

Jezikovna orodja za slovenščino v Italiji: jezikovni asistent Loris 1.1

Damjan Popič, David Bordon, Jasmin Franza, Luka Bezgovšek

Izvleček

V prispevku predstavimo orodje Loris 1.1, ki je nastalo v sodelovanju med Slovenskim raziskovalnim inštitutom (SLORI) in Centralnim uradom za slovenski jezik Avtonomne pokrajine Furlanije - Julijske krajine, temelji pa tudi na orodjih Centra za jezikovne tehnologije Univerze v Ljubljani. Loris je zasnovan kot spletna platforma za samodejno pregledovanje besedil v slovenščini, v prvi vrsti pa je namenjen govorcem in govorkam slovenščine na območju jezikovnega stika med slovenščino in italijanščino. V različici 1.1 zajema Loris primere jezikovnega stikanja, pravopisne iztočnice (SP 2001), Sloleks, vejice in črkovalnik, v prihodnosti pa načrtujemo še obsežne in kontinuirane nadgradnje, vključno z izdatnejšim opiranjem na umetno inteligenco.

Ključne besede: jezikovne tehnologije, jezikovna orodja, manjšinski jeziki, slovenščina v Italiji, Loris

1 Uvod

Kot primer jezikovnega orodja za slovenščino v Italiji v prispevku predstavimo orodje Loris 1.1¹ za pregledovanje besedil v slovenščini, ki je nastalo v sodelovanju med Slovenskim raziskovalnim inštitutom (SLORI) in Centralnim

1 Orodje je bilo v različici 1.0 javnosti predstavljeno 15. 2. 2023, v času priprave prispevka pa potekajo sklepne priprave za različico 1.1.

uradom za slovenski jezik Avtonomne dežele Furlanije - Julijske krajine (CU),² ki deluje v okviru spletišča Jezik na klik³ in je namenjeno predvsem govorcem in govorkam slovenščine na območju jezikovnega stika med slovenščino in italijanščino, čeprav je po zasnovi uporabno za bistveno širšo skupino pišočih v slovenščini.

Kot osnova orodja Loris služijo obsežne podatkovne baze o rabi slovenščine na območju jezikovnega stika, ki se nenehno posodablajo, zajemajo pa predvsem paronime, kalke in druge zaznamovane jezikovne pojave; tem se pridružujejo tudi drugi dostopni normativni podatki, obenem pa tudi sodobna jezikovno-tehnološka orodja, ki temeljijo predvsem na umetni inteligenci. Pri uporabi zaznamovanih jezikovnih elementov dobijo uporabniki in uporabnice povratne informacije o tem, da gre za (morebiten) odstop od osrednjega standarda slovenščine, obenem pa prejmejo tudi informacije o tem, kateri jezikovni elementi so morda v posameznem primeru ustrežnejši, nezaznamovani in standardni. Orodje je tako namenjeno predvsem uporabnikom in uporabnicam slovenščine na območju jezikovnega stika med italijanščino in slovenščino, pri katerih je slovenščina sicer prvi oz. materni jezik, vendar njihova sporazumevalna zmožnost v slovenščini v različnih govornih položajih ne ustreza vselej pragmatično-sporočajskim potrebam.

V prispevku predstavljamo tako gradivno osnovo kot vsebinsko ogrodje Loris, obenem pa tudi tehnično ozadje delovanja aplikacije in možnosti za njen nadaljnji razvoj. Pri tem poskušamo vlogo in pomen Loris vpeti v širši kontekst, ne zgolj na njegovo tehnično podstat, saj smo prepričani, da so (specializirane) jezikovne tehnologije, usmerjene v ohranjanje in razvijanje slovenščine v Italiji, izjemnega pomena. Zaradi globalizacije in jezikovne homogenizacije sta promocija in spodbujanje manjšinskih jezikov vse pomembnejša (Fishman, 1991), saj jezik ni le sredstvo sporazumevanja, temveč tudi bistvena sestavina kulturne identitete in dediščine (Edwards, 2009), ki se lahko po našem mnenju v digitalni dobi in prav zaradi nje razvija bolje (Bird, Chiang, 2012) ali pa – ob umanjkanju digitalne podpore (ali z njeno napačno rabo (Bird, 2020)) – tudi povsem propade (gl. Krek, 2012).

2 Ob pripravi pričujočega prispevka (poleti 2024) je bilo na orodju opravljenih 4970 iskanj, od tega 3456 različnih (orodje privzeto predlaga demonstracijski povedi »Je šla avtistka v Duino da bi podala davčno prijavo. Pa vzame avtobus ker ima abonma in zato ne potrebuje avto« kot zgled, kar je tudi najpogostejša iskalna poizvedba, s katerim lahko uporabniki in uporabnice preizkusijo delovanje orodja).

3 www.jeziknaklik.it

Orodje Loris je torej zasnovano kot povsem modularen in univerzalen, prosto dostopen pripomoček za slovenščino, zlasti v stiku z italijanščino, ki bo v nadaljnjih iteracijah vedno boljši, želimo pa si, da bi bil ob vsaki izboljšavi jezikovnotehnološke infrastrukture ta napredek samodejno preslikan tudi v samo orodje. Namen orodja nikakor ni, da bi govorcem in govorkam vsiljeval določeno varianto slovenščine kot lepšo ali boljšo, temveč zgolj to, da opozarja na odstopanja od jezikovnega standarda, s tem pa ozavešča jezikovno skupnost in pripomore k njenemu opolnomočenju.

2 Jezikovnotehnološka opremljenost skupnosti v jezikovnem stiku

V tem poglavju obravnavamo jezikovnotehnološko opremljenost skupnosti v jezikovnem stiku, pri čemer jezikovnotehnološko opremljenost razumemo kot nujno za ohranitev in razvoj jezika manjšinske skupnosti.

2.1 Jeziki v stiku in digitalna tehnologija

Jeziki so se skozi zgodovino razvijali v zapletenem prepletanju kulturnih, družbenih in geografskih dejavnikov, pri tem pa je bil jezikovni stik pogosto vzrok in gonilo jezikovne raznolikosti in prilagajanja. V današnjem povezanem svetu dobiva ta pojav nove razsežnosti, saj dinamiko jezikovnega stika narekujeta globalizacija in tehnološki napredek (Crystal, 2008).

Napredek na področju jezikovnih tehnologij ponuja številne obetavne možnosti za podporo in razvoj manjšinskih jezikov. Za večinske in velike jezike so na voljo številne jezikovne tehnologije za preverjanje črkovanja in slovnice ter aplikacije za učenje jezikov in samodejne pregledovalnike jezikovnih napak, vendarle pa tovrstne rešitve niso vedno ustrezne – predvsem z vidika »nearhetipskih« govorcev in govork teh jezikov, kamor lahko zaradi svojstvenih potez, ki jih je razvila slovenščina v stiku z italijanščino (Grgič, 2017), prištevamo tudi njene govorce in govorko, čeprav raziskave (prim. Leacock, 2014) po navadi zajemajo tiste, ki se jezika učijo. Vendarle pa lahko tehnologije na manjšinske (oz. ogrožene) jezike, tudi take z izjemno malo viri, uspešno prenesemo, (Anastasopoulos, Chiang, 2018; Scannell, 2007), zlasti če upoštevamo dejstvo, da je slovenščina v stiku z italijanščino gradivno in infrastrukturno relativno dobro (oz. vedno boljše) zastopana (gl. Popič, Grgič, 2023).

Poleg tega, da potrebujemo za jezikovnotehnološko podporo jezikom v stiku specializirane ali prilagojene vire, pa obstaja še znatna ovira, da so jezikovni viri na splošno nezadostno povezani in fragmentarni, kar je pogosta posledica projektnega financiranja, to pa kljub številnosti virov, tudi denimo za slovenščino, omejuje njihovo dostopnost in učinkovitost. Povezovanje teh virov v enotno platformo je bistvenega pomena za povečanje njihove uporabnosti in učinkovito podporo jezikovnim skupnostim (Scannell, 2007). Pri tem procesu povezovanja imajo ključno vlogo odprtokodni viri in tehnološka dostopnost.

Še posebej pomembno je presečišče jezikov v stiku z jezikovno tehnologijo, ki lahko pomeni ohranitev in opolnomočenje manjšinskih jezikov in njihovih skupnosti (Hinton in drugi, 2008), po drugi strani pa lahko predstavlja tudi velik izziv, če jezik ni (jezikovnotehnološko) opremljen za digitalno dobo (Krek, 2012: 4). V 21. stoletju jezikovni stik ni več omejen na fizično interakcijo, temveč se razširi v digitalno sfero. Ljudje z različnim jezikovnim ozadjem se povezujejo tudi prek spleta in digitalnih platform, zaradi česar se jeziki, ki so bili prej izolirani, zdaj srečujejo na spletu, medtem ko ogroženi jeziki tako ali tako po navadi sobivajo v večjezičnih kontekstih. To je privedlo do nastanka hibridnih jezikovnih različic in razvoja manjšinskih jezikov v digitalnem prostoru.

Jezikom manjšinskih, marginaliziranih ali domorodnih skupnosti grozi nazadovanje ali celo (digitalno) izumrtje, obenem pa jim jezikovne tehnologije ponujajo nove priložnosti (Grenoble, Whaley, 2006). Orodja za obdelavo naravnega jezika (NLP), strojno prevajanje, prepoznavanje govora in sintezo besedila v govor lahko premostijo jezikovne vrzeli in opolnomočijo manjšinske jezikovne skupnosti na različne načine. Te tehnologije namreč omogočajo komunikacijo, spodbujajo ohranjanje kulture in izboljšujejo dostopnost za govorce in govorke manjšinskih jezikov (Wilks, 2009), pri čemer pa moramo biti izjemno previdni, da jezika oz. kulture ne ohranimo zgolj za digitalne repozitorije, medtem ko jeziki in kulture dejansko izginejo (prim. Bird, 2020).

