

# Šolar 3.0: Korpus šolskih pisnih besedil

Poročilo projekta Razvoj slovenščine v digitalnem okolju

*Aktivnost DS1.6*

**Avtorji: Špela Arhar Holdt<sup>◦,□</sup>, Iztok Košem<sup>◦</sup>, Eva Pori<sup>□</sup>, Tina Munda<sup>◦</sup>, Mojca Stritar Kučuk<sup>□</sup>,  
Ines Voršič<sup>▽</sup>, Tomaž Petek<sup>◇</sup>, Polonca Šek Mertük<sup>▽</sup>,  
Luka Krsnik<sup>◦</sup>**

- Fakulteta za računalništvo in informatiko Univerze v Ljubljani
  - Filozofska fakulteta Univerze v Ljubljani
  - ▽ Pedagoška fakulteta Univerze v Mariboru
  - ◇ Pedagoška fakulteta Univerze v Ljubljani

Ljubljana: Center za jezikovne vire in tehnologije, Univerza v Ljubljani, 2023

## Vsebina

1	Špela Arhar Holdt: Projektni okvir	1
2	Špela Arhar Holdt, Tina Munda: Opis projektnih aktivnosti	1
3	Špela Arhar Holdt, Iztok Kosem, Tina Munda, Ines Voršič: Korpus Šolar 3.0	4
	3.1 Sestava korpusa Šolar 3.0	5
	3.2 Metodologija označevanja jezikovnih popravkov	6
	3.3 Popravljanje napak	8
4	Špela Arhar Holdt, Eva Pori, Iztok Kosem, Tomaž Petek, Polonca Šek Mertük: Kontinuirano zbiranje korpusnega gradiva	9
	4.1. Pravne rešitve za dolgoročno zbiranje gradiva	9
	4.2 Portal za preprosto in kontinuirano korpusno gradnjo	10
	4.2.1 Motivacija za učiteljsko sodelovanje	10
	4.2.2 Pričetek dela s portalom	10
	4.2.3 Učiteljska evalvacija portala	11
	4.2.4 Izgled in funkcionalnosti portala	13
5	Špela Arhar Holdt, Iztok Kosem, Mojca Stritar Kučuk, Luka Krsnik, Tina Munda: Program Svala in novi format	19
	5.1 Program Svala	19
	5.1.1 Izvor programa Svala	19
	5.1.2 Izgled in funkcionalnosti programa CJVT-Svala	20
	5.1.3 Evalvacija programa	23
	5.2 Novi korpusni format	23
	5.3 Konkordančnik	24
6	Špela Arhar Holdt, Iztok Kosem: Prioritete za nadaljnji razvoj	26
7	Priloge	28
8	Literatura	28

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

## 1 Špela Arhar Holdt: Projektni okvir

Poročilo oz. kazalnik *Šolar 3.0: Korpus šolskih pisnih besedil* je nastalo pod okriljem projekta Razvoj slovenščine v digitalnem okolju, ki sta ga med leti 2020 in 2023 sofinancirali Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj (Operacija se je izvajala v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014–2020). Cilj projekta je bil zadovoljiti potrebe po računalniških izdelkih in storitvah s področja jezikovnih tehnologij za slovenski jezik za raziskovalne organizacije, podjetja in širšo javnost. Spletna stran projekta z dostopom do rezultatov: <https://slovenscina.eu/>.

Kazalnik se umešča v prvi projektni delovni sklop z naslovom *Jezikovni viri*. Namen delovnega sklopa je bil nadgraditi slovenske besedilne korpuse in leksikon besednih oblik. Prenovili smo učne množice in postopke za strojno označevanje sodobne slovenščine. Rezultat so osveženi in povečani jezikovni viri, ki so na voljo tako uporabniški skupnosti kot za strojno rabo. Z razvitimi postopki in orodji bo posodabljanje slovenskih korpusov v prihodnosti hitrejše in preprostejše.

Med besedilne korpuse, ki smo se jim na projektu posebej posvetili, sodi tudi specializiran korpus, ki ga opisujemo v tem kazalniku, tj. korpus šolskih izdelkov, ki so nastali na slovenskih osnovnih in srednjih šolah Šolar. Korpus ponuja pomemben uvid v jezikovno rabo določenega tipa, poseben pa je tudi zato, ker zahteva dodatne korake pri pripravi korpusnega gradiva: vsebuje namreč informacije o učiteljskih popravkih jezikovnih napak. Da bi s pripravo in nadgradnjo tovrstnih korpusov lahko učinkovito nadaljevali, je bilo treba vzpostaviti protokole za kontinuirano zbiranje in procesiranje korpusnega gradiva, razvili pa smo tudi nova orodja za ročno označevanje in kategoriziranje jezikovnih popravkov. Nova orodja so odprto dostopna za nadaljnjo rabo, že med projektom pa smo jih uporabili za izboljšavo in dopolnitev obstoječih korpusov.

## 2 Špela Arhar Holdt, Tina Munda: Opis projektnih aktivnosti

Eden izmed ciljev projekta RSDO je bil pripraviti načrt za kontinuirano nadgrajevanje korpusov usvajanja slovenščine ter vzpostaviti delotoke in orodja za redno zbiranje in procesiranje besedil.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014–2020.

Aktivnost se je posvečala korpusom usvajanja slovenščine, ki v primerjavi z ostalimi korpusnimi viri vsebujejo oznake jezikovnih napak oz. popravkov. Tako označena besedila so izjemnega pomena, saj omogočajo empirični uvid v opismenjevanje šolajoče se populacije, pripravo didaktičnih gradiv ter orodij, ki opismenjevanje in pismenost podpirajo in razvijajo. V projektu smo se koordinirano posvečali tako usvajanju slovenščine kot prvega kot tudi učenju slovenščine kot drugega/tujega jezika. V povezavi s prvim je bil pred projektom RSDO na voljo korpus Šolar, ki v različici 2.0 vsebuje skoraj 5.500 besedil, ki so jih učenci starosti 13–19 let samostojno tvorili pri pouku (Kosem et al. 2016). 2.094 besedil vsebuje tudi učiteljske popravke jezikovnih napak, ki so kategorizirani s trinivojskim sistemom po jezikovnih ravneh glede na tip jezikovnega problema. Ta korpus smo nadgradili, kot opisujemo v tem kazalniku. Na drugi strani smo pripravili korpus KOST (korpus slovenščine kot tujega in drugega jezika), ki vsebuje besedila, ki jih pišejo tujci v okviru programov Leta plus in različnih dejavnosti Centra za slovenščino kot drugi in tuji jezik (Stritar Kučuk 2022). Tudi v ta besedila so bili ročno pripisani in kategorizirani jezikovni popravki. V tem kazalniku se posvečamo korpusu Šolar 3.0, KOST pa je opisan v ločenem kazalniku.

