovič Bahtin je v prvi polovici 20. stoletja pisal o večglasnosti in večideološkosti diskurza in jezika, torej o heteroglosiji in polifoniji. Po Bahtinu ideološki pomen nekega jezikovnega sredstva (čemur Silverstein reče »popolno jezikoslovno dejstvo« oziroma v angleščini *total linguistic fact*) ni enodimenzionalen, pač pa je vpet v raznolik diskurz in se ga lahko mobilizira z več ideoloških vidikov. Tovrstno večdimenzionalnost sistemska definicija jezika zelo težko zajame, kar lepo pokaže primer razprave o rabi slovničnega spola na Filofaksu, ki jo je omenila že kolegica Monika. V tej razpravi je bil z dela slovenistične stroke podan sistemski argument, da slovenščina preprosto že ima jezikovno sredstvo za izražanje spolne nevtralnosti, to je moški slovnični spol, in da predlagani ukrep fakultete (da naj se v uradnih besedilih uporablja nevtralni ženski slovnični spol) zaradi tega ni sprejemljiv, saj naj bi posegal v nespremenljive notranje zakonitosti jezika. Tak sistemski argument preprosto ne zmore zajeti dejstva, da se očitno v diskurzu to sredstvo, torej slovnični spol, opomenja z več ideoloških zornih kotov, da za neke uporabnike pomeni nekaj, za druge pa ima drugo ideološko razsežnost. Tudi zato v zelo specifičnem kontekstu rabe jezika (uradna besedila na FF) premik od rabe enega sredstva k drugemu ne more pomeniti posega v statičen jezikovni sistem, pač pa je del naravne geneze diskurza oziroma razvoja rabe jezika v družbi. Vse to nam pove, da jezik s sociolingvističnega vidika ni enodimenzionalen in ni statičen. Hkrati s sociolingvističnega vidika jezik tudi ni avtonomen oz. ločen od drugih ravni tvorjenja pomena, saj je na voljo veliko raziskav, v katerih se jezikovna raba obravnava kot del raznolikega nabora semiotičnih sredstev, ki so na voljo članom neke dane družbe. Lep primer je delo o pisni rabi jezika v javnih prostorih, po slovensko jezikovna krajina (angleško *linguistic landscape*), kjer se raziskuje skupno opomenjanje jezikovnih in drugih vizualnih sredstev, na primer rabe barv, umeščanja elementov v prostor, velikosti in tako dalje.

Še ena pomembna točka je, da je sodobna sociolingvistika veda, ki raznolikost in večjezičnost – zdaj se v evropskem kontekstu govori o raznojezičnosti

(angleško *plurilingualism*) – pojmuje kot naravno stanje v vseh družbah, ne pa kot izjemo. Vedno več se nasploh razmišlja o tem, da sama ideja poimenovanih jezikov – na primer slovenščine, hrvaščine, italijanščine – in varietet – na primer knjižni jezik ali pogovorni jezik – ni nevtralna, ni samo znanstveno določena, pač pa je produkt ideologije in služi interesom moči. Torej vprašanja, kaj je slovenščina, kaj je dobra oziroma prava slovenščina, niso nevtralna, niso neko naravno dejstvo, ki ga sociolingvistika opiše, pač pa so družbeno konstruirana in odsevajo ideologije, kot je recimo nacionalizem in elitizem. To je zelo relevantno predvsem pri manjšinskih govorcih, ki jih je omenjala že Matejka Grgič, saj pri njih ta tradicionalni model nacionalne države, polne enojezičnih govorcev, ne deluje, posledica tega pa je, da je njihova raznojezičnost pogosto marginalizirana, ko pride do izobraževalnih kontekstov. Tukaj se je sicer potrebno vprašati, ali nismo vsi raznojezični na različne načine, saj nihče od nas ne govori in piše vedno na enak način. V tej smeri se zdaj v sociolingvistiki veliko razmišlja ne samo o odmikanju od sistemskega razumevanja jezikov, pač pa tudi o odmiku od same ideje jezikov kot zamejenih sistemov, ki so ločeni od kontekstov rabe ter tudi od uporabnikov. Namreč, kot je omenila Karmen Kenda-Jež, pri sistemskem pojmovanju se hitro zapade v obravnavanje jezikov oziroma varietet kot »sistemov brez uporabnikov«, torej kot nečesa, kar posamezniki samo pasivno realizirajo, ne da bi imeli možnost v ta sistem posegati, ga definirati. Tu s sociološkega vidika govorimo o vprašanju, kje je jezik v razmerju med strukturo in agentnostjo. O tem razmišlja na primer zelo sodobna literatura okoli koncepta *translanguaging* (morda bi lahko slovenili kot »čezjezičenje«), ki je poskus konceptualizacije rabe jezikovnih sredstev kot dinamičnega procesa, ki je vpet v zelo specifične družbene okoliščine in kontekste in kjer zgodovinske, politične meje poimenovanih jezikov ne štejejo vedno absolutno, tako kot to določajo ideologije nacionalne države.

Mislim, da sociolingvistiki slovenščine ta razvoj vede, ki ga jaz tu zelo površinsko povzemam, postavlja nekaj specifičnih izzivov. Glavni izziv za mene kot raziskovalca je nujen odmik od ideologije nacionalnega standardnega jezika in nanjo vezane tradicije predpisovanja, ki danes slovenistiko še naprej definira. Tukaj je, kot so povedale govorke pred mano, treba predvsem kultivirati znanstveno kulturo, ki se avtentične, prave jezikovne rabe ne boji, ampak jo zanima, ne glede na standardno oziroma nestandardno rabo in ne glede na

osebne jezikovne sodbe ali znanstvene predpostavke. Potrebno je spoznanje, da v vsakem kontekstu jezikovne rabe dihotomija standardno/nestandardno (oziroma po domače prav/narobe) ni relevantna. Tam, kjer pa je relevantna, pa tudi ni statična, pač pa je pomensko umeščena v kontekst, z njo pa se torej ta prostor med knjižnim in pogovornim jezikom, ki ga je omenila Karmen Kenda-Jež, tudi premika. Treba je tudi spoznati, da dihotomija standardno/nestandardno ni samo dinamična, ampak je tudi vpeta v razmerja družbene neenakosti in je zato ne moremo obravnavati kot nevtralno znanstveno dognanje, pač pa jo je treba obravnavati kritično. To tudi pomeni, da je potrebno relativizirati znanstveno vedenje o jeziku in razumeti, da so jezikoslovci sami tudi del govorne skupnosti ter da njihova spoznanja odsevajo družbene razmere. Nadalje je treba več razmišljati o tem, da je slovenska jezikovna skupnost izjemno raznolika, raznojezična in vpeta v procese kulturne in jezikovne izmenjave, ki jih ne moremo opisati zgolj glede na pripadnost nacionalnim skupinam ali jezikom. Angleščine na primer ne moremo obravnavati preprosto kot vpliv »anglo-saksonske« kulture na slovensko, saj so lahko angleški jezikovni elementi nosilci mnogih različnih kulturnih pomenov, tudi slovenskih.

Za zaključek, predvsem je treba spoznati, da slovenščina ni ena sama, ampak da slovenščine so, da vse obstajajo v rabi, da se vse spreminjajo in prepletajo ter, na koncu, da so vse vredne enakovredne, ozaveščene sociolingvistične analize. Hvala lepa.

Maja Bitenc: Najlepša hvala! In za konec še redni profesor Marko Stabej z Oddelka za slovenistiko Filozofske fakultete, tudi predsednik Centra za slovenščino kot drugi in tuji jezik.

Marko Stabej: Lep pozdrav. Jaz bom tako rekel: ni čudno, da je Krištof s takimi in tako izraženimi stališči našel službo na Tajskem, ker je tam dovolj daleč, da se zdijo nenevarna. Pravzaprav bi najraje kar obmolknil, ampak čisto ne bom. Samo nekaj opomb.

Jaz sem pritaval v sociolingvistiko iz stilistke, ker nisem mogel razumeti, kako je v 19. stoletju lahko veljalo za kvalitetno literarno delo nekaj, kar je sodobnemu bralcu popolnoma nezanimivo, in ni bilo drugačne možnosti, kot to gledati kontekstualno. Ena od osrednjih pomanjkljivosti slovenskega jezikoslovnega prostora je to, da je v resnici jezikoslovje sluga v rokah ideologije, in to ne

nujno strankarske ideologije, ampak globoke ideologije nacionalnega, in da ima, čim hoče izstopiti iz tega, hude težave, saj se sicer zdi popolnoma nepotrebno. In ne gre samo za razmerje družba – jezik, tovrstno prepletanje in součinkovanje, ampak tudi na splošno, s kontekstom. Slovensko jezikoslovje, vsaj v *mainstreamu*, je še vedno pretežno dekontekstualizirano oziroma si konteksta jemlje samo toliko, kolikor ga nujno rabi, da lahko ohranja videz nujne potrebnosti za nadaljnje življenje neke nacionalne in državne skupnosti, ki je domnevno trdna in od nekdaj je in vedno bo. Pa čeprav vsi vemo, da se spreminja, ampak marsikdo noče niti pomisliti, da bi to spreminjanje sprejel, kaj šele, da bi ga resnično raziskoval. Zato je sociolingvistika, kot je bilo tu malo nakazano, v slovenskem prostoru pravzaprav slabšalna oznaka, če že ne zmerljivka. Deloma tudi upravičeno: kot smo že večkrat slišali, je sociolingvistika težko področje za kvalitetno raziskovanje. To kaže tudi primer omenjene štiri leta stare raziskave, ki je bila ultra pomanjkljiva, v marsičem tudi zavajajoča, v zvezi z njo pa je bil problematičen tudi marketing. Kar je morda najhujše: če misliš, da dobiš rezultate, pa sploh ne veš, kaj imaš. In v tem smislu je najbrž tudi ideja kolegice Albine Nećak Lük, da šele longitudinalno raziskovanje prinese zanesljive rezultate, čeprav je ob spremenljivosti situacije ta longitudinalnost in njena zasnovanost najbrž zelo težek izziv. A sam mislim, da je ravno v ta izziv treba ugrizniti in po mojem je predlog, da enkrat poskusimo prijaviti nek skupni projekt, vsaj neke parametre, dober. To bi bil nek korak. Ker sicer samo govorimo, kako bi bilo lepo, da se povezujemo, v resnici pa vsak hoče samo preživeti sebe in svojo institucijo. In to je tudi težava slovenskega prostora, zato ker nenehno govorimo, kako bi bilo dobro, v resnici pa se sodeluje tako rekoč nič, še znotraj oddelkov na fakultetni ravni ne, kaj šele na medinstitucionalni.

Jaz bi rekel, da imamo dve možnosti. Če bi bili kritično in znanstveno uspešni, smo prave zgage v slovenskem prostoru, saj bi zamajali vse, kar se zdi vsem, od levih pa do najbolj desnih politikov, samoumevno. »Jezik je identiteta, brez knjižne slovenščine ni naroda« in tako naprej, mi pa začnemo nekaj šumeti, nato pa uspemo mogoče celo dokazati, da to sploh ne drži. Nihče ne potrebuje tega. Oziroma jaz mislim, da vsi potrebujemo to refleksijo, ampak zato bi rekel, je težava biti sociolingvist, ker je težko empirično raziskovati in potem to modelirati in nekako komunicirati, čeprav se strinjam s kolegico Matejko, da

so šele takrat aplikativne rešitve zares možne. Ker če načrtuješ nekaj, kar je nerealno in neizvedljivo, lahko načrtuješ nov projekt, pa še en nov projekt, pa tega nikoli ne boš mogel uresničiti. In takšnih projektov imamo tu vse polno, recimo iskanje idealnega govorca za normiranje govorne slovenščine za naslednjih 30 let. Takšnega idealnega govorca, kot smo tukaj skoraj vsi prepričani, ni in ga nikoli ne bo. In še vedno se ga išče. Ne samo šest oseb, ampak cela stroka išče tega govorca.

Če pa sociolingvisti in sociolingvistke nismo kritični in uspešni, smo pa tako ali drugače samo v službi te ali one ideologije. Da malo še z druge strani govorimo o statusu, o jezikovni krajini, o takšnih in drugačnih zadevah, a pravzaprav samo ozaljšamo neko že obstoječo ideologijo. In dober primer tega je recimo nesorazmerje v novi resoluciji o nacionalnem programu jezikovne politike, kjer uvodno poglavje o jezikovni krajini govori o jezikovni lojalnosti, torej da jo je treba razvijati. Nič se pa ne vpraša, kateri lojalnosti, čigavi, do katerega jezika, do vseh jezikov ali samo do avtohtonih. Izgleda pa zelo sociolingvistično, na prvi pogled. To je takšno prekletstvo.

Če smo humanisti – in jaz mislim, da smo – ne smemo samo govoriti, kaj vse jezik je, ampak moramo pravzaprav razumeti, s tem da raziskujemo – ne nekaj, kar je pomembno za slovensko državo ali slovenski narod ali slovenski jezik, ampak nekaj, kar je pomembno za človeškost. Ampak to je pedagoško in raziskovalno težko. Ne samo glede tega, da je zahtevno narediti raziskovalni projekt s terenskimi raziskavami, z modeliranjem in tako naprej. Težko je tudi učiti, težko je mentorirati diplome, če študenti niso pripravljeni prebrati niti dveh do treh knjig. Če lahko v bolj *hardcore* jezikoslovju nekaj precej hitro razvrstiš, umestiš in imaš potem občutek, da si nekaj novega dodal veliki stavbi slovenskega sistema, čeprav ne veš, koliko slovenščin je, kakšne so, kako jih dojemamo. Tukaj se mi zdi nekakšna groza početi nekaj, kar bo destabiliziralo splošno predstavo o jeziku, in zato tudi tega ne moremo skomunicirati. Ker imamo v bistvu premalo podatkov in vedenja, hkrati pa tudi okolje ni naklonjeno temu, da bi razvili nek narativ, kjer bi različnost in jezikovno ranljivost začeli obravnavati kot normalno od vrtca naprej. Ampak se nenehno perpetuirajo ideološka vprašanja, kaj je potrebno, kaj ne … Ni čudno, da se potem pojavljajo takšne metafore: predstojnik Inštituta za slovenski jezik je imel v Mladini intervju in je javno izjavil, da jezikoslovci nismo več psi čuvaji, ampak

smo psi vodniki. Jaz se ne počutim kot pes vodnik. Pes vodnik koga? Ali so govorci slepi? V tem smislu obstaja neka dominantnost znanstvenega jezikoslovja, ki ga lahko primerjamo s tem, da bi nam dobri mehaniki ali pa celo fiziki urejali prometno situacijo, prometno infrastrukturo. To ne gre. Tisti, ki nam je do tega, ki nas to zanima, moramo to zanimanje združiti, napisati en skupni, večji projekt. Seveda z njim ne bomo vsega rešili, bi pa tukaj lahko bi en konkreten korak. Hvala lepa.

RAZPRAVA

Maja Bitenc: Nadaljujemo torej z razpravo. Marko, ti si odločno rekel, da bi pripravili projekt, potem upali na dolgotrajno financiranje in longitudinalno raziskavo, obširno in celovito in polno stvari obsegajočo, ampak se mi zdi, da smo vseeno govorili o zelo različnih stvareh. Eno bi bil projekt v smislu spremljanja avtentične jezikovne rabe in jezikovnih stališč. Potem, kar je izpostavila Monika, bi morali raziskovati širše družbeno zelo relevantne teme, kot je slovenščina v šoli, jezik in razne ideologije. Ali ima to dvoje nek skupni imenovalec oziroma kako bi bilo lahko prvo mogoče uporabno za drugo oziroma ali so to povsem ločene stvari in ločeni projekti? Glede sklopa o jeziku in šoli se zastavlja vprašanje, kje začeti. Meni se zdi pomenljivo, ko otrok sprašuje, zakaj se nekaj uči. Ali ko maturanti tarnajo, da znajo vse mogoče stvari, niso pa pismeni, se ne čutijo suverene pri rabi slovenščine. Kaj je smiselno v šoli početi in kako je to povezano s študijem slovenščine in vlogo jezikoslovja, stanjem duha v jezikoslovju, kot si ga zelo dobro opisal? Ne vem, ali imamo pri vseh teh idejah res nekaj, za kar bi rekli: »To bi bilo potrebno najprej.« Ali oziroma koliko in kako so te različne pobude združljive in povezane med sabo?

Nataša Gliha Komac: Jaz bi se tu oglasila. Naša raziskava pred štirimi leti je imela vrsto pomanjkljivosti, a smo se iz nje tudi marsikaj naučili. Zdi se mi povezano s tvojim vprašanjem, zlasti zaradi tega, ker smo se mi takrat lotili resnično velike in kompleksne raziskave, v katero so bili in smo bili vključeni raziskovalci in strokovnjaki z zelo različnih ustanov, zraven pa smo poskušali vključiti tudi jezikovne uporabnike, tako predstavnike različnih raziskovalnih, izobraževalnih in strokovnih ustanov in služb kot čisto konkretne jezikovne uporabnike, specializirane in splošne. Kot je bilo prej izpostavljeno, so se porajala zelo različna vprašanja in soočili smo se s težavo, kako naenkrat, v eni

raziskavi, pokriti različna področja in različna vprašanja ter hkrati priti do celostnega vpogleda. Raziskava je potekala tako kratek čas, da je bilo to skoraj nemogoče, a ta izkušnja je vsakemu posebej dala kar nekaj spoznanj, vzpostavila se je mreža raziskovalcev in mislim, da smo si takrat zastavili številna vprašanja, ki se zdaj skozi čas razvijajo naprej in dobivajo odgovore oz. nova vprašanja. Preizkušen je bil kompleksen metodološki aparat in zelo različne tehnike zbiranja podatkov. Tako da jaz na to, da je imela ta raziskava samo pomanjkljivosti, nikakor ne bi pristala, hkrati bi pa še opozorila, da je bila ta raziskava zelo večplastna. Najprej smo se po delovnih skupinah dobivali poznavalci zelo različnih področij, se pogovarjali o problematikah, soočali zelo različna mnenja in vprašanja. Drugo je bila izvedba spletne ankete ter nato iskanje povzemalnih odgovorov. Ob tem je bilo opravljenih še več manjših, izrazito ciljno naravnanih raziskav, npr. med uporabniki slovenskega znakovnega jezika, med uporabniki z motnjo v duševnem razvoju, anketna raziskava med načelniki vseh 58 slovenskih upravnih enot in še kaj bi se našlo. Gradivo še čaka na obdelavo, saj je bil to ciljni projekt in je bilo treba priti do rezultatov in smernic v zelo kratkem času, ki smo jih zato zaradi hude časovne stiske vlekli dobesedno iz rokava, pri čemer pa so se o posameznem področju vedno odločali strokovnjaki, poznavalci področja, in sicer tako na osnovi specializiranih znanj in izkušenj kot tudi na konkretnih rezultatih in spoznanjih opravljenega raziskovalnega dela. To je bil zelo zahteven in kompleksen poskus celostne raziskave, kjer smo se vsi naučili, da takšna raziskava potegne za sabo garanje in zelo veliko različnih znanj in združevanje zelo različnih področij. Kar pa je vse prej kot enostavno. In to je nesporno konkreten primer empirične izkušnje in poskusa sodelovanja in raziskovanja celotnega slovenskega jezikovnega prostora, za katerega mislim, da je bil tudi uspešen. Sicer pa sem res vesela današnjega srečanja, saj so izmenjave mnenj, pogledov, tudi zgolj »slišanje« danes povedanega dragoceni za vse nas, ki se tako ali drugače ukvarjamo z jezikoslovjem.

Albina Nećak Lük: Zelo se strinjam z vsemi, ki so opozorili na dinamično naravo družbe in na jezikovne posledice te dinamike. Vendar mislim, da to ni zadržek za skupno raziskavo, ki bi pokazala, kakšno je pravzaprav stanje v posameznih panelnih rezih; to bo namreč pokazalo tudi jezikovne posledice družbene dinamike. V prispevku sem navedla nekaj tradicionalnih teoretskih konceptov, omenila pa sem še, da so na razpolago tudi drugi: to je stvar

načrtovanja takšne raziskave. Pomembno je, ali obstaja volja, želja prepoznati stanje, ugotoviti, kakšni procesi potekajo v družbeni komunikaciji, kaj se ob tem dogaja z jezikom in kaj pravzaprav to dogajanje pomeni za prihodnost te dinamične, razvijajoče se skupnosti. Ampak če se odpovemo spremljanju tega razvoja, tega spreminjanja, potem smo se odpovedali tudi eksaktnemu pogledu na jezikovno stanje, na jezikovna razmerja in dinamiko ter s tem na jezikovno spremembo. Zakaj toliko poudarjam eksaktnost? Nekajkrat ste omenili reševanje. Jaz ne govorim o reševanju. Kaj bomo reševali? Mi smo znanstveniki, mi moramo stvari prepoznavati, ugotavljati, ali stvarnost neke predpostavke, ki jih preverjamo, potrjuje ali ne. Rešujejo naj drugi ob upoštevanju izsledkov naših raziskav, mi moramo dati odločujočim v politiki, tudi jezikovni, verodostojno preverjene podatke o stanju, trenutnem, prihodnjem. Ta ambicija nekaj reševati me rahlo moti, ker je to naloga nekoga drugega. Res pa je, da morajo biti rezultati naših raziskav verodostojni, vedno znova preverljivi in omogočati morajo posplošitve. Povrh tega je treba podatke in izsledke pripraviti tako, da bodo berljivi tudi za tiste, ki so odgovorni tako za družbeno spremembo kakor tudi jezikovno spremembo znotraj nje. In še nekaj: Ker se zavzemam za longitudinalni projekt, moram poudariti še, da je za dober raziskovalni projekt potrebno tudi financiranje načrtovanja projekta. To se mi zdi bistveno. Projekt, ki naj bi bil longitudinalen, sploh dolgoročen in bi segal v neko neznano prihodnost, za katero upam, da jo boste doživeli, je treba pripraviti z vidika današnjega znanja in s predvidevanjem o vključevanju novih konceptov. Bistvena značilnost longitudinalne raziskave je namreč tudi to, da vključuje znanja in spoznanja, ki nastajajo med samim raziskovanjem. To pa ni majhno delo, zato je treba že v predpripravah predvideti čas in denar za načrtovanje projekta.

Marko Stabej: Albina, tudi ta tvoja misel o pregledu sociolingvističnih ugotovitev v revijah …

Albina Nećak Lük: Predlagala sem, da se kakšna jezikoslovna revija posveti tudi temu, da recimo vključi letni pregled domače sociolingvistične raziskovalne produkcije. Takšni pregledi bi bili koristni tudi za katerokoli drugo strokovno področje.

Marko Stabej: Ja, samo nekako nihče tega ne dela več. Mogoče moramo vzpostaviti nek forum, kjer bi lahko to tudi nekako pisno oblikovali, mogoče

tudi moderirali … Dobro, saj bo ta razprava tudi zapisana, ampak če bi še malo naprej razmišljali. Seveda ne more biti vse v eni raziskavi, to je gotovo, da ne. Jaz nisem želel biti kritičen, da nič ni dobrega, seveda je bilo narejenega ogromno dela. Ampak, to je približno tako, kot če bi s sredstvi za hišo želel zgraditi stolpnico, ker ne moremo zanikati, da tu ni bilo nekega namena ustvarjati največje, najboljše in tako naprej. Skratka, po mojem nek forum ali pa mogoče načrtujemo in poskušamo dobiti posebno tematsko, sociolingvistično številko, ker, roko na srce, najbolj reprezentativna *Slavistična revija* ima bolj malo sociolingvistike, milo rečeno, in najbrž so tudi neki razlogi za to. Je pa več možnosti, da bi mogoče do česa takšnega prišli kje drugje, da bi na kakšen tak način poskušali naprej razvijati.

Albina Nećak Lük: To ni ključno, to je bil pač eden od mojih predlogov. Jaz se niti ne bi zavzemala za to, da se naredi kakšna posebna revija. Društvo za uporabno jezikoslovje je nekaj časa izdajalo revijo *Uporabno jezikoslovje*, ampak zdaj to spi. Če bi se kdo potrudil, bi tudi *Uporabno jezikoslovje* lahko opravljalo tudi takšno funkcijo. Razumem, da je to zelo, zelo težko. Če za tem ne stoji institucija in je treba računati na to prostovoljno delo, potem je brez izrazite notranje motivacije, kot pravi Sv. Avguštin, težko izvesti kakršenkoli projekt. Časi prostovoljstva so, kot je videti, minili. Na internetu je danes res marsikaj dosegljivo. Vseeno pa ne bi bilo odveč k pripravam takšnih pregledov pritegniti študente v okviru seminarjev ali kako drugače.

Nataša Gliha Komac: Mislim, da je imela profesorica v mislih mednarodno sociolingvistično revijo *Sociolinguistica*, kjer vsako leto ena institucija v posamezni državi poroča o temeljnih objavah, delih s področja sociolingvistike. Pri nas, kolikor vem, je nekaj časa za to skrbel Inštitut za narodnostna vprašanja.

Monika Kalin Golob: Mogoče bi predlagali uredništvu, da se ena tematska številka *Slavistične revije* na to temo naredi.

Tina Lengar Verovnik: Jaz nisem sociolingvistka, sociolingvistiko samo spremljam, že odkar me je profesorica Albina Nećak Lük navduševala med študijem splošnega jezikoslovja, to je zdaj že skoraj 30 let. Ampak zelo všeč mi je zamisel o nekem repozitoriju sociolingvističnih člankov, raziskav, zato ker sem vedno znova navdušena nad tem, kaj vse v našem prostoru nastaja, kaj vse najdem, ko usmerjeno iščem za potrebe svojega predvsem pedagoškega

dela, včasih tudi raziskovalnega, ko kaj sežem proti sociolingvistiki. Res bi bilo za vse nas, ki se manj razgledujemo po tem področju, ki ga manj sproti spremljamo, koristno, če bi imeli na enem mestu zbrano vse za zadnjih mogoče 20 let, ko je to področje veliko bolj živo, ko nastajajo raziskave tudi v okviru doktorskih in magistrskih del, kakovostne stvari, zanimive stvari, ki jih je vredno brati, če seveda veš, da obstajajo.

Marko Stabej: V tem smislu je bila zasnovana knjiga *Sociolingvistično iskrenje*, ki pa seveda ni niti približno pokrila vse produkcije, ampak tisto, kar je bilo mogoče na ta način, v tem času.

Krištof Savski: Lahko še jaz tukaj nekaj dodam? Mislim, da je iskanje prostora za sociolingvistično razpravljanje in razmišljanje vsekakor potrebno. Še zdaj smo namreč vsi kar nekako veseli, da lahko končno govorimo o sociolingvistiki, kakor da je to neka marginalna stvar, o kateri bog ne daj, da bi se preveč odprto razpravljalo. To je na slovenski jezikoslovni sceni žal realnost, saj prevladuje formalističen, sistemski pogled na jezik. Moja izkušnja iz Anglije je, da raziskave stimuliraš z rednim srečevanjem in debato. Mogoče bi bilo treba imeti recimo sociolingvistični krožek, kjer bi se lahko pogovarjali o delu na tem področju. Vsekakor ni nujno, da vsakič nekdo predstavi neko svojo lastno raziskavo, ker jih seveda toliko na leto ni. Kdaj se lahko skupaj prebere kakšen članek, ki je ravno izšel in se o njem predebatira. To je zelo pomembno zato, da se tudi študenti recimo priključijo, da se tudi njih motivira in se ta prostor tudi malo odpre. Je pa to tudi čisto praktična rešitev: takšne govorjene stvari so večinoma manj intenzivne, kot pa recimo pisanje in urejanje knjig, člankov, zbornikov in podobnih stvari. Torej to je praktična rešitev, da porabiš za debato recimo tri ure namesto 500 ur, kot jih potrebuješ, da pripraviš in izdaš neko knjigo. Taka knjiga bo seveda takoj zgolj ena na deset let. To bi samo dodal, da bi bilo takšno bolj redno srečevanje tukaj res potrebno.

Albina Nećak Lük: Zelo pozdravljam predlog kolega Savskega in mislim, da zdaj imamo mlade moči, ki bi mogoče bile voljne spet spraviti v pogon takšna srečanja, mogoče prav v okviru Društva za uporabno jezikoslovje. To društvo še vedno živi, čeprav na papirju, in plačuje redno članarino, tako da je Slovenija ostala članica Mednarodne zveze za uporabno jezikoslovje. Ta prireja srečanja s sociolingvistično tematiko. Ugodno je tudi to, da za njihove dogodke

lahko dobite znižano kotizacijo in mogoče celo kakšno podporo za svojo aktivnost. Če je med vami oziroma v tem krogu kdo, ki bi bil pripravljen pomagati poživiti delovanje in organizirati takšna srečanja, bi to bilo mogoče. Jaz bi to podprla in bi tudi spodbudila odgovorne, da pripravijo srečanje, mogoče občni zbor, da se spet vzpostavi program in oživi delovanje društva. Je pa to vedno težko. Tudi dokler je društvo še delovalo, smo zelo agitirali, kot se je takrat lepo reklo, da so ljudje bili aktivni, vsaj kolikor toliko.

Nataša Logar: Zahvaljujem se vam za predstavitve. Odprli ste številne vidike sociolingvistike, ki jih do zdaj sploh nisem poznala. Omenili ste nek večji kongres. Katerega že?

Albina Nećak Lük: 24. sociolingvistični simpozij v Gentu.

Nataša Logar: Moj predlog je naslednji: naj gre kdo od vas ali od kolegov, ki se nam danes niso utegnili pridružiti, v Gent in po kongresu na temu podobnem srečanju poroča o aktualnih temah. Tako bi hitro ugotovili, kako se tudi pri nas lahko s čim manj zamude pridružimo evropskim trendom.

Maja Bitenc: Kot je bilo rečeno, je že samo načrtovanje projekt in bi moralo biti proces. Zdi se mi, da tu ne gre drugače kot z nadaljnjim tovrstnim srečevanjem in pogovarjanjem in da tudi enkratna številka *Slavistične revije* ne bo prinesla življenja, če ne bo odnosov in idej, ki se bodo razvijale in enkrat mogoče meso postale. Mislim pa, da je v vsakem primeru potrebno začeti od spodaj in na terenu, kvalitativno, pri tem pa upoštevati, kar je že bilo narejeno. Zdi se mi namreč, da je bila težava omenjene raziskave in še mnogokatere druge o jezikovnih stališčih in jezikovni rabi, da so rezultati na podlagi samoocen in njihova verodostojnost res zelo vprašljivi. Po tem, ko intervjuvaš ljudi in vidiš, s katerimi izrazi operirajo za katere varietete, ki jih uporabljajo oni ali njihovi bližnji in ki jih imaš na posnetku v avtentični, vsakodnevni komunikaciji, vidiš, da so tukaj res lahko prepadi vmes in da potem taka, pa še tako kvantitativna in obsežna raziskava ne pove veliko – res prazni so lahko ti rezultati. Tako se mi zdi, da mora biti izhodišče pri posamezniku, pri vsakodnevni komunikaciji, pri avtentični jezikovni rabi, pri dejanski analizi posnetkov, pa kakorkoli je mučna. In v kombinaciji različnih pristopov, saj vsaka metoda nekaj doprinese, samo v prepletanju različnih pa lahko pridemo do verodostojnih, merodajnih podatkov o stanju stvari.