2.2 Jezikovnotehnološka opremljenost govorcev in govork slovenščine v Italiji

Kot je že bilo izpostavljeno, je Loris primarno namenjen skupnosti govork in govorcev slovenščine, ki živijo na območju jezikovnega stika med slovenščino

in italijanščino, zlasti na italijanski strani meje,⁴ in so na splošno opredeljeni kot (zamejski) Slovenci in Slovenke.⁵ Podatki kažejo, da prihaja na območju poselitve avtohtone slovenske skupnosti v Italiji do gradualnega razhajanja jezikovnega koda od centralnega standarda slovenščine, tako da se »lokalne rabe včasih že močno distancirajo od slovenskega jezikovnega kontinuuma« (Grgič, 2017: 95). Pretekle raziskave izpostavljajo predvsem dva vidika odmika zamejske slovenščine od centralnega standarda, in sicer: a) odmik same rabe in nastanek novega, lokalnega lekta ter b) omejenost govornih položajev, v katerih se slovenščina v tem okolju uporablja (pri čemer je verjetno slednji vidik tudi (vsaj delni) povzročitelj prvega):

[N]a naselitvenem območju slovenske manjšine v Italiji [so se] že uveljavili procesi, ki imajo nekatere značilnosti pidžinizacije, folklorizacije, okamnitve, oslabitve in opuščanja (zamenjave) slovenskega jezika. Ti procesi niso posledica nezadostne pravno-formalne zaščite ali nižjega statusa oz. prestiža manjšinskega jezika, ampak predvsem pomanjkanja pasivne (input) in aktivne (output) izpostavljenosti različnim rabam slovenskega splošnosporazumevalnega jezika in drugih idiomov slovenskega jezikovnega kontinuuma. (Grgič, 2017: 97)

Slovenska skupnost v Italiji je tako razvila specifične jezikovne poteze, ki so bile predvsem v zadnjih dveh desetletjih predmet številnih raziskav (Grgič, 2017; Jagodic in drugi, 2017; Mezgec, 2016). Te analize so potrdile prisotnost tovrstnih pojavov, niso pa je mogle kvantitativno meriti, razen na sorazmerno majhnem (in posledično manj reprezentativnem) vzorcu.

S sodobnimi metodami korpusnega jezikoslovja lahko na širšem segmentu besedil preverimo, katere izbire govorcev in govork slovenščine v Italiji se razlikujejo od izbir govork in govorcev slovenščine v Sloveniji in posledično od besedil, ki nastajajo v primerljivih okoliščinah in s primerljivim namenom. Na podlagi tega lahko spremljamo stopnjo oddaljenosti dveh ali več kodov, variant oz. idiomov v danem jezikovnem kontinuumu. To najbolj opazimo pri besedilih, v katerih je pričakovana variantnost minimalna – to so strukturirana besedila z ustaljenimi sporazumevalnimi vzorci in visoko koncentracijo

4 Številni vplivi italijanskega jezika so seveda močno vidni tudi pri govoricah in govorkah slovenščine na slovenski strani meje, vendar niso pod vplivom slovenskega šolskega sistema, v katerem se perpetuirajo še restriktivnejši jezikovni vzorci in elementi (npr. tintenblic, ‚brisalec‘, gumica ‚radirka‘, kažištran ‚zaznamek‘ ipd.).

5 Še enkrat lahko seveda poudarimo, da je aplikacija namenjena vsem govorcem in govorkam slovenščine, vendarle pa je primarno zasnovana za uporabnike in uporabnice na območju slovenščine kot jezika v stiku (z italijanščino).

terminov za poimenovanje iste predmetnosti (eno-/istoreferenčni termini oz., drugače rečeno, termini z eno/isto referenco na obeh straneh meje).

S korpusnojezikoslovno raziskavo terminologije, vezane na epidemijo covid-19 (Grgič, Popič, 2023), smo preverili, do kolikšne stopnje variantnosti prihaja v naj sodobnejših zamejskih javnih besedilih, in sicer smo zgradili korpus iz dveh letnikov Primorskega dnevnika (2020 in 2021; cca. 20 milijonov besed). Korpusnojezikoslovna analiza je pokazala, da prihaja tudi na področju terminologije v publicistični, torej javni rabi do znatnega razhajanja med jezikovnim standardom in zamejskim kodom. Jezikovno (in specifično leksikalno) variantnost običajno pričakujemo pri manj formalnih sporazumevalnih kodih, pri referiranju na različno predmetnost/pojavnost in, seveda, pri jezikovnih različicah (idiomih), ki jih sami govorci in govorce prepoznavajo kot lokalne/regionalne (Fought, 2006: 37). Korpusna analiza pa je pokazala, da je leksikalna variantnost v besedilih, ki nastajajo na območju poselitve slovenske skupnosti v Italiji, znatna tudi takrat, ko gre za strokovno terminologijo (in bi torej pričakovali večjo doslednost), za ubesedovanje predmetnosti/pojavnosti, ki se na obeh straneh meje ne razlikuje, ter za besedila, ki jih govorci in govorce dojemajo kot standardna.

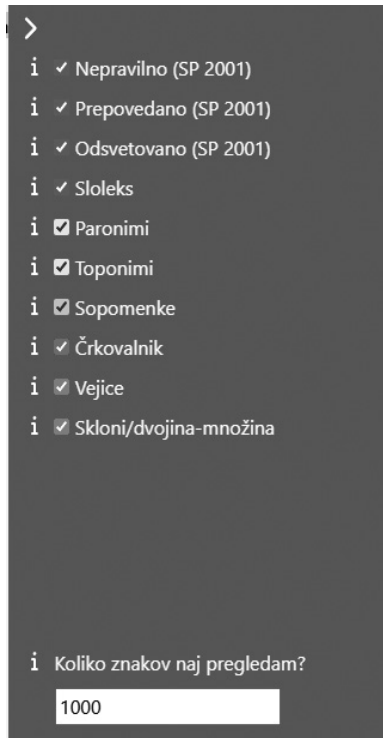
Podobne trende kažejo tudi preliminarne raziskave, opravljene na korpusu besedil v slovenskem jeziku, ki so nastali v javni upravi neposredno pred vzpostavitev Centralnega urada za slovenski jezik pri Avtonomni pokrajini Furlaniji – Julijski krajini in Mreže za slovenski jezik v javni upravi Furlanije – Julijske krajine.

3 Jeziko(slo)vna zasnova orodja Loris

Kot smo že izpostavili, je orodje Loris po zasnovi izjemno modularno in združuje ročno sestavljene, izvožene in z dodatnimi informacijami opremljene sezname leksike na eni strani, na drugi pa računalniške, strojno naučene modele, pri tem pa želimo biti pri razvoju orodja čim bolj odprti in povezujoči tudi v prihodnje. V trenutni inačici zajema orodje naslednje podatkovne zbirke:

1. podatke, vezane na jezikovni stik med slovenščino in italijanščino;
2. interno zbirko toponimov z območja slovenske poselitve v Italiji;
3. normirano besedje v *Slovenskem pravopisu 2001* (leksika, kvalificirana s kvalifikatorji nepravilno, prepovedano in odsvetovano);
4. normirano besedje iz Sloleksa, slovenskega oblikoslovnega leksikona;

5. sopomenke za nekatere najpogostejše lekseme, pridobljene iz slovarja Sopomenke 1.0 (CJVT);
6. sistem Vejica 1.0 (CJVT);
7. sistem za popravljanje sklonov in števila (dvojina – množina) EssayHelper;
8. nevronske črkovalnik (CJVT).



Slika 1: Izbirni meni za prikaz povratnih informacij v orodju Loris

Z izjemo internih seznamov SLORI so bili vsi preostali pridobljeni iz odprtopodatkovnih in prosto dostopnih zbirk, v nadaljevanju pa jih podrobneje predstavljamo.

3.1 Podatki, vezani na jezikovni stik med slovenščino in italijanščino

Loris 1.1 vsebuje dve vrsti izvirnih jezikovnih podatkov, ki jih je nalašč za to orodje obdelala skupina SLORI: jezik: pojave jezikovnega stikanja (paronime in kalke) ter toponime. Oboje predstavljamo v nadaljevanju.

Za potrebe orodja Loris 1.1 je skupina jezikoslovcev in jezikoslovk testno izbrala in obdelala 159 leksemov in leksemsko-skladenjskih enot, ki jih lahko razvrščamo med pojave italijansko-slovenskega jezikovnega stikanja – to so paronimi (t. i. lažni prijatelji) in kalki. Pri opredelitvi in izboru je skupina upoštevala naslednja merila, ki sicer niso enoznačna, so se pa izkazala za funkcionalna glede na cilje orodja:

- kot pojavi jezikovnega stikanja se obravnavajo lokalno specifične rabe jezikovnih sredstev, torej tiste, ki odstopajo od standarda (domnevnega povprečja), kakršnega izkazujejo primerljivi korpusi sodobne slovenščine;
- kot paronimi se obravnavajo eno- ali (redkeje) večjezični leksemi, ki imajo v italijanščini in slovenščini podobno glasovno (izrazno) podobo, a delno ali popolnoma različen pomen;
- kot kalki se obravnavajo stalne ali proste besedne zveze, torej enote pomena, ki so večje od enobesednega leksema; za potrebe tega orodja so se upoštevali zlasti terminološki kalki, medtem ko je frazeoloških manj.

IŠČETE PREVOD ZA:

abbonamento

STE MORDA MISLILI:

vozovnica; naročnina

NA PRIMER:

mesečna vozovnica za javni promet; kupiti letno vozovnico za smučišče; plačati naročnino

KAJ BESEDA PÖMENI V SLOVENŠČINI:

abonma = vnaprej plačano naročilo za predstave, koncerte (ita: "abbonamento a teatro, ai concerti")

Slika 2: Prikazno okno ob nasvetu Loris pri leksiki kot posledici stika med slovenščino in italijanščino (leksem *abonma*)

Različica Loris 1.1 ne vključuje drugih pojavov jezikovnega stikanja; nadgradnja je predvidena pri nadaljnjih različicah orodja.