Za učinkovito pripravo in nadgradnjo korpusov z jezikovnimi popravki smo vzpostavili ustrezne delotoke in razvili dve novi orodji. V prvem koraku smo pripravili spletno mesto z informacijami, pogodbami in repozitorijem za zbiranje korpusnega gradiva, kar je učiteljem in lektorjem omogočilo enostavno sodelovanje. Na ravni procesiranja pridobljenih korpusnih besedil smo izbrali uporabniku prijazno in zmogljivo orodje – za slovenščino smo lokalizirali in nadgradili orodje Svala, ki omogoča (a) transkripcijo besedil, ki so napisana na roko, (b) označevanje jezikovnih popravkov, ki je bilo doslej najbolj zamudno in programsko izredno slabo podprto in (c) anonimizacijo. Ob razvoju programa smo na novo definirali format korpusnih besedil in ga dali na voljo tudi v ostalih formatih, da je kompatibilen z ostalimi slovenskimi korpusi.

Na projektni aktivnosti je sodelovalo več raziskovalcev z Univerze v Ljubljani in Univerze v Mariboru, pa tudi zunanjih sodelavcev, kot so člani ekipe, ki je korpus Šolar razvijala predhodno, programer, razvijalec čelnega sistema, oblikovalci, učiteljski evalvatorji, lektorji, študenti. Avtorstvo je navedeno pri vsakem posameznem projektnem rezultatu.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Delo je potekalo prek celotnega časa trajanja projekta po naslednjih korakih:

- Priprava spletnega mesta, na katerem lahko učitelji dobijo vse informacije glede sodelovanja pri gradnji korpusa, vključno s pogodbami za sodelovanje in možnostjo, da zbrana besedila enostavno in ustrezno dokumentirano naložijo v repozitorij. Stran je namenjena tudi vzpostavljanju skupnosti sodelujočih besedilodajalcev, npr. s prikazovanjem sodelujočih šol, statistik zbranega gradiva itd.
- Aktivacija mreže učiteljev, lektorjev in drugih, ki so sodelovali pri testnem zbiranju gradiva v času trajanja projekta. Sodelovanje je služilo za evalvacijo delotokov in preglednosti, razumljivosti ter prijaznosti spletnega mesta. V sodelovanje smo vključili 13 šol, 10 iz vzhodne in 3 iz zahodne regije, ter 24 sodelavcev s programov, povezanih s slovenščino kot drugim oz. tujim jezikom.
- Izbira programa za anonimizacijo, transkripcijo ročno napisanih besedil in označevanje ter kategorizacijo jezikovnih napak. Izbrali smo odprto dostopen program Svala, ki omogoča pregledno in učinkovito delo s podatki. Uporabniška prijaznost je zagotovljena tudi na ravni uvoza in izvoza podatkov. Projektno delo je zajemalo lokalizacijo programa, ki smo ga prilagodili tako, da vsebuje predpripravljene nabore kategorij oznak za korpusa Šolar in KOST. Skupaj z razvojem programa smo nadgradili format predhodnega korpusa Šolar 2.0, da je združljiv z ostalimi slovenskimi korpusi in označevalnim programom. Za testiranje programa smo označili besedila obeh korpusov v količini 2.500 besed in glede na izkušnjo označevalcev implementirali dodatne popravke. Uporabnost orodja smo preizkusili na skupini študentov pedagoških smeri in ocenili, do kolikšne mere je mogoče v delotoke vključiti študentsko delo.
- Priprava obstoječih korpusnih gradiv za vključitev v konkordančnike. Obstoječi korpus Šolar 2.0 smo nadgradili v Šolar 3.0, ki je po koncu projekta dostopen v konkordančnikih in s tem na voljo za analize. Pri urejanju formata so bile odpravljene številne težave, ki smo jih identificirali v predhodnih projektnih korpusne gradnje. Opravljena je bila tudi nadgradnja pripisovanja dvojnih oznak napakam in nadgradnja na ravni segmentacije napak. Za korpuse, ki vsebujejo jezikovne popravke, smo razvili specializirani format XML TEI, ki je skladen z ostalimi slovenskimi besedilnimi korpusi in povezljiv s formatom JSON orodja Svala.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

## REZULTATI AKTIVNOSTI

- Spletno mesto z informacijami za besedilodajalce in rešitvami za dolgoročno zbiranje gradiva: <https://rsdo.slovenscina.eu/zbiranje-besedil-za-korpus-solar>.
- Program za označevanje jezikovnih napak oz. popravkov v korpusih usvajanja slovenščine: <https://orodja.cjvt.si/svala/>.
- Korpus Šolar 3.0 kot baza na repozitoriju CLARIN.SI: <http://hdl.handle.net/11356/1589> in vključen v konkordančnike na CLARIN.SI:
  - KonText:
    - korpus učenci: [https://www.clarin.si/kontext/query?corpname=solar30\\_orig](https://www.clarin.si/kontext/query?corpname=solar30_orig)
    - korpus učitelji: [https://www.clarin.si/kontext/query?corpname=solar30\\_corr](https://www.clarin.si/kontext/query?corpname=solar30_corr)
  - NoSketch Engine Bonito:
    - korpus učenci: [https://www.clarin.si/noske/sl.cgi/first?corpname=solar30\\_orig&reload=1&iquery=](https://www.clarin.si/noske/sl.cgi/first?corpname=solar30_orig&reload=1&iquery=)
    - korpus učitelji: [https://www.clarin.si/noske/sl.cgi/first?corpname=solar30\\_corr&reload=1&iquery=](https://www.clarin.si/noske/sl.cgi/first?corpname=solar30_corr&reload=1&iquery=)
  - NoSketch Engine Crystal:
    - korpus učenci: [https://www.clarin.si/ske/#dashboard?corpname=solar30\\_orig](https://www.clarin.si/ske/#dashboard?corpname=solar30_orig)
    - korpus učitelji: [https://www.clarin.si/ske/#dashboard?corpname=solar30\\_corr](https://www.clarin.si/ske/#dashboard?corpname=solar30_corr)
- (Pričujoče) specifikacije za nadgradnjo korpusa Šolar z opredelitvijo želene vsebine, pravnih vprašanj, motivacije besedilodajalcev za sodelovanje pri korpusni gradnji, metodologijo in delotoki transkribiranja, označevanja ter kategoriziranja jezikovnih napak.