Karmen Kenda-Jež: Samo po drugi strani sva pa tudi midve ves čas ugotavljali, da dokler ni to delo timsko, taka raziskava ves čas ostaja nekakšen sondažni projekt. Torej, dokler nimaš možnosti, da jo opraviš v večjem obsegu.

Nataša Gliha Komac: Meni je prišlo na misel, da imamo zdaj na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU odličen fonolaboratorij in s tem možnost za kontinuirano spremljanje govorjenega jezika, zbiranje posnetkov in tako naprej – tudi to je pravzaprav ideja profesorice Albine Nećak Lük. Ampak tu je res treba, da se združi skupina strokovnjakov, da razmisli, kaj se da narediti, kako. Če so na voljo tako dobra tehnološka orodja, je škoda, da se jih ne uporabi.

Zato se mi zdi, da bi bilo dragoceno, da se posveti, kot je današnji, na nek način nadaljujejo. Mogoče vsebinsko bolj zamejeni: kaj koga posebej zanima, da se pač združi, pokliče, povabi na posamezno srečanje ljudi, raziskovalce, ki se ukvarjajo s točno določeno tematiko. Zgolj ideja.

Kot je omenila profesorica Albina Nećak Lük, pri japonskem projektu kontinuirano spremljajo stanje govorjenega jezika že od 50. let dalje. Tudi za slovenski jezik bi bilo nekaj takega izjemno dragoceno in zdi se mi, da imamo nekatere možnosti že na voljo.

A pri takih stvareh je izjemno pomembno sodelovanje. Zame je (bila) zelo dragocena izkušnja tako pedagoškega dela na različnih fakultetah, Podiplomski šoli ZRC SAZU kot raziskovanja na ISJ FR ZRC SAZU. In za organizacijo tako velikih in dolgoročnih projektov je potrebno veliko ljudi in znanja, a tudi veliko dobre volje in predanosti, vrsta drobnih stvari in dejavnosti, nekdo mora iskati finance in načrtovati ipd. Na fakultetah se npr. srečujete z mladimi ljudmi, ki jih zanimajo različne stvari in še iščejo svojo pot, na inštitutih oz. pri nas so interesi že precej izoblikovani in raziskovalne poti zastavljene, zato včasih težje najdemo mlade ljudi, ki jih zanimajo določena področja. Npr. kdaj veš, da neko raziskovalno vprašanje je, je zelo zanimivo in pomembno, želel bi ga (dobro) raziskati, a moraš preprosto slediti zastavljenim projektom, omejen si z že zastavljenimi temeljnimi dejavnostmi in projekti, ki se trenutno izvajajo. In morda ne veš, da nekoga zanimajo enaka vprašanja, problematike, da ima neka znanja, da ima sposobnosti in morda ne možnosti za raziskovanje. Skupaj je včasih lažje najti pot. Tudi tovrstno sodelovanje je po moje pomembno.

Marko Stabej: Jaz mislim, da kar omogoča Filozofska fakulteta, je pa to ravno tako možno na Inštitutu, kjer imate tudi vrsto različnih inštitutov, je to, da bi v ta forum ob določenih vprašanjih že v tem snovanju vključili tudi ljudi z drugih področji. Ne samo zato, da od njih nekaj izveš, ampak da se jih vključi v proces. Ker tukaj se mi zdi, da je dejansko zelo slabo poskrbljeno za to, tudi prej, ko je kakšna stvar bolje delovala, konkretno recimo Društvo za uporabno jezikoslovje, ki je bilo zelo živahno nekaj časa, ni bilo te izmenjave med različnimi strokami. Tukaj pa sta tako Filozofska fakulteta kot Inštitut možnost prostora, kjer ti ciljno vabiš določene ljudi v določene debate.

Marija Sotnikova Štravs: Jaz bi se na ta pogovor navezala samo kot nekdo, ki se ukvarja s sociolingvistiko in jezikovno politiko iz ljubezni do področja. Potem ko se poklicno ukvarjaš z nečim drugim, ostaneš izven vseh teh okvirjev in ti vse skupaj ostane bolj v smislu nekih občasnih debat ali pa kakšnih razprav na Facebooku. Samo toliko, kot mnenje z drugega pola – tudi za nekoga takšnega, kot sem jaz, bi ta prostor debate prišel prav, da nisi potem sam in se ne počutiš izoliranega.

Maja Bitenc: Očitno nek interes po pogovarjanju, načrtovanju in raziskovanju je, tako da bi morda res dorekli, da se srečamo še kdaj. Mogoče bi si kot izhodišče vsakič izbrali določeno temo, lahko kakšno delo, članek, monografijo – da bi vnaprej pregledali gradivo in potem skupaj razpravljali. Tisti, ki ste zainteresirani, pošljite na moj elektronski naslov sporočilo, da vas vključimo v skupino.

Marko Stabej: Najlepša hvala vsem prisotnim na vse načine. V upanju na nadaljnje sodelovanje …

Albina Nećak Lük: … na dolgo, čudovito prijateljstvo.

COLLOCATION RANKING: FREQUENCY VS SEMANTICS

Nikola LJUBEŠIĆ
Jožef Stefan Institute; Faculty of Computer and Information Science, University of Ljubljana

Nataša LOGAR
Faculty of Social Sciences, University of Ljubljana

Iztok KOSEM
Faculty of Arts, University of Ljubljana; Jožef Stefan Institute

Ljubešić, N., Logar, N., Kosem, I.: Collocation ranking: frequency vs semantics. Slovenščina 2.0, 9(2): 41–70.

DOI: https://doi.org/10.4312/slo2.0.2021.2.41-70

Collocations play a very important role in language description, especially in identifying meanings of words. Modern lexicography's inevitable part of meaning deduction are lists of collocates ranked by some statistical measurement. In the paper, we present a comparison between two approaches to the ranking of collocates: (a) the logDice method, which is dominantly used and frequency-based, and (b) the fastText word embeddings method, which is new and semantic-based. The comparison was made on two Slovene datasets, one representing general language headwords and their collocates, and the other representing headwords and their collocates extracted from a language for special purposes corpus. In the experiment, two methods were used: for the quantitative part of the evaluation, we used supervised machine learning with the area-under-the-curve (AUC) ROC score and support-vector machines (SVMs) algorithm, and in the qualitative part the ranking results of the two methods were evaluated by lexicographers. The results were somewhat inconsistent; while the quantitative evaluation confirmed that the machine-learning-based approach produced better collocate ranking results than the frequency-based one, lexicographers in most cases considered the listings of collocates of both methods very similar.

Keywords: collocations, word embeddings, logDice, general language, academic language

1 INTRODUCTION

The importance of the notion of collocation has been acknowledged by linguists for a long time, ever since J. R. Firth's famous statement: "You shall know a word by the company it keeps" (Firth, 1957). In fact, collocations themselves are considered by many as lexical units with different levels of semantic transparency (Singleton, 2000). As a result, even transparent collocations (and not only idioms, phrases and other more fixed multiword units) have started to receive more attention in dictionaries.

Collocation identification requires a computational approach. Several statistics for measuring collocation have been proposed in the past decades, for example t-score, MI, MI3, the log-likelihood ratio, the Dice coefficient, etc. (see Manning and Schütze, 1999, for an overview). In fact, collocation has been the pervasive driving force behind the development of tools for analysing and describing language in general. However, with progress also new challenges arose. Problematic aspects of different statistical approaches for measuring collocation have often been discussed (cf. Kilgarriff and Kosem, 2012), which led to the proposals of new measures such as logDice (Rychlý, 2008), which has been developed with lexicographic use in mind, and has been used by a large number of dictionary projects.

Nowadays, new, non-statistical methods are slowly finding their way into dictionary-making (and language) analysis. We thus decided to test one popular and up-to-date language modelling technique, namely word embeddings (Levy and Goldberg, 2014; Li and Jurafsky, 2015; Camacho-Collados and Pilehvar, 2018; etc.).

1.1 The aim and the scope of the paper

As Levy and Goldberg (2014, p. 302) explain, in the embeddings, distributional semantics word embeddings are vector representations of all the contexts in which a word occurred, and "enable efficient computation of word similarities through low-dimensional matrix operations". Recent uses of word embeddings for identifying collocations are well recorded (cf. Section 2). Various experiments proved the method to be moderately to highly successful in various tasks. We decided to find out how well it performs when given one other task, that is a task of collocate ranking. Since our research was lexicographically

oriented, we were especially interested in how well the method performs in comparison to the lexicographically highly popular logDice metric (Rychlý, 2008), which uses heuristics (i.e. a set of fixed rules).

Broadly speaking, we also wanted to find out whether a dictionary-making process (in our case a Slovene dictionary-making process) could become less time consuming and more efficient, if complemented with collocate ranking data acquired by the semantic-based method of word embeddings.

In order to establish how well word embeddings tackle the task of collocate ranking for lexicographic purposes we set a two-part experiment. It consisted of:

1. the quantitative analysis of
    a) heuristic-based vs machine-learning-based approach to collocate ranking, and
    b) frequency-based vs semantics-based machine-learning approach to collocate ranking;
2. the qualitative analysis of different collocate ranking results, which was performed by lexicographers.

In both analyses, two datasets were used:

- a general Slovene language dataset named KOLOS (Kosem et al., 2018), and
- a Slovene for special purposes (LSP) dataset named KAS (Erjavec et al., 2020).

Namely, we also wanted to draw some initial conclusions about the two approaches to collocation ranking with regards to differences in text type, and monosemy/polysemy of words.

All in all, the experiment arose from an actual dictionary-making process, and is described here with the purpose of bringing possible benefits to similar endeavours elsewhere as well.

2 MEASURING COLLOCATIONS: ASSOCIATION MEASURES, AND MORE RECENT – WORD EMBEDDINGS

An extensive body of research exists on measuring collocation strength or collocativity (e.g. Berry Rogghe, 1973; Church and Hanks, 1990; Church et al., 1991; Biber, 1993; Manning and Schütze, 1999; Evert, 2004; Gries, 2013), and different statistical methods (i.e. association measures) have been used up to this day. Association measures have also been regularly compared, and new ones proposed. Two good overviews of association measures are Wiechmann (2008) who compared 47 different association measures, and Pecina (2009), who conducted a comparison of more than 80 measures for collocation extraction. General observations of the majority of such studies were aptly summarized by Evert (2009), namely that "different association measures will produce entirely different rankings of the collocates" (ibid., p. 1218) and that "there is no ideal association measure for all purposes" (ibid., p. 1236).

A recent study by Evert et al. (2017) inspected the role of variables such as corpus size, context span, and frequency threshold in collocation identification. Using two different dictionaries as gold standards, it proved that "very large Web corpora and small co-occurrence contexts produce the best results" (ibid., 543). Moreover, in terms of co-occurrence span, researchers concluded that syntactic dependency was the best choice in most cases.

There is some literature on association measures used on Slovene corpus data as well (e.g. Gorjanc and Vintar, 2000; Gorjanc and Fišer, 2010), however there are no studies that would comprehensively compare the effectiveness of various association measures for identifying collocations in Slovene. As far as language description is concerned, in recent years most Slovene lexicographical and terminological projects have started using the Sketch Engine (Kilgarriff et al., 2004) and rely on association measures provided by this tool, especially logDice which is used by the well-known Word sketch function. However, as Gantar et al. (2015) and Gantar et al. (2016) observed, logDice often misses, or attributes very low ranking to certain important collocates, which is why researchers started combining logDice and raw frequency rankings when extracting and analysing collocates for dictionary purposes.

All association measures have one shortcoming in common: even if they are limited by predefined syntactic relations (such as in word sketches), they rely solely on co-occurrence frequencies and do not consider semantic aspects of words. And precisely this type of information is contained in word embeddings.

Word embeddings have been used extensively in the field of natural language processing (NLP) in the last decade. For example, Rodríguez-Fernández et al. (2016a) followed the well-known association approach early identified in Mikolov et al. (2013), where *king* to *man* is the same as *queen* to *woman*. They applied the same technique to collocation extraction, hoping to obtain the proper headword for the collocate *suggestion*, related to the known *take a walk* collocation. In their approach they hoped to be able to remove the *walk* information from *take* and add the *suggestion* information, ending up with *make* being a near-neighbour of the resulting vector, that is they calculated *vec(take) – vec(walk) + vec(suggestion)* with the goal of the result being close to *vec(make)*. This approach, evaluated in follow-up work, obtained a mean reciprocal rank (MRR) score between 0.01 and 0.47.[1]

Another piece of work by the same group of authors (Rodríguez-Fernández et al., 2016b) found that a linear transformation of the headword embedding can be used to predict the optimal collocate word embedding, learning this transformation per Mel'cuk semantic typologies (Mel'cuk, 1996). They did not compare this approach to the basic frequency-based one, nevertheless they achieved promising, but varying results, with the mean reciprocal rank (MRR) of the best-performing system between 0.3 and 0.9. This methodology was followed by Enikeeva and Mitrofanova (2017), who applied it to Russian data. They reported slightly higher MRR scores, ranging from 0.48 to 0.9. Again, they did not compare their results to the traditional frequency-based methods.

Liu and Huang (2017) showed that using the cosine distance between the distributional word representations of headwords and collocates as a function for

1 Mean reciprocal rank (MRR) is a relative score meant for ranked results that calculates the average of the inverse of the ranks at which the first positive instance occurs. MRR ranges between 0 and 1, and an MRR of 1 is obtained if in each ranking the positive instance is ranked in the first position, an MRR of 0.5 is obtained if in each ranking the positive instance occurs in second position, 0.33 for the third position and so on.

ranking collocation candidates yielded just slightly better results measured by F1 than the chi-square and mutual information co-occurrence statistics. Additionaly, Wanner et al. (2017) used distributional word representation to classify collocations into semantic classes, and Garcia et al. (2017) used multilingual word embeddings to find collocation translations in other languages.

Examining related literature, we can conclude that regardless of the fact that word embeddings are a very popular source of semantic information and that their usage as input features for making predictions in NLP has been considered a standard approach for years now, they have not yet been tested in a supervised learning setting on the task of general collocation ranking.

3 RESEARCH

3.1 Methodology

3.1.1 Research questions

In order to establish how well word embeddings tackle the task of collocate ranking for lexicographic purposes in the case of Slovene, we compared the embeddings results to the results obtained using the logDice method. The comparisons were made in a quantitative and qualitative way and were led by the following three research questions:

Q1: Which approach produces lexicographically more relevant rankings of collocates: the one that uses machine learning over manually annotated data, or the one that uses heuristics?

Q2: Which approach is a more useful source of information for the rankings of collocates: the word embeddings approach, which encodes distributional semantics of words, or the logDice approach, which encodes frequency information?

Q3: Which ranking of collocates is preferred by lexicographers: the embeddings ranking, or the logDice ranking?

As questions imply, we wanted to know whether the currently still dominant approach of using heuristics for collocate ranking is really better than the machine learning approach, which implicitly learns the underlying rules from examples. The second question was aimed at comparing two sources of

information – frequency, which is used in a heuristic way in logDice, and distributional semantics, which is exploited from word embeddings via machine learning. Finally, our third question put potential users of the two compared approaches (i.e. lexicographers) into focus and examined their preferences in actual cases.

3.1.2 Collocation datasets

3.1.2.1 KOLOS dataset

The KOLOS dataset contained a carefully selected set of 333 headwords, consisting of 154 nouns, 73 verbs, 81 adjectives, and 25 adverbs. The selected headwords were as heterogeneous as possible in terms of word class subcategories (e.g. plural nouns, countable nouns, transitive vs intransitive verbs etc.), corpus frequency, level of polysemy (number of different meanings), semantic characteristics (e.g. abstract vs concrete senses; qualitative vs classifying adjectives), etc. For each headword, we used collocations extracted for the purposes of the Collocations Dictionary of Modern Slovene (Kosem et al., 2018; Kosem et al., 2019). It should be noted that we already had a set of validated collocations from the Slovene Lexical Database (Gantar et al., 2016), and in order to devise a training dataset of good and bad collocation candidates, we decided to annotate only new ones (i.e. not yet validated collocations). This meant that we were often annotating the collocations slightly lower down the logDice-ordered list for each grammatical relation.

In the annotation task, the annotators were presented with a collocation, the information of its grammatical relation, and a corpus example of its use. The annotation of collocations was conducted in the Pybossa tool,[2] with each collocation being annotated by three annotators-linguists. The examples were extracted with the GDEX tool (Kosem et al., 2013) in the Sketch Engine (Kilgarriff et al., 2008), using the Slovenian configuration. The annotators were presented with three main answer groups – YES ('yes, this is a valid collocation'), NO ('no, this is not a collocation') and I DON'T KNOW ('I don't know if this is a collocation or not') (the YES and NO groups had additional sub-options, but they were not used in this experiment).

2 https://mnozicenje.cjvt.si/

Taking YES, NO and I DON'T KNOW answers, the agreement was analysed and the final decision for the training dataset, which could only be YES or NO, was made on the basis of the agreement (e.g. total agreement was YES or NO), while in borderline cases the final decision was made by making additional annotation or after joint discussion by the annotators.

The whole KOLOS dataset consisted of 17,540 collocation candidates belonging to 260 different grammatical relations. For the experiments performed in this paper we organised collocation candidates under 7,460 headwords (those being any of the two lexical parts of a bidirectional grammatical relation, so for *take a walk* we would have two collocations, once under the headword *take*, once under the headword *walk*). Experiments were done only on headwords that (1) had at least 10 collocation candidates for a specific grammatical relation as our evaluation was headword-based (this was the only data organisation that allowed evaluation of frequency-based statistics), and that (2) covered both the positive and the negative class so that discriminative machine learning (distinguishing between good and bad examples) can be performed. With these selection criteria the KOLOS dataset was shrunk to the most frequent 8 grammatical relations (actually 4 bidirectional relations), 212 headwords and 2,671 collocation candidates.

3.1.2.2 KAS dataset

The KAS dataset is a set of academic Slovene headwords, such as *analiza* (*analyses*), *tabela* (*table*), *razlikovati* (*to distinguish*), *relativno* (*relatively*), accompanied by collocations and examples of use (Logar et al., 2019). The set was built from a one-billion-word corpus KAS (Erjavec et al., 2020). The corpus was harvested from the Open Science Portal of Slovenia (2000–2015). For the most part (71% of tokens), it consists of BSc and BA theses, followed by MSc and MA theses (20%), and PhD theses (4%). Firstly, the initial list of candidates for the vocabulary of academic headwords was built by using the method of frequency profiling (Rayson and Garside, 2000). With this method we extracted lemmas that most differentiated the KAS corpus from a fiction part of the general corpus Kres (Logar et al., 2012, p. 79–97). Secondly, we inspected each lemma on the list in the KAS corpus concordances, and also checked its typical context in the Sketch Engine tool. In this manner we

determined whether the word in question belonged to a common expert discourse or not (the latter were excluded as it meant they were either grammatical words or technical terms). And thirdly, the final list of 463 headwords identified as typical of academic Slovene was supplemented by collocations and three examples of use for each collocation. The extraction of data was automatic; we used the same methodology as in the case of the KOLOS dataset (Kosem et al., 2011; Krek, 2012; Gantar et al., 2015; Kilgarriff and Kosem, 2012; Logar et al., 2014).

Automatically extracted data was then reviewed. We corrected the most obvious tagger performance mistakes, rearranged not ideally semantically grouped collocates, and deleted personal proper names, deixis, modal verbs and verbs with very broad meaning (e.g. *to be*, *to be about (sth)*). Nevertheless, all deletions remained part of the dataset, but were labelled as NEGATIVE collocation candidates.

Content-wise, the KAS dataset was heterogeneous with regards to its meaning and text function, but was either obviously or indirectly related to three roughly defined segments (Logar and Erjavec, 2019, p. 212–213): (a) the formal structure and the writing of academic texts (e.g. in English *bibliography*, *introduction*, *conclusions*; *empirical*, *defined*, *mentioned*; *to define*, *to cite*); (b) the methodology of academic texts (e.g. *method*, *hypothesis*, *respondent*; *to analyse*, *to identify*, *to classify*); or (c) the presentation and interpretation of the research data (e.g. *number*, *portion*, *dependence*; *measured*, *calculated*, *accurate*; *to result from*, *to indicate*, *to cause*; *subsequently*, *relatively*, *successfully*). With regard to word class, out of 463 headwords 226 were nouns, 119 adjectives, 86 verbs, and 32 adverbs (Logar et al., 2019). As far as the use in the KAS corpus is concerned, all words in the KAS dataset were monosemous.

In total, the KAS dataset consisted of 70,254 collocation candidates belonging to 342 different grammatical relations, organised under 5,220 headwords. By applying the same selection criteria as on the KOLOS dataset, our final KAS dataset on which we performed experiments shrunk to 8 grammatical relations (gramrels hereafter), 525 headwords and 14,722 collocation candidates.

3.1.3 Corpus information

The frequency and semantic information for our collocation candidates was obtained from the Gigafida 2.0 corpus (Krek et al., 2020). For calculating the frequency and logDice information as representatives of the frequency signal we used the Sketch Engine API. For calculating the (head)word embeddings as representatives of the semantic signal we used the fastText tool (Bojanowski et al., 2016) – in skip-gram mode with default parameters – and the lemma and part-of-speech annotations present in Gigafida 2.0, KAS and other large corpora of Slovene (Ljubešić and Erjavec, 2018).

3.2 Experiment

As explained in the Introduction section, our experiment consisted of two main parts:

1. the quantitative analysis, and
2. the qualitative analysis.

In both, we compared two approaches to collocate ranging, i.e. the logDice method and the word embeddings method. In the quantitative analysis, we performed two parts of the experiment, and in the qualitative part one more followed. Each of the three parts of our experiment was directly related to one of the research questions formulated at the beginning of the research.

3.2.1 Quantitative analysis

3.2.1.1 Experimental setup

In the quantitative part of the experiment, our goal was to compare traditional statistic-based approaches to collocate ranking with approaches based on machine learning. Since the only organisation that we can obtain through traditional approaches are ranked results (collocation candidates with higher frequency or higher logDice score are ranked higher), we set up our machine-learning experiments also in the way that enabled us to obtain ranked results. To evaluate traditional methods in their regular usage scenario, we performed evaluation on a per-gramrel and per-headword basis.

For our evaluation metric, we used the AUC (area-under-the-curve) ROC (receiver operating characteristic) score, which is considered to be the go-to

evaluation metric for ranking results, especially if the classes (positive and negative collocation candidates) are not balanced. Precisely this was the case in our datasets as in our original KOLOS dataset we had 13,812 positive candidates and 3,728 negative ones. The situation in the KAS dataset was similar, with 53,150 positive and 8,811 negative collocation candidates.

The AUC ROC score quantifies the quality of a ranking result, with the worst-possible ranking (all negative collocation candidates being ranked higher than all positive candidates) obtaining the result of 0.0, a perfect ranking (all positive collocation candidates being ranked higher than all negative collocation candidates) obtaining the result of 1.0, and a random ranking (positive and negative candidates being randomly mixed) obtaining the result of 0.5.

For performing supervised machine learning experiments, we used support-vector machines (SVMs), a regular go-to algorithm in traditional machine learning. We did not use more recent neural-network approaches as (1) their parameters are harder to interpret, and (2) initial experiments on our datasets had shown very similar results regardless of the machine-learning approach used. We had to be able to predict continuous values to be used for ranking candidates, thus we trained SVM regressors. All our implementations are written in the scikit-learn toolkit (Pedregosa et al., 2011).

Given that we obtained AUC ROC scores per each ranking (i.e. for each gramrel and headword we got a score), we had to set up a way to average all scores on some defined level. We aimed at averaging on the gramrel and overall level. As (1) different headwords under specific gramrels had a different number of candidates, and (2) different gramrels had a different number of candidates, we decided to normalise our results given the number of candidates, that is each collocation candidate would have the same impact on the final score of a method.

Supervised machine learning required two sets of data: training data (the data the model is built on) and testing data (the data the built model is evaluated on). Therefore, we performed a five-fold cross-validation, that is we split our training data into five groups, running five iterations of using four groups for training and one group for testing. By doing so we managed to evaluate the

model on each data point available, which is directly comparable to the output of the statistic-based ranking methods where we do not require training data. Furthermore, we made sure that headwords were sampled into groups, so that there was no spillage between training and testing data (e.g. training on some collocations of a headword and testing on other collocations of that headword). This makes the machine-learning approach quite challenging and measures to what extent the model can generalise regularities on the gramrel level, but not on the level of specific headwords present in our dataset.

3.2.1.2 Results

As explained, we obtained results on two datasets, KOLOS and KAS, by comparing four different approaches to collocation candidate ranking:

- freq: ordering via decreasing frequency of the collocations;
- logDice: ordering via decreasing logDice statistic of the collocations (using the frequencies of the headword, collocate and collocation);
- SVM_freq: machine learning the ranking from the frequency of the collocation, the headword, the collocate and the logDice statistic (all frequencies being represented on the logarithm scale);
- SVM_emb: machine learning the ranking from the embeddings of the headword, the collocate, and a sum of the two embeddings (to represent in a basic fashion the interaction between the two embeddings).

In Table 1, we present our results on the KOLOS dataset, together with the statistics on the size of the dataset for each gramrel. In Table 2, we give a similar description and results on the KAS dataset. Focusing first on the overall results on each dataset (the TOTAL row), the depicted picture is quite simple. The answer to our first research question, namely whether machine learning approach produces more relevant rankings of collocates than the approach based on heuristics, is positive. On the KOLOS dataset the two statistic-based approaches yielded scores of 0.52 and 0.47, while the two machine-learning-based approaches obtained scores of 0.58 and 0.71. On the KAS dataset the statistic-based approaches achieved scores of 0.58 and 0.63, while the machine-learning-based approaches obtained scores of 0.76 and 0.87.

With our second research question regarding the usefulness of both ambeddings approach and the logDice approach we again favoured the former. On the KOLOS dataset the frequency-based learning obtained the score of 0.58, while the semantic-based approach achieved the score of 0.71. On the KAS dataset the numbers obtained were 0.76 and 0.87, aiming at the same conclusion. Even more, there was only one gramrel (among 16) on which the machine-learning approach based on semantic information did not score the best results among the four approaches evaluated here (namely, the logDice score 0.65 for the *VERB + noun (accusative)* gramrel, see italics in Table 2).

An interesting, if not troubling observation is that ranking results via heuristics are quite close to the random baseline, with an average result on the KOLOS dataset of around 0.5 and on the KAS dataset of around 0.6. This suggests that their ranking is actually quite incapable of pushing the negative candidates as far down as possible. However, it still might be that the overall order of candidates via these two heuristics is useful for human use. In our experiments, we were aware only of the positive vs negative collocation candidate distinction and not of all subtle differences that collocations bring in a ranking scenario.

Table 1: *KOLOS dataset: the ranking results of the machine learning approach**

gramrel	# heads	# collos	freq	logDice	SVM_freq	SVM_emb
adjective + NOUN	38	576	0.526	0.405	0.56	**0.653**
ADJECTIVE + noun	54	983	0.503	0.463	0.534	**0.692**
NOUN + noun (genitive)	22	481	0.698	0.353	0.712	**0.78**
noun + NOUN (genitive)	47	967	0.517	0.501	0.631	**0.723**
VERB + noun (accusative)	13	231	0.468	0.443	0.432	**0.64**
verb + NOUN (accusative)	13	242	0.444	0.405	0.472	**0.737**
ADVERB + adjective	12	261	0.368	0.677	0.602	**0.802**
adverb + ADJECTIVE	13	221	0.584	0.62	0.515	**0.669**
TOTAL	212	3962	**0.523**	**0.469**	**0.577**	**0.71**

* Capital items: the headword and the starting point of the collocation (also from here forward, i.e. in Table 2, Table 4, etc.).

Table 2: *KAS dataset: the ranking results of the machine learning approach*

gramrel	# heads	# collos	freq	logDice	SVM_freq	SVM_emb
ADJECTIVE + noun	53	1737	0.537	0.563	0.665	**0.738**
adjective + NOUN	118	3045	0.58	0.689	0.8	**0.932**
NOUN + noun (genitive)	46	1677	0.559	0.534	0.603	**0.866**
noun + NOUN (genitive)	72	1999	0.565	0.556	0.623	**0.878**
VERB + noun (accusative)	18	828	0.619	*0.651*	0.59	0.556
verb + NOUN (accusative)	77	1947	0.632	0.597	0.913	**0.922**
ADVERB + adjective	52	1468	0.745	0.709	0.802	**0.894**
adverb + ADJECTIVE	89	2021	0.431	0.706	0.915	**0.954**
TOTAL	525	14722	**0.576**	**0.628**	**0.757**	**0.871**

For the different gramrels we also performed a correlation analysis to measure to what degree the results through gramrels and applied methods are stable between the two datasets. We calculated the Pearson correlation coefficient between the 8 results for each of the four methods on the KOLOS and on the KAS dataset. For the frequency method, we obtained a significant ($p = 0.043$) strong negative result ($r = -0.722$), and for the logDice method we again obtained a significant ($p = 0.029$), but strong positive result ($r = 0.758$). For the SVM_freq method our result was not significant ($p = 0.36$) and was moderately negative ($r = -0.375$), while for the SVM_emb method the result was also not significant ($p = 0.183$), but was moderately positive ($r = 0.524$). These results show that in the machine learning scenario achievements on specific grammatical relations differ quite a lot between datasets, while the logDice method was similarly (un-)successful on different gramrels. Nevertheless, the samples we obtained these calculations on are very small and one should take these results with caution. The only claim that could be made here is that in most cases the per-gramrel results are quite inconsistent.

3.2.2 Qualitative analysis

3.2.2.1 Experimental setup

We expected that lexicographers, too, would prefer the machine-learning results to those of heuristics, hence we tested our third hypothesis by presenting

them with two side-by-side columns for each headword in a specific grammatical relation, one column representing logDice ranking and one column representing embeddings ranking of collocates (see an example in Table 3). Lexicographers were asked to evaluate which column was more informative to them (column A or B), but they could also choose an answer *Both columns are similarly (un)informative*. This meant that either (a) both measures were equally informative or useful, or that (b) none of the measures was informative or useful. In addition, participants were alerted to the fact that they were evaluating results of the two aforementioned collocation extraction methods, but did not know which column was the result of which method. We also instructed them to pay more attention to top halves of lists in both columns. No other instructions for the evaluation process were given.