Pojavi jezikovnega stikanja, obravnavani v tej različici orodja Loris, so bili izbrani na podlagi analize, priredbe in tehnične obdelave:

- gradiva iz dveh namenskih slovarjev, ki sta izšla v devetdesetih letih prejšnjega stoletja in predstavljata doslej edini leksikografski deli, specifično namenjeni tem pojavom: *Lažni prijatelji / I falsi amici* (Fabjan Bajc, 1994) in *Dve muhi na en mah / Due piccioni con una fava* (Fabjan Bajc, 1995);
- gradiva, ki ga je zbrala in obdelala skupina SLORIjezik v sklopu svojih lektorskih in svetovalnih storitev;⁶
- gradiva, ki ga je terminografsko obdelal Centralni urad za slovenski jezik v sodelovanju z nekaterimi strokovnjaki in strokovnjakinjami skupine SLORIjezik;⁷
- gradiva, ki ga je skupina SLORIjezik pridobila od uredništva Primorskega dnevnika in ga korpusnojezikoslovno procesirala, in sicer tako, da je lahko identificirala najbolj očitne leksikalne razlike med slovenščino, rabljeno v časniku Primorski dnevnik, in slovenščino, zajeto v korpusu Gigafida;
- gradiva, nastalega na območju avtohtone naselitve slovenske skupnosti v Italiji in zajetega v korpusu Gigafida (do 30. 6. 2021);
- iskalnih poizvedb uporabnikov in uporabnic orodja Loris po izdaji različice orodja 1.0;
- predlogov uporabnikov in uporabnic, zbranih prek portala Jezik na klik.

Tako zbrano gradivo smo primerjali z domnevno standardno rabo, in sicer zlasti na podlagi dveh korpusov: Gigafide in KAS. V Loris so bili vključeni tisti primeri, ki so bili v obravnavanem gradivu pogostejši in pri katerih so uporabniki in uporabnice, s katerimi smo prišli v stik, izkazovali nižjo stopnjo ozaveščenosti o odstopanju od standarda.

Ti primeri so bili nazadnje leksikografsko obdelani. Pri tem smo upoštevali predvsem uporabniško izkušnjo, kar pomeni, da smo skušali razlage čim bolj skrajšati, poenostaviti ter približati izkušnjam in potrebam ciljne skupine. Definicije smo uporabili v čim manjši meri; tam, kjer smo se zanje odločili, smo jih povzeli po SSKJ², a jih pri tem nekoliko prilagodili. Pomen smo skušali ponazoriti predvsem s sopomenkami (povzeli smo jih po Slovarju sopomenk sodobne slovenščine), rabo pa s kolokacijami (po Kolokacijskem slovarju sodobne slovenščine).

6 Zaradi specifik teh besedil (relativno omejenega obsega, slogovne in žanrske raznolikosti ter strukturne neenotnosti) gradivo ni bilo obdelano korpusno. Posamezni primeri so se evidentirali ročno.

7 Dostopno na: <https://www.jeziknaklik.it/zvezek/>.

V Loris 1.0 je bilo tako vključenih 206 izrazov,⁸ vezanih na jezikovni stik med slovenščino in italijanščino, za različico 1.1 pa načrtujemo vključitev nadaljnjih 157 izrazov;⁹ v sprotnih opombah so ti izrazi navedeni v osnovni obliki,

- 8 abonma, administracija, administrator, agent, akantonacija, akt, aktivnost, alikvota, anagrafski, argument, asistenca, asistent, asistentka, atlet, atletinja, avdicija, avla, avtist, avtistka, avtonomija, avtoriteta, avtorizacija, bankomat, beli teden, bilanca, brati v CC, bulizem, cikel, cikel, cirkularka, civilen, civilna, črpati iz lestvice, črtalo, davčna prijava, dekadenca, deklaracija, dekret, delati grdo figuro, didaktika, didaktika na daljavo, diploma, diplomirati, disciplina, disciplinski, dispozicija, dobro delo, docent, doktor, doktorica, določen, domicil, dopolnilna blagajna, dublirati, ekografija, farmacist, farmacistka, finančen, fiskalen, forfetaren, formacija, funkcionar, funkcionarka, galerija, garancija, grda figura, grdo vreme, gumica, imeti glavo na ramenih, imeti silo, imponirati, incident, informativa, iniciativa, inštitut, inštrukcija, inštruktor, inštruktorica, interesent, interesentka, izgubiti avtobus, izgubiti letalo, izgubiti trajekt, izgubiti tramvaj, izgubiti vlak, izvenšolski, javna dela, jestvina (jestvine), kamomilica, kanon, katedra, kiosk, kodeks, kolegij, kolokvij, komercialist, komercialistka, koncesija, konfekcija, konvencija, kostim, kredit, kriterij, kritičnost, krivda, kurikulum, kurz, kvalifikacija, kvota, laboratorij, lekcija, lesenka, lestvica, likvidacija, limit, lista, listek, lomiti škatle, magazin, mapa, menjati, monitoraža, monitoriranje, motivacija, multimedij-ska tabla, narediti grdo figuro, natečaj, natečajni razred, ne imeti volje, ne vidim ure, neprekinjena blagajna, nesti, norma, normativa, obrniti mail, odvisni delavci, operater, operaterka, organik, ostati nazaj, ostati slabo, ovrednotiti, palača, parcela, patent, penzija, personalen, plan, pobuda, podbradnik, pol, potreben, pravni zastopnik, prekvalificirati, prevzeti si, procedura, protokol, prva pomoč, rabiti, raketa, razbivati škatle, razsuvati škatle, register, registracija, rekurz, rekvizit, sedež, selekcija, servis, sheda, sigla, siglirati, sile javnega reda, škandal, škatla, šolabus, sončno leto, spektakel, stati slabo, statist, statistka, stolica, struktura, študent, študentka, suplenca, taksa, tehničen, telematski, teritorij, termin, terminska pogodba, toponomastika, upati, uprava, upravitelj, uspeti, valorizirati, ven iz glave, vzeti (avtobus, letalo, taksi, trajekt, tramvaj, vlak), zainteresiran, zapuščanje šole, zdaj kot zdaj, zloraba moči, zunajšolski
- 9 dobro delo, določen, ne vidim ure, zapreti luč, nadobuden, odpreti luč, en tubo, brez drugega, narediti odmor, menjati, refleks, težka droga, težko mamilo, na štiri roke / štiročno, oklepaj, tintenblic/tintenbliz, postrežnik, sluga, prijeti, imeti na grbi, zmenek, zeleni teden, komaj, prosojnost, tastatura, okrivčen, priti v manj, videti se v osebi, v prisotnosti, kažištran, agriturizem, se pustiti it, potreben, delati odbojko, privatist, zmrzniti se, ne imeti volje, ne vidim ure, prevzeti si, sile javnega reda, sončno leto, javna dela, zapuščanje šole, ostati nazaj, ven iz glave, zunajšolski, registrirati, zelena številka, meja dolga, delo na zahtevo, napredki, znati, slediti, biti sredi med ljudmi, namesto, odstop, plačevalci, tišina, usluga, frontalna učna oblika, svoj, doneč, predno, z zagotovostjo, vzgojno leto, stalež, suplent, aretacija, vprašati, dividend, ceremonija, poveriti, dobiti, ažurnirati, spomniti, asikuracija, log (logi), spet, honorarij, nonprofit, kratek spomin, slide, file (fileta), ravnatelj, biti žalosten / vesel zanj/zanjo/zanje, neplačevanje, plačevanje, dobrobit, dolgoobstojen, prostovoljno dejanje, evidentirati, zaprositi, položaj do nečesa (npr. do dolžnosti služenja vojaškega roka), anticipirati, (naj)več, preveritev, z upanjem, dati/staviti se na kolena, s težavo, komodat, registrator, ne hočem, večerni izhod, zatorej, (super) avtonomen, za dopust, gozdna policija, neizostaven, današnjica, natečajnik, selvijeta, pogovarjati/govoriti, doktoriranje, antenat, katod(o), anod(o), peškati, proč, zdole, zdol, samozadosten, intestirani papir, priposestevanje, sponka, patent, hospitacija, dojeman, soklopnik, rabi mi, abuzivno poslopje, namesto (kot veznik), konzulta, davčna koda, davek, teza, cel/ves, razbiti, agonističen, agonist, upravitelj, upraviteljica, pooblaščen upravitelj, pooblaščen upraviteljica, čelna pisarna, tekoči račun, , reklamacija, slediti, trio, obrat, lokal, estetski, oproščati, v vidiku, v sklopu, v okviru, odstavka, protagonist.

medtem ko jih Loris seveda »lovi« v vseh oblikah, pri čemer poskušamo predvideti tudi morebitne nestandardne oblike (npr. *Duinotu).

3.2 Toponimi

Jezikovni svetovalec Loris vsebuje tudi izbor slovenskih izvirnih krajevnih imen, endonimov, ki so tradicionalno v rabi na območju poseljenosti avtohtone slovenske narodne skupnosti v Italiji, pa tudi nekaj eksonimov. Ta seznam je bil vključen v orodje že v prvi fazi, saj v praksi večkrat naletimo na rabo italijanskih zemljepisnih imen, čeprav obstajajo in so izpričani slovenski endonimi. Raba italijanskih krajevnih imen je pogosto polcitatna: gremo v *Monfalcone, smo v *Duinu, se kopamo v *Sistiani, javna raba italijanskih krajevnih imen za območje avtohtone slovenske poselitve v Italiji pa je sicer pogosta tudi v Republiki Sloveniji. Loris uporabnike in uporabnice opozarja na ustrezno rabo slovenskih zemljepisnih in še posebej krajevnih imen za območje poselitve slovenske skupnosti v Italiji.