### 3 Špela Arhar Holdt, Iztok Kosem, Tina Munda, Ines Voršič: Korpus Šolar 3.0

V tem razdelku predstavljamo sestavo korpusa Šolar 3.0, sistem označevanja jezikovnih popravkov in popravljanje gnezdenih popravkov.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

### 3.1 Sestava korpusa Šolar 3.0

Na projektu je bila pripravljena različica 3.0 korpusa Šolar<sup>1</sup>. Sestavlja ga 5.485 pisnih izdelkov, ki so jih pri pouku samostojno tvorili učenci slovenskih osnovnih in srednjih šol. Večinoma gre za besedila učencev 7.–9. razreda osnovne šole – vključen pa je tudi manjši vzorec besedil iz 6. razreda – in dijakov vseh letnikov srednje šole. S korpusom torej spremljamo razvoj pisne kompetence šolajoče se populacije starosti 13–19 let.

Vsako besedilo je opremljeno z metapodatki, in sicer: vrsta šole (osnovna ali srednja), predmet, pri katerem je bilo besedilo tvorjeno, razred oz. letnik tvorca besedila, regija, v katero je šola umeščena, in datum nastanka besedila. Po tipologiji v korpusu prevladujejo šolski spisi oz. eseji, sestavljajo pa ga še druga besedila, tvorjena med poukom, kot so na primer obnove, opisi, praktična besedila idr.

Del korpusa (2.094 besedil) je označen z učiteljskimi popravki po sistemu oznak, ki ga podrobneje opisujemo v nadaljevanju tega razdelka. Popravki učiteljev so del izvornih pisnih izdelkov učencev, kar pomeni, da odsevajo realno sliko popravljanja šolskih spisov v izobraževalnem procesu.

Korpus je bil jezikoslovno označen s cevovodom CLASSLA v1.1.1<sup>2</sup>, in to na ravneh tokenizacije, stavčne segmentacije, lematizacije, oblikoskladenjskih oznak po sistemu MULTEXT-East v6<sup>3</sup>, odvisnostne skladnje po sistemu JOS<sup>4</sup> in imenskih entitet (NER). Oznake na nivoju odvisnostne skladnje in imenskih entitet predstavljajo novost v primerjavi z različico 2.0, izboljšana pa je tudi natančnost oznak na ostalih nivojih, saj so bile pripisane z izboljšanim označevalnim orodjem.

Trenutna različica korpusa ima v primerjavi s prejšnjo tudi izboljšano strukturo. Verzijo 3.0 sestavljata različica z nepopravljenimi besedili učencev (Šolar - učenci) in različica s popravljenimi besedili (Šolar - učitelji), datoteki pa povezujejo medbesedilne povezave na ravni pojavnice z oznako popravka. Korpus je na voljo v XML TEI, ki smo ga razvili posebej za korpuse, ki vsebujejo jezikovne popravke. To omogoča enostavno primerjavo korpusnih podatkov z drugimi slovenskimi besedilnimi korpusi in je pomemben korak naprej od prejšnje različice; slednja je sicer bila na voljo v formatu XML, ki pa se je razlikoval od

---

<sup>1</sup> <http://hdl.handle.net/11356/1589>

<sup>2</sup> <https://github.com/clarinsi/classla/>

<sup>3</sup> <http://nl.ijs.si/ME/V6/msd/html/msd-sl.html>

<sup>4</sup> <http://nl.ijs.si/jos/bib/jos-skladnja-navodila.pdf>

XML TEI, ki je v vmesnem času v slovenskem prostoru postala standard in referenčna točka za pripravo in pretvorbo korpusnih datotek. Del nadgradnje je še ročna kuracija tipov popravkov v približno 350 besedilih, pomemben doprinos pa je tudi več različnih formatov korpusa, kar omogoča njegovo širšo uporabo (gl. razdelek 5.3).

### 3.2 Metodologija označevanja jezikovnih popravkov

V korpusu usvajanja tujega in maternega jezika so pogosto vključene oznake jezikovnih težav in popravkov (Granger 2008). Tako označeni korpusi so dragocena osnova za raziskovanje na področju didaktike, pa tudi za pripravo učnih gradiv, vaj, testov in navsezadnje za pripravo učnih množic za naravno procesiranje jezika (Arhar Holdt et al. 2022: 24). Pri procesu odločanja, v katero kategorijo popravkov spada določena težava, so nepogrešljive jasne smernice. Za korpus Šolar 3.0 smo uporabili sistem oznak, ki je bil razvit v različici korpusa 2.0, ki smo jih na projektu RSDO dodatno uredili in nadgradili<sup>5</sup>. Pomemben vidik pripisovanja oznak je konsistentnost, ki smo jo skušali dosegati s pregledom že označenega gradiva, pri čemer smo potencialne nedoslednosti s pomočjo novih orodij tudi učinkovito popravili. V nadaljevanju predstavljamo osnovno kategorizacijo oznak za jezikovne popravke. Glavnih kategorij je sedem, te pa se, odvisno od kompleksnosti, delijo naprej na več nivojev podkategorij.

**Črkovanje:** na to raven uvrščamo učiteljske popravke, ki se nanašajo na zapis glasu ali glasovnega sklopa v besedi. Lahko gre za odvečne, izpuščene ali zamenjane glasove (*polen\** [laži] namesto *poln* [laži]; [*je*] *vrjel\** namesto [*je*] *verjel*; *vstrajen\** namesto *vztrajen*) ali črkovne sklope (*zberejejo\** namesto *zberejo*; *sprej\** namesto *sprejel*; *zastojn\** namesto *zastonj*), pa tudi za vzglasje besed na u- oz. v- (*Vsedla\** namesto *Usedla*; *uzamejo\** namesto *vzamejo*) in variantne predloge (*k\** [*koncu*] namesto *h*).