Table 3: *KOLOS dataset: headword* belina *(whiteness), grammatical relation: NOUN + noun (genitive) (the whiteness of ___)*

ranking	logDice (A)	embeddings (B)
1.	**zob (*tooth*)****	stena (*wall (interior)*)
2.	**sneg (*snow*)**	pokrajina (*landscape*)
3.	marmor (*marble*)	oblačilo (*clothes*)
4.	**polt (*complexion*)**	perilo (*washing*)
5.	perilo (*washing*)	kamen (*stone*)
6.	platno (*linen*)	marmor (*marble*)
7.	**papir (*paper*)**	obleka (*dress*)
8.	stena (*wall (interior)*)	koža (*skin*)
9.	zid (*wall*)	platno (*linen*)
10.	kamen (*stone*)	zid (*wall*)
11.	nebo (*sky*)	**sneg (*snow*)**
12.	obleka (*dress*)	nebo (*sky*)
13.	oblačilo (*clothes*)	**papir (*paper*)**
14.	pokrajina (*landscape*)	**polt (*complexion*)**
15.	koža (*skin*)	**zob (*tooth*)**
16.	obraz (*face*)	obraz (*face*)

** Bold print = in the case of the embeddings method, a noticeable drop in the ranking; underlined words = in the case of the embeddings method, a noticeable increase in the ranking.

This part of the experiment was partially done via a set of .txt documents and partially via an online survey. First, a preliminary evaluation on a smaller set

of .txt documents was performed by two lexicographers; one familiar with the KAS database and the other familiar with the KOLOS database. During this phase, the lexicographer evaluating the KAS database favoured logDice as having better ranking results, while the second lexicographer in some cases preferred the embeddings and noticed that the performance of this method might have been gramrel dependent. Since preliminary evaluation was inconclusive, seven other lexicographers were later invited to participate in the study (that is the online survey part of it).

The questionnaire of the online survey only included headwords from the KOLOS dataset, while the KAS dataset was further inspected only by the lexicographer who conducted the preliminary analysis. The reason for this decision was that all lexicographers invited to the online survey had experience with general dictionary and general dictionary-like resources and they were all involved in the KOLOS project, while only one lexicographer participated in the KAS project, that is the part that focused on general academic discourse vocabulary. Since we wanted to keep the expectations and initial positions of all of the lexicographers homogeneous, we kept them separate, as well as the datasets they evaluated.

Further KAS dataset analysis that was performed, as mentioned, by one lexicographer was done on eight randomly chosen headwords in ten different grammatical relations (i.e. 80 headwords: 24 nouns, 8 adjectives, 32 verbs, 16 adverbs), which in total summed up to 2,095 collocations repeated in two columns. On average, this meant 26 collocates per headword in a specific gramrel (with the smallest number of 10 and the largest number of 93 collocates per headword). In this second phase of the evaluation, the lexicographer evaluating the KAS dataset paid a closer attention to top halves of collocate columns, as did the online survey participants.

The online survey consisted of 63 headwords (34 nouns, 18 adjectives, 11 verbs) and their collocates in seven different grammatical relations. Because we wanted to broaden the number of gramrels, only three of them were the same in both datasets. The survey was divided into seven separate grammatical relation subsurveys, which meant that each grammatical relation had its own survey link. This was done to keep the cognitive load manageable for participants (they could complete the survey for one grammatical relation

and continue with the next one on another day), and to facilitate the analyses. In total, there were 146 pairs of collocate lists (i.e. questions in the survey; see Table 4). It should be noted that due to various reasons (time constraints etc.) not all the participants completed all seven grammatical relation surveys.

Table 4: *Online surveys: number of headwords and number of lexicographers participating*

gramrel	number of headwords	number of participants
VERB + noun (accusative)	12	6
verb + NOUN (accusative)	26	8
ADJECTIVE + noun	19	6
adjective + NOUN	30	6
adverb + ADJECTIVE	11	7
NOUN + noun (genitive)	19	6
noun + NOUN (genitive)	29	6

3.2.2.2 Results

3.2.2.2.1 KAS collocates ranking

As Table 5 and Figure 1 show, the lexicographer evaluating the KAS database in the second phase of the study again did not find the embeddings rankings better than the logDice rankings. In almost two thirds of cases (51/80), she decided that both columns were very similar, and in almost all of the rest of them (26/80), in her opinion, the embeddings performed worse. Thus, a small number of only three cases of embeddings performing better can be perceived as exceptions.

A closer look at grammatical relations reveals that the success of both ranking methods differs according to the lexicographers' judgments. Collocate ranking according to logDice was preferred in grammatical relations *NOUN + noun (genitive)* and *VERB + noun (accusative)*, while the ranking results of both methods were very similar in four relations (right side of Figure 1): *NOUN + "for" + noun (accusative)*, *VERB + "and_or" + verb*, *ADVERB + adjective*, and *ADVERB + verb*.

Table 5: *KAS dataset: logDice ranking vs embeddings ranking of collocates per grammatical relation (in absolute numbers and percentage)*

gramrel	logDice better: number and (%)	embeddings better: number and (%)	very similar: number and (%)
ADJECTIVE + noun	4 (50)		4 (50)
VERB + adverb	3 (37)		5 (63)
NOUN + noun (genitive)	5 (63)	1 (2)	2 (25)
VERB + noun (genitive)	3 (38)	1 (2)	4 (50)
VERB + noun (accusative)	5 (63)		3 (37)
NOUN + preposition *v* (*in*) + noun (locative)	4 (50)		4 (50)
NOUN + preposition *za* (*for*) + noun (accusative)	1 (2)		7 (98)
VERB + conjunction *in_ali* (*and_or*) + verb			8 (100)
ADVERB + adjective	1 (2)		7 (98)
ADVERB + verb		1 (2)	7 (98)
TOTAL	**26 (32)**	**3 (4)**	**51 (64)**

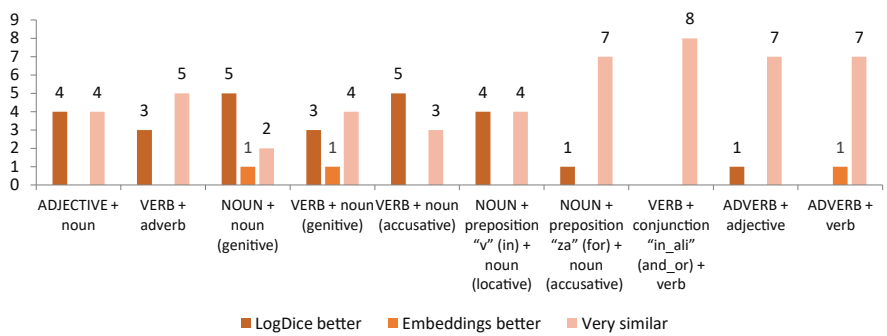


Figure 1: KAS dataset: logDice ranking vs embeddings ranking of collocates per grammatical relation (in absolute numbers).

3.2.2.2.2 KOLOS collocate ranking

Overall, the most popular answer in the online survey was *Both columns are similarly (un)informative* (45% of the answers, Table 6), which indicates that the participants having a general dictionary-like resource in mind did not, almost half of the time, consider one ranking better than the other.

Table 6: *KOLOS dataset: logDice ranking vs embeddings ranking of collocates per grammatical relation (in absolute numbers and percentage)*

gramrel	logDice better: number and (%)	embeddings better: number and (%)	very similar: number and (%)	TOTAL ANSWERS: number
VERB + noun (accusative)	16 (24)	22 (33)	28 (42)	66
verb + NOUN (accusative)	74 (37)	24 (12)	102 (51)	200
ADJECTIVE + noun	33 (31)	31 (29)	44 (41)	108
adjective + NOUN	73 (42)	21 (12)	80 (46)	174
adverb + ADJECTIVE	27 (39)	8 (11)	35 (50)	70
NOUN + noun (genitive)	23 (21)	36 (33)	50 (46)	109
noun + NOUN (genitive)	44 (26)	62 (37)	62 (37)	168
TOTAL	**290 (32)**	**204 (23)**	**401 (45)**	**895**

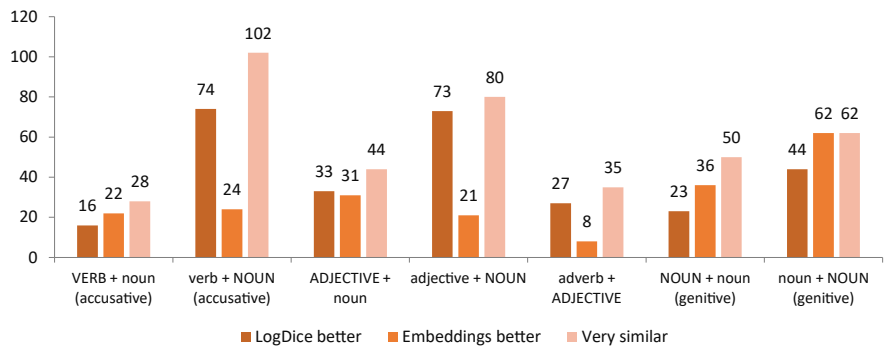


Figure 2: KOLOS dataset: logDice ranking vs embeddings ranking of collocates per grammatical relation (in absolute numbers).

Of the two measures, logDice was considered better more frequently than embeddings, with 32% vs 23% answers selected respectively. However, as Table 6 and Figure 2 show, this ratio between the two measures varied considerably according to the grammatical relation. Ranking of collocates according to logDice was more preferred in grammatical relations *verb + NOUN (accusative)*, *adjective + NOUN*, and *adverb + ADJECTIVE*. On the other hand,

embeddings ranking was preferred in *VERB + noun (accusative)*, *NOUN + noun (genitive)*, and *noun + NOUN (genitive)* grammatical relation.

We also searched for patterns in the results on a headword level, especially for headwords that featured in at least two different grammatical relations. We wanted to establish whether certain headwords prefer one of the measures across different grammatical relations. Similar to above mentioned findings, logDice was again preferred more often than embeddings, with the participants preferring it at 26 headwords in different grammatical relations, while embeddings results were preferred at only 14 headwords (for the remaining headwords no considerable differences in preferences were observed). There were also no clear patterns that the headwords identified had in common.

At the end of both evaluations, we made a numerical comparison of the results in total for both datasets (Figure 3).

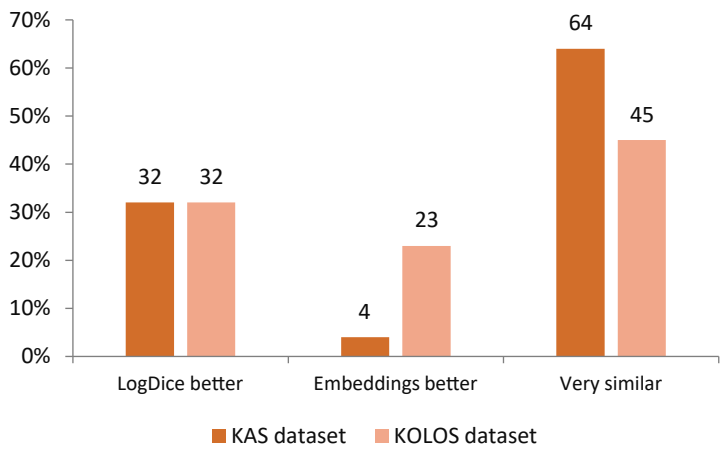


Figure 3: KAS and KOLOS dataset: logDice ranking vs embeddings ranking of collocates – both evaluations in total (in percentage).

Even though our deduction is limited due to the fact that only one lexicographer examined the KAS dataset, one feature in Figure 3 stands out: to a noticeably larger extent (23%) the embeddings rankings of collocates of the KOLOS dataset were recognised as more informative than those of the KAS dataset (4%). It is possible that this is a consequence of the KOLOS being

much more polysemous. If that is the case, at least this part of our qualitative analysis favours the semantic-based method to logDice metrics. Nevertheless, as a whole our third research question, namely which ranking of collocates is prefered by lexicographers, must be answered in the following way: lexicographers prefer the logDice ranking.

4 DISCUSSION

The main point that needs to be discussed is the difference between the results of quantitative and qualitative analyses. With the results of the quantitative analysis so convincingly in favour of the embeddings approach, it was somewhat surprising to learn that the lexicographers did not confirm this finding. In this section, we present some possible explanations for this discrepancy.

But first, let us turn our attention to the fact that in comparison to the KOLOS dataset, higher scores of the machine-learning-based approaches were consistently obtained on the data from KAS. It seems like this was influenced by two features: the (non)specialised content of the two corpora, and the monosemy or polysemy of selected headwords. As mentioned, all headwords in the KAS dataset are monosemous (but not technical), and secondly, the KAS corpus is domain- and genre-specific; on the other hand, more than half of the KOLOS headwords were polysemous and therefore used in various contexts, but they (and their collocates) also originated from a general, domain and genre diverse corpus of Slovene. The latter very probably limited the machine-learning process, while the first enhanced it. It is our belief this should be kept in mind in follow-up testings of the embeddings method and its use in dictionary-making projects.

When answering our first and second question using the AUC ROC score and the SVM learning algorithm, the machine-learning-based approaches ranked better than statistic-based ones (KOLOS scores on frequency information: 0.52 vs 0.58), and the semantic information given through word embeddings was more useful than frequency information (KOLOS scores on using machine learning on frequency and embeddings: 0.58 vs 0.71). Yet lexicographers' most frequent evaluation was a non-decisive one: to them in half or more cases (45% for KOLOS and 64% for KAS) both rankings of collocates seemed very similar. In fact, the survey participants' comments suggest that

the task of deciding which ranking was better even proved frustrating at times. Many KOLOS survey participants mentioned that they often deliberated on monosemous or polysemous characteristic of the headword, similarity of collocates and their broad meaning, while the lexicographer evaluating KAS dataset disfavoured columns that had too general or too technical words among approximately top ten collocates. Nevertheless, the votes of all of them were given with considerable uncertainty and were very diverse.

Our survey instructions were intentionally non-explicit, in other words: instruction-wise, we did not address the aforementioned differences. We wanted to learn in general, whether the semantic nature of the embeddings collocation extraction method could be recognised and found advantageous for lexicographic work. Unfortunately, our conclusions suggest that higher algorithm scores, though numerically significant, were in most part not obvious to humans. Just one segment of KOLOS vs KAS evaluation results confirmed that there is indeed some potential in the semantic nature of the embeddings collocate rankings; namely 23% of the much more polysemic KOLOS dataset was recognised as more informative than the logDice ranking, while this was the case for only 4% of the KAS database. However, since KAS data was evaluated solely by one lexicographer, further studies should examine this indication in more detail.

With regard to the embeddings method being gramrel dependent, i.e. that it is more successful for some grammatical relations, but not the others, nothing can be concluded. By choosing a set of 17 various relations (KAS: 10, KOLOS: 7), with only three of them overlapping, gramrel-wise we were able to get a broader view, but the number of headwords per each grammatical relation was thus reduced (in total KAS: 80, KOLOS: 63). Subsequently, none of the relations was analysed comprehensively. Even with gramrels that overlapped in the datasets (*ADJECTIVE + noun*; *NOUN + noun (genitive)*; *VERB + noun (accusative)*), the survey results were not uniform and do not allow for any obvious inference. The question of gramrel importance for the task of embedding-based collocation extraction is in fact rather questionable as initial experiments on training one single model for collocation extraction on all gramrels showed very similar results to those of training separate models for each gramrel. For the sake of a better control over the process and a more

interesting analysis, in this research we opted for keeping gramrel data and gramrel experiments separate, but other scenarios are, of course, possible for future fine tunings of the method.

Finally, we must consider the part human intuition, or rather lexicographers' knowledge, experience, and past and present project involvement played in our experiment. Lexicographers' evaluation, though an expert one, played a crucial role not once, but twice. Firstly, during the annotation of collocations before the quantiative part of the experiment; and secondly, after it in the form of lexicographers' judgments of the informativeness of the collocate rankings. Machine learning was, of course, performed on the pre-annotation dataset taken as a kind of gold standard, which actually meant that the lexicographers' preferences in the post-ranking phase primarily reflected annotators' preceding decisions. Here, it is important to stress that both groups of experts consisted of almost the same people, though the time that passed between the two phases of the experiment was about five months. Also, since the pre-treatment of the KAS datasets was not identical to the pre-treatment of the KOLOS dataset, and the same goes for the evaluation part of the experiment, the comparison between the results of both datasets is far from optimal. In this respect, our conclusions need to be treated as just preliminary.

5 CONCLUSIONS

Recent trends in lexicography have focused on automating certain aspects of language description, especially those related to collocations and examples (e.g. Kilgarriff and Rychlý, 2010; Rundell and Kilgarriff, 2011). As Cook et al. (2013, p. 50) point out, a "striking outcome of the work done so far in this area is that automation not only delivers efficiency savings but also leads to improvements in quality".

Lexicographers are used to inspecting long lists of collocates, separating the wheat from the chaff, but when automatically produced language resources are in question, different results of different extraction tools matter, and improvements in quality are always possible. In our research, we used a supervised machine-learning approach to collocation extraction and ranking with the aim of establishing how advantageous it is when compared to heuristic frequency-based logDice metrics. We found that while supervised approaches

do improve over the unsupervised baseline in an automation setting, in most cases the lexicographers did not appreciate this "improvement".

Nevertheless, the results are not discouraging. They prove (and confirm) that, ideally, a good collocation extraction tool is one that combines computational measurements and lexicographers' input. Obviously, modern lexicography is still an inherently multidisciplinary endeavour with the never justly answered question of how to measure what is informative, relevant, and significant – this seems even more so for language resources of the digital era.

Acknowledgments

The research was conducted as part of the project Collocation as a basis for language description: semantic and temporal perspectives (J6-8255), funded by the Slovenian Research Agency, and within the national research programme Slovene language – basic, contrastive, and applied studies (P6-0215), and the national research programme Language resources and technologies for Slovene language (P6-0411), also funded by the Slovenian Research Agency.

REFERENCES

Berry-Rogghe, G. L. (1973). The Computation of Collocations and their Relevance in Lexical Studies. In A. J. Aitken, R. W. Bailey, and N. Hamilton-Smith (Eds.), *The Computer and Literal Studies* (pp. 103–112). Edinburgh, New York: University Press.

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, *8*(4), 243–57.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching Word Vectors with Subword Information. In H. Schütze (Ed.), *Transactions of the Association for Computational Linguistics 5* (pp. 135–146).

Camacho-Collados, J., & Pilehvar, M. T. (2018). From Word to Sense Embeddings: A Survey on Vector Representations of Meaning. *Journal of Artificial Intelligence Research 63*, 743–788.

Church, K. W., Gale, W., Hanks, P., & Hindle, D. (1991). Using Statistics in Lexical Analysis. In U. Zernik (Ed.), *Lexical Acquisition: Exploiting On-line Resources to Build a* Lexicon (pp. 116–164). Erlbaum, Hillsdale, NJ.

Church, K., & Hanks, P. (1990). Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics*, *6*(1), 22–29.

Cook, P., Lau, J. H., Rundell, M., McCarthy, D., & Baldwin, T. (2013). A Lexicographic Appraisal of an Automatic Approach for Detecting New Word Senses. In I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets & M. Tuulik (Eds.), *Electronic Lexicography in the 21st Century: Thinking Outside the Paper, Proceedings of the eLex 2013 Conference, 17–19 October 2013, Tallinn, Estonia* (pp. 49–65). Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.

Enikeeva, E. V., & Mitrofanova, O. A. (2017). Russian Collocation Extraction Based on Word Embeddings. In V. Selegey et al. (Eds.), *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2017"* (pp. 52–64). Moscow: The Computational Linguistics and Intellectual Technologies.

Erjavec, T., Fišer, D., & Ljubešić, N. (2020). The KAS Corpus of Slovenian Academic Writing. *Language Resources & Evaluation 55*, 551–583.

Evert, S. (2004). The Statistics of Word Cooccurrences: Word Pairs and Collocations, PhD Thesis. University of Stuttgart.

Evert, S. (2009). Corpora and Collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook 2* (pp. 1212–1248). Berlin/New York: Mouton de Gruyter.

Evert, S., Uhrig, P., Bartsch, S., & Proisl, T. (2017). E-VIEW-alation – a Large-scale Evaluation Study of Association Measures for Collocation Identification. In I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek & V. Baisa (Eds.), *Electronic Lexicography in the 21st Century, Proceedings of eLex 2017 Conference* (pp. 531–549). Leiden, Netherlands/Brno: Lexical Computing CZ s.r.o.

Firth, J. R. (1957). *Modes of Meaning: Papers in Linguistics*: 1934–1951. London: Oxford University Press.

Gantar, P., Kosem, I., & Krek, S. (2016). Discovering Automated Lexicography: the Case of the Slovene Lexical Database. *International Journal of Lexicography*, *29*(2), 200–225.

Gantar, P., Krek, S., Kosem, I., & Gorjanc, V. (2015). Collocation Dictionary for Slovene: Challenge for Automatic Extraction of Data and Crowdsourcing. In

G. Corpas Pastor, M. Buendía Castro & R. Guttiérrez Florido (Eds.), *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives (Fraseologı´a computacional y basada en corpus: perspectivas monolingu¨es y multilingu¨es), Europhras, 2015* (pp. 84–86). Malaga: Lexytrad, Research Group in Lexicography and Translation.

Garcia, M., García-Salido, M., & Alonso-Ramos, M. (2017). Using Bilingual Word-embeddings for Multilingual Collocation Extraction. In S. Markantonatou, C. Ramisch, A. Savary & V. Vincze (Eds.), *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)* (pp. 21–30). Valencia: Association for Computational Linguistics.

Gorjanc, V., & Fišer, D. (2010). *Korpusna analiza*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Gorjanc, V., & Vintar, Š. (2000). Iskanja po korpusu slovenskega jezika FIDA. In T. Erjavec & J. Gros (Eds.), *Jezikovne tehnologije: Zbornik konference* (pp. 20–27). Ljubljana: Institut Jožef Stefan.

Gries, S. (2013). 50-something Years of Work on Collocations. *International Journal of Corpus Linguistics*, *18*(1), 137–165.

Kilgarriff, A., Rychlý, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6–10, 2004* (pp. 105–116). Lorient: Université de Bretagne – sud.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., & Rychlý, P. (2008). GDEX: Automatically Finding Good Dictionary Examples in a Corpus. In E. Bernal & J. DeCesaris (Eds.), *Proceedings of the Thirteenth EURALEX International Congress* (pp. 425–432). Barcelona, Spain: Institut Universitari de Linguistica Aplicada, Universitat Pompeu Fabra.

Kilgarriff, A., & Rychlý, P. (2010). Semi-automatic Dictionary Drafting. In G.-M. de Schryver (Ed.), *A Way with Words: A Festschrift for Patrick Hanks* (pp. 299–312). Kampala: Menha Publishers.

Kilgarriff, A., & Kosem, I. (2012). Corpus Tools for Lexicographers. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography* (pp. 31–56). Oxford: Oxford University Press.

Kosem, I., Gantar, P., & Krek, S. (2013). Automation of Lexicographic Work: an Opportunity for Both Lexicographers and Crowd-sourcing. In I. Kosem,

J. Kallas, P. Gantar, S. Krek, M. Langemets & M. Tuulik (Eds.), *Electronic Lexicography in the 21st century: Thinking Outside the Paper, Proceedings of the eLex 2013 Conference, 17–19 October 2013, Tallinn, Estonia* (pp. 32–48). Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.

Kosem, I., Husak, M., & McCarthy, D. (2011). GDEX For Slovene. In I. Kosem & K. Kosem (Eds.), *Electronic Lexicography in the 21st century: New Applications for New Users, Proceedings of eLex 2011, 10–12 November 2011, Bled, Slovenia* (pp. 150–159). Ljubljana: Trojina, Institute for Applied Slovene Studies.

Kosem, I., Krek, S., Gantar, P., Arhar Holdt, Š., Čibej, J., & Laskowski, C. (2018). Collocations Dictionary of Modern Slovene. In J. Čibej, V. Gorjanc, I. Kosem & S. Krek (Eds.), *Proceedings of the 18th EURALEX International Congress: Lexicography in Global Contexts, 17–21 July 2018, Ljubljana* (pp. 989–997). Ljubljana: Ljubljana University Press, Faculty of Arts.

Kosem, I., Gantar, P., Krek, S., Arhar Holdt, Š., Čibej, J., Laskowski, C., Pori, E., Klemenc, B., Dobrovoljc, K., Gorjanc, V., & Ljubešić, N. (2019). *Collocations Dictionary of Modern Slovene KSSS 1.0*. Ljubljana: Slovenian Language Resource Repository CLARIN.SI. Retrieved from http://hdl.handle.net/11356/1250 (26. 8. 2021)

Krek, S. (2012). New Slovene Sketch Grammar for Automatic Extraction of Lexical Data: Presentation given at SKEW3, Brno, Czech Republic, 21–22 March 2012. Retrieved from https://trac.sketchengine.co.uk/attachment/wiki/SKEW-3/Program/Krek_SKEW-3.pdf?format=raw (26. 8. 2021)

Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I., & Dobrovoljc, K. (2020). Gigafida 2.0: the Reference Corpus of Written Standard Slovene. In N. Calzolari (Ed.), *LREC 2020: Twelfth International Conference on Language Resources and Evaluation: May 11–16, 2020, Palais du Pharo, Marseille, France, Conference Proceedings* (pp. 3340–3345). Paris: ELRA – European Language Resources Association.

Levy, O., & Goldberg, Y. (2014). Neural Word Embedding as Implicit Matrix Factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 27 (NIPS 2014)* (pp. 1–9).

Li, J., & Jurafsky, D. (2015). Do Multi-sense Embeddings Improve Natural Language Understanding?. In L. Màrquez, C. Callison-Burch & J. Su (Eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1722–1732). Lisbon: Association for Computational Linguistics.

Liu, X., & Huang, D. (2017). Translation Oriented Sentence Level Collocation Identification and Extraction. In D. Wong & D. Xiong (Eds.), *Machine Translation, CWMT 2017: Communications in Computer and Information Science 787* (pp. 78–89). Singapore: Springer.

Ljubešić, N., & Erjavec, T. (2018). *Word Embeddings CLARIN.SI-embed.sl 1.0*. Ljubljana: Slovenian Language Resource Repository CLARIN.SI. Retrieved from http://hdl.handle.net/11356/1204 (26. 8. 2021)

Logar, N., Grčar, M., Brakus, M., Erjavec, T., Arhar Holdt, Š., & Krek, S. (2012). *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.

Logar, N., Gantar, P., & Kosem, I. (2014). Collocations and Examples of Use: a Lexical-semantic Approach to Terminology. *Slovenščina 2.0*, *2*(1), 41–61.

Logar, N., & Erjavec, T. (2019). Slovene Academic Writing: a Corpus Approach to Lexical Analysis. In I. Simonnæs (Ed.), *New Challenges for Research on Language for Special Purposes: Selected Proceedings from the 21st LSP-Conference, 28–30 June 2017, Bergen, Norway* (pp. 205–217). Berlin: Frank & Timme.

Logar, N., Kosem, I., & Erjavec, T. (2019). *Collocation Lexicon of Slovene Academic Discourse Aleks*. Ljubljana: Slovenian Language Resource Repository CLARIN.SI. Retrieved from http://hdl.handle.net/11356/1245 (26. 8. 2021)

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing, Chap. 5: Collocations*. Cambridge, Massachusetts: The MIT Press.

Mel'cuk, I. (1996). Lexical Functions: a Tool for the Description of Lexical Relations in a Lexicon. *Lexical Functions in Lexicography and Natural Language Processing, 31*, 37–102.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Retrieved from https://arxiv.org/abs/1301.3781 (26. 8. 2021)

Pecina, P. (2009). Lexical Association Measures and Collocation extraction. *Language Resources and Evaluation*, *44*(1–2), 137–158.

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

Rayson, P., & Garside, R. (2000). Comparing Corpora using Frequency Profiling. In *WCC'00, Proceedings of the Workshop on Comparing Corpora, 9*, 1–6.

Rodríguez-Fernández, S., Carlini, R., Espinosa Anke, L., & Wanner, L. (2016a). Example-based Acquisition of Fine-grained Collocation Resources. In N. Calzolari et al. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 2317–2322). Portorož: ELRA.

Rodríguez-Fernández, S., Carlini, R., Espinosa Anke, L., & Wanner, L. (2016b). Semantics-driven Recognition of Collocations Using Word Embeddings. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 499–505). Berlin: Association for Computational Linguistics.

Rundell, M., & Kilgarriff, A. (2011). Automating the Creation of Dictionaries: Where Will It All End?. In F. Meunier, G. Gilquin & M. Paquot (Eds.), *A Taste for Corpora: in Honour of Sylviane Granger* (pp. 257–282). Amsterdam: John Benjamins.

Rychlý, P. (2008). A Lexicographer-Friendly Association Score. In P. Sojka & A. Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008* (pp. 6–9). Brno: Masaryk University.

Singleton, D. (2000). *Language and the Lexicon: an Introduction*. New York: Oxford University Press.

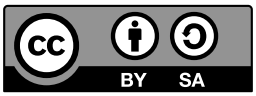
Wanner, L., Ferraro, G., & Moreno, P. (2017). Towards Distributional Semantics-Based Classification of Collocations for Collocation Dictionaries. *International Journal of Lexicography*, *30*(2), 167–186.

Wiechmann, D. (2008). On the Computation of Collostruction Strength. *Corpus Linguistics and Linguistic Theory, 42*, 253–290.

RAZVRŠČANJE KOLOKATORJEV V SEZNAM: POGOSTOST *PROTI* SEMANTIKI

Kolokacije imajo v opisu jezika zelo pomembno vlogo. Še zlasti to velja za prepoznavanje pomena besed. Zato so postali v moderni leksikografiji neobhoden del pomenske členitve prav seznami kolokatorjev, razvrščeni po eni od statističnih mer povezovalnosti. Prispevek prikazuje primerjavo med dvema pristopoma k razvrščanju kolokatorjev: (a) metodo logDice, ki je zelo uveljavljena in temelji na pogostosti, ter (b) metodo besednih vložitev, ki je nova in temelji na strojnem učenju ter besedni semantiki. Primerjavo med rezultati obeh pristopov smo naredili na dveh zbirkah podatkov za slovenščino, eno z iztočnicami in njihovimi kolokacijami iz splošnega jezika, drugo z iztočnicami in njihovimi kolokacijami iz strokovno-znanstvenega jezika. Pri ocenjevanju rezultatov smo uporabili dve metodi: v kvantitativnem delu preizkusa smo izvedli nadzorovano strojno učenje z AUC ROC evalvacijo algoritma podpornih vektorjev (SVM); v kvalitativnem delu pa so rezultate obeh pristopov k razvrščanju kolokatorjev ocenili še leksikografi. Ugotovitve niso enoznačne; medtem ko je kvantitativno ocenjevanje pokazalo, da je pristop s strojnim učenjem in semantično razpršenostjo dal boljše razvrstitve kolokatorjev kot pristop, ki izhaja iz pogostosti, pa so leksikografi večinoma ocenili, da so seznami kolokatorjev obeh pristopov med sabo zelo podobni.