The screenshot shows the Loris linguistic tool interface. On the left, the logo 'loris' is displayed with the text 'jezikovni svetovalec' below it. The main content area is divided into two columns. The left column lists various linguistic categories with their corresponding Slovenian equivalents: ITALIJANSKO IME: Duino; SLOVENSKO IME: Devin; PREBIVALEC: Devinčan; PREBIVALKA: Devinčanka; PRIDEVNIK NA -SKI/-ŠKI: devinski; OBČINA: Devin - Nabrežina; KAJ JE (VAS, MESTO, ČETRTO, OBMEŠTJE, OBČINA, DEŽELA, DRUGO): mesto. The right column shows a search input field with the text 'Ste morda mislili Devin?' and a dropdown menu with the text 'Je šla **avtistka** v **Duino** da bi **podala** **dav**'.

Slika 3: Pojavno okence ob toponimskem napotilu v orodju Loris na primeru Duino → Devin

Različica Loris 1.1 vsebuje le izbor slovenskih zemljepisnih imen na Tržaškem in Goriškem, ki je nastal na podlagi skrbno zbranega gradiva, za kar je prvotno poskrbel Centralni urad za slovenski jezik Avtonomne dežele Furlanije - Julijske krajine. V Loris so bila vključena samo imena, pri katerih je raba ustaljena in dvomljive različice niso izpričane. Za številna krajevna imena predvsem na Videmskem namreč obstajata po dve ali več različic, zapis imen in oblike niso poenotene, kar je razvidno tudi iz zapisov na kašiputih in drugih javnih napisih (gl. Vidau, 2021). Ob tem je bilo smiselno vključiti tudi starejše slovenske eksonime za večja italijanska mesta (npr. Rim, Firenze, Neapelj), v prihodnjih iteracijah Lorisa pa nameravamo vključiti tudi imena pomembnih zgradb, spomenikov in stvaritev, pri čemer bomo kot osnovno gradivo uporabili tudi pretekle vsebine iskalnih poizvedb na Lorisu, da bi tako čim bolj ažurno odgovarjali na potrebe uporabnikov in uporabnic.¹⁰

Orodje ne vsebuje imen vseh krajev v slovenskem zamejstvu v Italiji predvsem zato, ker številna imena še niso standardizirana: »večji« kraji imajo po večini (eno) uradno slovensko poimenovanje, nekatere vasi, mestni predeli, deli vasi pa pogosto po več različic. Kot težavni se kažejo tudi pridevniki oz. imena prebivalcev in prebivalk, saj različni viri navajajo različna poimenovanja. Pri iskanju in zbiranju različnih variant smo želeli biti čim bolj celoviti,¹¹ pri tem pa smo preverjali tudi statute občin.

Tudi avtorji priročnikov navajajo, da glede slovenskih toponimov v Italiji vlada zmeda (Merkù, 1999), težave s poimenovanjem pa segajo v preteklost: v Trstu so v latinskih kodeksih do 15. stoletja prevladovale izvirne slovenske oblike krajevnih imen, ki so jih šele takrat začeli latinčiti, iz polatinjenih pa so se kasneje pojavile poitaljanjene oblike slovenskih krajevnih imen. Iz časa Marije Terezije izhajajo zapisi z nemškim pravopisom, pri čemer lahko opazimo razliko med zapisovalci, ki so slovenščino razumeli, in tistimi, ki je niso. Potem je nastopil isti problem z italijanščino, ko se je leta 1923 začelo nasilno poitaljančevanje slovenskih krajevnih imen. Veliko imen se je tako izgubilo, Merkù (1999) pa je za rekonstrukcijo slovenskih poimenovanj intervjuval domačine in domačinke, preučeval zgodovinske vire in pregledal topografske karte, pri

10 Pri tem želimo poudariti, da bi bili prosto dostopni prečiščeni sezname zgradb, spomenikov in drugih stvaritev, ki jih prevajamo, tako z vidika normativistike kot tudi priprave jezikovnih tehnologij izjemno dobrodošli.

11 Za Tržaško in Goriško smo kot osrednja priročnika uporabili obsežna leksikona (Bufon, Kalc, 1990; Rupel, 1995), za Videmsko pa nekoliko priročnike oziroma slovarčke (Merkù, 1999; Campigotto, Merkù, 2005; Merkù, 2006).

čemer se je soočal s težavami predvsem v primerih, ko ni bilo živih prič oz. spomina in so bili zapisi stari in težko berljivi.

Krajevna imena na Videmskem tako še niso vključena v bazo Lorisa, ker delo komisije za standardizacijo slovenskih krajevnih imen, ki jo je sklical Centralni urad za slovenski jezik Avtonomne dežele Furlanije - Julijske krajine, še poteka. Orodje pa ne svetuje le ustreznega slovenskega krajevnega imena: uporabnikom in uporabnicam so na voljo še druge informacije, kot so prebivalska imena, izpeljani pridevnik na -ski/-ški in podatek o vrsti imena (vas, mesto, občina idr.), kot je ponazorjeno v tabeli 1, v kateri so podani tako primeri toponimov, ki so zaradi usklajenosti že vključeni v orodje, kot tudi primeri, ki zaradi neusklajenosti (še) niso del podatkovne baze. V tabeli so pri neusklajenih poimenovanjih prav tako navedene različne rešitve, ki se pojavljajo v različnih virih, tako da lahko vidimo, katere rešitve so pri posameznih poimenovanjih na voljo. Pri zasnovi orodja smo se odločili, da bomo v končni izdelek vključili zgolj dorečene in usklajene informacije, tako da uporabnikov in uporabnic ne zmedemo z neusklajenostjo in navedenjem različnih virov, čeprav je posledično seznam (oz. število) vključenih toponimov bistveno skromnejši (64 vključenih poimenovanj : 479 zbranih poimenovanj).¹²

12 Vključeni so naslednji toponimi: Duino/Devin, Trieste/Trst, Barcola/Barkovlje, Roiano/Roijan, San Giovanni / Sveti Ivan, San Giacomo / Sveti Jakob, Coloncovez/Kolonkovec, San Vito / Sveti Vid, Guardiella/Vrdela, Miramare/Miramar, Conconello/Ferlugi, Cattinara/Katinara/Četnara, Longera/Lonjer, Banne/Bani, Basovizza/Bazovica, Gropada/Gropada, Santa Croce / Križ, Villa Opicina / Opčine, Padriciano/Padriče, Prosecco/Prosek, Trebiciano/Trebče, Malchina/Mavhinje, Aurisina/Nabrežina, Aurisina Stazione / Nabrežina Postaja, Aurisina Cave / Nabrežina Kamnolomi, Villaggio del Pescatore / Ribiško naselje, San Pelagio / Šempolaj, Visogliano/Vižovlje, Bagnoli della Rosandra / Boljunec, Grozzana/Gročana, Caresana/Mačkoviče, Pesek/Pesek, San Giuseppe della Chiusa / Ricmanje, Monrupino/Repentabor, Sales/Salež, Gorizia/Gorica, Lucinico/Ločnik, Oslavia/Oslavje, Doberdò del Lago / Doberdob, Valone/Dol, Jamiano/Jamlje, Marcottini/Poljane, Gabria/Gabrje, Savogna d'Isonzo / Sovodnje ob Soči, San Floriano del Collio / Števerjan, Cormons/Krmin, Ronchi dei Legionari / Ronke, Staranzano/Štarancan, Monfalcone/Tržič, Vencò/Jenkovo, Malborghetto/Naborjet, Collio/Brda, Cividale del Friuli / Čedad, Firenze/Firence, Val Canale / Kanalska dolina, Napoli/Neapelj, Aquileia/Oglej, Sottolungera/Podlonjer, Roma/Rim, Val Rosandra / Dolina Glinšči-ce, Sella Nevea / Na Žlebeh, Duino Aurisina / Devin - Nabrežina, Malborghetto Valbruna / Naborjet - Ovčja vas.

Tabela 1: Primeri izbranih toponimov in njihovih izpeljank, opisanih v besedilu

Italijansko ime	Slovensko ime	Ime prebivalca/-ke	Pridevnik
Že vključena poimenovanja			
Trieste	Trst	Tržačan/-ka	tržaški
Barcola	Barkovlje	Barkovljan/-ka	barkovljanski
Roiano	Rojan	Rojančan/-ka	rojanski
Gorizia	Gorica	Goričan/-ka	goriški
Cormons	Krmin	Krminec/-ka	krminski
Monfalcone	Tržič	Tržičan/-ka	tržiški
Nedorečena poimenovanja (še ne vključena v orodje)			
Sistiana	Sesljan	Sesljanec/-ka (Bufon, Kalc, 1990)	sesljanski
		Sesljančan/-ka (Merkù, 1999)	
Slivia	Slivno	Slivenc/-ka (Bufon, Kalc, 1990; Merkù, 1999)	slivenski
		Slivenčan/-ka (Amebis)	
Redipuglia	Redipulja (Merkù, 1999), Sredipolje (Rupel, 1995)		
Medeazza	Medjevas / Medja vas	Medvejec/-ka	medvejski
Draga Sant'Elia	Draga	Dražan/-ka (Bufon, Kalc, 1990), Dražca (Merkù, 1999)	draški (Bufon, Kalc, 1990), dražanski (Merkù, 1999)

Toponimi, ki jih Loris predlaga kot ustrežnejšo varianto, torej niso posledica (naših) poskusov podomačevanja, temveč so izvirna slovenska imena na območju, kjer je slovenska narodna skupnost zgodovinsko prisotna.