**Oblika:** na ravni oblike označujemo (a) težave na ravni izbire sklona, števila, spola, recimo [*o dekletu, ki je*] *zanosila\** namesto *zanosilo*, in kategorij drugih besednih vrst, (b) popravke besednih oblik, ki niso del standardnih paradigem, recimo *poprimiti\** namesto *poprijeti* in (c) dodatne oznake, ki jih pripisujemo le v primeru, da osnovna vsebinska oznaka popravka že obstaja, dodatna oznaka pa omogoča združevanje podatkov po drugem kriteriju ali pripis dodatne (poljubne) informacije. Tukaj so

---

<sup>5</sup> <https://wiki.cjvt.si/books/razvojni-korpus-solar/page/oznacevalne-smernice>



zajete variantne oblike besede, recimo obliki leksema grad za 2. sklon ednine *grada* in *gradu*, pri čemer sta legitimni obe, a je ena nevtralnejša, na kar želi učitelj učenca opozoriti.

**Besedišče:** na to raven umeščamo popravke besedišča, kar vključuje predvsem menjavo ene besede z drugo, pri čemer se lahko besedna vrsta in/ali besednozvezna struktura ohrani ali spremeni. Podkategorije so razdeljene na besedne vrste, pri samostalniku recimo ena izmed oznak obeležuje napačno lastno ime (*Lovrenc\** namesto *Lovro [Kuhar]*), pri glagolu menjavo z drugim glagolom (*[ne bi] moral\** [*opisati*] namesto *mogel*) ipd., sledi podkategorija menjav prek meja besedne vrste, zadnja pa tudi tu obeležuje dodatne oznake, v tej kategoriji popravkov so to zaznamovane besede, recimo *faks\** namesto *fakulteta*.

**Skladnja:** na tej ravni označujemo popravke, ki posegajo na raven besednozvezne, stavčne in povedne sklanje, npr. popravke besednega reda (*[prepričan je da] generalove ukaze je potrebno\** [*upoštevati*] namesto *je generalove ukaze potrebno*), skladenjskih struktur (*truplo matere\** namesto *materino truplo*), medstavčnih razmerij (*Herod je nekega dne priredil slavje. Salomi slavje\** ni bilo preveč po godu. namesto *Herod je nekega dne priredil slavje, ki Salomi ni bilo preveč po godu.*) itd. Podkategorija dodatne oznake tukaj zaobjema pleonazme, recimo *[Ko se je] vrnila nazaj\** namesto *[Ko se je] vrnila*, odvečne, pomensko prazne ali vsebinsko napačne dele.

**Zapis:** na ravni zapisa označujemo predvsem popravke začetnic (*[v] Nemškem\** [*jeziku*] namesto *nemškem*) in pisanja skupaj ali narazen (*Nažalost\** namesto *Na žalost*). V korpusu so označena tudi mesta napačne stave ločil, kjer prevladuje raba vejice. Skupina ločil je edina, ki ni bila v celoti ročno pregledana in kategorizirana, in sicer zaradi razširjenosti pojava.

**Povezani popravki:** v to kategorijo uvrščamo vse primere, ki niso samostojen popravek, ampak so posledica primarnega jezikovnega popravka (povezan popravek je recimo neustrezen sklon odnosnice predloga, pri čemer je napačen predlog primarna težava). Pomembno je, da so ti posegi v besedilo ločeno od dejanskih popravkov učenčevih jezikovnih izbir, da lahko označene podatke ustrezno interpretiramo. Povezani popravki sledijo obstoječi tipologiji, le da del oznake ponazarja, da gre za povezan popravek.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

**Nečitljivi in sumljivi primeri:** sem spadajo primeri, kjer je pri učenčevi napaki ali učiteljskem popravku oznaka za nečitljivo ali anonimizirano lastno ime, recimo *šššmorššš\** namesto *moraš* – tem pripišemo oznako za nečitljivo; in primeri, kjer so popravki nenavadni, kjer recimo sumimo, da je prišlo do napake pri transkripciji ali je popravek enak napaki – tem pripišemo oznako za preverbo.

Uvrščanje popravkov v kategorije ni vedno enoznačno, zato smernice vsebujejo še poglavje Odločanje med mejnimi primeri, kjer so obravnavani težji primeri.

### 3.3 Popravljanje napak

Na projektu smo se posebej posvetili razpletanju gnezdenih popravkov v korpusu Šolar 2.0 oz. razvozlanju težjih primerov – problematičnih odstavkov, tj. tistih z gnezdeno napako. Skupaj smo pregledali 339 spisov z različnim številom daljših in krajših odstavkov z gnezdeno napako, nekateri od spisov so lahko imeli samo en problematičen odstavek, mnogi so jih imeli več.

Delo je potekalo v aplikaciji Svala (gl. razdelek 5.1), in sicer tako, da smo uvozili problematičen primer, pregledali učenčevo besedilo in učiteljski popravek in identificirali, kje je prišlo do označevalne zadrege. Napačno grupirane elemente smo razvezali in jih (v skladu s smernicami) znova ustrezno grupirali. Pri kategorizaciji učiteljskih popravkov sem upoštevala Smernice za označevanje korpusa Šolar 2.0<sup>6</sup>. Napake smo preverjali tudi s pomočjo izvornih digitaliziranih in skeniranih spisov.

Rešljive so bile skorajda vse težave oz. napake, in sicer s spremenjenim segmentiranjem problema. Pri nekaterih popravkih je bilo treba spremeniti besedilo učenca ali učitelja, saj je bila težava v napačno pretipkanem spisu, kar smo ugotavljali s pomočjo skeniranih originalov. Popravljen dokumente (ustrezno označene) smo sproti nalagali na strežnik in jih v zadnjem koraku vključili v prenovljeni korpus Šolar 3.0.