Ključne besede: kolokacije, besedne vložitve, logDice, splošni jezik, strokovno-znanstveni jezik



STALNOST, VARIANTNOST IN MODIFICIRANA RABA FRAZEMOV V SLOVENSKEM JEZIKU IN SLOVARJIH

Eva TRIVUNOVIĆ
Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU

Trivunović, E.: Stalnost, variantnost in modificirana raba frazemov v slovenskem jeziku in slovarjih. Slovenščina 2.0, 9(2): 71–99.

DOI: https://doi.org/10.4312/slo2.0.2021.2.71-99

Prispevek prinaša pregled variant in modifikacij sedmih (iz)biblijskih frazemov v sodobni slovenščini ter njihove prisotnosti v sodobnem jeziku. Ugotovitve so primerjane z obravnavo frazemov v obstoječih slovarjih, kjer se kaže velik razkorak med slovarskim prikazom in stanjem, ki ga izkazuje korpusno gradivo. Za zanesljivejše ugotavljanje, v katerih primerih lahko govorimo o že ustaljeni variantnosti, so bili v raziskavi uporabljeni trije zvrstno različni korpusi: *Gigafida 2.0*, *Janes* in *slWaC*. Poleg ustaljenih variant so predstavljene neustaljene modifikacije, poseben poudarek je na prenovitvah, vendar se je jasno zastavljena tipologija mestoma izkazala za preveč togo, saj pri nekaterih mejnih primerih ni bilo mogoče nedvoumno ločiti ustaljenih variant od neprenovitvenih modifikacij ter neprenovitvenih modifikacij od prenovitvenih. Vsi izbrani frazemi in njihove prenovitve so najpogostejši v korpusu *Janes*, kar dokazuje nujnost vključevanja večjega števila raznovrstnih korpusov v jezikoslovne raziskave.

Ključne besede: frazeologija, variante, modifikacije, prenovitve, korpusni pristop

1 UVOD

Frazeološka teorija je najbolj zapletena in neenotna prav na področju definicijskih lastnosti, ki so obravnavane z različnih vidikov, zato obstaja več nasprotujočih si pojmovanj in teorij, kaj je za frazeme bistveno in kaj frazeologija vključuje (Gantar, 2007, str. 79). Tradicionalno se kot eno od definicijskih lastnosti frazemom pripisuje stalnost oblike in pomena, vendar je predvsem

z uporabo korpusnega pristopa postalo jasno, da je ta pogosto razrahljana in ohlapnejša, kot se zdi na prvi pogled, saj se frazemi v besedilih uresničujejo na zelo različne načine ter obstajajo v različnih variantah in oblikah.

V članku[1] so s pomočjo zvrstno različnih korpusov zbrane variante in modifikacije sedmih (iz)biblijskih[2] frazemov v sodobni slovenščini.[3] Poleg analize variantnosti je prikazano, kako prisotni so nekateri najbolj poznani (iz)biblijski frazemi v sodobni pisni slovenščini ter kako ti frazemi, ki so nastali na podlagi tako starega besedila in so (vsaj nekateri) v slovenskem jeziku v rabi že stoletja (prim. Trivunović, 2019), ostajajo aktualni za izražanje sodobnih problemov, dilem in družbenih okoliščin še danes.[4] Ugotovitve iz korpusov so nato primerjane s stanjem v obstoječih slovarjih, o pri čemer je upoštevano, da je razlika med stanjem v korpusih in stanjem v slovarjih lahko posledica dejanskega jezikovnega razvoja in premikov v jeziku, ki so se zgodili po nastanku slovarjev, ali pa pomanjkljive slovarske obravnave frazemov v slovarjih, ki izhaja predvsem iz nesistematičnega vključevanja in prikaza variant in omejenosti gradiva. Pregledane so bile pojavitve izhodiščnih frazemov in njihove prenovitve, ustaljenost variante se je dokazovala predvsem s pogostostjo rabe, prisotnostjo uvajalnih sredstev in neprenovitvenim učinkom v besedilu. Ker je za uspešno prenovitev bistveno, da naslovnik izhodiščni frazem pozna, se postavlja vprašanje, ali so prenovitvam bolj podvrženi frazemi, ki so tudi sicer (v svoji izhodiščni obliki) pogostejši.

1 Članek je razširjena verzija konferenčnega prispevka Trivunović, 2020.

2 V diahronem raziskovanju se je izkazalo nujno ločevanje biblijskih (frazemi, ki so bili prevzeti v slovenski knjižni jezik iz jezikov biblijskih predlog kot že izoblikovane frazeološke enote) in izbiblijskih frazemov (frazemi, ki so svojo stalnost in/ali preneseni pomen dobili šele s procesom frazeologizacije v slovenskem jeziku) (Trivunović, 2019, str. 50–51), vendar je za to raziskavo to zanemarljivo. Zapis »(iz)biblijski« zajema oba tipa frazemov: tako biblijske (*Oko za oko, zob za zob*) kot izbiblijske frazeme (vsi ostali), ki so obravnavani v prispevku.

3 Redki zadetki v korpusih iz starejših besedil zato v raziskavo niso bili vključeni (npr. v korpusu *slWaC* najdemo tudi primer: *Bogu, kar je božjega, cesarju, kar je cesarjevega in narodu, kar je narodovega*, ki je iz časopisa *Novice gospodarskih, obrtniških in narodnih stvari* iz leta 1882).

4 O specifikah razvoja (iz)biblijskih frazemov v slovenskem jeziku predvsem o izgubljanju verske konotacije z ustaljenimi sestavinskimi variantami, kjer so besede iz verskega diskurza nadomeščene z versko nezaznamovanimi besedami, gl. Trivunović, 2020.

2 FRAZEOLOŠKA VARIATNOST V TEORIJI IN LEKSIKOGRAFIJI

Stalnost oziroma ustaljenost frazemov pomeni predvsem to, da so frazemi del tvorčevega mentalnega leksikona kot celota in niso tvorjeni sproti. Vendar prav zaradi obstoja različnih variant in modifikacij ne moremo govoriti o popolnem priklicu iz spomina. Ob uporabi v konkretnem besedilu se frazemi namreč obnašajo podobno kot druge besedne zveze: večinoma se pregibajo, vse sestavine niso vedno skupaj in z izjemo določenih (predvidljivih) omejitev podlegajo običajnim pretvorbam (Jakop, 2006, str. 31). Ugotavlja se tudi, da govorci nekatere tipe variant in modifikacij procesirajo podobno kot osnovno obliko, branje in razumevanje jim ne vzame več časa (Geeraert idr., 2017, str. 88). To kaže, da govorci frazemov ne razumevajo in obravnavajo bistveno drugače od dobesednega jezika (Geeraert idr., 2017, str. 88).

Variantnost je stalnosti nasproten pojem, ki »opozarja na lastnost večbesednih leksikalnih enot, da v okviru bolj ali manj omejene možnosti zamenjav sestavnih elementov in parafraz (še vedno) zastopajo eno samo leksikalno enoto« (Gantar, 2007, str. 80). Variante so torej pretežno pogoste in ustaljene zamenjave, ki na pomen frazema ne vplivajo, zato ne moremo govoriti o dveh različnih frazemih,[5] se pa variante istega frazema med seboj lahko razlikujejo v stilni zaznamovanosti in v tipu besedila, za katerega so značilne (Gantar, 2007, str. 230–233). Vendar pa so med frazemi zelo velike razlike, nekateri dopuščajo le malo variant in modifikacij, drugi pa bistveno več (Geeraert idr., 2017, str. 88). Število variant za posamezen frazem ni neomejeno, variante tudi niso poljubne, večinoma so vzete iz omejenega nabora leksemov. Večje število frazemov dopušča dokaj odprte vzorce, ki bi jih po intuiciji opredelili kot stalne. Leksemi teh variant so si pomensko blizu, pomen frazema ostaja enak, vendar so dejanske besede, s katerimi je realiziran frazeološki pomen, drugačne (Hanks idr., 2017, str. 95).

S pojmom stalnosti je povezan pojem ustaljene zgradbe frazema, ki zajema omejitve v izbiri sestavin, njihovi razporeditvi in razmerju med njimi, vendar se predvideva in upošteva tudi določeno mero variantnosti in (omejeno) pretvorbenost (Jakop, 2006, str. 35). Ker ima le malo frazemov popolnoma ustaljeno zgradbo, se raje govori o relativni ustaljenosti zgradbe frazemov, ki

5 E. Kržišnik (2004, str. 205) poleg vidika ohranjanja pomena izpostavlja tudi, da imajo variante del skupnih sestavin in enako skladenjsko vlogo.

je razrahljana s sestavinskimi variantami in pretvorbnimi možnostmi zaradi rabe frazema v določenem besedilu. Kljub temu ima vsak frazem osnovno, invariantno obliko (Jakop, 2006, str. 40). Zaradi različnih realizacij frazema pa ostajata določanje te invariantne oblike in ugotavljanje stalnega jedra frazema najtežji nalogi frazeologije (Gantar, 2001, str. 214), zato je tudi leksikografska obravnava frazeoloških enot izjemno zahtevno delo (Jakop, 2014). Identifikacija slovarske oblike je vedno vsaj deloma subjektivna, saj je variantnost stvar interpretacije raziskovalca (Moon, 1996, str. 252).

2.1 Pomen korpusov za raziskovanje frazeoloških variant in njihov slovarski prikaz

Sodobna frazeologija vse bolj izpostavlja, da frazemi niso tako trdne enote, kot jih prikazuje tradicionalna leksikografija (npr. Kržišnik, 1996, str. 133; Moon, 1996, str. 246; Gantar, 2003, str. 212; Jesenšek in Ulčnik, 2014, str. 282; Geeraert idr., 2017, str. 87; Hanks idr., 2017, str. 95; Meterc, 2019, str. 34–35). Predvsem z uporabo korpusnega pristopa se je namreč pokazalo, da so frazemi veliko bolj nestabilni, saj korpusno gradivo izkazuje veliko raznolikost in spremenljivost teh jezikovnih enot (Jakop, 2014; Jesenšek in Ulčnik, 2014, str. 282). V sodobnem pojmovanju je variantnost frazemov razumljena kot njihova inherentna lastnost in ne kot napaka (Jesenšek in Ulčnik, 2014, str. 285). Ustaljenost sicer še vedno velja za ključno lastnost frazemov, vendar dejanska raba kaže na večjo svobodo pri izbiri posameznih sestavin frazema, »kar postavlja pod vprašaj ne samo do sedaj registrirano in uslovarjeno slovensko frazeologijo, pač pa tudi pojem frazemske stabilnosti in ustaljenosti sploh.« (Gantar, 2003, str. 212; prim. tudi Gantar, 2007, str. 99–101)

Stanje in prikaz frazeologije v slovenskih slovarjih sta dolgo bila nezadovoljiva,[6] saj so kljub teoretično dobro opredeljeni tipologiji variant v mnogih slovarjih in zbirkah izpostavljene le naključne variante, nekatere variante v *Slovarju slovenskega knjižnega jezika* (v nadaljevanju: *SSKJ*) pa verjetno niso bile relevantne niti v času nastanka slovarja (Meterc, 2019, str. 33–34;

6 Tudi tuji raziskovalci opozarjajo, da korpusi kažejo na veliko variantnost, ki v slovarjih in teoriji ni predstavljana v ustreznem obsegu. Uskladiti zahteve teoretičnih spoznanj, slovarskih načel in korpusnih podatkov je ena izmed težjih nalog slovarskega prikaza frazemov (Moon, 1996, str. 245). Dejstvo je, da so frazemi vedno zahtevni za slovarsko obravnavo (Moon, 1996, str. 254).

prim. tudi Jakop, 2014; Gantar, 2006, str. 152). Nesistematičnost obravnave frazeološkega gradiva v *SSKJ* je eden od razlogov, da med uslovarjenimi variantami in stanjem v korpusih pogosto prihaja do velikih razlik (Meterc, 2017, str. 89; Gantar, 2007, str. 33), slovarju se pozna tudi omejenost v gradivu, saj se opira predvsem na leposlovje in pisna besedila iz 1. polovice 20. stoletja, kar pomeni, da v njem ni zajeta živa, novejša frazeologija (Jakop, 2014). Vendar je treba poudariti, da nekatere razlike izvirajo iz dejanskih sprememb v jeziku in frazeologiji, ki so nastale po izidu slovarja (Jesenšek in Ulčnik, 2014, str. 284; Meterc, 2017, str. 216).

Pri preučevanju frazeologije so postali nepogrešljivi korpusi, saj velika količina avtentičnih besedil povečuje verodostojnost raziskav (Gantar, 2007, str. 22–23); zlasti so pomembni pri »ugotavljanju regularnosti pretvorbenih procesov in variantnosti« (Gantar, 2006, str. 154), ki je skupaj s frazeološkimi prenovitvami osrednja tema prispevka. Korpusne analize omogočajo, da se z raziskovanjem dejanskih besedil opazuje in beleži tako regularno jezikovno rabo kot tudi ustvarjalno in nepričakovano (Justin idr., 2015, str. 33). Elektronski korpusi zagotavljajo podrobne in sistematične analize variantnosti, ki ne prinašajo le pregleda nad različnimi variantami določenega frazema, temveč tudi ugotavljanje pogostnostnega razmerja med njimi (Meterc, 2019, str. 37). Njihova uporaba je nujna tudi v sodobni leksikografski praksi, saj sta le tako zagotovljeni konsistentnost in zanesljivost informacij v slovarju (Gantar, 2014). Slovarji, ki so nastali na podlagi korpusne analize, z večjo zanesljivostjo ugotavljajo pomene leksikalnih elementov in njihovo pogostost, novo težo so s korpusi dobili tudi zgledi rabe (Gantar, 2006, str. 155). Ker frazeologija ni vezana le na pisna knjižna besedila, ki bi jih pokrival en referenčni korpus, je pri poglobljenih analizah frazemov in variantnosti nujno vključiti več različnih korpusov (Meterc, 2019, str. 35–36).

Novosti in izboljšave pri slovarski obravnavi frazemov pa niso le posledica razvoja frazeologije kot vede, uporabe korpusnega pristopa in drugih virov, temveč je pomemben dejavnik tudi dejstvo, da novejši slovarji izhajajo v digitalnem okolju, kjer ni prostorskih omejitev, zato je mogoče navajanje več ali vseh variantnih možnosti in njihovo komentiranje (Jesenšek in Ulčnik, 2014, str. 285). To velja tudi za tretjo izdajo *Slovarja slovenskega knjižnega jezika* (v nadaljevanju: *eSSKJ*), ki z empirično preverjenimi podatki o frazemih

in njihovih variantah (Meterc, 2019, str. 34) izboljšuje pomanjkljivo stanje frazeologije v slovarjih slovenščine. eSSKJ, kot je pričakovano za splošne slovarje, navaja najpogostejše oziroma najrelevantnejše variante, specializirani frazeološki in/ali paremiološki slovarji pa naj bi prikazovali tudi redkejše variante (Meterc, 2019, str. 34; prim. tudi Moon, 1996, str. 253). Kljub vsemu pri slovarski obravnavi frazemov še vedno prihaja do dilem, kje je meja med variantami enega frazema in kdaj že lahko govorimo o samostojnih enotah (Gantar, 2007, str. 100). Ker ta meja ni vedno nedvoumno določljiva in ker je v frazeologiji več različnih pristopov in pojmovanj variantnosti, ni neobičajno, da različni slovarji mejo postavljajo različno. Tako se lahko skupino variant razume in v slovarju prikaže kot več različnih frazemov ali pa kot variante enega frazema, med njimi pa ni ostre meje, tudi teoretična spoznanja še zdaleč niso jasna (Moon, 1996, str. 253). Pogosto so v slovaropisni praksi variante obravnavane kot samostojni frazemi, če se zelo razlikujejo v sestavinah (Moon, 1996, str. 253).

Vendar ima tudi korpusni pristop svoje omejitve in pomanjkljivosti. Omejevanje zgolj na korpusno metodologijo je lahko problematično, zato je pomembno pritegniti več obstoječih virov. Zgolj velika pogostost določenega pojava v korpusu namreč ni zadosten kriterij za ugotavljanje splošnega poznavanja in razširjenosti v vsakdanji rabi (Jesenšek in Ulčnik, 2014, str. 281), poleg tega je ključnega pomena tudi sprotno nadgrajevanje jezikovnih orodij in korpusov (Gantar, 2014; Jakop, 2014). Frazeolog se pri analizi variantnosti z uporabo korpusa ne sme zanašati le na lastno jezikovno intuicijo in tudi ne le na eno iskalno metodo, saj bi s tem zaobjel le del variant. Korpusne raziskave frazemov zahtevajo določeno mero zavedanja o variantnosti analiziranih enot, sicer bo zunaj iskanja ostalo več relevantnih pojavitev, zato je potrebna primerna stopnja odprtosti iskanja. Do določene mere lahko pomagajo korpusna orodja, vendar ne vedno (Moon, 1996, str. 252), zato je bistveno uporabiti različne korpusne metode, poleg tega se vse bolj poudarja vključevanje anketiranja rojenih govorcev, saj lahko med korpusnimi podatki in rezultati anket prihaja do pomembnih razlik (Meterc, 2019, str. 42–43; prim. tudi Kržišnik, 1996, str. 136–137). Eden od razlogov za to razhajanje je dejstvo, da je veliko frazemov pogosteje rabljenih v govornem sporazumevanju kot pisnem, hkrati pa nimamo dovolj velikega korpusa, kjer bi lahko preverili govorno

rabo (Meterc, 2017, str. 186). Še posebej je pridobivanje frazeološkega gradiva iz različnih virov pomembno za temeljni splošni slovar (Kržišnik in Jakop, 2015, str. 419). Zato avtorji *eSSKJ* pri reševanju frazeoloških dilem korpusni pristop dopolnjujejo z drugimi viri: »Težave, ki se pojavljajo ob korpusni analizi frazeoloških enot v skrajnih frekvenčnih legah in v povezavi s specifičnimi lastnostmi korpusa, rešujemo z vzporedno analizo sekundarnih jezikovnih virov, npr. slovarjev, spletnih portalov, jezikovnih svetovalnic, anketiranja jezikovnih uporabnikov.« (Gliha Komac idr., 2016, str. 26). Predvsem pri daljših enotah, kot so pregovori in reki, je poseganje po sociolingvističnih podatkih o poznanosti enot med govorci nujnejše, saj korpusna orodja pogosto niso dovolj učinkovita za daljše ustaljene enote, hkrati pa gre pri teh enotah tudi za večjo stopnjo kombinatorike (Gliha Komac idr., 2016, str. 26).

2.2 Frazeološke variante in modifikacije

Bistveno je ločevati med različnimi pojmi: izhodiščna oblika, oblike rabe, normirane variante in modificirane besedilne rabe. Izhodiščno obliko tvorijo vrsta in zaporedje sestavin, njihovo minimalno število in razmerje med njimi (Kržišnik, 1996, str. 133, po Čermák, 1985, str. 184). Oblike rabe vključujejo oblikoslovne prilagoditve in dopustne pretvorbe (Kržišnik, 1996, str. 133–134). Celostne pretvorbe frazemov zajemajo celotno skladenjsko strukturo frazema, s čimer se spremeni njegova skladenjska vloga, vendar pa je frazeološki pomen pri tem ohranjen. Celovite pretvorbe so potrpnjenje, pooziralјenje in posamostaljenje (Kržišnik, 2018, str. 38). Od izhodiščne oblike je treba ločiti normirane variante, ki so v jeziku ustaljene, med seboj so lahko enakovredne ali stilno zaznamovane. Od variant ločeni pojav je modificirana raba, ki je za razliko od variant neustaljena, modifikacije pa so lahko prenovitvene ali neprenovitvene. Neprenovitvene modifikacije so nefunkcionalne spremembe in so z vidika normirane stalnosti neke vrste napake (Kržišnik, 1996, str. 133–134). Delitev modifikacij na prenovitvene in neprenovitvene je upravičena z različnim učinkovanjem v besedilu. Pri neprenovitveni modifikaciji se realizira samo frazeološki pomen zveze, pri prenovitvi pa vedno pride do interakcije med pomenom frazeološke in proste besedne zveze, to razmerje pa je odvisno od konkretne besedilne uresničitve in je lahko vsakokrat različno (Kržišnik, 2006, str. 269).

Prenovitvene modifikacije se tako od neprenovitvenih ločijo po besedilni vlogi. »Prenovitve so namerne in sobesedilno pogojene spremembe pomena in/ali oblike frazeologemov« (Kržišnik, 1990, str. 399; prim. tudi Kržišnik, 1987; Kržišnik, 1996; Kržišnik, 2006). Vezane so na konkreten kontekst oziroma situacijo, v kateri jih avtor tvori, zato za razliko od izhodiščnega frazema obstajajo le na besedilni ravni, ne pa tudi na jezikovni (Kržišnik, 1990, str. 401). Prenovitve so v besedilu stilno opazne, pogosto postanejo aktualni pomenski prenosi, ki so v frazemu postali že bolj ali manj prikriti (Kržišnik, 1996, str. 140–141, po Pogorelec, 1976), nov učinek je lahko dosežen s »podobesedenjem« prenesenega pomena (Gantar, 2007, str. 243), velikokrat se nanašajo na aktualne družbene, politične ali osebne dogodke (Justin idr., 2015, str. 37). Prenovitve niso oblikovane le za poimenovalne potrebe, temveč predvsem z namenom učinkovati, prepričati, vrednotiti oziroma doseči nov pomen ali stilni učinek (Gantar, 2007, str. 241).

Čeprav se prenovitve kot jezikovne inovacije pojavljajo na vseh jezikovnih ravneh (prim. Kržišnik, 2006, str. 265–266), so zelo pogoste prav pri frazemih zaradi večsestavinskosti, ki v primerjavi z besedo omogoča več možnosti za izrabo semantične potence, in zaradi visoke stopnje prepoznavnosti frazemov (Kržišnik, 2006, str. 266). Predpogoj za prenovitve je tako ustaljenost jezikovnih enot, frazemi pa se zdijo posebej primerni za kreativno jezikovno rabo tudi zaradi prisotnih pomenskih prenosov med sestavinami in pomenom zveze kot celote (Gantar, 2007, str. 230). Čeprav so izbira frazemov za prenovitev in možnosti za prenovitve neomejene (Kržišnik, 2006, str. 271), so nekateri frazemi prenovljeni veliko pogosteje kot drugi, hkrati so tudi nekatere sestavine frazemov bolj podvržene zamenjavi kot druge, a hipotetično je možna prenovitev vsake sestavine kateregakoli frazema (Meterc, 2016, str. 125).

Ali bo prenovitev uspešna ali ne, ni toliko odvisno od tipa prenovitve kot od tega, kako je vzpostavljeno pomensko razmerje med izhodiščnim frazemom in prenovitvijo, ter od umeščenosti prenovitve v kontekst in funkcioniranja v njem (Kržišnik, 1990, str. 418). Da naslovnik prepozna modifikacijo in da je ta uspešna, morata biti izpolnjena vsaj dva pogoja: naslovnik mora poznati in razumeti izhodiščni frazem ter vsaj intuitivno poznati postopek tvorjenja jezikovnih inovacij (Kržišnik, 2006, str. 262–263). Med pomenom prenovitve in pomenom izhodiščnega frazema mora biti vzpostavljeno razvidno razmerje,

za uspešno prenovitev je bistveno ustvarjanje napetosti med obema pomenoma (Kržišnik, 1990, str. 401).

Kljub jasno zastavljeni tipologiji in terminologiji je pomembno zavedanje, da je vsako postavljanje mej umetno, saj je jezik kontinuum z mnogimi mejnimi primeri (prim. Meterc, 2019, str. 37). Frazeološke variante so področje, kjer korpusno gradivo narekuje ponovni premislek o slovarski predstavitvi frazemov. Leksikografi in leksikologi se morajo prilagoditi in sprejeti dejstvo, da se jezikovnih pojavov vedno ne da kategorizirati v jasne skupine ter da morajo razviti nove pristope, ki bodo to predstavili na primeren način, ki ne bo zavajal uporabnikov. Korpusno gradivo tako narekuje potrebo po novem definiranju določenih jezikovnih enot z bolj ohlapno zastavljenimi skupinami (Moon, 1996, str. 254). Ločevanje med različnimi tipi pojavitev in razvrščanje zgledov iz avtentičnih besedil je tudi ob jasnih definicijah jezikovnih pojmov pogosto zahtevno ali celo nemogoče. Še posebej težavno je ločevanje med variantami in neprenovitvenimi modifikacijami (Kržišnik, 1996, str. 134). Čeprav je neprenovitvena modifikacija neke vrste napaka, to ne privede vedno do razpada frazeološkega pomena, predvsem zaradi uveljavljanja t. i. potencialne norme, znotraj katere se gibljeta naslovnik in tvorec besedila ter je hkrati tudi vzrok na nastalo napako (Kržišnik, 1996, str. 146–149). Poleg tega lahko neprenovitvene modifikacije napovedujejo tudi nove variante in spremembo norme (Kržišnik, 2006, str. 262–264). Dokumentiranih je več frazemov, ki so bili prvotno nenormativne napake oziroma modifikacije, a so se z rabo toliko ustalile, da so bile sprejete v slovar kot variante (Kržišnik, 1996, str. 137). Podobno lahko z diahronega vidika tudi prenovitev postane ustaljena varianta ali nov frazem (Kržišnik, 1996, str. 137, 141–142; Justin idr., 2015, str. 36).

3 IZBOR FRAZEMOV IN METODOLOGIJA

3.1 Izbor frazemov

V raziskavo je vključenih sedem (iz)biblijskih frazemov[7] (šest pregovorov, ki so spodaj zapisani z veliko začetnico, in en glagolski frazem, ki je zapisan z

7 Zamejitev frazeologije in enot, ki jih preučuje, ni enotna; problematika obsega in (delnega) prekrivanja frazeologije in paremiologije je natančno obravnavana npr. v Meterc, 2017, str. 23–29. V tem prispevku je termin frazem rabljen v širšem smislu in zajema tudi pregovore in reke.

malo začetnico). Izhodiščne oblike vseh frazemov so v tem prispevku vseskozi zapisane tako, kot so predstavljene v *Slovarju slovenskega knjižnega jezika, druga izdaja* (v nadaljevanju: *SSKJ2*),[8] saj je to edini uporabljeni slovarski vir, ki obravnava vse frazeme. V raziskavo je bil vključen tudi *Slovar slovenskih frazemov* (v nadaljevanju: *SSF*), v katerega pa so vključeni le nekateri od izbranih frazemov.[9] Ker je za uspešno prenovitev bistveno poznavanje izhodiščnega frazema, so bili štirje frazemi (*Oko za oko, zob za zob*; *Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega*; *Kdor je brez greha, naj prvi vrže kamen vanjo*; *Beseda je meso postala*) izbrani s seznama najbolj poznanih in pogostih slovenskih pregovorov in sorodnih paremij (Meterc, 2017, str. 237–259). Ostali trije frazemi (*Človek obrača, Bog obrne*; *Duh je sicer voljan, ali meso je slabo*; *metati bisere svinjam*) so bili izbrani glede na lastno jezikovno kompetenco (poznavanje in predvidevanje pogostosti) z obsežnega seznama različnih biblijskih stalnih besednih zvez (Galer, 2001, str. 56–82). Ker je analiziran zelo ozek in specifičen nabor frazemov, je pri posploševanju ugotovitev treba biti previden.

3.2 Uporabljeni korpusi

Za ugotavljanje, v katerih primerih gre za ustaljeno variantnost in v katerih za neprenovitvene modifikacije, kjer frazeološki pomen ne razpade zaradi uveljavljanja potencialne norme, je ključnega pomena količina in raznovrstnost gradiva, ki jo imamo na voljo. Da bi bilo umeščanje pojavitev v zastavljene skupine zanesljivejše, so bili v raziskavo vključeni trije korpusi, ki prinašajo različne tipe besedil. To, katera besedila imajo na jezikovni razvoj največji vpliv, se je skozi desetletja spreminjalo. V devetdesetih letih so bila zelo vplivna publicistična besedila, radio in televizija, saj so dosegli veliko število naslovnikov. Zanje je značilen vdor prvin govorjenega jezika in rahljanje norme, za doseganje želenega učinka pogosto izrabljajo prenovitve frazemov. Pred

8 Razlika v zapisu izhodiščne oblike v primerjavi s *SSKJ2* je le v sestavini *bog/Bog*, ki je v slovarju v vseh primerih zapisana z malo, tukaj pa se uporablja zapis z veliko začetnico. V uporabljenih korpusih se pojavljata oba zapisa, a prevladuje zapis z veliko. Tudi drugi raziskovalci opozarjajo na pravopisno variantnost (prim. Meterc, 2017). Glede na biblijski izvor frazemov se zdi utemeljen zapis z veliko začetnico.

9 Novi *eSSKJ* ne vsebuje nobenega od analiziranih frazemov in zato ni vključen v to raziskavo. Podobno tudi *Slovar pregovorov in sorodnih paremioloških izrazov* (Meterc, 2020–) v prvem prirastku ne prinaša nobenega od obravnavanih frazemov.

tem so bila vplivna in normotvorna predvsem umetnostna besedila, ki imajo danes bistveno manjši vpliv zaradi majhnega kroga naslovnikov (Kržišnik, 1996, str. 138–139). V sodobnem času je opazen velik pomen spletnih besedil, povečuje se vpliv družbenih omrežij, za katere je značilen bistveno drugačen jezik od pisnega standarda (Fišer idr., 2016, str. 67–68). Sodobnejše raziskave kažejo, da so frazemi zelo pogosti v tipično spletnih besedilih, vendar je v tem tipu besedil bistveno več neopredeljivih primerov, saj zanje (predvsem tvite in spletne komentarje) velja kratkost in specifičnost izražanja (Justin idr., 2015, str. 34–35). Ker je frazeologija vezana predvsem na neformalne govorne položaje in nestrokovna besedila (Gantar, 2001, str. 212), predvidevam, da bodo v spletnih besedilih frazemi in njihove prenovitve pogostejše kot v znanstvenih besedilih in leposlovju (prim. tudi Kržišnik-Kolšek, 1988 v Gantar, 2007, str. 241–242). Kot relevantne za frazeološke raziskave se poleg *Gigafide* kot največjega referenčnega korpusa pogosto izpostavlja korpuse spletnih besedil (*slWaC*), korpuse, ki zajemajo slovenske tvite in druga družbena omrežja (*Janes*), korpuse govorjenega jezika ter druge specializirane korpuse (Meterc, 2019, str. 37; Gantar, 2014). Zato so bili za to raziskavo uporabljeni trije prej omenjeni korpusi: *Gigafida 2.0*, *slWaC* in *Janes*,[10] za iskanje je bilo uporabljeno orodje noSketch Engine infrastrukture CLARIN.SI.[11] Vsa iskanja so bila izvedena v vseh treh korpusih.