3.3 Normirano besedje iz Slovenskega pravopisa 2001 in leksikona Sloleks

Orodje Loris vsebuje tudi normativne informacije (poleg takojšnjih povratnih informacij o rabi vejice, sklonov in dvojine/množine, ki jih zagotavljajo strojno naučeni modeli), primarni razlog za to pa je, da smo želeli uporabnikom in uporabnicam podati tudi informacije o jezikovnem standardu. Naša želja je, da bi Loris tudi v vseh prihodnjih iteracijah na čim bolj dinamičen način uporabnike in uporabnice obveščal o potencialnih odmikih od jezikovnega standarda, pri tem pa besedila ne bi samodejno popravljaj, temveč bi zgolj svetoval oz. opozarjal.

Tako tehnični kot tudi vsebinski izziv pri zajemanju podatkov o jezikovnem standardu ostaja pridobivanje podatkov, saj je v trenutni obliki jezikovnega predpisa (*Slovenski pravopis 2001* na portalu Fran) treba normativne podatke ročno izluščiti, pretvoriti in pripraviti za vnos v orodje Loris. Vse iztočnice iz *Slovenskega pravopisa 2001* z normativnimi kvalifikatorji, ki so vključene v orodje Loris, so podane v tabeli 2, dodane pa so tudi iztočnice iz leksikona Sloleks, ki imajo pripisano normativno informacijo. Ta je bila pripisana v času priprav za *Slogovni priročnik slovenskega jezika* (Krek in drugi, 2013),¹³ informacijo pa lahko v leksikonu vidimo pri iztočnicah, ki spadajo v katero od obdelanih kategorij in pri katerih je v orodju pri posameznih variantah zapisan kvalifikator »nestandardno«.

13 Kazalo vseh obdelanih iztočnic v okviru projekta Sporazumevanje v slovenščini je na voljo na spletnem naslovu <http://slogovni.slovenscina.eu/Kazalo/Kazalo>. Pri tistih iztočnicah, pri katerih je bilo to glede na naravo vira smiselno, so bili napravljeni spiski problematičnih leksemov. Pri teh leksemih je bil v leksikonu Sloleks nato kot podatek pripisan tudi metapodatek za kategorijo »norma«, in sicer bodisi *variantno* bodisi *nestandardno*.

Tabela 2: Iztočnice z normativnimi informacijami v orodju Loris iz Slovenskega pravopisa 2001 in Sloleksa

Slovenski pravopis 2001	
Prepovedano	Nepravilno
Agin, Agripin, Ajatolin, akademkinja, Akindžijin, Alasijin, Alipašin, Anaksagorin, anglosaksonska, anglosaksonski, anglosaksonsko, Atilin, atin, Badjurin, Baragin, Barbarosin, barbin, Belin, Birendrin, Birollin, Boljkin, Bonaventurin, Borgiū, Bothin, Brahmin, Bratinin, Bratušin, Březinin, Budin, Burgibin, čačin, Cadornin, Cafutin, Capablancin, Casanovin, čičin, čikin, Cimarosin, Ciuhin, Costin, Croniin, dahijin, dalajlamin, Dalmatin, delibašin, Detelin, Dobidin, Dobrilin, Eljijin, Falladin, Fekonjin, Fidijin, Fondin, Fontanin, Gavellin, gazdin, Gezin, Glinkin, Godinin, Goljin, Gomufkin, Gongorin, Gonzagin, Goyin, Gregin, Guevarin, Hankin, harambašin, Hasanagin, hodžin, Hvalin, iskrin, Izajjin, Jakin, Jamin, Janšin, Jehovin, Jeremijin, Judin, Jurgurtin, Jurčin, kadijin, Kafkin, Kajfin, Kalidasin, Kaligulin, kamikazin, Karakalin, Katilinin, Keleminin, Kendin, Kobilčin, kolegin, kolovodjin, Kozinin, Kozmin, Krišnin, Krležin, Kuzmin, lamin, Lukin, Lumumbin, Machin, Magajnin, maharadžin, Malahijin, Mandelin, Matijin, Mesijin, Mihin, Missijin, Mitjin, Mitrin, Montezumin, Možinin, mulin, Mundin, nadvojvodin, Nehemijin, Nemanjin, Nerudin, očin, Okudžavin, oprodin, Palestrinin, pančenlamin, Parmin, pašin, Pavlihin, Petrarkin, pevovodjin, Pitagorin, Planinin, Podreccin, Poldin, Pribinin, Protagorin, protin, Pulcinellin, Radžin, Ramin, Rebulin, Robbin, Rogeljnin, Savonarolin, ščitonošin, Šegin, Segovijin, Senekin, Seničin, Šenoin, šerpín, Šerugin, Sinatrin, Šivin, Škodin, skupinovodjin, Slanin, slugin, Smetanin, Soyinkan, spahijin, Spinozin, starešinín, starikin, Stašin, Strnišin, strojevodjin, Stupičin, Sulin, Švarin, Svetinin, tatin, Tobijin, Tomizzin, Trdinin, Trohin, Tumin, Ulemin, Vanjin, Vargin, Vegin, vladikin, vodjin, vojvodin, Voltin, Vončinin, Vugin, Wulfflin, Zaharijin, Zamorin, Zapatín, Zaratuštrin, Žigin, Žižkin	Aldota, Benkota, Benota, bimbota, Božota, Bračkota, Brankota, Bratkota, Brečkota, Brunota, bučkota, črnkota, Cvetkota, Cvetota, Dajnkota, Danilota, Darkota, debelinkota, debelkota, debeluškota, dečkota, Dragota, Duškota, Edota, Ferdota, Fickota, fičkota, Frankota, Gojkota, Gvidota, Hinkota, Ivota, Jankota, Jelkota, Jenkota, Jesenkota, Jožkota, jugota, kapotta, Karlota, Kodalyta, krušta, Ladota, liskota, Liskota, Ljubota, Lovrota, mačota, Marinkota, Mariota, Markota, Maticota, Matkota, Meškota, Metelkota, Mihata, Milenkota, Milkota, Mirkota, Mirota, miškota, Miškota, Mišota, Mohorkota, Murkota, Nedeljkota, Nikota, Ninota, Otota, Perkota, polota, Radkota, Radota, Rafkota, Rajkota, rdečkota, Rojkota, Samota, Sardenkota, Sašota, Savota, Šeligota, Šerkota, Silvota, sinkota, sivkota, Slavkota, srečkota, Srečkota, Stankota, stričkota, Tomota, Veljkota, Venota, Vilkota, Vinkota, Vitota, Vladkota, Vladota, Vojkota, volvota, Vrečkota, Vučkota, Žarkota, Zdenkota, Zdravkota, Željkota, Živkota, Zlatkota, Zmagota, Zorkota, Zvonkota

Odsvetovano	Sloleks / Slogovni priročnik
	Normativna oznaka
<p>bagerist, beležiti, bojen, borba, bremeniti, cepidlačiti, cepidlaštvo, čigavosten, ciljati, cinobarit, compact disc, cross-country, deaktivirati, deliti, demonstrirati, dinama, diplomatičen, dkg, DNa, dobrostoječ, dohajati, dol, dr.-roman, edinica, elektrifikacijski, elektrotehničen, elektrotehniški, ensilaža, ensilirati, Euratom, flanša, flanšati, flumaster, fundacija, funkcija, gazificirati, gazifikacija, gotov, granatirati, instanten, interesirati, izdvajati, izdvojitve, izdvojiti, izgovorjava, iznenađen, iznenaditi, izneveriti, izogib, iztrošen, iztrošiti, izvežban, izvežbati, izvršen, izvrševati, izvršilec, izvršitev, izvršiti, izvršljiv, javljati, kakao, kolebati, kolebav, kolumen, komandirati, končati, končevati, kontrola, kontrolirati, merodajen, merodajno, nadaljevati, nadmodriti, nadoknaditi, nahajati, nameščati, namestiti, nasipan, nastopiti, nastrojen, nehati, nehavati, nehovati, neizbežen, neizkrivljen, neiztrošen, neizvežban, neizvršen, neopran, neopravičen, neosveščen, nepogrešljiv, nepogrešljivo, nepokreten, nepomirljiv, nepotrošen, nerazdvojen, nerazdvojno, nevpadljiv, nevpadljivo, nezaključen, nezaželjen, neželjen, nezgrajen, nočevati, nočitev, nočitven, nositi, nuditi, obeležba, obeleževati, obeležiti, objasniti, objaviti, območje, obnašati se, obratovati, obrazložiti, obstati, obstoječ, odbijati, odbiti, oddvajati, oddvojitve, oddvojiti, odgovarjajoč, odgovarjati, odgovoren, odigravati, odkloniti, odklonljiv, odkriti, odločiti, odnesti, odpadati, odpasti, odpor, odprema, odpremen, odpremiti, odpremljati, odpremnica, odpremnik, odrediti, odstopati, ograditi, ograjevati, oklopen, oklopnica, oklopnik, okorel, okoriščati se, okoristiti se, omejevati, omejiti, onostranski, onostranstvo, ONU, opran, oprati, opravičen, opravičevati, opravičiti, opremljen, organizirati, osigurati, oskubiti, oskubljen, ostali, ostvariti, ostvarjati, osvajati, osveščati, osveščen, osveščeno, osveščevalen, osvestiti, osvoboditi, osvojiti, otežkočati, otežkočen, otežkočiti, pažnja, petokrak, pociljati, podhranjen, podleči, podlegati, podvaliti, podvzemat, podvzeti, pogajati, pogojen, pogojevati, pogojiti, pogrešen, pogrešiti, polagati, položiti, ponuditi, porušiti, posluževati, poslužiti, postaviti, postavljati, postojati, potegniti, potrebno, potrošen, potrošiti, poveriti, povleči, pozivati, predpostaviti, predpostavljen, pregaziti, prelamljati, prelòm, prelomiti, prenehati, prenehavati, prenehovati, prenešen, prepričan, prepričujoč, preseči, presežen, preuranjen, preuranjeno, prevaliti, prinešen, pristajati, pristati, pristopati, pristopiti, prizanešen, prizanešeno, proglas, proglašati, proglašen, proglasitev, proglasiti, proizvajati, propuščati, propust, propusten, protioklopen, protiviti se, razčistiti, razdelan, razdelati, razdelava, razmak, razreševati, razrešiti, realizirati, regrut, regrutirati, rušiti, škoda, skubiti, sledeč, slučaj, smatrati, smer, soočati, soočiti, sovrstnik, sovrstništvo, spektakel, špiker, špikerski, špinel, špinet, špirala, špiralast, splovitev, sploviti, spoznati, spoznavati, sprovajati, sproveden, sprovesti, sprožiti, štadij, štadion, starocerkvenoslovanščina, statut, tangirati, trošiti, učenec, uFO, ugriz, ukinitev, ukiniti, umeten, UN, unešen, UNO, uporen, upravljati, upropaščati, upropastiti, USA, ustmen, ustmeno, uvežban, uvežbati, vezalec, vežba, vežbati, višek, visokomorski, visokovreden, visokozmogljiv, vklopiti, vklopljen, vladin, vodič, voditi, vodostaj, vpadati, vpasti, vpricho, vršiti, zabeležiti, začeti, začeti, zadolžen, zadržati, zagotoviti, zaključevati, zaključiti, zakolebati, zatečen, zavariti, zavarovati, zavesljaj, zaželjen, željen, zemlja, zmrzniti, zmrzovalen, zrušiti</p>	<p>Applea, Bruca, Die-sela, FBIja, Geoga, Houstna, Jemena, OZNa, OZNja, Pierrea, Powella, Shakespearea, SMSa, SMSja, Stallonea</p>