---

<sup>6</sup> <https://wiki.cjvt.si/books/razvojni-korpus-solar/page/oznacevalne-smernice>

## 4 Špela Arhar Holdt, Eva Pori, Iztok Kosem, Tomaž Petek, Polonca Šek Mertük: Kontinuirano zbiranje korpusnega gradiva

### 4.1. Pravne rešitve za dolgoročno zbiranje gradiva

Pomemben korak pri izgradnji korpusa je pravno urejeno zbiranje besedil, saj brez tega zbrano gradivo ne more biti odprto dostopno za nadaljnjo rabo. Za vsa vključena besedila mora obstajati pogodba, ki opredeljuje prenos avtorskih pravic in načine hranjenja in procesiranja besedil. V predhodnih projektih so pravne rešitve zajemale obdobje trajanje specifičnega projekta, npr. točno določeno šolsko leto, kar je onemogočalo kontinuirano in širše zbiranje besedil. Na projektu RSDO smo zato pravne rešitve v sodelovanju s pravniki, specializiranimi za avtorsko pravo, ustrezno nadgradili in pripravili: pogodbo o sodelovanju s šolami, v kateri šola opredeli namen, da bo omogočila zbiranje in posredovanje besedil, in to pod pogojema, da bo učenec (ali zanj zakoniti zastopnik) podpisal pogodbo o prenosu ustreznih avtorskih pravic in da bodo osebni podatki učenca anonimizirani in varovani v skladu z evropsko zakonodajo; in dve različici pogodbe o prenosu pravic z dovoljenjem za uporabo pravic – ena za polnoletne učence, druga pa za mladoletne, ki vsebuje privoljenje staršev oz. zakonitih zastopnikov –, v kateri učenec (ali zanj zakoniti zastopnik) odstopa Univerzi v Ljubljani vse »avtorske pravice, avtorski sorodne pravice in ostale pravice avtorja v skladu z Zakonom o avtorski in sorodnih pravicah (ZASP) na besedilih, ki jih je v sklopu šolanja napisal/-a pri pouku« (Pogodba o prenosu pravic in dovoljenje za uporabo pravic) za namen izgradnje korpusa Šolar. Podpisane pogodbe (dva izvoda pogodbe s šolo in dva izvoda pogodbe z vsakim avtorjem oz. avtorico šolskih besedil oz. njegovim zakonitim zastopnikom) učitelj oz. učiteljica pred pričetkom sodelovanja pri zbiranju pošlje raziskovalni enoti, ki gradi korpus, kjer jih podpišemo še druga stranka in po en izvod vrne na šolo. Na projektu smo raziskali možnosti digitalnega podpisovanja pogodb, vendar se izkaže, da zlasti učenci in starši v trenutku projekta še niso opremljeni za takšno rešitev, zato šolski sistem preferira fizično podpisovanje. Vse tri pogodbe so v prilogi tega poročila.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

## 4.2 Portal za preprosto in kontinuirano korpusno gradnjo

### 4.2.1 Motivacija za učiteljsko sodelovanje

Do sedaj je zbiranje besedil za korpus Šolar potekalo dokaj počasi in nesistematično. Pri zbiranju besedil za prvo različico so učitelji in učiteljice pošiljali fizične kopije besedil učencev, njihova kakovost pa se je razlikovala glede na uporabljen kopirni stroj (Arhar Holdt et al. 2022: 26). Dodatno oviro je predstavljalo to, da so bili kopirani dokumenti črno-beli, kar je velikokrat oteževalo razlikovanje med popravki, ki jih je opravil učitelj, in tistimi, ki je zabeležil učenec sam (Arhar Holdt et al. 2017: 98). V postopku zbiranja besedil za drugo različico korpusa smo to težavo naslovili – prešli smo na zbiranje skeniranih besedil, po možnosti barvnih, in s tem na posredovanje PDF-datotek prek spleta (Arhar Holdt et al. 2022: 26). Da bi oddajanje besedil – in vse korake, potrebne za sodelovanje – olajšali tako skupnosti sodelujočih učiteljev kot raziskovalcem, ki besedila za korpus pripravljajo, smo v projektu RSDO razvili portal za avtomatsko zbiranje pisnih izdelkov učencev. Portal, ki ga predstavljamo v nadaljevanju, pomeni pomemben korak naprej v metodologiji gradnje korpusa usvajanja slovenščine.

Pri razvoju portala je bila v ospredju želja, da bo njegova uporaba enostavna in intuitivna, hkrati pa bo vsebovala vse uporabne funkcionalnosti. Uporabniku prijazen vmesnik je osnovni pogoj za sodelovanje čim večjega števila učiteljev, katerih večšina dela z računalnikom in s tem odnos do njega se lahko močno razlikuje. Učiteljsko skupnost smo za sodelovanje želeli dodatno motivirati s pridobitvijo točk za napredovanje v nazive. Njihovo sodelovanje v projektu še šteje kot dodatno strokovno delo in se skladno s *Pravilnikom o napredovanju zaposlenih na področju vzgoje in izobraževanja v nazive*<sup>7</sup> ovrednoti z eno točko za sodelovanje in z dvema za koordiniranje/vodenje, in to za obdobje enega šolskega leta. Dodatna motivacija je možnost, da prek portala lahko sodeluje ekipa učiteljev z iste šole in da lahko sproti spremljajo napredek pri zbiranju besedil, tudi primerjalno glede na druge šole v regiji.

### 4.2.2 Pričetek dela s portalom

Pri prvi uporabi portala za zbiranje besedil se mora uporabnik registrirati. Na spletnem mestu <https://zbiranje.cjvt.si/solar/login/> se pod razdelkom s prijavo (za že obstoječe uporabnike) nahaja

---

<sup>7</sup> Uradni list RS, št. 54/02, 123/08, 44/09, 18/10, 113/20  
<http://www.pisrs.si/Pis.web/pregledPredpisa?id=PRAV4272>.

gumb 'Registracija', ki pelje na stran za registracijo. Tam uporabnik vpiše svoje ime in priimek, naziv institucije, na kateri je zaposlen, e-naslov in geslo za vstop v portal, določi pa še svojo vlogo pri zbiranju, bodisi je to vloga mentorja/-ice bodisi koordinatorja/-ke. Vlogo mentorja izbere, kdor bo zbiral in oddajal šolska besedila, vlogo koordinatorja pa tisti, ki bo poleg zbiranja in oddajanja skrbel še za komunikacijo z vodstvom šole in znotraj skupine mentorjev, če je sodelujočih učiteljev z iste šole več. Učiteljeva vloga se beleži v zavihku Ekipa pod tekoče šolsko leto, administrator institucije pa lahko vloge tudi spremeni. Na izbiro je še 'Druga vloga', ki pokriva opazovalce ali stranske deležnike. Po kliku na 'Registracija' bodoči uporabnik dobi obvestilo, da je podatke vnesel uspešno in da bo o potrditvi uporabniškega računa obveščen po e-pošti.