Gigafida 2.0[12] je referenčni korpus standardne slovenščine, saj so bila iz prejšnje različice korpusa odstranjena besedila z nestandardnimi jezikovnimi značilnostmi. V primerjavi s prejšnjo različico vsebuje večji delež leposlovnih in šolskih besedil, dodanih je tudi več besedil spletnih besedilodajalcev z večjo produkcijo (npr. novičarski portali, dnevni časopisi). Največji delež besedil v korpusu predstavljajo časopisi (47,8 %), sledijo spletna besedila (28,0 %), revije (16,5 %), stvarna besedila (3,8 %) in leposlovje (3,5 %). Besedila drugih tipov predstavljajo le 0,3 % (Krek idr., 2019, str. 1–7; prim. tudi Krek idr., 2016). Korpus vsebuje 1.333.360.653 pojavnic in je največji od uporabljenih korpusov.

10 Raziskava ni zajela govorjenega jezika.

11 https://www.clarin.si/noske/index.html

12 https://www.clarin.si/noske/run.cgi/corp_info?corpname=gfida20_dedup&struct_attr_stats=1

slWaC[13] je korpus slovenskega spleta. V prvo različico korpusa so bila vključena besedila samo z domene.si, v novejšo različico pa so vključene tudi spletne strani drugih domen (Erjavec in Ljubešić, 2014, str. 50). Korpus vsebuje 895.903.321 pojavnic.

Korpus *Janes*[14] sestavljajo besedila slovenskih družbenih omrežij. Najnovejša različica korpusa vsebuje pet zvrsti javno objavljenih uporabniških spletnih vsebin: tvite, forume, komentarje na spletne novice, bloge in komentarje nanje ter uporabniške in pogovorne strani na *Wikipediji* (Fišer idr., 2020, str. 225; prim. tudi Fišer idr., 2016, str. 68–69). Največji je delež tvitov, ki predstavljajo več kot polovico celotnega korpusa, najmanj pa je komentarjev z *Wikipedije* (Fišer idr., 2020, str. 232). Korpus vsebuje 252.904.238 pojavnic.

3.3 Metode iskanja variant

Kot posebej relevantne frazeologi izpostavljajo naslednje metode iskanja (povzeto po Meterc, 2019, str. 38–43): iskanje po dveh sestavinah z izločanjem določenih besed iz okolice, iskanje po domnevno najpogostejših sestavinah, iskanje s kombinacijo dveh sestavin z dodajanjem pozitivnega in negativnega filtra, iskanje s kombinacijami leksikalnih sestavin (problem lahko nastane pri sestavinah istega tematskega polja, saj se lahko pojavljajo v več različnih frazemih) in iskanje po polovicah pregovorov. V primerih izjemno variantnih frazemov, predvsem pri prepletanju skladenjske in sestavinske variantnosti, je najboljša rešitev kombiniranje različnih metod iskanja, saj nekatera zajamejo predvsem sestavinske, druge pa predvsem skladenjske variante (Meterc, 2019, str.43).

Ker je prenavljanje frazemov kreativen proces z neomejenimi možnostmi, se ne da predvideti vseh možnih prenovitev. V raziskavi so obravnavani le naslednji tipi prenovitev, drugi tipi prenovitev, ki tem vzorcem ne sledijo (npr. *Kdor še nikoli ni nič privatnega stiskal v službi, naj prvi vrže printer vame*)[15], niso obravnavani.

13 https://www.clarin.si/noske/run.cgi/corp_info?corpname=slWaC&struct_attr_stats=1

14 https://www.clarin.si/noske/run.cgi/corp_info?corpname=Janes&struct_attr_stats=1

15 Primeri so predstavljeni tako, kot so zapisani v korpusu, popravljeni so le presledki ob ločilih.

1. *Oko za oko, zob za zob*[16]
   - x za x, oko za oko / oko za oko, x za x
   - x za x, zob za zob / zob za zob, x za x
2. *Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega*
   - dati cesarju, kar je cesarjevega, in (dati) x, kar je x-ovega[17]
   - dati Bogu, kar je božjega, in (dati) x, kar je x-ovega
   - dati x, kar je x-ovega
3. *Človek obrača, Bog obrne*
   - človek obrača ...
   - ... Bog obrne
   - x obrača, y obrne
4. *Kdor je brez greha, naj prvi vrže kamen vanjo*
   - ... naj prvi vrže kamen / naj prvi vrže kamen ...
   - kdor je brez greha ... / ... kdor je brez greha
5. *Duh je sicer voljan, ali meso je slabo*
   - duh je voljan ... / ... duh je voljan
   - ... meso je slabo / meso je slabo ...
6. *Beseda je meso postala*
   - beseda je x postala
   - x je meso postal
7. *metati bisere svinjam*
   - metati x svinjam
   - metati bisere x

16 Glede na predhodno objavo (Trivunović, 2020) so na novo upoštevani nekateri tipi prenovitev: dodajanje novega dela pred ustaljenim delom (npr. *Tu gre kg za kg, oko za oko* (*slWaC*)) in tip prenovitev, kjer sta v novem delu dve različni besedi (npr. *Zob za zob, oko za oko ... in rit za povišico* (*Janes*)).

17 Upoštevani so tudi primeri, kjer se samostalnik in pridevnik ne ujemata oziroma pridevnik ni tvorjen iz samostalnika (npr. *Daj cesarju kar je cesarjevega in državi kar je sosedovega* (*slWaC*); *Daj bogu, kar je božjega, in Cerkvi, kar je slovenskega* (*Gigafida 2.0*; podčrtala E.T.)), kar je bilo v predhodni objavi že upoštevano (Trivunović, 2020).

Najprej so bila izvedena ločena iskanja za pojavitve izhodiščnega frazema in prenovitev, vendar se je kasneje pri večjem številu frazemov izkazalo, da je bolje iskati samo po dveh ali treh najmanj variantnih sestavinah ter nato pregledati rezultate in s pomočjo različnih opcij filtriranja rezultatov ločiti pojavitve izhodiščnega frazema od prenovitev. Za ponazoritev vzemimo frazem *Oko za oko, zob za zob*. Za iskanje prenovitev sta bili izvedeni dve iskanji, ki sta zajeli nadaljevanje dela *oko za oko* (prvo iskanje) oziroma *zob za zob* (drugo iskanje). Iskanje ni zajelo izhodiščnega frazema, temveč samo prenovitve, ki nadaljujejo vzorec in kjer se v prenovljenem delu dvakrat pojavi enaka beseda. Z iskalnimi pogoji so bile izločene že uslovarjene variante s samostalniki *oko*, *zob* in *glava* v drugem delu.

1. [lemma_lc="oko"] [word="za"] [lemma_lc="oko"] []{0,2} 1:[tag="S.*" & lemma_lc!="oko" & lemma_lc!="zob" & lemma_lc!="glava"] [word="za"] 2:[tag="S.*" & lemma_lc!="oko" & lemma_lc!="zob" & lemma_lc!="glava"] & 1.lemma_lc=2.lemma_lc

2. [lemma_lc ="zob"] [word="za"] [lemma_lc ="zob"] []{0,2} 1:[tag="S.*" & lemma_lc!="oko" & lemma_lc!="zob" & lemma_lc!="glava"] [word="za"] 2:[tag="S.*" & lemma_lc!="oko" & lemma_lc!="zob" & lemma_lc!="glava"] & 1. lemma_lc =2. lemma_lc

Za iskanje pojavitev izhodiščnega frazema sta bili prav tako izvedeni dve iskanji: [word="oko"] [word="za"] [word="oko"] in [word="zob"] [word="za"] [word="zob"]. Pri drugem iskanju so bili iz okolice izločeni zadetki z besedo *oko*, saj je te primere zajelo že prvo iskanje. Iz zadetkov so bile izločene že prej zabeležene prenovitve in nefrazeološke pojavitve zveze, hkrati pa so bile izpisane prenovitve, ki jih iskanje prenovitev ni zajelo, a ustrezajo tipom, ki so predmet raziskave. To so primeri, kjer se v prenovljenem delu ne pojavi samostalnik, temveč beseda pripada drugi besedni vrsti (npr. *zob za zob oko za oko prav za prav* (*Janes*)), primeri z več kot samo *za* med besedama (npr. *oko za oko, zob za zob, šnops za še en šnops* (*Janes*)) ter dodatno še primeri, kjer je prenovljeni del pred ustaljenim delom frazema (npr. *Udarec za udarec, oko za oko* (*Gigafida 2.0*)), in primeri, kjer sta v prenovljenem delu dve različni besedi (npr. *zob za zob, oko za oko*. *Denar za trpljenje, bi dodali danes* (*Gigafida 2.0*)).

Za zelo učinkovito se izkazalo tudi iskanje po na novo odkritih polovicah pregovorov v kombinaciji z izločanjem ali dodajanjem besed v okolici. Ta metoda

je bila še posebej relevantna pri frazemu *Kdor je brez greha, naj prvi vrže kamen vanjo*, saj se je prvi del frazema izkazal za bistveno bolj variantnega, kot se je predvidevalo. Drugi del frazema je glede na pregledano korpusno gradivo manj varianten, najbolj trdno so povezane sestavine *prvi*, *vreči* in *kamen*. Z iskanjem po teh treh sestavinah sem nato preverjala variantnost prvega dela, ki izkazuje sestavinsko variantnost z besedami besednih družin *greh*, *kriv* in *nedolžen*, hkrati so variante tega dela strukturno zelo različne, pogoste so tudi razne pretvorbe. Najpogosteje se pojavljajo: *kdor je brez greha* (in pretvorba: *tisti, ki je brez greha*), *kdor (nič) ne greši*, *kdor (še) ni grešil*, *kdor ni kriv*, *kdor je brez krivde* (in pretvorba: *ta, ki je brez krivde*), *kdor je nedolžen* (in pretvorba: *tisti, ki je nedolžen*). Izpisu različnih variant prvega dela so sledila iskanja po novo odkritih polovicah ali sestavinah novo odkritih polovic z namenom hkrati zajeti čim več zgledov prenovitev in morebitnih variant drugega dela.

Proces iskanja je bil podoben za vse frazeme, pogosto je zahteval veliko iznajdljivosti, nekaterim naključno najdenim primerom je sledilo popravljanje iskalnih pogojev, pogosto v smer večje odprtosti, dopuščanja več prostih mest ter odstopov od pravopisne in frazeološke norme. Predvsem je bilo težavno predvidevati, koliko prostih mest naj se pusti v zapisu in med katerimi sestavinami, saj se sestavine frazemov ne pojavljajo vedno skupaj (npr. *Besede iz Avstralije so meso postale* (*Gigafida 2.0*)). Tako so se za učinkovitejša pogosto izkazala bolj odprta iskanja, npr. iskanje [lemma="biser"] []{0,2} [lemma="svinja"] je prineslo več relevantnih zadetkov kot [lemma="metati"] [lemma="biser"] [lemma="svinja"], saj je poleg predvidevanja potencialnih mest med sestavinami frazema zajelo tudi glagolske variante.

Za iskanje po korpusih ni zanemarljiv pravopisni vidik. Prvo za to raziskavo relevantno pravopisno vprašanje so ločila: opuščanje ločil oziroma redkeje dodajanje in nadomeščanje enega ločila z drugim (npr. vezaj namesto pogostejše vejice) (prim. Jesenšek in Ulčnik, 2014, str. 282–285). Predvsem za objave na družbenih omrežjih je značilno, da pogosto odstopajo od norme, zato so bila iskanja zastavljena tako, da so zajela zapise z različnimi ločili in tudi brez njih. Izhodiščni frazemi in njihove prenovitve se brez vejic pojavljajo v vseh treh korpusih v različnih tipih besedil, npr. *Cesarju kar je cesarjevega* (*Gigafida 2.0*, publicistično besedilo); *Kaj ti drugega preostane kot oko za oko zob za zob* (*slWaC*, spletno besedilo); *Dal je cesarju kar je cesarjevega in cestarju*

kar je cestarjevega (*Janes*, spletno besedilo). Redkeje se namesto vejice pojavi drugo ločilo, npr. vezaj (*človek obrača - trg obrne* (*Janes*)) ali tri pike (*zob za zob … oko za oko … rit pa za kariero* (*Janes*)). Kot je bilo že omenjeno, je pri nekaterih frazemih zaznana variantnost v veliki začetnici pri samostalniku *Bog/bog*, zato je bilo treba z iskanjem zajeti obe možnosti zapisa.

V nekaterih primerih je bilo treba določen tip iskanja opustiti, ker se iskalnih pogojev ni dalo oblikovati tako, da bi našli obvladljivo število zadetkov. Tako v raziskavo niso bile vključene prenovitve *x za x*, ki se pojavljajo brez vsaj enega dela izhodiščnega frazema, saj je pojavitev, ki ustrezajo tem pogojem, preveč. Podobno tudi v primeru (*dati) x, kar je x-ovega*. Izhodiščni frazem se pojavlja tudi brez glagola *dati*, vendar se je iskanje [lemma="dati"]? [tag="S.*" & lemma_lc!="bog" & lemma_lc!="cesar"] []{0,1} [word="kar"] [word="je"] [tag="P.*"] within <s/> izkazalo za preveč odprto. Zato je bilo iskanje prenovitev omejeno na pojavitve z glagolom *dati* in prenovitve, ki nadaljujejo enega od delov izhodiščnega frazema.

4 REZULTATI

4.1 Neujemanja ustaljenih variant v slovarjih in korpusih

Od obravnavanih frazemov je v slovarjih in literaturi največ variant zabeleženih za frazem *Oko za oko, zob za zob*. *SSKJ2* prinaša sicer le eno varianto frazema, v *SSF* pa so zabeležene naslednje: 1) *oko za oko, zob za zob*; 2) *oko za oko*; 3) *zob za zob*; 4) *zob za zob, glava za glavo* in 5) *glava za glavo*, vendar so nekatere navedene le pri eni iztočnici, nekatere pa pri več: pri iztočnici *oko* sta obravnavani 1) in 2), pri iztočnici *zob* 1), 3) in 4), pri iztočnici *glava* pa samo 5); vse variante niso nikjer zbrane na enem mestu. Takšne nedoslednosti v obravnavi frazeoloških variant v *SSF* so dokaj pogoste (prim. Konicka, 2012, str. 172–175). Poleg naštetih variant je v vseh treh korpusih pogosto tudi zaporedje *Zob za zob, oko za* oko, ki ga *SSF* ne navaja. Meterc (2017, str. 90) temu dodaja še variante s *kri za kri* ter opozarja, da je za ta frazem značilno podaljševanje z ustaljenimi in neustaljenimi sestavinami, kar je povezano z dejstvom, »da je paremija že v izvornem biblijskem odlomku del daljše verige« (Meterc, 2017, str. 90).[18] Po pregledu primerov v vseh treh korpusih bi med ustaljene variante lahko šteli še variante z *življenje za življenje*.

18 V *Bibliji* (www.biblija.net) se zveza pojavi v štirih odlomkih, od tega trikrat kot del daljše verige: »Če pa se nesreča pripeti, daj življenje za življenje, oko za oko, zob za zob,

Kot je bilo opozorjeno že pri prikazu metodologije, je za frazem *Kdor je brez greha, naj prvi vrže kamen vanjo* značilna bistveno večja variantnost, kot jo prikazujejo slovarji. Že Meterc (2017, str. 247) opozarja, da zaimek *vanjo* ni obvezna sestavina frazema, temveč fakultativna, kar potrjuje tudi pregled primerov v korpusu. Po pregledu primerov v vseh treh korpusih so bile kot ustaljene variante prvega dela upoštevane različice s sestavinami iz besednih družin *greh*, *kriv* in *nedolžen*.

Frazem *Človek obrača, Bog obrne* je v *SSKJ2* obravnavan samo v obliki *človek obrača, bog obrne*, vendar je pregledano korpusno gradivo pokazalo, da je samostalnik *Bog* pogosto nadomeščen s samostalnikoma *življenje* in *usoda*, redkeje tudi s *čas* v taki meri, da lahko govorimo o ustaljeni variantnosti in ne o modificirani rabi. Da so to že ustaljene variante izhodiščnega frazema, dokazujejo tudi uvajalna sredstva: *kot pravi pregovor »človek obrača, usoda obrne«* (*Gigafida 2.0*); *Kako gre že tisti rek … »Človek obrača, življenje obrne«* (*Janes*; podčrtala E. T.). V nekaterih primerih se skupaj pojavi več variant, npr. *Pravijo, da človek obrača, življenje ali bog pa obrne* (*Gigafida 2.0*). Pogosto je frazem rabljen tudi v množini *ljudje obračajo* ali *ljudje obračamo*. Opazne so različne slogovne uresničitve (prim. Jesenšek in Ulčnik, 2014, str. 284), saj se poleg slogovno nevtralnega izraza *življenje* v korpusu *Janes* pojavi tudi bistveno bolj pogovorna sopomenka *lajf* (npr. *Človk obrača, lajf obrne*), pri več sestavinah je pogost neknjižni zapis (npr. *Človek obrača, življeje obrne*; *človuk obrača, buh obrne*).

Že v obstoječih slovarskih virih se frazem *Duh je sicer voljan, ali meso je slabo* pojavi v nekoliko različnih variantah: *duh je sicer voljan, ali meso je slabo* (*SSKJ2*) in *duh je voljan, a meso je slabo* (*SSF*), vendar korpusno gradivo kaže še večjo variantnost. Samostalnik *duh* pogosto nadomeščata samostalnika *volja* in *um*. Veznik *sicer* se pojavlja pogosto, a ne vedno. Pridevnik *voljan* je pogosto nadomeščen s pridevniki *močan*, *dober* in *čvrst*. Od veznikov se pojavijo: *a*, *ampak*, *toda*, *le*, *samo*, *ali*, *pa*. V drugem delu se namesto pridevnika

roko za roko, nogo za nogo, opeklino za opeklino, rano za rano, modrico za modrico!« (2 Mz 21,23–25) »Kdor svojemu rojaku prizadene poškodbo, naj mu storijo, kakor je storil: zlom za zlom, oko za oko in zob za zob; kakor je on prizadel človeka, tako naj prizadenejo njega.« (3 Mz 24, 19–20) »Tvoje oko bódi brez usmiljenja: življenje za življenje, oko za oko, zob za zob, roka za roko, noga za nogo!« (5 Mz 19,21) »Slišali ste, da je bilo rečeno: Oko za oko in zob za zob.« (Mt 5,38)

slab pojavita tudi pridevnika *slaboten* in *šibek*. Težko se določi mejo, kaj je ustaljena variantnost in kaj je neprenovitvena modifikacija, saj se variante prepletajo med sabo, zagotovo pa niso prenovitve, saj spremembe niso hotene, v besedilu nimajo prenovitvenega učinka in tudi pomen ostaja nespremenjen.

V *SSKJ2* je frazem *metati bisere svinjam* obravnavan samo v obliki *metati bisere svinjam*, *SSF* pa navaja še pretvorbo z glagolnikom *metanje* in opozarja, da se v tem primeru frazem rabi tudi v krajši obliki *metanje biserov*, torej brez samostalnika *svinja*. Glede na pregledano korpusno gradivo bi kot varianti lahko opredelili pojavitve s predlogoma *pred* in *med*: *metati bisere pred svinje* in *metati bisere med svinje*. Namesto glagola *metati* se po večkrat pojavljajo: *vreči*, *dati*, *dajati*, *trositi*, po enkrat tudi *deliti* in *sipati*, ki nista ustaljeni varianti, temveč neprenovitveni modifikaciji.

Pri frazemih *Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega* in *Beseda je meso postala* v korpusu niso bile potrjene nove ustaljene variante, vendar je opazna velika raznolikost v oblikah rabe. Prvi se tako ne rabi samo v velelniku 2. osebe množine, temveč v različnih oblikah npr. *Naša oblast nas je obvestila, da je čas, da damo cesarju, kar je cesarjevega* (*Gigafida 2.0*); *Daš Bogu kar je božjega, Cesarju, kar je cesarjevega* (*slWaC*); *Naj da Bogu, kar je božjega, in cesarju, kar je cesarjevega* (*Gigafida 2.0*). Drugi frazem se pojavlja v različnih časih in številih (*besede so meso postale*, *beseda bo meso postala*, *beseda meso postane*), zanikano (*besede niso meso postale*), pooziralјeno (*besede, ki so meso postale*) ipd.

Štirje frazemi *Oko za oko, zob za zob*; *Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega*; *Kdor je brez greha, naj prvi vrže kamen vanjo* in *Duh je sicer voljan, ali meso je slabo* se pogosto pojavljajo tudi z obratnim vrstnim redom delov, torej tudi: *Zob za zob, oko za oko*; *Dajte Bogu, kar je božjega, in cesarju, kar je cesarjevega*; *Naj prvi vrže kamen, kdor je brez greha* in *Meso je slabo, a duh je močan*. Na osnovi prvih treh frazemov so se uveljavili tudi krajši frazemi. Pri frazemu *Oko za oko, zob za zob* se pojavlja več krajših frazemov s samo enim delom, pri frazemu *Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega* se osamosvajata in rabita samostojno oba dela. Iz frazema *Kdor je brez greha, naj prvi vrže kamen vanjo* je nastal glagolski frazem *prvi vreči kamen* s pomenom ‘obsojati napake drugih’ (npr. *Valič je velik v razumevanju šibkosti in zadržanem opozarjanju na to, da poglejmo vase,*

preden prvi vržemo kamen (*Gigafida 2.0*); *Posebno še ker biti »na oblasti« prinaša vse manjši ugled in je v slovenski družbi vse več »pravičnikov«, ki prvi vržejo kamen* (*Janes*)). Krajšanje frazemov in osamosvajanje posameznih delov je v frazeologiji dokaj pogost pojav, v *SSF* je npr. zabeleženo krajšanje frazema: *umiti si roke (kot Pilat)*. Pri teh frazemih bi bilo smiselno razmisliti o uslovarjenju novonastalih krajših frazemov kot samostojnih izhodiščnih oblik ali vsaj opozoriti na možnost samostojne rabe posameznega dela.

4.2 Mejni primeri

Predvsem pri frazemih *Duh je sicer voljan, a meso je slabo* in *Kdor je brez greha, naj prvi vrže kamen vanjo* je iz zgledov v korpusih težko povsem enoznačno določiti, kaj je že ustaljena variantnost in kaj neprenovitvena modifikacija. Veliko manj je dvoma o tem, v katerih primerih gre za prenovitve, vendar je v nekaterih primerih konteksta premalo, da bi lahko razbrali natančen pomen prenovitve (npr. *želja je vroča, meso je šibko* (*Janes*)).

Posebej težaven za ločevanje med različnimi tipi pojavitev – tako med ustaljenimi variantami in neprenovitvenimi modifikacijami kot neprenovitvenimi in prenovitvenimi modifikacijami – je frazem *Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega*, saj večina sestavinskih zamenjav ostaja znotraj pomenskega polja posvetne in cerkvene oblasti, zato tudi konotativni pomen ostaja blizu izhodiščnemu frazemu, poleg tega pogosto zaradi pomanjkanja konteksta ni razvidno, ali gre za hoteno spremembo (prenovitveno modifikacijo), napako (neprenovitveno modifikacijo) ali pa celo za že ustaljeno variantnost. V nadaljevanju so podrobneje predstavljene relevantne pojavitve iz korpusov.

V korpusih se pojavita dve podaljšavi izhodiščnega frazema, tako da ima prenovitev tri dele, pri čemer ena od njiju ostaja znotraj pomenskega polja verske oblasti: *Bogu kar je božjega, cesarju kar je cesarjevega, papežu kar je papeževega* (*Gigafida 2.0*), druga pa ne *Vendar, bogu kar je božjega, cesarju, kar je cesarjevega, in kulturi, kar je kulturnega* (*Gigafida 2.0*). Prvi so podobne štiri dvodelne prenovitve (*Gigafida 2.0*: 2, *slWaC*: 2), kjer ostaja konotativni pomen skoraj nespremenjen, saj je kljub zamenjavi sestavin ohranjeno razmerje med posvetno oblastjo (*cesar*, *gospod*, *gospodar*) in Bogom oziroma cerkveno oblastjo (*Bog*, *papež*): *daj cesarju, kar je cesarjevega, in papežu, kar je papeževega* (*Gigafida 2.0*); *srednjeveške mantre o Bogu, kar*

je božjega, gospodu, kar je gospodovega (*slWaC*); *dati Bogu, kar je božjega, in gospodarju, kar je gospodarjevega* (prenovitev se pojavi v dveh korpusih: *Gigafida 2.0* in *slWaC*). Poleg tega se pojavijo še štirje primeri enodelnih zvez (*Gigafida 2.0*: 2, *Janes*: 1, *slWac*: 1), vendar ni mogoče z gotovostjo trditi, ali so to prenovitvene ali neprenovitvene modifikacije ali pa morda že ustaljene variante. Z največjo gotovostjo lahko o neprenovitveni modifikaciji govorimo v primeru tvita, kjer je samostalnik *cesar* nadomeščen s samostalnikom *car*, saj gre za sopomenki s podobnim slušnim vtisom. Drugače je v primerih, kjer je *cesar* nadomeščen z *državo*, *Bog* pa s *papežem*, npr. *Dajo državi kar je državnega* (*slWaC*); *Dajmo papežu, kar je papeževega* (*Gigafida 2.0*).

Po definiciji so prenovitve enkratne in namerne tvorbe, vendar se nekatere zveze pojavijo večkrat (predvsem z *država* in *papež*) tako v samostojnih enodelnih zvezah kot v dvodelnih in trodelnih zvezah. V nekaterih kontekstih delujejo prenovitveno (npr. kot tretji del), v nekaterih ne ali pa je konteksta premalo, da bi lahko ugotovili, ali gre za nameren odstop od izhodiščnega frazema ali napako ali morda že za (skoraj) ustaljeno variantnost. V teh primerih gre za odmik v smeri sodobne stvarnosti, predvsem pri pogostem nadomeščanju samostalnika *cesar* s samostalnikom *država*.[19] Za potrebe te raziskave so bile vse zgoraj navedene pojavitve obravnavane kot prenovitve, vendar bi jih lahko obravnavali tudi drugače.

4.3 Primerjava med korpusi

V Tabeli 1 je v prvi vrstici pri vsakem frazemu prikazano približno število pojavitev izhodiščnega frazema in natančno število prenovitev, ki so bile vključene v raziskavo. V drugi vrstici so prikazane relativne frekvence[20] izhodiščnih frazemov in prenovitev v korpusih.

19 Tem prenovitvam sta podobni dve pojavitvi, kjer je skupaj navedenih več izrazov, kar bi lahko kazalo na njihovo izenačitev: *daj cesarju kar je cesarjevega in bogu, papežu, kleru kar je božjega* (*Gigafida 2.0*); *daj cesarju (državi), kar je cesarjevega in bogu, (cerkvi) kar je božjega* (*slWaC*).

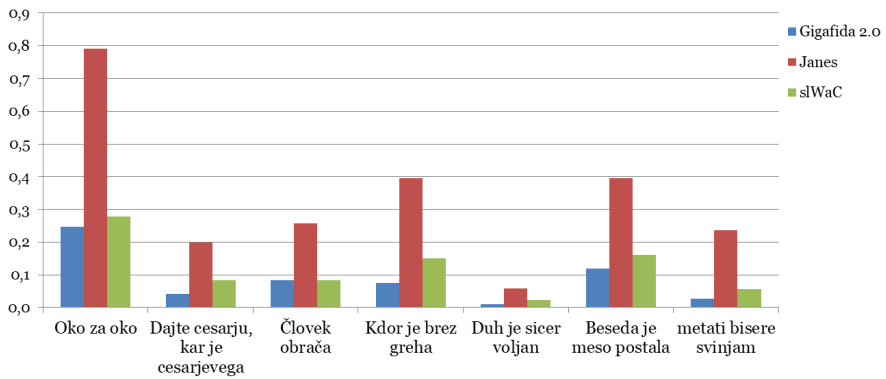
20 Relativna frekvenca pomeni število pojavitev na milijon pojavnic. Ta podatek omogoča primerjavo med korpusi različnih velikosti.

Tabela 1: *Zastopanost izhodiščnih frazemov in prenovitev v korpusih*

	Izhodiščni frazem				**Prenovitve**			
	GF 2.0	*Janes*	*slWaC*	**Skupaj**	*GF 2.0*	*Janes*	*slWaC*	**Skupaj**
Oko za oko, zob za zob	330	200	250	**780**	19	13	9	**41**
	0,2476	0,7905	0,279	**0,3143**	0,0143	0,0514	0,01	**0,2476**
Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega	55	50	75	**180**	8	4	5	**17**
	0,0413	0,1976	0,0837	**0,0725**	0,006	0,0158	0,0056	**0,0068**
Človek obrača, Bog obrne	110	65	75	**250**	18	9	12	**39**
	0,0825	0,2569	0,0837	**0,1007**	0,0135	0,0356	0,0134	**0,0157**
Kdor je brez greha, naj prvi vrže kamen vanjo	100	100	135	**335**	12	36	22	**70**
	0,075	0,3953	0,1507	**0,135**	0,009	0,1423	0,0246	**0,0282**
Duh je sicer voljan, ali meso je slabo	15	15	20	**50**	1	1	1	**3**
	0,0113	0,0593	0,0223	**0,0201**	0,0008	0,004	0,0011	**0,0012**
Beseda je meso postala	160	100	145	**405**	25	14	11	**50**
	0,12	0,3953	0,1618	**0,1632**	0,0188	0,0553	0,0123	**0,0201**
metati bisere svinjam	35	60	50	**145**	2	0	1	**3**
	0,0263	0,2372	0,0558	**0,0584**	0,0015	0	0,0011	**0,0012**
Skupaj po korpusih	**805**	**590**	**750**	**2145**	**85**	77	**61**	**223**

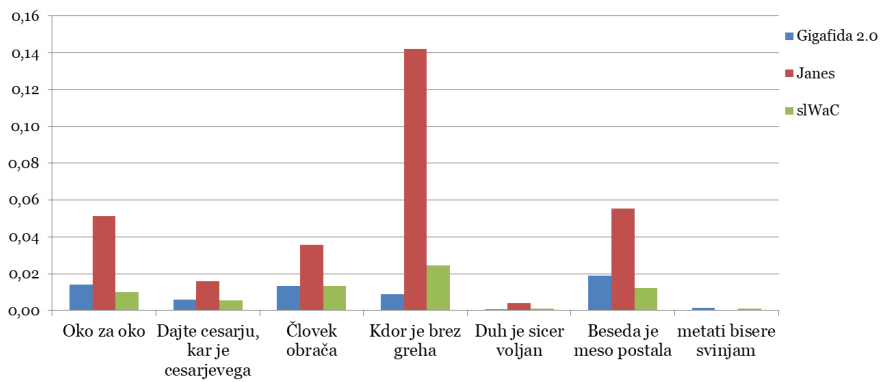
Zanimalo me je, ali obstaja povezava med pogostostjo izhodiščnega frazema in številom prenovitev. Štirje frazemi z največ pojavitvami izhodiščnega frazema (*Oko za oko, zob za zob*; *Beseda je meso postala*; *Kdor je brez greha, naj prvi vrže kamen vanjo*; *Človek obrača, Bog obrne*) imajo največ prenovitev in trije frazemi z najmanj pojavitvami izhodiščnega frazema (*Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega*; *metati bisere svinjam*; *Duh je sicer voljan, ali meso je slabo*) imajo tudi prenovitev najmanj, vendar so med njimi

pomembne razlike. Za frazem *metati bisere svinjam* je tako približno 100 pojavitev več kot za frazem *Duh je sicer voljan, ali meso je slabo*, prenovitev pa imata oba enako (le 3). Izstopata tudi frazema *Oko za oko, zob za zob* in *Kdor je brez greha, naj prvi vrže kamen vanjo*, kjer bi več prenovitev pričakovali pri prvem, vendar je prenovitev bistveno več pri drugem. Torej le delno drži, da so pogostejši frazemi pogosteje prenovljeni.