V tabeli 2 so prikazane vse osnovne oblike oz. iztočnice, ki vsebujejo normativno napotilo. Kot lahko vidimo, je z naskokom najboljše kategorija odsvetovane leksike v SP 2001, najskromnejša zbirka pa je tista iz Sloleksa. Pri tem oba vira, SP 2001 in Sloleks, podajata podatke na različne načine. Na sliki 4 je prikazan način podajanja informacij za informacije, pridobljene iz SP 2001.



Slika 4: Prikazno okno v orodju Loris z napotilom Agripin → Agripov (SP 2001)

Normativne informacije iz SP 2001 so podane brez dodatnih razlag, zgolj z možnostjo, da uporabnik ali uporabnica besedilo s klikom na predlog takoj zamenja in torej upošteva napotilo. Napotki iz Sloleksa so nekoliko obsežnejši, kot je prikazano na sliki 5.



Slika 5: Prikaz napotila v orodju Loris za nestandardno pregibanje imena Apple

Kot vidimo, uporabnik ali uporabnica Loris pri tovrstnih nasvetih ne dobi zgolj hitre informacije o standardni rabi, temveč tudi kratko razlago in uporabno povezavo do nadaljnjih virov. Pri tem smo želeli uporabiti vire, izdelane pri projektu Sporazumevanje v slovenščini, ki so ponujali ekosistem medsebojno povezanih virov, vendar ti niso več aktivno podprti in smo se zato odločili, da uporabnike in uporabnice zaenkrat usmerimo na druga spletna mesta z relevantnimi informacijami. V prihodnje želimo tudi tovrstne dodatne informacije bodisi vezati na trajna, institucionalna spletna mesta bodisi jih vključiti v notranjo infrastrukturo spletišča Jezik na klik.

3.4 Podatki, pridobljeni s strojno naučenimi modeli: vejica, skloni in število (dvojina – množina)

V Loris smo vključili tudi več strojno naučenih modelov, ki uporabniku in uporabnici nudijo hipne povratne informacije, prav vsi pa temeljijo na nevronskega modelu BERT oz. sloBERTa.¹⁴

- sistem Vejica 1.0 (Božič in drugi, 2020),
- sistem EssayHelper za samodejno popravljanje sklonov in glagolskega števila (dvojina – množina) (Petrič, 2021),
- nevronske črkovalnik (Klemen in drugi, 2024).

3.4.1 Orodje Vejica 1.0¹⁵

Sistem Vejica 1.0 je strojno naučeni model za stavljenje vejic v slovenščini, ki vejice stavi s 94-odstotno natančnostjo. Spletno orodje preverja postavitev vejic v slovenskih besedilih s pomočjo jezikovnega modela BERT, izpopolnjena za problem postavljanja vejic. Za izpopolnjevanje modela je bila zgrajena učna množica iz dela besedil korpusa sodobne standardne slovenščine Gigafida 2.0. Učna množica je obsegala 907.870 stavkov, ki so v povprečju vsebovali vsak po dve vejici. V različico orodja Loris 1.0 smo vključili lastni API,¹⁶ ki je bil zgrajen iz odprto dostopnih virov CJVT; v različico Loris 1.1 pa je vključen novi API, kar pomeni, da bodo v Loris hipno vključene vse morebitne izboljšave modela Vejica 1.0.

14 <https://huggingface.co/EMBEDDIA/sloberta>

15 <https://orodja.cjvt.si/vejice/>

16 Application programming interface (angl.), vmesnik, s katerim lahko dostopamo do podatkov nekega vira, ne da bi ta vir podvajali, temveč lahko do njegovih podatkov preprosto dostopamo prek »vrat«
API-ja.

3.4.2 *Model Essay Helper*

S pomočjo jezikovnega modela tipa BERT (sloBERTa), naučenega na slovenskem jeziku, je bil na osnovi dela Timoteja Petriča (2021) razvit sistem za popravljanje sklonov in števila, pri čemer je največ pozornosti deležna zaznava pravilnosti rabe sklona tožilnik : rodilnik in števila množina : dvojina, pri za-znavi pa je, podobno kot Vejica 1.0, okoli 95-odstotno uspešen.

3.4.3 *Nevrnski črkovalnik*

V okviru projekta PROP je bil razvit nov nevronske črkovalnik SloNSpell, ki je prvi (odprtodostopni) nevronske črkovalnik za slovenščino in predstavlja velik korak naprej od klasičnih črkovalnikov; obenem pa je bil razvit tudi klasični morfološki črkovalnik (SloSpell; Klemen in drugi, 2024: 86). V prvo različico orodja Loris smo vključili klasični odprti črkovalnik tipa hunspell, ki je primeren za morfološko bogate jezike, vendar pa ima z vidika integracije v orodje Loris številne pomanjkljivosti: ker ni komercialno podprt, je nabor vključene leksike izjemno skop, nadgradnje pa časovno potratne, obenem pa ni možnosti dodajanja besedišča v neki centralni leksikalni fond, ki bi se tako lahko posodabljal.

Novi nevronske črkovalnik teh pomanjkljivosti nima, saj temelji na sodobnem gradivu in se lahko tudi prilagaja (oz. ga je mogoče posodobiti), posodobitve pa so takoj vidne tudi v orodju Loris (in vseh ostalih potencialnih orodjih, ki so na nevronske črkovalnik vezana prek vmesnika API).

3.5 *Sopomenke*

Kot osnovo za nabor sopomenk smo izluščili sopomenke iz jezikovnega vira Sopomenke 1.0,¹⁷ pri čemer smo primerjali podatke iz korpusa *Primorskega dnevnika* z referenčnim korpusom Gigafida 2.0. Primerjali smo pogostnost posameznih lem in tistih, ki v zamejski rabi izkazujejo bistveno višjo frekvenco, pripisali sopomenke, izluščene neposredno iz vira.¹⁸ Ker pa je za tovrstno integracijo jezikovnega vira potrebnega veliko dela, obenem pa spremembe in nadgradnje izvornega vira niso samodejno prenesene tudi v orodje

17 <https://viri.cjvt.si/sopomenke/slv/about>

18 Krek et al., 2023. <https://www.clarin.si/repository/xmlui/handle/11356/1916>

Loris, želimo v prihodnje tudi sopomenke (in kolokacije)¹⁹ v orodje vnesti z uporabo vmesnika API.

4 Sklep in nadaljnje delo

Razvoj Lorisa predstavlja pomemben napredek pri podpiranju jezikovnih in kulturnih potreb slovenske manjšine v Italiji. Ključna značilnost Lorisa je njegova obsežna podatkovna zbirka, ki združuje podatke, pridobljene s strojno naučenimi modeli, ročno zbrano vsebino in, kar je najpomembneje, leksikografsko obdelan seznam leksike, na katero je vplival jezikovni stik z italijanščino. Ta ročno pripravljene seznam obravnava subtilne, a vseprisotne spremembe v rabi besed, do katerih pride ob tesnem soobstoju jezikov in medsebojnem vplivu. Ena glavnih značilnosti Lorisa je torej vključitev različnih podatkovnih virov, od povsem ročno zbranih in pripravljenih informacij do strojno naučenih modelov Centra za jezikovne tehnologije in vire. Kljub temu da so za slovenščino na voljo številna jezikovna orodja, so številna še vedno nepovezana in fragmentirana, zato smo obsojeni na dolge sezname povezav in zaznamkov za sicer komplementarne vire. Pri zasnovi Lorisa si prizadevamo združiti te vire in čim več orodij združiti v enotno platformo. Pri tem je pomembno, da so jezikovni viri odprti in tehnološko dostopni.

Ustvarjanje jezikovnih tehnologij za manjše in regionalne jezike, jezike v manjšinskem in migrantskem položaju ter za izpostavljene in ogrožene jezike je v našem vse bolj povezanem svetu izjemnega pomena. Orodja, kot je Loris, imajo ključno vlogo pri ohranjanju kulturne dediščine in razvijanju jezikovne raznolikosti. Manjšinskim skupnostim omogočajo hitro, uporabnikom in uporabnicam prijazno dostopanje do osnovnih elementov in informacij za udejstvovanje v informacijski družbi – črkovalnike, slovnične popravjalnike, bralnike, prepoznavalnike govora ipd.). Namen Lorisa je omogočiti, da se govorniki in govornice samozavestno vključijo v pisno komunikacijo, s čimer krepijo svojo kulturno identiteto, hkrati pa krmarijo po sodobnih digitalnih krajinah in pri tem niso prikrajšani pri dostopu do informacij in učinkovitem sporazumevanju.