Ob uspešni prijavi se uporabnik znajde na začetni strani portala. Več o sodelovanju lahko izve s klikom na zavihek 'Več informacij o sodelovanju' desno zgoraj na modri pasici. Preusmerjen je na spletišče <https://rsdo.slovenscina.eu/zbiranje-besedil-za-korpus-solar>, kjer najde kratek opis projekta, spodbudo za sodelovanje, preprosta navodila za sodelovanje po korakih ter pogosta vprašanja in odgovore nanje.

Po odločitvi za sodelovanje sledi pravna ureditev sodelovanja, in sicer med Univerzo v Ljubljani na eni strani ter šolo in učenci na drugi. S šolo se sklene pogodba o sodelovanju, z učenci oz. njihovimi zastopniki pa pogodbo o prenosu ustreznih avtorskih pravic (gl. razdelek 3.2). Tretji korak je zbiranje besedil. Podanih je nekaj smernic: besedila naj bodo dolga vsaj 40 besed in v digitalni obliki – lahko so natipkana in popravljena v Wordu ali optično prebrana z barvnim skenerjem (najmanj zaželen format besedila so fotografije). Zaželeni so besedila, ki vsebujejo jezikovne popravke učitelja, vendar pa to ni nujno za sodelovanje.

#### 4.2.3 Učiteljska evalvacija portala

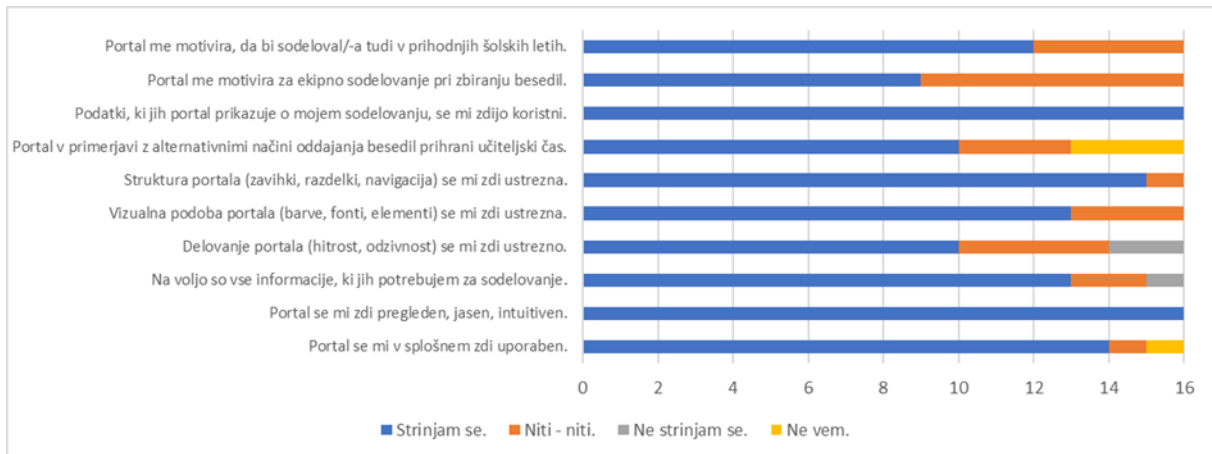
Ob koncu projektne zbiranja besedil v šolskih letih 2020/21 in 2021/22, v katerem je sodelovalo 16 učiteljic in učiteljev s 13 šol, smo med njimi izvedli evalvacijsko anketo. Povratno informacijo o uporabi portala je podalo 16 učiteljev. Velika večina jih meni, da so informacije izčrpne in odgovorijo na vsa njihova vprašanja. Prav tako se je večina evalvatorjev uspešno registrirala, le dva sta imela težave, pri čemer jima je podporna ekipa raziskovalcev priskočila na pomoč. Dva izmed vprašanih sta pri prijavi v portal opozorila na počasnost portala – tu je šlo za dva ločena problema: nalaganje oz. odpiranje

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

spletne strani in čakanje na potrditev registracije. Prvi je bil tehnične narave, ki smo ga naslovili, na drugega pa smo opozorili z obvestilom, da je v primeru neuspešne prijave treba počakati na potrditev uporabniškega računa. Pri določanju metapodatkov za besedila so bili uspešni vsi, le da so štirje naleteli na oviro ali vprašanje. Dva sta opozorila na ime spustnega seznama za izbiro letnika oz. razreda, ki je bil 'Letnik'; to smo upoštevali in razdelek preimenovali v 'Letnik/razred', tako da ime ne izključuje ene ali druge vrste šole. Dva udeleženca ankete pa sta opozorila na tehnično napako, in sicer eden pri možnostih izbire vrste besedila, drugi pa pri zaporednem vnosu; obe težavi smo po evalvaciji odpravili. Pri oddaji besedil večina ni izkusila težav, podan pa je bil smiseln predlog za izboljšavo, tj. da bi bilo mogoče datoteko z besedilom še enkrat odpreti in preveriti, če je prava, ko je že naložena, a še ni oddana.