Slika 1*:* Relativne frekvence izhodiščnih frazemov v korpusih.

Če upoštevamo relativne frekvence (Slika 1), so vsi izhodiščni frazemi najpogostejši v korpusu *Janes* in najredkejši v korpusu *Gigafida 2.0*. Razlika v pogostosti je največja pri frazemu *Oko za oko, zob za zob*, najmanjša pa pri frazemu *Duh je sicer voljan, ali meso je slabo*, ki je tudi po številu pojavitev najredkejši.



Slika 2: Relativne frekvence prenovitev v korpusih.

Iz Slike 2 je opazno, da so tudi prenovitve[21] najpogostejše v korpusu *Janes*. Največja razlika z ostalimi korpusi je pri prenovitvah frazema *Kdor je brez greha, naj prvi vrže kamen vanjo*, najmanjša pa pri prenovitvah frazema *Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega*. Prenovitve frazema *Človek obrača, Bog obrne* so v korpusih *Gigafida 2.0* in *slWaC* približno enako pogoste. Prenovitve frazemov *Oko za oko, zob za zob*; *Dajte cesarju, kar je cesarjevega, in Bogu, kar je božjega* in *Beseda je meso postala* so najredkejše v korpusu *slWaC*, prenovitve frazema *Kdor je brez greha, naj prvi vrže kamen vanjo* pa so najredkejše v korpusu *Gigafida 2.0*. Pri prenovitvah tega frazema je presenetljivo, da je tudi številčno prenovitev največ v korpusu *Janes* (36), sledi *slWaC* (22), najmanj prenovitev je v korpusu *Gigafida 2.0* (12). Pri frazemih *Duh je sicer voljan, ali meso je slabo* in *metati bisere svinjam* je prenovitev tako malo, da težko govorimo o razlikah med korpusi.

5 SKLEP

Glavna prednost uporabe treh korpusov je, da je bilo zajetih veliko raznovrstnih primerov iz različnih tipov besedil in diskurzov, saj je za ugotavljanje mej med ustaljenostjo in neustaljenostjo frazeoloških variant poleg velike količine podatkov ključnega pomena tudi heterogenost virov. Če bi za raziskavo uporabili le korpus *Gigafida 2.0* kot največji in referenčni korpus, bi spregledali mnogo relevantnih pojavitev. Kot je pokazala ta raziskava, so spletna besedila zelo bogata s frazemi in frazeološkimi prenovitvami, zato je bila njihova vključitev v analizo ključnega pomena. Samo zelo velik referenčni korpus splošne slovenščine za res kvalitetno jezikoslovno raziskovanje ni dovolj, temveč potrebujemo tudi več specializiranih korpusov.

Za celovitejši odgovor glede razmerja pogostosti izhodiščnega frazema in števila prenovitev bi bilo treba pregledati veliko večje število izvorno in oblikovno raznovrstnih frazemov, saj so bili vključeni le (iz)biblijski frazemi, večinoma pregovori. Na podlagi analiziranih frazemov lahko predpostavko, da se pogosteje prenavljajo frazemi, ki so tudi v svoji izhodiščni obliki pogostejši, potrdimo le delno. Verjetno imajo večji vpliv na prenovitvene možnosti frazema njegove pomenske lastnosti in dejstvo, da so frazemi kolektivne metafore, skupne celotni jezikovni skupnosti, in zato predstavljajo dobro izhodišče za

21 Za natančen prikaz in analizo prenovitev po frazemih gl. Trivunović, 2020.

tvorbo novih metafor (prim. Justin idr., 2015, str. 37). Predvsem pri pojavitvah tvitov v korpusu *Janes* je pogosto premalo konteksta, da bi primere z gotovostjo umestili v enega od tipov pojavitev. To je še posebej problematično pri zamenjavi sestavin znotraj istega pomenskega polja, kjer je meja med ustaljeno variantnostjo in neprenovitveno modifikacijo zabrisana, kar otežuje slovarski prikaz frazema in njegovih variant. Pri nekaterih frazemih tako zgolj na podlagi korpusnega gradiva ni moč z gotovostjo potrditi ustaljenosti variant, zato bi bilo za res zanesljivo slovarsko obravnavo treba upoštevati še druge vire, govorjeno slovenščino in mnenje rojenih govorcev. Za novonastajajoče slovarje je namreč bistveno, da skrbno obravnavajo frazeološke variante in spremembe, ki se kažejo v sodobnem gradivu, ter da prinašajo tudi podatek o pravopisni variantnosti (v našem primeru predvsem pri samostalniku *bog/Bog*), kjer to podpira korpusno gradivo.

Zahvala

Prispevek je nastal v okviru usposabljanja mladih raziskovalcev, ki ga financira ARRS pri programu P6-0038.

LITERATURA

Slovarski in drugi digitalni (korpusni) viri

BIBLIJA.net – Sveto pismo za vsakogar. Pridobljeno z www.biblija.net (26. 10. 2021)

eSSKJ: Slovar slovenskega knjižnega jezika 2016–. Ljubljana: Založba ZRC, ZRC SAZU. Pridobljeno z www.fran.si (9. 12. 2020)

Gigafida 2.0. Korpus pisne standardne slovenščine. Pridobljeno s https://www.clarin.si/noske/run.cgi/corp_info?corpname=gfida20_dedup&struct_attr_stats=1 (9. 12. 2020)

Janes. Korpus slovenskih spletnih uporabniških vsebin. Pridobljeno s https://www.clarin.si/noske/run.cgi/corp_info?corpname=Janes&struct_attr_stats=1 (9. 12. 2020)

Keber, J. *Slovar slovenskih frazemov* (2011/spletna različica: 2015). Ljubljana: Založba ZRC, ZRC SAZU. Pridobljeno z www.fran.si (20. 4. 2020)

Meterc, M. *Slovar pregovorov in sorodnih paremioloških izrazov 2020–*. Ljubljana: Založba ZRC, ZRC SAZU. Pridobljeno z www.fran.si (9. 12. 2020)

Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja (2014/spletna različica: 2014). Ljubljana: Založba ZRC, ZRC SAZU. Pridobljeno z www.fran.si (20. 4. 2020)

slWaC. Pridobljeno s https://www.clarin.si/noske/run.cgi/corp_info?corpname=slWaC&struct_attr_stats=1 (9. 12. 2020)

Drugo

Erjavec, T., & Ljubešić, N. (2014). The *slWaC* 2.0 Corpus of the Slovene Web. V T. Erjavec & J. Žganec Gros (ur.), *Jezikovne tehnologije: Zbornik 17. mednarodne multikonference Informacijska družba – IS 2014, Zvezek G*: 50–55. Ljubljana: Institut Jožef Stefan.

Fišer, D., Erjavec, T., & Ljubešić, N. (2016). *JANES* v0.4: Korpus slovenskih spletnih uporabniških vsebin. *Slovenščina 2.0*, *4*(2), 67–99.

Fišer, D., Ljubešić, N., & Erjavec, T. (2020). The *Janes* project: language resources and tools for Slovene user generated content. *Language resources and evaluation*, *54*(1), 223–246.

Galer, M. M. (2001). *Biblične stalne zveze v slovenskem jeziku, diplomska naloga*. Ljubljana: Filozofska fakulteta.

Gantar, P. (2001). Slovenska frazeologija v dosedanjih slovarjih glede na aktualna slovaropisna načela. *Jezikoslovni zapiski*, *7*(1–2), 207–223. Opole: Uniwersytet Opolski, Instytut Filologii Polskiej; Ljubljana: Univerza v Ljubljani, Filozofska fakulteta.

Gantar, P. (2003). Stalnost in spremenljivost frazema v slovarju. V S. Gajda & A. Vidovič Muha (ur.), *Współczesna polska i słoweńska sytuacja językowa (Sodobni jezikovni položaj na Poljskem in v Sloveniji*) (str. 209–223).

Gantar, P. (2006). Korpusni pristop v frazeologiji in slovarske aplikacije. V A. Vidovič Muha (ur.), *Slovensko jezikoslovje danes* (str. 151–163). Ljubljana: Slavistično društvo Slovenije.

Gantar, P. (2007). *Stalne besedne zveze v slovenščini: korpusni pristop*. Ljubljana: Založba ZRC, ZRC SAZU.

Gantar, P. (2014). Moč in nemoč korpusnega pristopa k analizi pomenov. V I. Grahek & S. Bergoč (ur.), *Novi slovar za 21. stoletje: e-zbornik s Posveta o novem slovarju slovenskega jezika na Ministrstvu za kulturo, 12. februar 2014*. Ljubljana: Ministrstvo za kulturo.

Geeraert, K., Newman, J., & Baayen, R. H. (2017). Understanding idiomatic variation. *Proceedings of the 13th Workshop on Multiword Expressions* (str. 80–90). Valencia: Association for Computational Linguistics.

Gliha Komac, N., Jakop, N., Ježovnik, J. Klemenčič, S., Krvina, D., Ledinek, N., Meterc, M., ..., & Žele, A. (2016). Novi slovar slovenskega knjižnega jezika – predstavitev temeljnih konceptualnih izhodišč. V. F. Marušič, P. Mišmaš, & R. Žaucer (ur.), *Zbornik prispevkov s simpozija 2015 (Škrabčevi dnevi 9)* (str. 16–33). Nova Gorica.

Hanks, P., El Maroof, I., & Oakes, M. (2017). Flexibility of Multi-Word Expressions and Corpus Pattern Analysis. *Multiword Expressions: Insights from a Multi-lingual Perspective*. (PARSEME countries: Great Britain). 93–119.

Jakop, N. (2006). *Pragmatična frazeologija*. Ljubljana: Založba ZRC, ZRC SAZU.

Jakop, N. (2014). Frazeologija v sodobnem slovarju slovenskega jezika. V I. Grahek & S. Bergoč (ur.), *Novi slovar za 21. stoletje: e-zbornik s Posveta o novem slovarju slovenskega jezika na Ministrstvu za kulturo, 12. februar 2014*. Ljubljana: Ministrstvo za kulturo.

Jesenšek, V., & Ulčnik, N. (2014). Spletni frazeološko-paremiološki portal: redakcijska vprašanja ob slovenskem jezikovnem gradivu. V Jesenšek, V. & Babič, S. (ur.), *Več glav več ve: frazeologija in paremiologija v slovarju in vsakdanji rabi*. Maribor: Oddelek za germanistiko Filozofske fakultete Univerze v Mariboru; Ljubljana: Inštitut za slovensko narodopisje ZRC SAZU.

Justin, M., Hirci, N., & Gantar, P. (2015). Rana ura, slovenskih fantov grob: analiza frazeoloških prenovitev v spletni slovenščini. V D. Fišer (ur.), *Zbornik konference Slovenščina na spletu in v novih medijih, Ljubljana, 25. –27. november 2015* (str. 33–37). Ljubljana: Znanstvena založba Filozofske fakultete.

Konicka, J. (2012). Prvi slovenski frazeološki slovar. *Jezikoslovni zapiski, 18*(2), 169–185.

Krek, S., Gantar, G., Arhar Holdt, Š., & Gorjanc, V. (2016). Nadgradnja korpusov Gigafida, Kres, ccGigafida in ccKres. V T. Erjavec & D Fišer (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika, 29. september–1. oktober 2016, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana, Slovenija* (str. 200–202). Ljubljana: Znanstvena založba Filozofske fakultete.

Krek, S., Arhar Holdt, Š., Čibej, J., Repar, A., & Ljubešić, N. (2019). *Specifikacije izdelave korpusa Gigafida 2.0*. Pridobljeno s https://www.cjvt.si/gigafida/wp-content/uploads/sites/10/2019/06/Gigafida2.0_specifikacije.pdf (20. 4. 2020)

Kržišnik, E. (1987). Prenovitev kot inovacijski postopek. *Slava*, *I*(1), 49–56.

Kržišnik, E. (1990). Tipologija frazeoloških prenovitev v Cankarjevih proznih delih. *Slavistična revija*, *38*(4), 399–420.

Kržišnik, E. (1996). Norma v frazeologiji in odstopi od nje v besedilih. *Slavistična revija*, *44*(2), 133–154.

Kržišnik, E. (2004). Poskusni zvezek slovenskega frazeološkega slovarja. *Slavistična revija*, *52*(2), 199–208.

Kržišnik, E. (2006). Izraba semantične potence frazemov. V A. Vidovič Muha (ur.), *Slovensko jezikoslovje danes* (str. 259–279). Ljubljana: Slavistično društvo Slovenije.

Kržišnik, E. (2018). Skladnja v frazeologiji med sistemom in rabo. *Jezik in slovstvo*, *63*(2–3), 35–47.

Kržišnik, E., & Jakop, N. (2015). Konceptualna izhodišča za vključitev frazeologije v splošni enojezični slovar slovenskega jezika. V M. Smolej (ur.), *Slovnica in slovar – aktualni jezikovni opis (1. del)* (str. 417–424). Ljubljana: Znanstvena založba Filozofske fakultete.

Meterc, M. (2016). Tematsko-konstrukcijski vzorci nastanka in prenovitve stavčnih frazemov. *Slavistična revija*, *64*(2), 125–138.

Meterc, M. (2017). *Paremiološki optimum: najbolj poznani in pogosti pregovori ter sorodne paremije v slovenščini*. Ljubljana: Založba ZRC, ZRC SAZU.

Meterc, M. (2019). Analiza frazeološke variantnosti za slovarski prikaz v eSSKJ-ju in SPP-ju. *Jezikoslovni zapiski*, *25*(2), 33–45.

Moon, R. (1996). Data, Description and Idioms in Corpus Lexicography. *Euralex 1996 Proceedings* (str. 245–256).

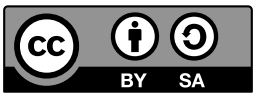
Trivunović, E. (2019). Diahrono raziskovanje biblijskih in izbiblijskih frazemov. *Jezikoslovni zapiski*, *25*(2), 47–61.

Trivunović, E. (2020). Variante in modifikacije (iz)biblijskih frazemov. V D. Fišer & T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika, zbornik konference, 24.–25. september 2020, Ljubljana, Slovenija* (str. 158–166). Ljubljana: Inštitut za novejšo zgodovino.

STABILITY, VARIABILITY AND MODIFIED USE OF PHRASEOLOGICAL UNITS IN SLOVENE LANGUAGE AND DICTIONARIES

The paper analyses biblical phraseological units, their variants and modifications, especially phraseological innovations in the modern Slovene language. Although stability or invariability of form and meaning is traditionally considered as one of the fundamental characteristics of phraseological units, actual use is much more diverse and unstable than one would expect. Most phraseological units are realized in many different forms, variants and modifications, so much so that variability is nowadays better considered as their inherent property and not a mistake or rarity. This became particularly apparent with the growing use of corpora in phraseological research, which allowed for observing and analysing phraseological units in large amounts of texts. In addition to this gap between the theoretical stability of phraseological units and their diverse use in texts there are often also differences between fixed variants recorded in existing dictionaries and those present in corpora. This can be the result of the deficient lexicographic presentation of phraseological units or actual changes in the language which took place after the release of the dictionary. A large amount of diverse texts is needed for a more reliable distinction between fixed variants and non-fixed modifications. This is why the paper uses three very different corpora: *Gigafida 2.0* (reference corpus of written standard Slovene), *Janes* (corpus of Slovene user-generated content) and *slWaC* (corpus of the Slovene web). Besides variants of phraseological units the second focus of this paper is non-fixed modifications with special attention on phraseological innovations and creative use of phraseological units. While the distinction between variants, noninnovative and innovative modifications appears clear in theory, the typology proved rigid in some cases. It is particularly difficult to draw the line between fixed variants and noninnovative modifications because of the so-called potential norm. All the examined phraseological units and innovations are most common in the *Janes* corpus, which supports the need for including diverse corpora in linguistic studies and not only relying on one large reference corpus.

Keywords: phraseology, variants, modifications, innovations, corpus linguistics



SPLETNA ORODJA ZA SLOVENŠČINO IN TUJI ŠTUDENTI UNIVERZE V LJUBLJANI

Mojca STRITAR KUČUK
Filozofska fakulteta, Univerza v Ljubljani

Stritar Kučuk, M.: Spletna orodja za slovenščino in tuji študenti Univerze v Ljubljani. Slovenščina 2.0, 9(2): 100–125.

DOI: https://doi.org/10.4312/slo2.0.2021.2.100-125

Redno vpisani tuji študenti Univerze v Ljubljani, ki se v prvem letu študija v okviru modula Leto plus učijo slovensko, se v drugem semestru na posebni delavnici podrobneje spoznajo s spletnimi jezikovnimi viri in tehnologijami za slovenščino. V prispevku je opisana izvedba te delavnice v študijskem letu 2019/20, ko je zaradi pandemije koronavirusa potekala na daljavo, v obliki interaktivnih videoposnetkov z nalogami za preverjanje razumevanja snovi. Drugi del prispevka se osredotoča na mnenje študentov o tovrstnih jezikovnih virih. S spletno anketo sem analizirala stališča in izkušnje študentov dveh generacij: študenti generacije 2018/19 so spletna orodja spoznavali v razredu, študenti generacije 2019/20 pa na daljavo. Sodeč po rezultatih ankete, mlajša generacija študentov jezikovne vire na spletu uporablja pogosteje. Študenti obeh skupin najpogosteje uporabljajo Googlov Prevajalnik, ki mu sledijo Sloleks, pregibnik Besana, Fran in Pons. Kot argumente za uporabo teh virov izpostavljajo predvsem hitrost oz. enostavnost uporabe in navajenost na določen vir.

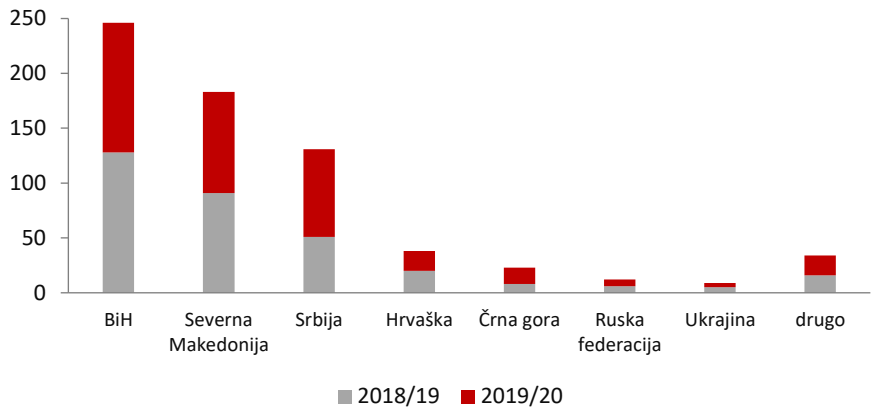
Ključne besede: spletni jezikovni viri, slovenščina, tuji študenti, spletno učenje

1 UVOD

Spletni jezikovni viri in jezikovne tehnologije si pot med uporabnike utirajo zlagoma in ob precejšnjem trudu jezikoslovcev.[1] Precej je bilo storjenega za ozaveščanje učiteljskega osebja v slovenskih šolah (prim. Arhar Holdt idr., 2021; Stritar Kučuk, Dobrovoljc, 2013; Portal jezikovnih virov), še večji izziv

1 Moška oblika je v tem prispevku uporabljena kot slovnično nevtralna.

pa jih je širiti med populacijo, ki ni več vključena v redni jezikovni pouk. V tem prispevku predstavljam, kako se s spletnimi jezikovnimi viri za slovenščino sistematično seznanja specifična skupina govorcev slovenščine, namreč tuji študenti na Univerzi v Ljubljani. Redno vpisani tuji študenti so v prvem letu študija upravičeni do udeležbe v modulu Leto plus, namenjenem usposabljanju za študij v Sloveniji. Med drugim vključuje dva semestra brezplačnega lektorata slovenščine kot drugega jezika. Vsako leto se v Leto plus vpiše med 300 in 350 študentov. Večina jih prihaja iz južnoslovanskega govornega področja (Grafikon 1) in se slovenščine pred prihodom na študij v Ljubljani ne uči.



Grafikon 1: Študenti Leta plus po državi izvora.

Vsak semester lektorata slovenščine pomeni 60 ur pouka slovenščine kot drugega jezika, in sicer 45 kontaktnih ur z učiteljem v razredu, poleg tega pa še 15 ur dodatnega programa, med katerega sodijo dejavnosti z učiteljem, kot so fonetične vaje ali konverzacija, pa tudi udeležba na različnih kulturnih dogodkih in podobno. V drugem semestru, ko so študenti že bolj suvereni pri samostojnem pisanju v slovenščini, je eden od dodatnih projektov delavnica uporabe spletnih jezikovnih virov in tehnologij za slovenščino.

To za večino študentov sicer ni prvo srečanje z uporabo jezikovnih virov in tehnologij. Večina je navajena, da si pri slovenščini pomaga z Googlovim Prevajalnikom, ki pa ga po izkušnjah lektorjev slovenščine pogosto uporabljajo nekritično. Zato jih od uvodnih ur lektorata naprej skušamo opozarjati na pomembnost kritične uporabe in jih hkrati seznanjamo tudi z drugimi

jezikovnimi viri. Ko je to pedagoško smiselno, jim npr. pokažemo slovarski portal Fran, korpus Gigafida 2.0 ipd. V večje podrobnosti pri uporabi pa se usmerimo prav na posebej temu namenjeni delavnici. Namen tega prispevka je torej predstaviti omenjeno delavnico in prek analize rabe posameznih jezikovnih virov in tehnologij, ki temelji na samoevalvaciji študentov, oceniti njeno uspešnost.

2 SPLETNA DELAVNICA UPORABE JEZIKOVNIH VIROV

Delavnica uporabe jezikovnih virov je v prvih letih modula Leta plus, ki je bil prvič izveden v študijskem letu 2017/18, potekala fizično. V študijskem letu 2018/19 so si denimo študenti izbrali enega od desetih terminov, v katerem so v računalniški učilnici v skupinah po 30 opravili triurno usposabljanje.[2] V študijskem letu 2019/20 pa smo bili delavnico zaradi pandemije covida 19 in prekinitve študijskega procesa na univerzi prisiljeni v celoti prestaviti na splet. Poimenovana je bila Spletna orodja za slovenščino (kratko SOS za slovenščino). Delavnica je bila izpeljana kot komplet interaktivnih videoposnetkov v spletni učilnici Filozofske fakultete.[3] Vsak posnetek je predstavil enega ali več sorodnih jezikovnih virov. Ker je bila delavnica namenjena vsem študentom, ne samo jezikoslovcem, so bile informacije podane splošno, brez zgodovinskih pregledov, bolj specifičnih informacij ali naprednih možnosti iskanja. Skupaj je bilo posnetih sedem videoposnetkov oz. okoli 40 minut gradiva. Študenti so si morali ogledati vse videoposnetke in rešiti z njimi povezane naloge, kar je bil eden od pogojev za pristop k izpitu.

2.1 Predstavljeni jezikovni viri in tehnologije

Ker je jezikovnih virov za slovenščino vedno več in se hitro razvijajo, je bilo treba izbrati tiste, ki jih je smiselno predstaviti po interesih in jezikovnem predznanju precej raznorodni skupini tujih študentov. Odločitev je bila sprejeta predvsem na podlagi naših predvidevanj o tem, kateri viri in tehnologije ponujajo podatke, ki so za te študente najbolj relevantni in uporabni (Tabela 1). Izbrani so bili viri, ki so prosto dostopni in predstavljajo predvsem sodobno standardno oz. knjižno slovenščino, pa tudi nekaj terminoloških oz. bolj

2 Delavnico je pripravila in izvajala dr. Tadeja Rozman.

3 Delavnico sem na osnovi gradiva prejšnjih let pripravila avtorica tega prispevka.

strokovnih virov. Pomemben dejavnik je bila tudi enostavnost uporabe samega vira.

Tabela 1: *Jezikovni viri in tehnologije, vključeni v delavnico spletnih orodij za slovenščino*

Tip izdelka	Jezikovni vir oz. tehnologija	Opis	Spletna povezava
Slovarji in slovarski portali	Pons	Večjezični spletni slovar	https://sl.pons.com/
	Slovarji: Spletni slovarji in prevajalski pripomočki	Spletna stran s povezavami na različne slovarje in prevajalske pripomočke	https://evroterm.vlada.si/slovarji
	Termania	Spletni slovarski portal s slovarji podjetja Amebis	https://www.termania.net
	Fran[4]	Spletni slovarski portal s slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU	https://fran.si/
	Kolokacije 1.0	Slovar kolokacij sodobne slovenščine (prim. Kosem idr., 2018)	https://viri.cjvt.si/kolokacije/
	Sopomenke 1.0	Slovar sopomenk sodobne slovenščine (prim. Krek idr., 2018; Arhar Holdt idr., 2018)	https://viri.cjvt.si/sopomenke/
Jezikovni korpusi	Gigafida 2.0	Korpus pisne standardne slovenščine (prim. Krek idr., 2020)	https://viri.cjvt.si/gigafida/
Pregibanje	Besana	Spletni pregibnik	https://besana.amebis.si/pregibanje/
	Sloleks	Slovenski oblikoslovni leksikon (prim. Dobrovoljc idr., 2015)	https://viri.cjvt.si/sloleks/
Ostali izdelki	Googlov Prevajalnik	Strojni prevajalnik	https://translate.google.com/
	črkovalniki (Microsoft, Google)	Strojni črkovalniki v urejevalnikih besedil ipd.	

Nekaj virov, ki so bili med fizično izvedbo delavnice še del predstavitve, je bilo leto kasneje v spletni delavnici samo omenjenih: govorni korpus GOS (prim. Verdonik, Zwitter Vitez 2011), korpus Kres (prim. Logar Berginc idr., 2012), korpus učečih se Šolar (prim. Kosem idr., 2012), vzporedni korpus Evrokorpus

4 V času priprave in izvedbe delavnice otroški slovarski portal Franček še ni bil dostopen.

(prim. Željko, 2003), aplikacija Igra besed (prim. Arhar Holdt idr., 2020). Predvidevati je namreč mogoče, da ti viri za večino naše študentske populacije niso zanimivi, so prezahtevni za njihovo jezikovno zmožnost v slovenščini ali pa so bili v času izvedbe delavnice uporabniški vmesniki premalo intuitivni za uporabo. Tisti študenti, za katere bi bili lahko koristni, pa so dovolj računalniško in jezikoslovno usposobljeni, da se lahko vanje samostojno poglobijo.

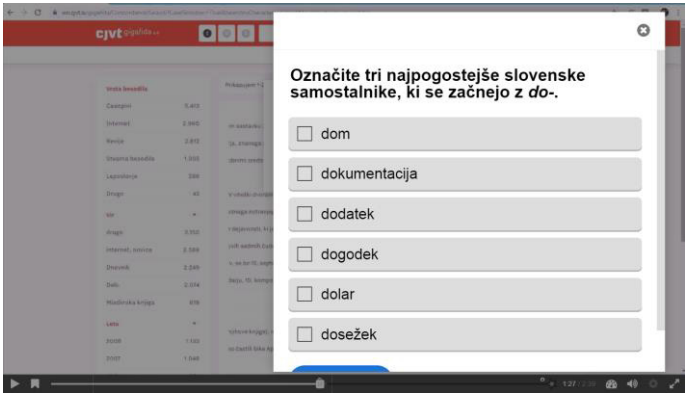
Nekaj virov, ki so bili v fizični delavnici predstavljeni, pa v spletno delavnico sploh ni bilo vključenih. Taka sta Pedagoški slovnični portal, ki je bolj kot tujcem namenjen domačim govorcem slovenščine in njihovim težavam (prim. Kosem idr. 2012, 108–122), in poskusni Spletni slovar slovenskega jezika kot testni prikaz gesel leksikalne baze za slovenščino (prim. Gantar, 2011), ki za tuje govorce vsebuje premalo gesel.

2.2 Predstavitev jezikovnega vira

Osnova vsakega videoposnetka je predstavitev določenega vira oz. tehnologije, ki je kratka, pripravljena v jeziku, za katerega predvidevamo, da je razumljiv tujim študentom, in osredotočena na informacije, ki so zanje pomembne. To sta predvsem podatka, kaj najdejo v viru in zakaj je to koristno. V videoposnetkih so izpostavljene tudi omejitve oz. slabosti vsakega vira: pri Ponsu in Termanii denimo nevključenost južnoslovanskih jezikov, pri Franu morebitni zavajajoči podatki iz manj relevantnih virov, npr. narečnih ali zgodovinskih slovarjev, pri jezikovnih virih, ki temeljijo na avtomatskem luščenju podatkov, nezanesljivost informacij ipd.

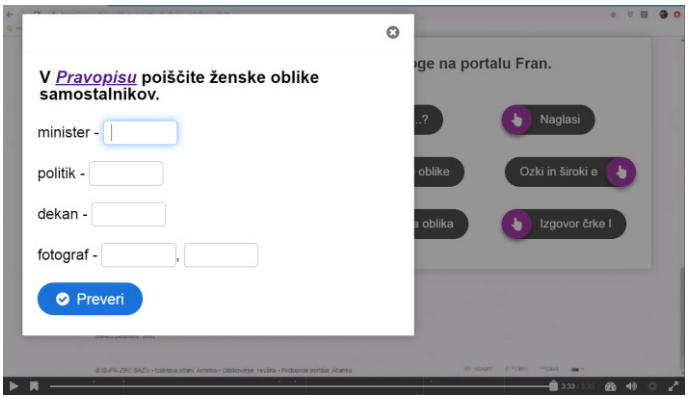
2.3 Interaktivne naloge

Da bi bil ogled posnetkov za študente bolj dinamičen in osmišljen, je v vsak posnetek vključenih nekaj interaktivnih nalog. Seveda v njih ne gre za preverjanje tega, ali študenti razumejo besedišče, ki se pojavlja v korpusnih, slovarskih in drugih primerih v posnetkih. Naloge preverjajo izključno to, kako dobro razumejo predstavljeno snov oz. ali se znajdejo pri iskanju po predstavljenih virih. Namen naloge na Sliki 1 torej ni, da se študenti naučijo, kateri so najpogostejši samostalniki, ki se začnejo z *do-*, pač pa, da znajo v korpusu poiskati podatek o njihovi pogostnosti.