19 <https://viri.cjvt.si/kolokacije/slv/>

Bibliografija

- ANASTASOPOULOS A., CHIANG D. (2017), *Tied Multitask Learning for Neural Speech Translation*, v R. Barzilay, M. Y. Kan (ur.), *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Vancouver, 82–91.
- BIRD S., CHIANG D. (2012) *Machine translation for language preservation*, v M. Kay, C. Boitet (ur.) *Proceedings of COLING 2012: Posters*, The COLING 2012 Organizing Committee, Bombaj/Mumbai, Indija/India, 125–134.
- BIRD S. (2020), *Decolonising Speech and Language Technology*, v D. Scott, N. Bel, C. Zong (ur.), *Proceedings of the 28th International Conference on Computational Linguistics*, International Committee on Computational Linguistics, Barcelona, 3504–3519.
- BIRD S., SIMONS G. (2003), *Seven Dimensions of Portability for Language Documentation and Description*, *Language*, 79, 3, 557–582.
- BUFON M., KALC A. (1990), *Krajevni leksikon Slovencev v Italiji: topografski, zemljepisni, zgodovinski, kulturni, gospodarski in turistični podatki o krajih v Italiji, ki jih naseljujejo Slovenci. Knj. 1, Tržaška pokrajina*, ZTT, Trst.
- CAMPIGOTTO S., MERKŪ P. (2005), *I país de patrie: dizionariut toponomastic: furlan, italiano, slovensko, deutsch*, *La Patrie dal Friül, Glemone*.
- CRYSTAL, D. (2008), *Txtng: The Gr8 Db8*, Oxford University Press, Oxford.
- EDWARDS J. (2009), *Language and Identity: An Introduction*, Cambridge University Press, Cambridge.
- FABJAN BAJC D. (1994), *Lažni prijatelji: slovensko-italijanski slovar paronimov/I falsi amici: vocabolario italiano-sloveni dei paronimi*, Mladika, Trst/Trieste.
- FABJAN BAJC D. (1995), *Dve muhi na en mah: slovensko-italijanski frazeološki slovar/Due piccioni con una fava: vocabolario fraseologico sloveno-italiano*, Zadruga Goriška Mohorjeva, Gorica/Gorizia.
- FISHMAN J. A. (1991), *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*, Multilingual Matters, Clevedon-Philadelphia.
- FOUGHT, C. (2006), *Language and ethnicity*, Cambridge University Press, Cambridge.
- GRENOBLE L. A., WHALEY L. J. (2006), *Saving Languages: An Introduction to Language Revitalization*, Cambridge University Press, New York.
- GRGIČ M. (2017), *Italijansko-slovenski jezikovni stik med ideologijo in pragmatiko*, *Jezik in slovstvo*, 62, 1, 89–98.
- HINTON L. (2013), *Audio-Video Documentation*, v L. Hinton in K. Hale (ur.), *The Green Book of Language Revitalization in Practice*, Brill, Leiden-Boston, 265–272.
- JAGODIČ D., KAUCIČ BAŠA M., DAPIT R. (2017), *Jezikovni položaj Slovencev v Italiji*, v N. Bogatec, Z. Vidau (ur.), *Skupnost v središču Evrope: Slovenci v Italiji od padca Berlinskega zidu do izživov tretjega tisočletja*, ZTT-SLORI, Trst, 66–88.
- KLEMEN M., BOŽIČ M., HOLDT Š. A., ROBNIK-ŠIKONJA M. (2024), *Neural Spell-Checker: Beyond Words with Synthetic Data Generation*, v E. Nöth, A. Horák, P. Sojka (ur.), *Text, Speech, and Dialogue. 27th International Conference, TSD 2024*, Lecture Notes in Computer Science, 15048, Springer, Cham.
- KREK S. (2012), *Slovenski jezik v digitalni dobi*, Zbirka Bela knjiga, Springer, Heidelberg.

- KREK S. in drugi (2023), *Thesaurus of Modern Slovene 2.0*, v M. Medved in drugi (ur.), *eLex: electronic lexicography in the 21st century (eLex 2023): proceedings of the eLex 2023 conference*, Lexical Computing, Brno.
- LEACOCK C., CHODOROW M., GAMON M., TETREAUULT J. (2014), *Automated Grammatical Error Detection for Language Learners*, Morgan & Claypool Publishers.
- MERKŪ P. (1999), *Slovenska krajevna imena v Italiji: priročnik/Toponimi sloveni in Italia: manuale*, Mladika, Trst/Trieste.
- MERKŪ P. (2006), *Krajevno imenoslovje na slovenskem zahodu*, Založba ZRC, Ljubljana.
- PETRIČ, T. (2022), *Predlogi jezikovnih popravkov v slovenščini z modelom SloBERTa*, diplomsko delo, Repozitorij Univerze v Ljubljani.
- RAVNIKAR A. (2010), *The Slovenian Minority in Italy: Linguistic Rights and Cultural Identity*, *Journal of Ethnopolitics and Minority Issues in Europe*, 9, 1, 45–65.
- RUPEL A. (1995), *Krajevni leksikon Slovencev v Italiji: topografski, zemljepisni, zgodovinski, kulturni, gospodarski in turistični podatki o krajih v Italiji, ki jih naseljujejo Slovenci ali sodijo v isto upravno enoto. Knj. 2, Goriška pokrajina*, NSK-SLORI, Trst.
- SCANNELL K. (2007), *The Crubadán Project: Corpus Building for Under-Resourced Languages*, v C. Fairon in drugi (ur.), *Building and exploring web corpora: proceedings of the 3rd Web as Corpus Workshop, volume 4*, Presses universitaires de Louvain, Louvain-la-Neuve, 5–15.
- Spletna različica pregibnika Amebis Besana 4.28.2. (<https://besana.amebis.si/pregibanje/>).
- TIEDEMANN J. (2012), *Parallel Data, Tools and Interfaces in OPUS*, v N. Calzolari in drugi (ur.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, ELRA, Istanbul, 2214–2218.
- WILKS Y. (2010), *Machine Translation: Its Scope and Limits*, Springer, Sheffield.

Povzetek

Prispevek obravnava jezikovno orodje za slovenščino v Italiji Loris 1.1, zasnovano kot spletna platforma za samodejno pregledovanje besedil v slovenščini. Orodje je v prvi vrsti namenjeno govorcem in govorkam slovenščine na območju jezikovnega stika med slovenščino in italijanščino. Nastalo je v sodelovanju med Slovenskim raziskovalnim inštitutom (SLORI) ter Centralnim uradom za slovenski jezik Avtonomne pokrajine Furlanije - Julijske krajine, opira pa se tudi na orodja Centra za jezikovne tehnologije Univerze v Ljubljani.

Orodje temelji na večjih podatkovnih bazah, ki vključujejo normativne podatke, kot so pravopisne norme (iz *Slovenskega pravopisa 2001*), oblikoslovni leksikon (Sloleks), CJVT vejice in nevronske črkovalnik, sopomenke za nekatere najpogostejše lekseme (iz slovarja Sopomenke 1.0), sistem za popravljane sklonov in števila (dvojina – množina), v prihodnosti pa so predvidene še obsežne nadgradnje, vključno z izdatnejšim opiranjem na umetno inteligenco. V različici Loris 1.1 so vgrajene funkcionalnosti za prepoznavanje pogostih napak, ki izhajajo iz jezikovnega stika, kot so napačne rabe paronimov

(t. i. lažnih prijateljev), kalkov (prevodov besednih zvez) in toponimov (krajevnih poimenovanj). Uporabnikom in uporabnicam omogoča dostop do relevantnih povratnih informacij glede odstopanj od standardne slovenščine, predvsem z vidika lokalno zaznamovanih jezikovnih pojavov.

Razvoj Loris predstavlja pomemben napredek pri podpiranju jezikovnih in kulturnih potreb slovenske manjšine v Italiji, saj globalizacija in digitalizacija povečujeta jezikovni pritisk večjih jezikov. Loris omogoča uporabnikom in uporabnicam, da se soočajo s tovrstnimi izzivi, saj z opozarjanjem na jezikovna odstopanja pripomore k večji ozaveščenosti in krepitvi jezikovnih kompetenc govork in govorcev obmejnega območja.

Ena glavnih značilnosti Loris je torej vključitev različnih podatkovnih virov, od povsem ročno zbranih in pripravljenih informacij do strojno naučenih modelov CJVT. Kljub temu da so za slovenščino na voljo številna jezikovna orodja, so številna še vedno nepovezana in fragmentirana, zato smo obsojeni na dolge sezname povezav in zaznamkov za sicer komplementarne vire. Loris si prizadeva združiti te vire in čim več orodij združiti v enotno platformo.

Jezikovni svetovalec Loris omogoča govorkam in govorcem samozavestno vključevanje v pisno komunikacijo, kar krepi njihovo kulturno identiteto ter zagotavlja enakopraven dostop do informacij in učinkovito sporazumevanje v sodobnem digitalnem okolju.

Abstract: Strumenti linguistici per lo sloveno in Italia: l'assistente linguistico Loris 1.1

Il presente contributo prende in esame l'app Loris 1.1, una risorsa linguistica per lo sloveno in Italia concepita come piattaforma online per la revisione linguistica automatica di testi prodotti in sloveno. L'app si rivolge in prima istanza ai parlanti e alle parlanti di lingua slovena nell'area di contatto linguistico tra lo sloveno e l'italiano. Nata dalla collaborazione tra l'Istituto sloveno di ricerche (SLORI) e l'Ufficio centrale per la lingua slovena della Regione Autonoma Friuli Venezia Giulia, si appoggia anche agli strumenti del Centro per le risorse e tecnologie linguistiche dell'Università di Lubiana (CJVT).