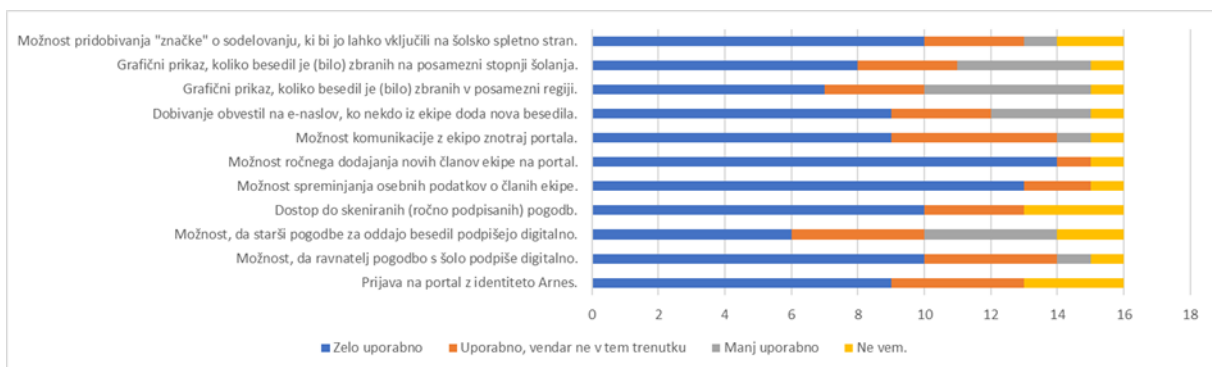
Razdelek 'Zgodovina sodelovanja', kjer se beležijo oddana besedila uporabnika, se zdi uporaben vsem razen enemu vprašanemu, ki razdelka ni uspel pregledati. Razdelek 'Ekipa' se zdi koristen trem četrtinam sodelujočih, dvema razdelka ni uspelo pogledati, dva pa sta pristavila, da razdelek za samostojno delo ni uporaben. Eden sodelujoči je na tem mestu podal komentar, da namen razdelka ni jasen – to je deloma razumljivo, saj v času evalvacije razdelek še ni vseboval vseh načrtovanih podatkov, tj. statistike oddanih besedil na uporabnika in vseh oddanih besedil, po regiji in vrsti šole; to smo do konca projekta implementirali. Naslednja povratna informacija zadeva zavihek 'Upravljanje z inštitucijo', za končno različico portala preimenovan v 'Administracijski meni', ki bo na voljo le administratorju na šoli (za namene ankete so imeli dostop do tega menija vsi sodelujoči). Ta razdelek, ki ima funkcijo dodajanja in odstranjevanja članov ekipe in pregleda nad njihovimi nalaganji, je pregledalo tri četrtine vprašanih – 69 % se je zdel uporaben, medtem ko je eden nasprotnega mnenja, ker sodeluje edini na šoli. Trije razdelka niso pregledali, predvidoma, ker ga niso opazili – glede na to, da gre za skoraj petino vprašanih, bi bilo v prihodnje smiselno administratorje nanj opozoriti. Na splošno so bili sodelujoči s portalom zelo zadovoljni:

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



Slika 1: Splošna ocena portala

Udeležence ankete smo vprašali še za mnenje o dodatnih funkcionalnostih, ki bi jih lahko dodali na portal. Rezultate ponazarja Slika 2:



Slika 2: Ocena dodatnih funkcionalnosti portala

Glede na rezultate in časovni okvir smo na portalu dodali možnost, da učitelj-koordinator ročno doda novega člana ekipe, ostale funkcionalnosti, ki so se izkazale kot potencialno smiselne, bomo upoštevali pri bodočih nadgradnjah vmesnika. (ANALIZA RAZISKAVE: Evalvacija Portala Šolar (RSDO), v prilogi).

#### 4.2.4 Izgled in funkcionalnosti portala

V nadaljevanju predstavljamo končni izgled portala in njegove funkcionalnosti.

**Zavihek Oddaja besedil:** Po uspešni prijavi se znajdete na strani z osrednjo funkcionalnostjo – oddajo besedil.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

### Korpus Šolar

Prosimo, določite podatke o besedilih, ki jih želite oddati, in nato naložite besedila. Če oddajate besedila, ki so nastala v različnih razredih, se razlikujejo glede učiteljskih popravkov in podobno, jih oddajte v ločenih paketih.

ODDAJA BESEDIL	ZGODOVINA SODELOVANJA	EKIPA
<p>REGIJA</p> <p>Celje (CE) ▼</p>		
<p>PROGRAM</p> <p>Osnovnošolski (OŠ) ▼</p>		
<p>PREDMET</p> <p>Slovenščina (SLO) ▼</p>		
<p>RAZRED/LETNIK</p> <p>1 ▼</p>		
<p>VRSTA BESEDILA</p> <p>Esej ali spis (E) ▼</p>		
<p>ŠOLSKO LETO</p> <p>2020/21 ▼</p>		
<p>JEZIKOVNI POPRAVKI</p> <p>Besedilo vsebuje učiteljske popravke in strinjam se z njihovo vključitvijo v korpus. (DD) ▼</p>		

**Slika 3:** Metapodatki za naložena besedila

Na tej strani učitelj določi metapodatke o besedilih, ki jih na koncu strani naloži. Določi regijo, v katero se šola učenca umešča; šolski program (Osnovnošolski (OŠ), Splošna in strokovna gimnazija (SSG), Mednarodni gimnazijski programi (MGP), Zasebne gimnazije (ZG), Nižje poklicno izobraževanje (NPI), Srednje poklicno izobraževanje (SPI), Srednje strokovno izobraževanje (SSI), Poklicno-tehniško izobraževanje (PTI)); predmet, pri katerem so besedila nastala; razred oz. letnik, v katerem so besedila nastala; vrsta besedila (esej ali spis; praktično besedilo (npr. vabila, prošnje ipd. pri pouku slovenščine), napisano za oceno; šolski test; delo v razredu, ne za oceno, pri čemer je treba vtiskati besedilno vrsto);

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



šolsko leto, in polje z informacijo o vsebnosti jezikovnih popravkov skupaj z dovoljenjem za vključitev besedil v korpus. Na dnu strani je okno z gumbom 'Kliknite to polje ali povlecite datoteke vanj', kar pomeni, da s klikom izberete datoteke na svojem računalniku, ali pa datoteke z besedili povlečete in spustite v to okno. Ob oknu je obvestilo s sprejemljivimi datotečnimi formati: txt, csv, pdf, doc, docx, xls, xlsx, ppt, pptx, jpg, jpeg, png, poleg tega pa še opozorilo, da je hkrati možno dodati več datotek, če le vsa ustrezajo istim, zgoraj izpolnjenim metapodatkom (besedila, ki so nastala v različnih razredih, se razlikujejo glede učiteljskih popravkov in podobno, je treba oddati ločeno).