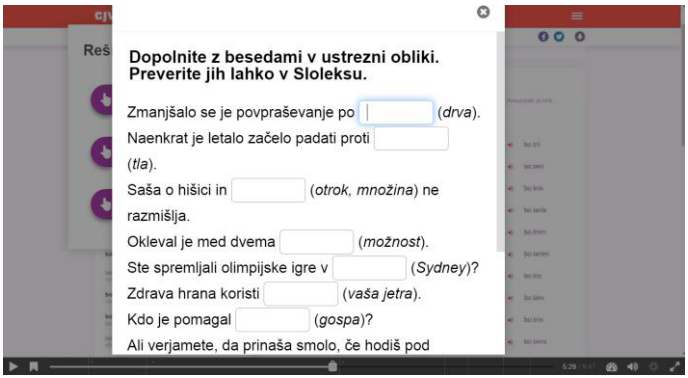


Slika 1: Naloga za iskanje po seznamih v korpusu Gigafida.

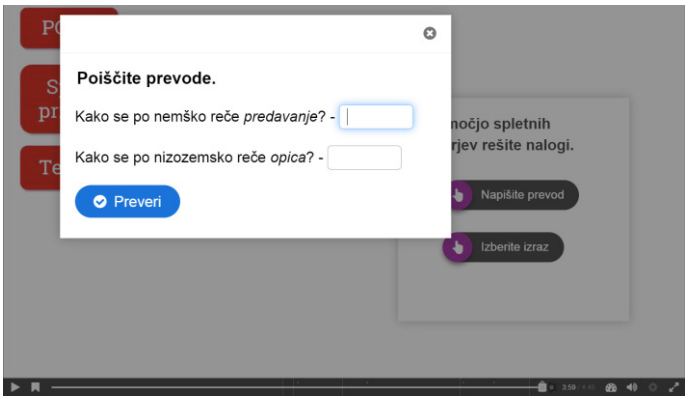
Vključene naloge so različnih vrst, kot to omogočata Moodle oz. odprtokodni okvir H5P. Vse so zaprtega tipa, torej takšne, da se odgovori lahko avtomatsko preverjajo. V nekaterih nalogah morajo študenti vpisovati odgovore; takšne so npr. naloga za iskanje ženskih oblik samostalnikov v Franu (Slika 2), za iskanje ustreznih sklonskih oblik v Sloleksu (Slika 3) ali za iskanje prevodov v različnih slovarskih portalih (Slika 4). Pri slednjem smo za iskanje prevodov v tuje jezike izbrali jezike, ki so južnoslovanskim študentom manj znani, npr. nemščino ali nizozemščino. Tako morajo za iskanje dejansko uporabiti splet.



Slika 2: Naloga za iskanje ženskih oblik samostalnikov v Franu.

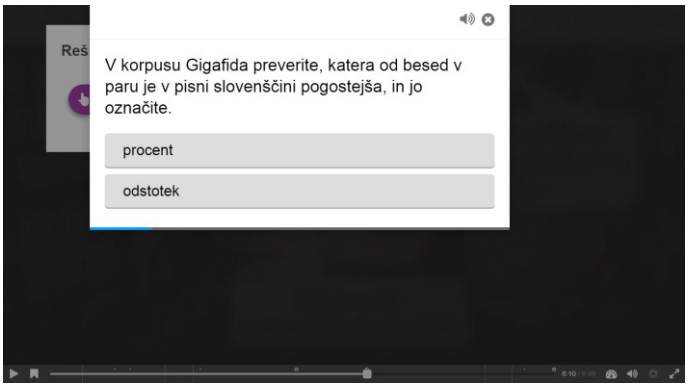


Slika 3: Naloga za uporabo pregibnika Sloleks.

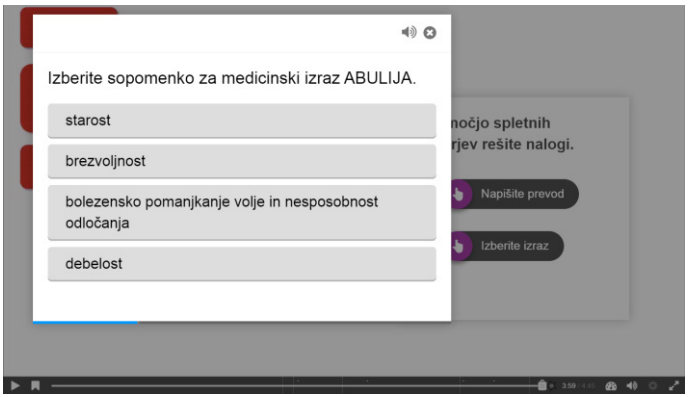


Slika 4: Naloga za iskanje prevodov v tuje jezike.

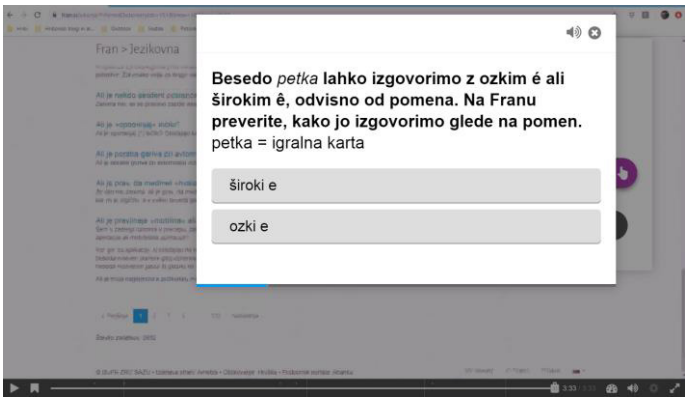
Večina nalog je izbirnega tipa. Študenti na podlagi podatkov v jezikovnih virih izberejo enega (ali več) od ponujenih odgovorov. Takšna je denimo naloga za osnovno iskanje (Slika 5) v korpusu Gigafida. Prikaz terminoloških slovarjev je popestren z nalogo za iskanje razlag besed oz. sopomenk (Slika 6). Z nalogo izbirnega tipa preverjamo tudi, ali znajo študenti najti informacije o izgovorjavi v Franu (Slika 7).



Slika 5: Naloga za osnovno iskanje po korpusu Gigafida.

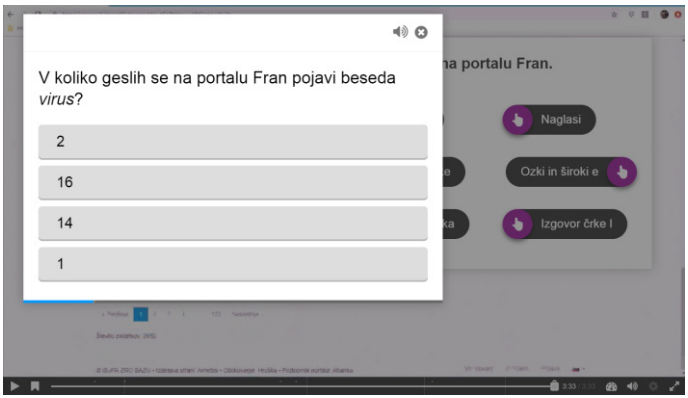


Slika 6: Naloga za izbor sopomenk strokovnega termina.



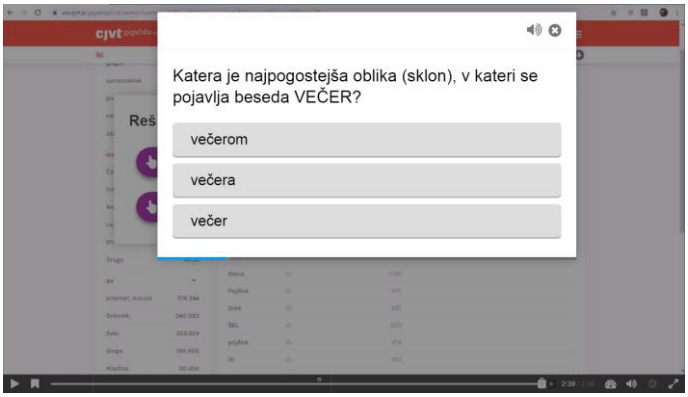
Slika 7: Naloga za iskanje podatkov o širini oz. ožini samoglasnikov v Franu.

Da je reševanje nalog bolj raznovrstno, je v odgovoru lahko zahtevana tudi številka (Slika 8).

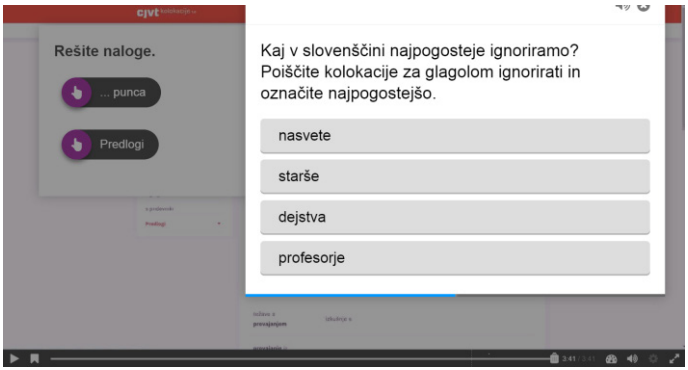


Slika 8: Naloga za določanje števila gesel v Franu.

V naloge so vključeni primeri, ki so za našo ciljno populacijo zahtevni (npr. v nalogo za iskanje po seznamu je vključen samostalnik *večer*, katerega oblike so zaradi nepodaljševanja osnove v odvisnih sklonih z *-j-* in podobnosti z besedo *večerja* za južnoslovanske govorce zahtevne, Slika 9), zanimivi ali pa, predvidoma, vsaj zabavni (Slika 10).



Slika 9: Naloga za iskanje po seznamih.



Slika 10: Naloga za iskanje najpogostejših kolokacij.

3 MNENJA ŠTUDENTOV

3.1 Ocena uporabnosti delavnice o jezikovnih virih in tehnologijah

Učitelji Leta plus od študentov ves čas dobivamo povratne informacije o njihovem učenju slovenščine. To vključuje tudi uporabo jezikovnih virov in tehnologij. Zapisov pogovorov s študenti, ki so delavnico rabe spletnih orodij na lektoratu ali izpitu večkrat izpostavili kot posebej koristno, zaradi anekdotičnosti nima smisla navajati. Omenim pa naj pisni izpit ob koncu letnega semestra, v katerem morajo študenti v slovenščini napisati daljše besedilo in komentirati preteklo leto, izobraževanje na daljavo, lektorat slovenščine, dodatne dejavnosti pri lektoratu, življenje v Ljubljani in podobno. Ena od v navodilu izpostavljenih točk je bila tudi, naj komentirajo dodatne dejavnosti. Od 34 študentov, ki so ob koncu letnega semestra 2019/20 opravljali pisni izpit, jih je 14 omenilo spletno delavnico jezikovnih virov in tehnologij. En študent je napisal, da tega ne potrebuje, ker »[ima] doma punco katera [mu] pomaga«,[5] preostalih 13 pa je tečaj pohvalilo kot koristen, npr.: »Od vsih dodatnih dejavnosti največ mi je bil zanimiv ta tečaj spletnih orodij za slovenščino kar sem se veliko naučil in še vedno uporabljam Sloleks ko imam težave z uporabo sklona (kar je nažalost pogosto).« Omenili so, da teh vsebin prej niso poznali: »Navduševljena sem z SOS in mi je žal ker za te korpuse nisem znala prej.« Trije študentje so predlagali, da bi morala biti delavnica spletnih orodij izvedena že v prvem semestru učenja slovenščine: »Priporočil bi da spletna orodja

5 Izjave študentov so navedene v izvirni obliki, z izvirnimi napakami.

za slovenski jezik (predvsem Sloleks) pokažete študentom v prvem semestru, ker so zelo zelo koristna in veliko pomagaju v učenju.«

Podobne povratne informacije smo dobili tudi v evalvaciji, ki jo ob koncu vsakega od lektoratov na Letu plus izpolnijo študenti. V njej ne sprašujemo eksplicitno po posameznih dodatnih dejavnostih, vendar jih nekateri samoiniciativno komentirajo. Ob koncu letnega semestra 2018/19, ko je delavnica potekala v fizični obliki, je anketo izpolnilo 265 študentov. Med komentarji na koncu so bili trije povezani z delavnico spletnih jezikovnih virov in tehnologij. Dva študenta sta napisala, da bi delavnica morala biti izvedena že v prvem semestru, en študent pa jo je ocenil negativno kot bistveno predolgo: »da bi imeli 3 ure tečaj za slovenske vire je res zgubljanje časa«. V študijskem letu 2019/20 je bila evalvacija, tako kot vse ostalo, izvedena na spletu. Odgovorov je bilo zato manj, 115. Tudi v tej evalvaciji sta dva študenta pohvalno ocenila delavnico in predlagala, da bi morala biti izvedena že v prvem semestru.

3.2 Anketa o uporabi jezikovnih virov in tehnologij

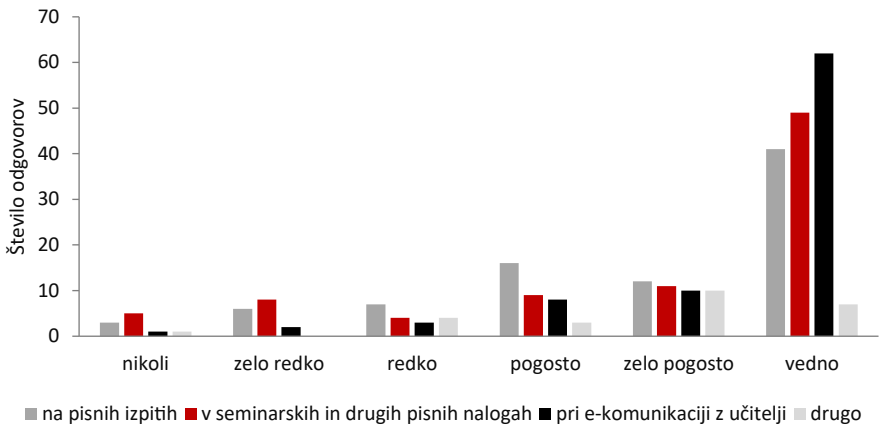
Da bi od študentov o uporabi jezikovnih virov pridobila neposredna mnenja, sem oblikovala še spletno anketo, ki je bila odprta od konca julija do konca septembra 2020. K reševanju so bili prek elektronske pošte pozvani vsi študenti Leta plus, ki so zaključili z lektoratom Slovenščina 2 v študijskih letih 2018/19 (okoli 250 študentov) in 2019/20 (okoli 190 študentov). Anketo je rešilo 99 študentov, 43 % iz študijskega leta 2018/19 in 57 % iz 2019/20. Glede na to, da so bili k izpolnjevanju povabljeni sredi poletja, ko je precejšen del študentov odmaknjen od študijskih obveznosti, in da je bil posebej za študente iz leta 2018/19 lektorat slovenščine časovno že oddaljen, sem bila z odzivom zadovoljna.

V skladu z zgornjima deležema je primerljiva delitev študentov glede na način opravljanja delavnice za uporabo jezikovnih virov: 43 se jih je delavnice udeležilo fizično v računalniški učilnici, 53 pa v spletni učilnici.[6] Obe skupini študentov, t. i. fizično in spletno, bom tudi primerjala, čeprav je treba upoštevati, da so določene razlike med njima nedvomno posledica sprememba medija poučevanja, druge razlike pa so lahko posledica drugih dejavnikov, zato na osnovi te ankete ni mogoče enoznačno sklepati o izvoru teh razlik.

6 Trije študenti so izbrali odgovor »drugo«, vendar v komentarju niso razumljivo pojasnili svojega odgovora.

Vsi anketiranci so kot prvi jezik govorili enega od južnoslovanskih jezikov: makedonščino (32 %), srbščino (28 %), bosanščino (19 %), črnogorščino (2 %), hrvaščino (1 %), ruščino (1 %). Izjema so trije študenti s prvim jezikom albanščino, madžarščino oz. turščino. Med odgovarjajočimi je bilo 56 žensk in 29 moških. 14 študentov prihaja z Ekonomske fakultete; 9 s Filozofske fakultete; po 5 s Fakultete za elektrotehniko, Fakultete za gradbeništvo in geodezijo, Fakultete za matematiko in fiziko, Fakultete za računalništvo in informatiko; po 4 s Fakultete za družbene vede in Naravoslovnotehniške fakultete; po 2 z Biotehniške fakultete in Fakultete za strojništvo; po 1 pa z Akademije za likovno umetnost in oblikovanje, Fakultete za arhitekturo, Fakultete za socialno delo, Medicinske fakultete, Pedagoške fakultete in Pravne fakultete. 57 jih študira na prvi stopnji, 28 na drugi, eden pa na doktorski stopnji. Večina študentov, ki so odgovarjali na anketo, študira v slovenščini (76), 10 v angleščini, eden pa »v italijanščini, španščini in slovenščini«.

Ker je dejavnost, pri kateri so jezikovni viri in tehnologije najlažje uporabni, pisanje, je zanimiv podatek, da anketirani študenti v slovenščini dejansko precej pogosto pišejo[7] (Grafikon 2) – največ pri spletnem sporazumevanju s pedagoškim osebjem – in da je torej uporaba jezikovnih virov zanje še kako relevantna. Med drugimi priložnostmi, ko uporabljajo pisno slovenščino, so



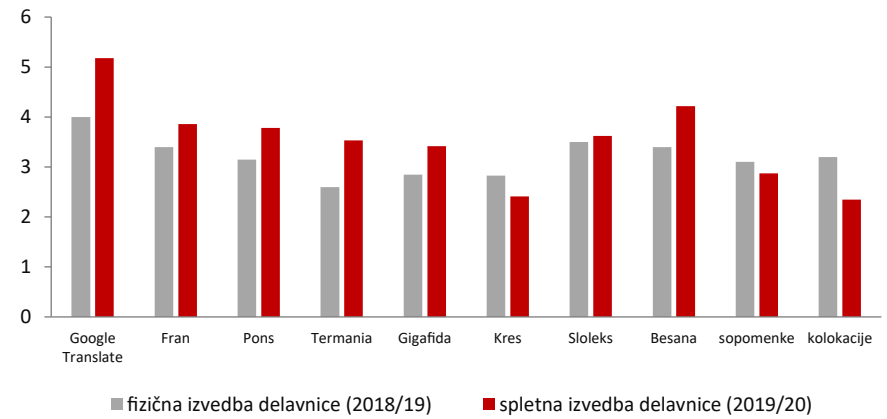
Grafikon 2: Pogostnost pisanja v slovenščini.

7 Anketa sicer ne daje odgovora na to, koliko je dejanske pisne produkcije študentov, temveč samo, kako pogosto pri tem – po lastni samooceni – uporabljajo slovenščino.

navedli vsakdanje življenje (3 odgovori), pogovor s slovenskimi kolegi oz. prijatelji (3 odgovori) in iskanje študentskega dela (1 odgovor).

3.2.1 Pogostnost uporabe jezikovnih virov pri študiju

Študenti so morali za vsakega od naštetih jezikovnih virov, ki so bili vsi predstavljeni v delavnici spletnih orodij za slovenščino, oceniti, kako pogosto jih uporabljajo pri študiju. Izbrali so eno od šestih ocen: nikoli, redkeje, enkrat na mesec, nekajkrat na mesec, enkrat na teden, večkrat na teden. Preračunana povprečja, prikazana na Grafikonu 3 in v Tabeli 2, nam dajo primerjavo, katere vire študenti uporabljajo pogosteje.



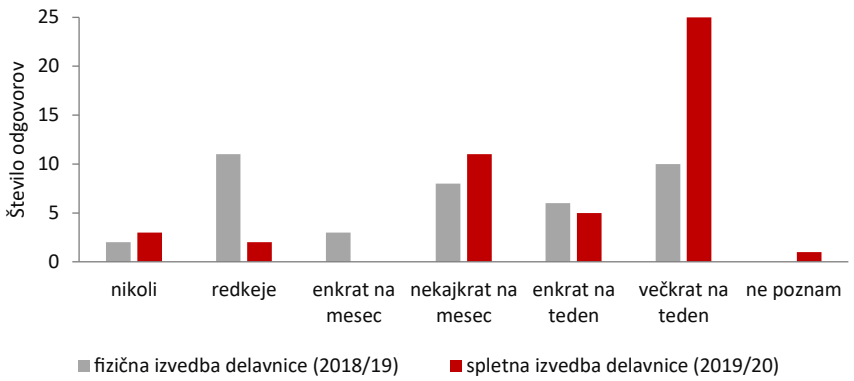
Grafikon 3: Povprečne ocene pogostnosti uporabe jezikovnih virov.

Pri skoraj vseh virih so ocene višje pri študentih spletne skupine. Verjetno na to vpliva spletni način študija, zaradi katerega študenti preživijo na spletu bistveno več študijskega časa, kot kadar se pedagoškega procesa udeležujejo fizično v predavalnicah, zato so jim spletni viri večino časa na dosegu miške. Po pričakovanjih študenti obeh skupin najpogosteje uporabljajo Googlov Prevajalnik oz. Google Translate. Pri tem izstopajo študenti spletne skupine. Sledijo Sloleks, pregibnik Amebis Besana, Fran in Pons. Ostale vire uporabljajo redkeje.

Tabela 2: *Preračunane povprečne ocene pogostnosti rabe posameznih jezikovnih virov in tehnologij, vključenih v delavnico spletnih orodij za slovenščino*

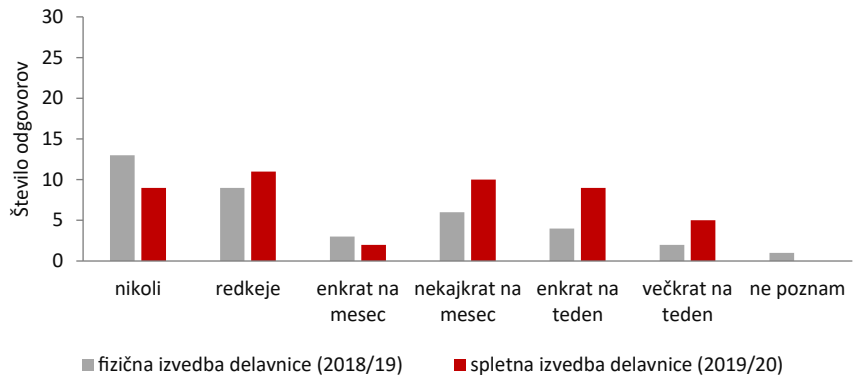
Jezikovni vir oz. tehnologija	Povprečna ocena	
	Fizična izvedba delavnice (2018/19)	Spletna izvedba delavnice (2019/20)
Google Translate	4,02	5,18
Fran	3,46	3,86
Pons	3,15	3,78
Termania	2,57	3,53
Gigafida 2.0	2,85	3,42
Kres	2,83	2,41
Sloleks 2.0	3,5	3,62
Pregibnik Amebis Besana	3,4	4,23
Sopomenke 1.0	3,06	2,87
Kolokacije 1.0	3,23	2,35

Oglejmo si še odgovore za najpogosteje uporabljane jezikovne vire. Pri strojnem prevajalniku Google Translate (Grafikon 4) je izrazita razlika med študenti fizične in spletne skupine. Po pričakovanjih največji del študentov spletne skupine prevajalnik uporablja večkrat na teden, večina odgovorov pa je zgoščena na pogostejši strani kontinuuma. Pri študentih fizične skupine so odgovori bolj razpršeni.

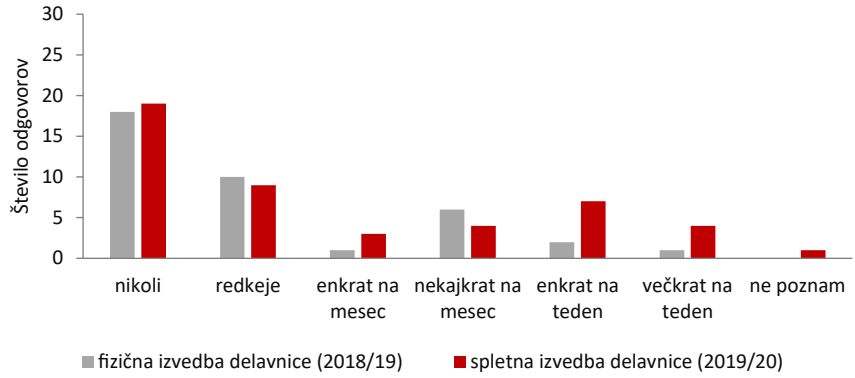


Grafikon 4: Pogostnost uporabe prevajalnika Google Translate.

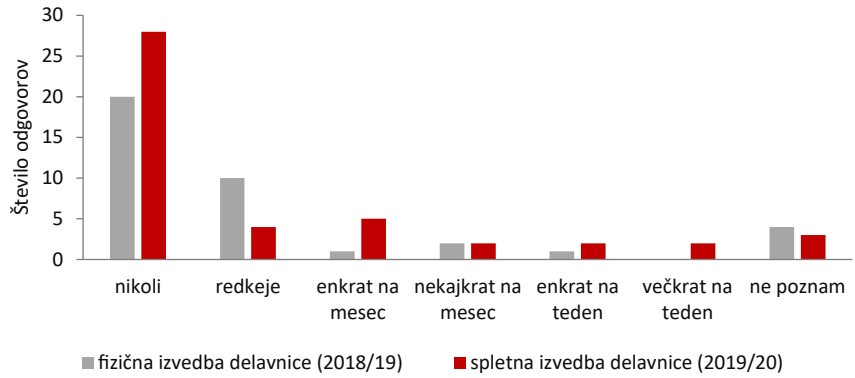
Med tremi slovarskimi portali tuji študenti najpogosteje uporabljajo Fran (Grafikon 5), sledi mu Pons (Grafikon 6). Termanio, pri kateri je največ študentov izbralo odgovor »nikoli«, uporabljajo najredkeje (Grafikon 7). Vse tri portale študenti spletne skupine uporabljajo večkrat kot študenti fizične skupine.



Grafikon 5: Pogostnost uporabe portala Fran.

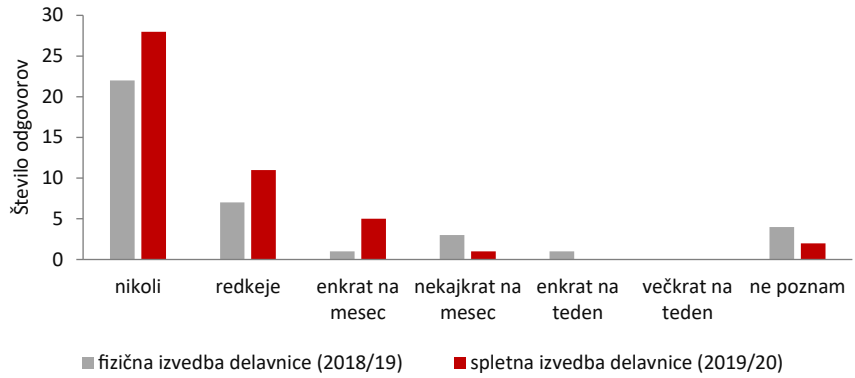


Grafikon 6: Pogostnost uporabe portala Pons.

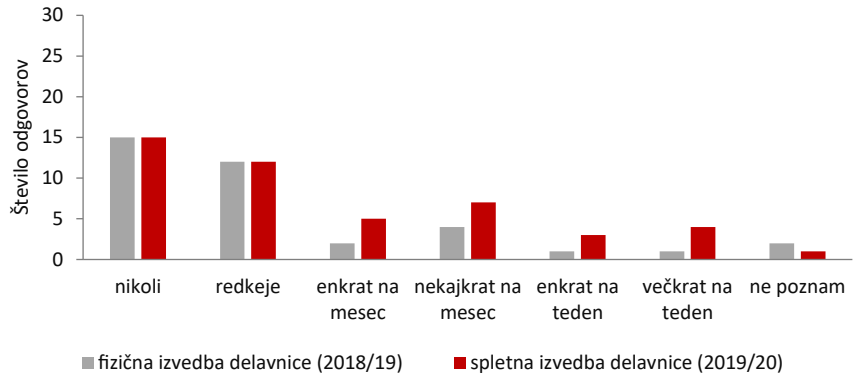


Grafikon 7: Pogostnost uporabe portala Termania.

V anketo sta bila vključena dva jezikovna korpusa, Gigafida in Kres. Kres (Grafikon 8) uporabljajo študenti najredkeje med vsemi obravnavami viri, kar je v skladu s tem, da jim je bil na usposabljanju predstavljen mimogrede kot manjša različica Gigafide. A tudi slednjo (Grafikon 9) uporabljajo razmeroma redko. Kot zanimivost naj omenim, da sta med petimi študenti obeh skupin, ki so izbrali odgovor »večkrat na teden«, le dva s Filozofske fakultete, torej morda jezikoslovca. Dva študenta sta s Fakultete za elektrotehniko, eden pa z Ekonomske fakultete. Odločitev za rabo spletnih jezikovnih virov torej ni odvisna od (jezikoslovne) stroke, pač pa od drugih subjektivnih dejavnikov, kakršni so osebni interesi in nagnjenja.

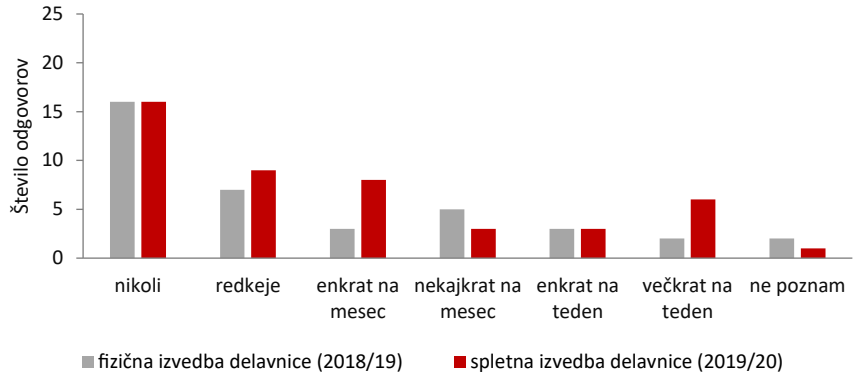


Grafikon 8: Pogostnost uporabe korpusa Kres.

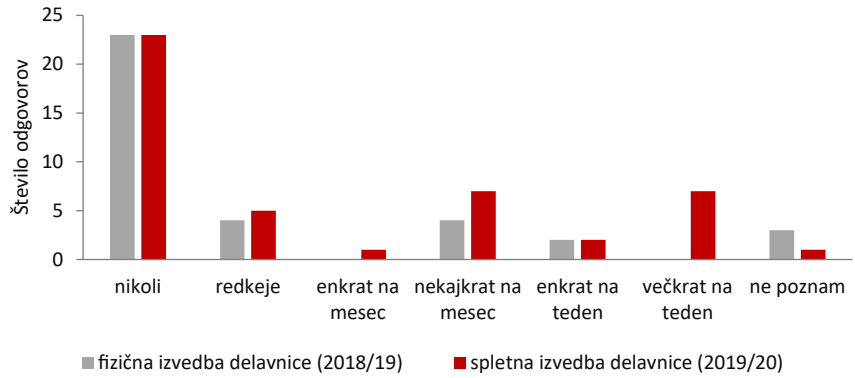


Grafikon 9: Pogostnost uporabe korpusa Gigafida.