Loris 1.1 si avvale di ampie basi di dati, che comprendono dati di tipo prescrittivo come le regole ortografiche (tratte dall'edizione 2001 dello *Slovenski pravopis*, opera di riferimento per l'ortografia slovena), un dizionario morfologico (*Sloleks*), le funzionalità del CJVT relative a punteggiatura e spell-check

neurale, un repertorio di sinonimi per alcuni lessemi di uso frequente (tratto dalla versione 1.0 del Dizionario dei sinonimi dello sloveno contemporaneo, CJVT), nonché una funzionalità per la correzione delle desinenze di casi e numero (duale vs plurale), cui seguiranno in futuro ulteriori sostanziali upgrade e un ricorso sempre più pervasivo all'intelligenza artificiale. Allo stato attuale, la versione 1.1 vede integrate le funzionalità per il riconoscimento degli errori più ricorrenti dovuti al contatto linguistico, come gli usi errati o impropri dei paronimi (i cosiddetti “falsi amici”), le specificità dei toponimi (nomi di oggetti geografici) e i calchi (traduzioni di sintagmi). Loris 1.1 consente agli e alle utenti di ricevere un feedback pertinente sulle divergenze dalla lingua slovena standard, soprattutto dal punto di vista dei fenomeni linguistici marcati in diatopia con riferimento alla realtà degli sloveni in Italia.

Il perfezionamento di Loris rappresenta un importante passo avanti nel sostenere le esigenze linguistiche e culturali della minoranza slovena in Italia, perché globalizzazione e digitalizzazione non fanno che accrescere la pressione esercitata dalle lingue a maggiore diffusione. Questa app mette gli e le utenti nelle condizioni di far fronte a simili sfide nella misura in cui, segnalando gli scostamenti dallo standard linguistico, aiuta ad acquisire consapevolezza e a rafforzare le competenze linguistiche dei e delle parlanti che vivono nell'area di confine.

Uno degli aspetti caratterizzanti di Loris è dunque il fatto di integrare diverse fonti di dati, da quelli raccolti e preparati in modo del tutto manuale fino ai modelli di machine learning del CJVT. Nonostante per lo sloveno siano disponibili numerose risorse linguistiche, altrettante si presentano tuttora slegate e frammentate, lasciando gli utenti in balia di lunghi elenchi di link e segnalibri per fonti che, di base, risultano complementari. Loris, in tal senso, si adopera per integrare tutte queste fonti e quante più risorse possibili in un'unica piattaforma.

L'app di consulenza linguistica Loris crea i presupposti perché i parlanti e le parlanti di sloveno possano inserirsi consapevolmente nella comunicazione in forma scritta, rafforzando l'identità culturale e garantendo non solo parità di accesso alle informazioni, ma anche un flusso di comunicazione efficace in un ambiente digitale moderno.

Summary: Language Tools for Slovene in Italy: The LORIS

1.1 Language Assistant

This chapter discusses the Loris 1.1 language tool for Slovene in Italy, designed as an online platform for automatic review of texts in Slovene. The

tool is primarily intended for Slovene speakers in the area of linguistic contact between Slovene and Italian. It was created as a collaboration between the Slovene Research Institute (SLORI) and the Central Office for the Slovene Language of the Autonomous Province of Friuli-Venezia Giulia, and it is based on the tools of the Center for Language Technologies of the University of Ljubljana (CJVT).

The tool is based on large databases that include normative data such as orthographic norms (from *Slovenski pravopis/Slovene Orthography 2001*), a morphological lexicon (Slolex), CJVT commas and a neural spell checker, synonyms for some of the most common lexemes (from the thesaurus *Sopomenke 1.0*), and a correction system for declensions and number (dual -- plural); in the future, further extensive upgrades are planned, including a more extensive reliance on artificial intelligence. Loris version 1.1 includes built-in functionalities to recognize common errors arising from language contact, such as incorrect use of paronyms (so-called false friends), calques (translations of phrases) and toponyms (place names). It provides users with access to relevant feedback regarding deviations from standard Slovene, especially from the point of view of locally marked linguistic phenomena.

The development of Loris represents an important advance in supporting the linguistic and cultural needs of the Slovene minority in Italy, as globalization and digitalization increase the linguistic pressure from majority languages. Loris enables users to face these kinds of challenges, since by drawing attention to language deviations, it helps to raise awareness and strengthen the linguistic competence of border area speakers.

One of the main characteristics of Loris is therefore the inclusion of various data sources, from completely manually collected and prepared information to machine-learned CJVT models. Although there are many language tools available for Slovene, a considerable number are still disjointed and fragmented, so we are subjected to long lists of links and bookmarks for otherwise complementary resources. Loris strives to combine these resources and integrate as many tools as possible into a single platform.

The Loris language consultant enables native speakers to confidently engage in written communication, which strengthens their cultural identity and ensures equal access to information and effective communication in the modern digital environment.

O avtorjih

Luka Bezgovšek je po končani srednji šoli nadaljeval študijsko pot na Univerzi v Ljubljani in si pridobil naziv inženirja multimedijskih komunikacij. Nato je diplomiral še na Univerzi v Mariboru, kjer si je pridobil naziv varstvoslovca informacijske varnosti.

V preteklosti je bil zaposlen kot programer in je v ekipi oblikoval in razvijal funkcionalnost spletnih strani. Karierno pot zdaj nadaljuje kot tehnični sodelavec Centra za jezikovne vire in tehnologije na Fakulteti za računalništvo in informatiko ter tudi kot samostojni podjetnik. Trenutno sodeluje predvsem s Slovenskim raziskovalnim inštitutom, kjer se uspešno spopada z nalogami razvijanja izvirnih jezikovnih virov in orodij, prilagojenih za slovensko manjšino v Italiji.

Na začetku poklicne kariere se je poglavitno ukvarjal s t. i. »front-endom«. Pristno zanimanje za izbrano strokovno pot ga je privedlo v raziskovanje drugih delov spletnega ekosistema. V zadnjem času vse več pozornosti posveča »back-end« sistemom, pri čemer pretvarja podatke v spletno obliko in optimizira sisteme za njihov prikaz. Prosti čas namenja predvsem izpopolnjevanju in pridobivanju novega znanja na področju učenja različnih programskih jezikov.

David Bordon je mladi raziskovalec (MR) na Filozofski fakulteti Univerze v Ljubljani. Primarno deluje na področjih digitalnega jezikoslovja, razvoja jezikovnotehnoloških orodij in sociolingvističnih analiz. Zagotavlja tehnično pomoč pri izvedbi spletnih konferenc, prenašanih dogodkov ter dogodkov v živo in je organizator glasbenega festivala in koncertov.

S SLORI-jem sodeluje že od leta 2019, in sicer kot del tehnične ekipe za razvoj jezikovnotehnoloških orodij: razvoj jezikovnega svetovalca LORIS, izdelava korpusa zamejskih besedil, razvoj arhitekture drugih jezikovnih orodij, priprava izobraževalnih in predstavitevnihih videoposnetkov.

Trenutno se ukvarja s pisanjem doktorske raziskave o slovenski parlamentarni diskurzni skupnosti z uporabo metod korpusnega jezikoslovja in sodobnih računalniških pristopov. Sodeloval je pri projektu RSDO, in sicer pri razvoju jezikovnih orodij, gradnji podatkovnih baz ter kot skrbnik spletnega portala. Aktivno sodeluje tudi s konzorcijem evropske raziskovalne infrastrukture CLARIN ERIC. Njegov prispevek z naslovom Govoriš nevronska? Kako ljudje razumemo jezik sodobnih strojnih prevajalnikov je bil izbran za najboljši

študentski prispevek na konferenci Jezikovne tehnologije in digitalna humanistika leta 2022.

Jasmin Franza je študirala medjezikovno posredovanje na Oddelku za prevajalstvo Filozofske fakultete Univerze v Ljubljani. Po diplomi je na matičnem oddelku tudi magistrirala iz prevajanja v kombinaciji slovenščina-angleščina-italijanščina. Med letoma 2018 in 2022 je bila zaposlena kot mlada raziskovalka na Oddelku za prevajalstvo, kjer je začela pisati doktorsko disertacijo na temo prepoznavanja in analize čustev v sovrážnem govoru na družbenih omrežjih. Pri raziskovanju posveča posebno pozornost korpusnemu jezikoslovju, sociolingvistiki in jezikovnotehnološkimi orodjem. Ukvarja se s prevajanjem, lektoriranjem in raziskovanjem. S Slovenskim raziskovalnim inštitutom sodeluje od leta 2017. Osredotoča se na razvoj, ohranjanje in promocijo slovenskega jezika na območju slovensko-italijanskega kulturnega stika. Sodeluje pri evropskih projektih ter razvoju orodij in glosarjev, ki uporabnikom omogočajo boljšo rabo slovenščine na območju na stičišču slovenskega in italijanskega kulturnega prostora in v digitalnem okolju.

Damjan Popič je študiral prevajalstvo na Oddelku za prevajalstvo Filozofske fakultete v Ljubljani. Po diplomi leta 2009 je na matičnem oddelku tudi doktoriral (2014), zdaj pa je tam zaposlen kot docent. Poučuje predmete, vezane na sistem slovenskega jezika, jezikovni standard in sodobne tehnologije.

Ukvarja se predvsem z raziskavami s področja rabe in analize naravnega jezika; sodeluje pri različnih projektih in v skupinah, ki razvijajo napredne tehnološke rešitve za opolnomočenje govork in govorcev slovenskega jezika.

Je avtor več znanstvenih in strokovnih publikacij, učbenikov in učnih gradiv ter korpusov in orodij za slovenski jezik. V okviru SLORI-ja je zadolžen za tehnološko pomoč pri jeziko(slo)vnih raziskavah ter za pripravo jezikovnih tehnologij za tvorjenje besedil v slovenskem jeziku.