Med viri s podatki o pregibanju besed ima Sloleks (Grafikon 10) nekaj prednosti pred pregibnikom Besana (Grafikon 11).

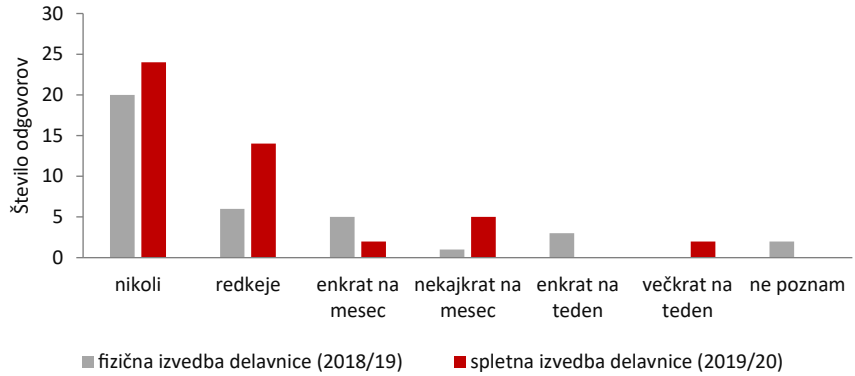


Grafikon 10: Pogostnost uporabe Sloleksa.

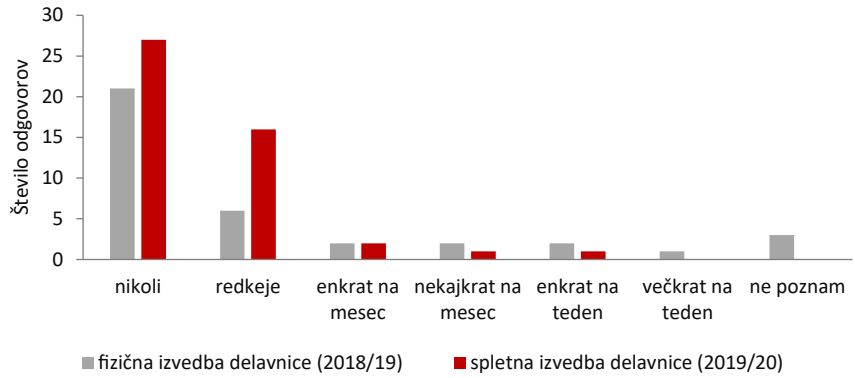


Grafikon 11: Pogostnost uporabe pregibnika Amebis Besana.

Poglejmo si še malo bolj specializirana slovarja, slovar sopomenk Sopomenke 1.0 (Grafikon 12) in kolokacij Kolokacije 1.0 (Grafikon 13). Oba študenti obeh skupin uporabljajo bolj ali manj samo izjemoma.



Grafikon 12: Pogostnost uporabe sopomenskega slovarja Sopomenke 1.0.

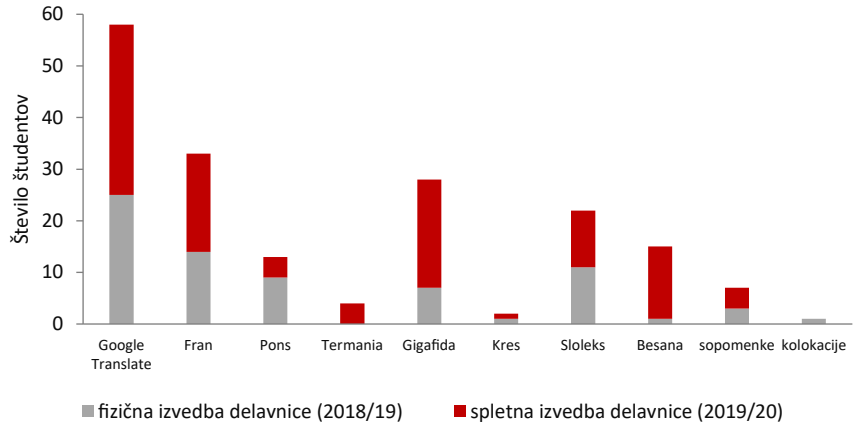


Grafikon 13: Pogostnost uporabe kolokacijskega slovarja Kolokacije 1.0.

Kot je razvidno iz grafikonov, so pri vsakem viru študenti lahko izbrali tudi odgovor »ne poznam«. Izvajalci lektorata smo zadovoljni, da so bile te številke v anketi razmeroma nizke. Nekoliko višje so pri študentih iz generacije 2018/19, kar je razumljivo, ker je delavnica spletnih virov od njih že bolj odmaknjena.

3.2.2 Najuporabnejši jezikovni viri

Študente sem prosila tudi, naj izberejo največ tri jezikovne vire, ki so zanje najuporabnejši (Grafikon 14). Dodatno so v komentarju lahko pojasnili, za kaj jih uporabljajo. Dobila sem 48 komentarjev, v katerih so poleg pojasnil uporabe navajali tudi argumente, zakaj se jim zdijo najuporabnejši.



Grafikon 14: Najuporabnejši jezikovni viri po oceni študentov.

Med splošnimi komentarji je zanimiva opazka študenta, ki spletna orodja uporablja, ker je njegova jezikovna zmožnost v slovenščini prenizka: »Zgoraj označena orodja uporabljam najpogosteje ker v mojemu vokabularju še vedno ni dovolj besed za pisno komunikacijo v slovenščini. Verjamem da bom jih vse uporabljal takrat ko bom moral več uporabljati pisno slovenščino.«[8]

Če podrobneje pogledamo najuporabnejša orodja po izboru študentov, je na prvem mestu seveda Googlov Prevajalnik, ki je bil najpogostejša izbira študentov obeh skupin. Uporabljajo ga za prevod besed, ki jih ne razumejo (4 odgovori), za pisanje različnih besedil za fakulteto (4 odgovori), pa tudi v vsakdanjem življenju (2 odgovora). Kot glavna argumenta so navedli hitrost (8 odgovorov) in enostavnost uporabe (8 odgovorov), npr. »Zelo so lahka za uporabo in vedno najdem tisto kar iščem.« Dva študenta sta omenila, da več od tega ne potrebujeta – »ima vse kaj potrebujem«. En študent pa je njegovo uporabnost relativiziral: »Zaradi hitre pristupačnosti, čeprav ni najboljše orodje za učenje slovenščine.« Nezanemarljiva prednost Prevajalnika, ki jo je izrecno navedlo 6 študentov, je dejstvo, da so ga večinoma že navajeni: »Ker sem ga že poznala preden sem začela študirati.«

Prevajalniku sledi slovarski portal Fran, ki so ga približno izenačeno izbirali študenti obeh skupin. Uporabljajo ga za iskanje besed, ki jih ne razumejo (3 odgovori), pri študiju ali v vsakdanjem življenju (2 odgovora), za »preverjanje besed, ki obstajajo tudi v bosanščini, vendar se zdi da imajo drugčen pomen«, in za popravljanje napak (1 odgovor). Tudi pri Franu sta kot glavni prednosti navedeni enostavnost uporabe (3 odgovori) – »lahko pridem do beseda ki jo iscem, ne zgubim dodaten cas dokler najdem na kateri gumb moram pritisniti« – in hitrost dostopa do podatkov (1 odgovor), izpostavili pa so tudi kakovost oz. podrobnost informacij (2 odgovora) in »zelo razumljive razlage pomena besed«. En študent pa je izrecno poudaril tudi, da je »navajan […] na uporabo google translate in frana od samega začetka študija v sloveniji«.

Približno enako pogosto kot Fran so študenti izbirali korpus Gigafida – študenti spletne skupine so ga izbrali celo večkrat kot Fran. Gigafido uporabljajo »zaradi lažjega in pravilnejšega pisanja v slovenščini« ali z zelo splošno utemeljitvijo, ki pa potrjuje koristnost delavnice spletnih orodij: »gigafida mi se

8 Če ni drugače navedeno, so vsi navedki odgovori iz ankete. Zapisani so v izvirni obliki, z izvirnimi napakami.

zdi zelo korisnom in je pogosto uporabljam od udeležbe na tečaju spletnih orodji.« Njena uporaba naj bi bila hitra in enostavna (3 odgovori), primerna pri prevajanju (2 odgovora) in za preverjanje pogovornih besed (1 odgovor) ali za izbiro ustreznejše besede, kar je študentka utemeljila s primerom iz spletne delavnice: »Gigafido sem uporabljala, če nisem bila zares, katero besedo bom uporabila, na primer odstotek oz. procent.« Drug študent pa je navedel, da Gigafido uporablja »za vse, kar [mu] pride na misel, čist da [preveri], če nekaj obstaja ali katera oblika je pogostejša«. Tretji študent uporablja Gigafido v sklopu Sloleksa kot pokazatelj, kje lahko uporablja določen sklon. Izpostavili so tudi, da »ima najboljše informacije« in – nekoliko nejasno oz. netočno – da so te informacije »najbolj uradne«.

Gigafidi sledi oblikoslovni leksikon Sloleks, ki so ga približno izenačeno izbirali študenti obeh skupin. V njem preverjajo predvsem sklonske oblike (5 odgovorov), posebej »nekih vsakdanjih nepravilnih besed«. Kot prednost so navedli hitrost in enostavnost uporabe (5 odgovorov), en študent pa je izpostavil, da dajeta Sloleks in Besana v primerjavi z Googlovim Prevajalnikom »veliko bolj konkretne in pravilne informacije«.

Pri pregibniku Besana je velika razlika med študenti obeh skupin – medtem ko ga študenti fizične skupine skoraj niso izbirali, so ga študenti spletne skupine izbrali celo večkrat kot prosto dostopni Sloleks. Pri tem žal ni razvidno, katero Besanino storitev dejansko uporabljajo. V spletni delavnici so študenti spoznali samo pregibnik, vendar je glede na odgovor dveh študentov, da Besano uporabljata za preverjanje besedila – »Lahko vidim napake, ki sem naredila.« –, mogoče sklepati, da poznajo tudi Besanino avtomatsko lektorico. Besano sicer uporabljajo pri pisanju (2 odgovora) in vsakdanjem govoru (1 odgovor), za preverjanje pregibanja besed (2 odgovora) in za prevajanje (1 odgovor). Po enkrat sta bili izpostavljeni tudi enostavnost uporabe in kakovost informacij.

Ostale jezikovne vire so študenti kot najuporabnejše izbirali precej redkeje. Nekoliko izstopa spletni slovar Pons, ki so ga večkrat izbrali študenti fizične skupine. Uporabljajo ga za prevajanje (2 odgovora), ker je »najlažje in najučinkovitejše«, en študent pa je prostodušno pojasnil, da je na študiju prevajalstva in rabi »vsa mogoča orodja za pomoč«. Med preostalimi viri je bila podana utemeljitev za slovarski portal Termania, ki da se uporablja »za prevajanje

nekaterih ekonomskih besed«, slovar sopomenk je eden od študentov izbral »kot prevajalec«, drugi pa je utemeljil: »slovar sopomenk pa mi omogoča da se izogibam ponovitvam in obogatim svoje besedišče.«

4 SKLEP

Učitelji slovenščine na Letu plus smo prepričani, da so delavnice spletnih orodij za slovenščino za tuje študente na Univerzi v Ljubljani koristne. K sreči to potrjujejo tudi povratne informacije, ki jih dobivamo od študentov. Seveda je treba upoštevati, da so v tem prispevku predstavljeno anketo verjetno reševali samo bolj motivirani študenti; gotovo je med tistimi, ki ankete niso rešili, več takšnih, ki smisla v seznanjanju z jezikovnimi viri in tehnologijami ne vidijo. Ne glede na to se bomo tudi v prihodnosti osredotočali na jezikoslovno bolj motivirane študente in jih v obliki, kakršno nam bodo dopuščale zunajjezikovne okoliščine poučevanja, še naprej usposabljali za uporabo najaktualnejših jezikovnih virov in tehnologij.

Glede na nekatere netočne informacije, ki so jih v anketi navajali študenti (npr. o tem, da so podatki v korpusu Gigafida in Sloleksu najbolj uradni), bomo v prihodnje še bolj opozarjali na to, na čem temeljijo posamezni jezikovni viri, in študente še bolj ozaveščali, da je podatke iz vseh virov in tehnologij treba uporabljati kritično. Vprašanje za prihodnost, ki terja nekaj premisleka, pa je, kdaj jim je smiselno predstavljati te vire in tehnologije. Kot je bilo razvidno iz prispevka, je bila glavnina rahlo negativnih komentarjev študentov, povezanih s temi vsebinami, dejansko povezana s predlogi, da bi to snov spoznali že v prvem semestru svojega študija na Univerzi v Ljubljani. Zaenkrat se za to (še) nismo odločili. Jezikovna zmožnost večine študentov je v prvem semestru namreč nizka, tako da bi lahko imeli težave pri razumevanju metajezika videoposnetkov. Predlog študentov pa nam vendarle služi kot opomnik, da jezikovnotehnološke vsebine še bolj vključujemo v poučevanje že med lektoratom. Z besedami enega od študentov: »[…] če imate vizijo da te vire uporabljamo pogostejše, bi bilo boljše učitelji leta plus na predavanjah uporabljajo z nami.«

LITERATURA

Digitalni slovarski in korpusni viri

Besana: Spletna verzija pregibnika Amebis Besana 4.24.1. Pridobljeno s https://besana.amebis.si/pregibanje/ (2. 9. 2021)

Evrokorpus. Pridobljeno s https://evroterm.vlada.si/evrokorpus (2. 9. 2021)

Fran. Pridobljeno s https://fran.si/ (2. 9. 2021)

Gigafida 2.0: Korpus pisne standardne slovenščine. Pridobljeno s https://viri.cjvt.si/gigafida/ (2. 9. 2021)

Google Prevajalnik. Pridobljeno s https://translate.google.com/ (2. 9. 2021)

GOS. Pridobljeno s http://www.korpus-gos.net/ (2. 9. 2021)

Igra besed. Pridobljeno s https://www.igra-besed.si/ (2. 9. 2021)

Kolokacije 1.0: Kolokacijski slovar sodobne slovenščine. Pridobljeno s https://viri.cjvt.si/kolokacije/slv/ (2. 9. 2021)

Kres. Pridobljeno s http://www.korpus-kres.net/ (2. 9. 2021)

Leto plus. Pridobljeno s https://www.uni-lj.si/studij/leto-plus/ (24. 11. 2021)

Pedagoški slovnični portal. Pridobljeno s http://slovnica.slovenscina.eu/ (2. 9. 2021)

PONS spletni slovar. Pridobljeno s https://sl.pons.com/prevod (18. 8. 2021)

Portal jezikovnih virov. Pridobljeno s https://viri.trojina.si/ (10. 12. 2021)

Sloleks 2.0: Slovenski oblikoslovni leksikon. Pridobljeno s https://viri.cjvt.si/sloleks/slv/ (2. 9. 2021)

Slovarji: Spletni slovarji in prevajalski pripomočki. Pridobljeno s https://evroterm.vlada.si/slovarji (18. 8. 2021)

Sopomenke 1.0: Slovar sopomenk sodobne slovenščine. Pridobljeno s https://viri.cjvt.si/sopomenke/ (2. 9. 2021)

Spletni slovar slovenskega jezika. Pridobljeno s http://ssj.slovenscina.eu/spletni-slovar (2. 9. 2021).

Šolar. Pridobljeno s http://korpus-solar.net/ (2. 9. 2021)

Termania. Pridobljeno s https://www.termania.net/ (18. 8. 2021)

Drugo

Arhar Holdt, Š., Čibej, J., Dobrovoljc, K., Gantar, A., Gorjanc, V., Klemenc, B., Kosem, I., Krek, S., Laskowski, C., & Robnik Šikonja, M. (2018). Thesaurus of Modern Slovene: By the Community for the Community. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global* Contexts (str. 401–410). Ljubljana: Znanstvena založba Filozofske fakultete. Pridobljeno s https://e-knjige.ff.unilj.si/znanstvena-zalozba/catalog/view/118/211/3000-1 (10. 12. 2021)

Arhar Holdt, Š., Logar, N., Pori, E., & Kosem, I. (2020). »Game of Words«: play the game, clean the database. *Lexicography for inclusion: EURALEX XIX: Congress of the European Association for Lexicography* (str. 41–49).

Arhar Holdt, Š., Kosem, I., & Pori, E. (2021). Jezikovni viri CJVT in njihova raba v izobraževalne namene. *Slovenščina na dlani 4* (str. 19–36). Maribor: Univerza v Mariboru, Univerzitetna založba.

Dobrovoljc, K., Krek, S., & Erjavec, T. (2015). Leksikon besednih oblik Sloleks in smernice njegovega razvoja. V V. Gorjanc, P. Gantar, I. Kosem & S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve* (str. 80–105). Ljubljana: Znanstvena založba Filozofske fakultete.

Gantar, P. (2011). Leksikalna baza za slovenščino: komu, zakaj in kako (naprej)? *Jezikoslovni zapiski, 17*(2), 77–92.

Kosem, I., Stritar, M., Može, S., Zwitter Vitez, A., Arhar Holdt, Š., & Rozman, T. (2012). *Analiza jezikovnih težav učencev: Korpusni pristop*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Kosem, I., Krek, S., Gantar, P., Arhar Holdt, Š., Čibej, J., & Laskowski, C. (2018). Kolokacijski slovar sodobne slovenščine. *Zbornik konference Jezikovne tehnologije in digitalna humanistika / Proceedings of the conference on Language Technologies & Digital Humanities, 20.–21. september 2018, Ljubljana* (str. 133–139). Ljubljana: Znanstvena založba Filozofske fakultete v Ljubljani. Pridobljeno s http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTDH-2018_Kosem-et-al_Kolokacijski-slovar-sodobne-slovenscine.pdf (10. 12. 2021)

Krek, S., Laskowski, C. A., Robnik Šikonja, M., Kosem, I., Arhar Holdt, Š., Gantar, P., Čibej, J., Gorjanc, V., Klemenc, B., & Dobrovoljc, K. (2018).

Thesaurus of modern Slovene 1.0. Ljubljana: Slovenian Language Resource Repository CLARIN.SI. Pridobljeno s http://hdl.handle.net/11356/1166 (10. 12. 2021)

Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I., & Dobrovoljc, K. (2020). Gigafida 2.0: the reference corpus of written standard Slovene. *LREC 2020: Twelfth International Conference on Language Resources and Evaluation: May 11–16, 2020, Marseille, France* (str. 3340–3345). Pariz: ELRA - European Language Resources Association. Pridobljeno s http://www.lrec-conf.org/proceedings/lrec2020/LREC-2020.pdf (10. 12. 2021)

Logar Berginc, N., Grčar, M., Brakus, M., Erjavec, T., Arhar Holdt, Š., & Krek, S. (2012). *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.

Stritar Kučuk, M., & Dobrovoljc, K. (2013). Korpusi na poti v šole. *Slovenščina 2.0, 1*(1), 181–194.

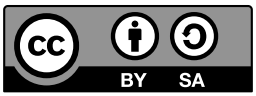
Verdonik, D., & Zwitter Vitez, A. (2011). *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Željko, M. (2003). Evroterm in Evrokorpus – terminološka baza in korpus prevodov. *Mostovi, 37*(1), 62–72.

ON-LINE SLOVENE LANGUAGE RESOURCES AND FOREIGN STUDENTS AT THE UNIVERSITY OF LJUBLJANA

In the first year of their studies at the University of Ljubljana, regularly enrolled foreign students are entitled to Slovene language classes within the Year Plus module. In the second semester of language classes, they learn about online Slovene language resources and technologies in a special workshop. This paper describes this workshop in the academic year 2019/20, when it was organized virtually due to the coronavirus pandemic, in the form of interactive videos with tasks to check students' understanding of the material. The second part of the paper focuses on students' perceptions of such language resources and technologies. With an online survey I analysed the views and experiences of two generations of students: students of the 2018/19 generation participated in this workshop in the classroom, while students of the 2019/20 generation participated on-line. According to the survey results, the younger generation of students uses online language resources more frequently. Students in both groups use Google Translate most often, followed by Sloleks, Besana, Fran and Pons. Their main arguments for using these resources are that they are fast and easy to use, or that students are the most accustomed to a particular resource.

Keywords: on-line language resources, Slovene language, foreign students, on-line learning



MEDNARODNI KONFERENCI ELEX (5.–7. JULIJ 2021) IN EURALEX (7.–9. SEPTEMBER 2021)

Magdalena GAPSA
Filozofska fakulteta, Univerza v Ljubljani

Gapsa, M.: Mednarodni konferenci eLex (5.–7. julij 2021) in EURALEX (7.–9. september 2021). Slovenščina 2.0, 9(2): 126–129.

DOI: https://doi.org/10.4312/slo2.0.2021.2.126-129

V poletnem času sta se na področju digitalne leksikografije odvili dve pomembni konferenci, in sicer sedma bienalna konferenca združenja Electronic lexicography in the 21st century (na kratko: eLex), ki je potekala med 5. in 7. julijem 2021, ter devetnajsta bienalna konferenca Evropskega leksikografskega združenja (European Association for Lexicography, EURALEX), ki je potekala med 7. in 9. septembrom 2021. Obe sta se zaradi epidemije koronavirusne bolezni 2019 ter z njo povezanih ukrepov in omejitev potovanj odvijali izključno na daljavo oz. na spletu. V okviru konference eLex so bile predstavitve vnaprej posnete in deljene z udeleženci, v živo pa so potekale razprave, medtem ko je za kongres EURALEX veljalo, da so bila predavanja udeležencem na voljo v obliki pretočnega prenosa v živo prek posebne platforme. Posnetki predavanj in predstavitev, predstavljenih v okviru kongresa EURALEX, so še vedno dostopni z uporabo podatkov, ki so bili udeležencem in slušateljem posredovani pred pričetkom konference, večina posnetkov kratkih in daljših predstavitev v okviru konference eLex pa je na voljo na spletni strani[1] in kanalu na portalu YouTube.[2]

Združenje eLex si prizadeva za interdisciplinarno povezovanje strokovnjakov s področja leksikografije, korpusne leksikografije, jezikovnih tehnologij, poučevanja in usvajanja jezika, prevajalstva, splošnega in uporabnega jezikoslovja

1 Program konference na spletni strani združenja eLex: https://elex.link/elex2021/conference-programme/

2 Posnetki predstavitev v okviru konference eLex: https://www.youtube.com/playlist?list=PLlR5sQQRZY8C_A8Ul9dSQaRP8O6f5lkPC

ter razvoja strojne in programske opreme. Konference, ki se od leta 2009 odvijajo vsake dve leti, so namenjene predstavitvi naprednih in sodobnih dosežkov na področju leksikografije.

Krovna tema letošnje konference eLex je bilo urejanje podatkov (ang. post-editing lexicography), kar dandanes predstavlja hkrati velik izziv in priložnost, saj so iz dovolj velikih in dobro označenih besedilnih korpusov lahko pridobljene oz. izluščene velike količine jezikovnih podatkov. Preverjanje in urejanje avtomatsko izluščenih podatkov bistveno pospeši delo, saj pomemben del opravijo računalniki, ki so pri manj zahtevnih nalogah že precej natančni, leksikografi pa se lahko osredotočijo na zahtevnejše naloge. Vendar tak pristop zahteva tudi prilagoditve delotoka in metodoloških pristopov. Tako so se v ospredju konference znašli prispevki, ki se ukvarjajo s teoretičnimi in praktičnimi izzivi, ki jih prinaša premik proti preverjanju in urejanju avtomatsko izluščenih podatkov, ki služijo kot priporočila in primeri dobrih praks na tem področju, se osredotočajo na znanja, spretnosti in izobrazbo, ki jih spremenjen pristop zahteva od urednikov, ter na orodja in tehnologije, ki ta premik omogočajo.

Razprave in predstavitev dosežkov je vsak dan odprlo vabljeno predavanje: v ponedeljek, 5. julija, sta Kris Heylen in Vincent Vandeghinste predstavila belo knjigo o prihodnosti akademskih slovarjev, v torek, 6. julija, je Pavel Rychlý govoril o nadgradljivosti matematike za leksikografske namene ter v sredo, 7. julija, je Pilar León Araúz govorila o načrtovanju in razvoju specializiranih virov na primeru EcoLexicona. Preostale prispevke so udeleženci predstavljali kot daljše predstavitve, združene v tematske sklope, ali pa kot krajše predstavitve oz. posterje. V skupno treh dneh smo udeleženci lahko spremljali 40 daljših predstavitev in 28 krajših predstavitev oz. posterjev. Tematski sklopi so obsegali teme, kot so večbesedni izrazi, avtomatizacija v slovaropisju, vključevanje uporabnikov, leksikografija in pedagogika, obogatitev podatkov, leksikografska orodja, zgodovinska leksikografija in digitalizacija ter modeliranje leksikografskih podatkov.

Zadnji dan konference je bila podeljena tudi Kilgarriffova nagrada (poimenovana po leta 2015 preminulem britanskem korpusnem jezikoslovcu, Adamu Kilgarriffu), ki jo je letos prejela Pilar León Araúz kot priznanje za vodilno vlogo, ki jo je odigrala pri pripravi in razvoju večjezičnega terminološkega vira

s področja okoljske znanosti, imenovanega EcoLexicon.[3] Nagrado, ki je bila ustanovljena leta 2016 kot priznanje za izvrstne dosežke na področju korpusnega in računalniškega jezikoslovja ter leksikografije, so podelili že tretjič.

Evropsko leksikografsko združenje (EURALEX) povezuje strokovnjake s celega sveta, ki se ukvarjajo z leksikografijo ali delujejo na povezanih področjih (npr. založnike, korpusne in računalniške jezikoslovce, znanstvenike, razvijalce strojne in programske opreme itn.) in hkrati zagotavlja prostor za izmenjavo idej. Združenje organizira tudi bienalne znanstvene konference oz. kongrese. Teme, ki so bile obravnavane v okviru letošnje devetnajste konference oz. kongresa, so obsegale med drugim postopek priprave slovarjev, raziskave s področja rabe slovarjev, leksikografske in jezikovne tehnologije, leksikografijo v povezavi s korpusnim jezikoslovjem, dvo- in večjezično leksikografijo, specializirane slovarje ter terminologijo, zgodovinsko in pedagoško leksikografijo, etimologijo, frazeologijo in kolokacije, poročila o leksikografskih in leksikoloških projektih in podobno.

Udeleženci smo lahko poslušali šest vabljenih predavanj, ki so bila povezana z diskusijo na določeno temo: v torek, 7. septembra, sta najprej predavala prof. Robert Lew in dr. Ana Frankenberg-Garcia, sledila je razprava na temo uporabe slovarjev in drugih leksikografskih izdelkov pri tvorjenju besedil, nato smo lahko poslušali prof. Christoforosa Charalambakisa, sledila je razprava na temo soočanja z miti in predsodki v leksikografiji na primeru *Slovarja sodobne grščine* (*Practical Dictionary of Modern Greek of the Academy of Athens*). V sredo, 8. septembra, je prvi osrednji govor imela prof. Anna Anastassiadis-Symeonidis, sledila je razprava na temo izdelave slovarja besednih družin, popoldne pa je predaval prof. Danie Prinsloo, razprava pa se je osredotočila na izstopajoče posebnosti tiskanih in elektronskih slovarjev. V četrtek, 9. septembra, je predavala prof. Janet DeCesaris, razprava pa je obsegala povezave med leksikografijo in morfologijo. Preostale prispevke so udeleženci, podobno kot v primeru konference eLex, predstavljali kot daljše predstavitve, združene v tematske sklope oz. sekcije, ki so večinoma potekale vzporedno, ali pa kot krajše predstavitve oz. posterje. V treh dneh smo v okviru kongresa EURALEX lahko spremljali preko 90 daljših predstavitev in 23 krajših predstavitev oz. posterjev.

3 Opis EcoLexicona: http://ecolexicon.ugr.es/en/index.htm

V okviru ene izmed sekcij je bila pozornost namenjena tudi nagradi ASHDRA, ki je bila ustanovljena leta 2018 in prvič podeljena leta 2019 in se od takrat podeljuje vsako leto. O nagradi je najprej spregovoril Michael Rundell, predsednik strokovne žirije, ki nagrado podeljuje. Najprej je slušateljem na kratko razložil, kdo je bil Albert Sidney Hornby, po katerem se nagrada tudi imenuje, spregovoril je o Hornbyevem skladu, nato pa je besedo predal štirim prejemnikom nagrade ASHDRA, ki so poudarili njen pomen ten opisali raznolike in napredne projekte, ki jih vodijo: Janine Knight, Aisling O'Boyle, Agus Riadi ter Yan Yan Teung.

V programu so bile predvidene tudi spremljevalne aktivnosti, ki so prav tako potekale na spletu. Prvi dan je bil posvečen glasbi, saj smo se naprej lahko udeležili delavnice tradicionalnih grških plesov *syrtos* ter *hasaposervikos*, sledil pa je koncert v živo. V sredo je bil na sporedu virtualni ogled Demokritove Abdere.

Med začetnimi nagovori v okviru konference EURALEX je Gilles-Maurice de Schryver udeležencem sporočil žalostno novico, da se je 3. septembra 2021 v 90. letu starosti poslovila leksikografinja Sue Atkins. Bila je začetnica na področju ustvarjanja dvojezičnih slovarjev iz korpusnih podatkov, ustanoviteljica združenja EURALEX, članica prvega sveta združenja in kasneje tudi njegova predstojnica. Bila je ena ključnih osebnosti, ki so oblikovale področje in usposabljale slovaropisce za področje afriških jezikov, zlasti preko delavnic *Afrilex-Salex*, ki jih je vodila skupaj z Michaelom Rundellom. Izjemno uspešni prvi delavnici sta se nato preobrazili v niz dogodkov, znanih kot *Lexicom workshop in lexicography and lexical computing* (na kratko: Lexicom). Na podlagi gradiva, ki je bilo uporabljeno za delavnico *Afrilex-Salex* in druga usposabljanja, sta Atkins in Rundell leta 2008 izdala odmeven priročnik *The Oxford Guide to Practical Lexicography*, ki še danes predstavlja temeljno branje za vse, ki se želijo seznaniti s praktičnimi rešitvami na področju sodobne leksikografije.

V okviru obeh konferenc smo lahko spremljali veliko dragocenih debat in idej ter se opredelili do trenutno najbolj perečih izzivov, s katerimi se srečujemo na leksikografski poti. Zahvaljujoč načinu izvedbe pa lahko do teh izsledkov in sklepov še vedno dostopamo ne samo v obliki prispevkov v zborniku, ampak tudi posnetkov predavanj in razprav. Imeli smo priložnost poslušati tudi prejemnike področnih nagrad, katerih delo lahko dojemamo kot primere dobrih leksikografskih praks.