**Slika 4:** Okno za oddajo datotek z besedili.

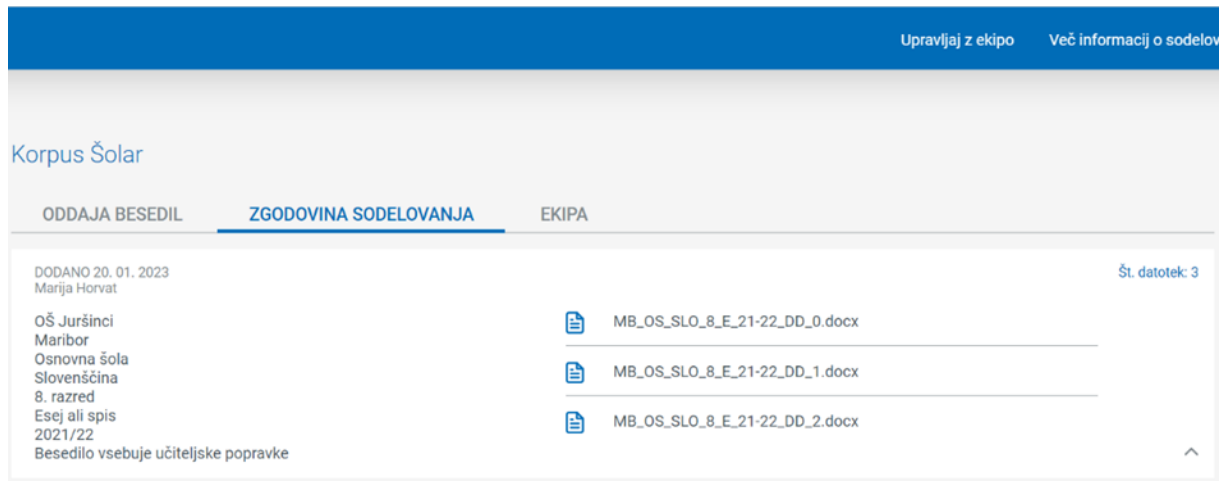
Ko datoteke naložite, jih lahko še enkrat pogledate in se prepričate, da so prave, ali jih zamenjate z drugimi – šele potem jih s klikom na gumb potrdite in oddate. Željo po tej možnosti so izrazili sodelujoči v evalvacijski anketi.

**Slika 5:** Pregled datotek pred potrditvijo in oddajo

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Po uspešni oddaji se izpiše obvestilo o številu oddanih datotekah.

**Zavihek Zgodovina sodelovanja:** V tem zavihku se beleži naložene in oddane datoteke uporabnika. Takoj so vidne osnovne informacije o oddaji, recimo datum oddaje, ime šole, predmet, za podrobne informacije in pogled na oddane datoteke pa lahko polje razširite.

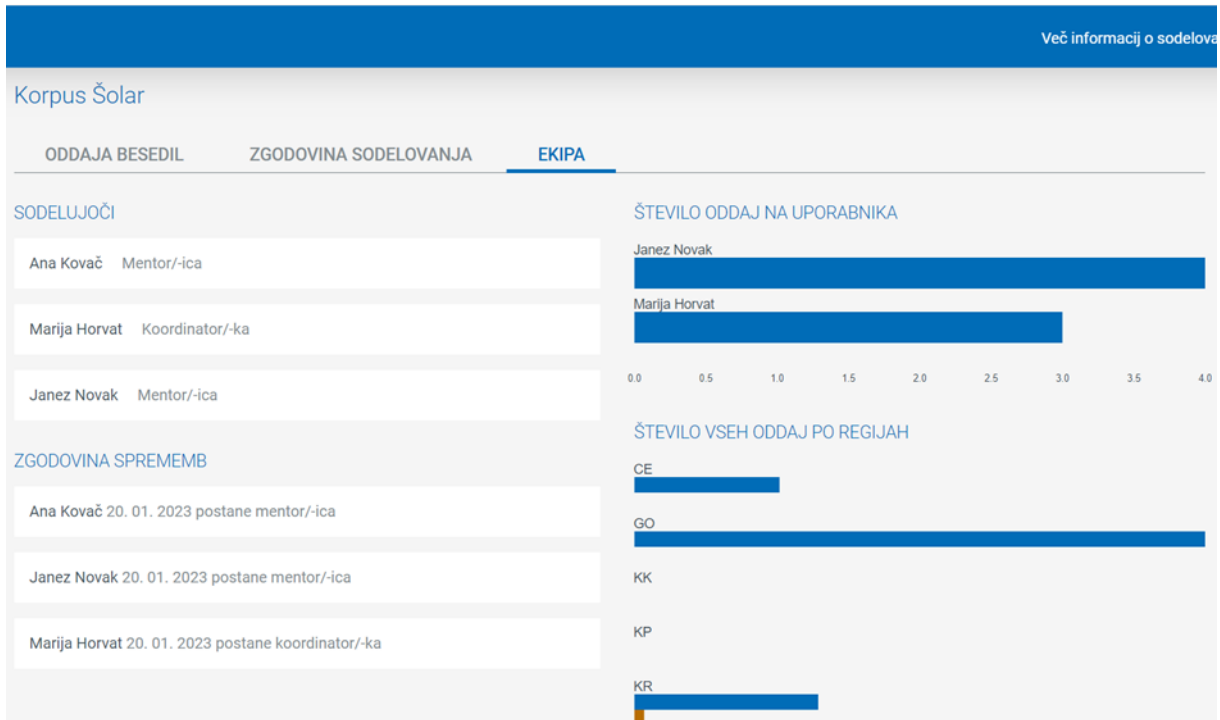


**Slika 6:** Razširjen pogled na paket oddanih besedil.

Imena naloženih datotek se tvorijo iz kod vseh izbranih metapodatkov o besedilih.

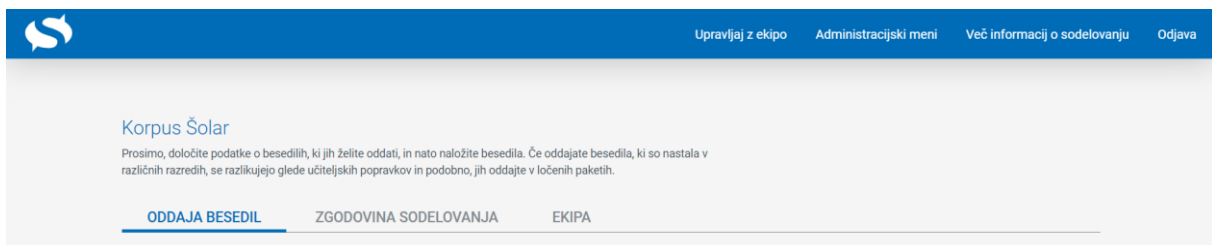
**Zavihek Ekipa:** Tu so shranjeni podatki o sodelujočih članih ekipe. Na levi strani zaslona so izpisana njihova imena skupaj z vlogo, pod tem pa se beleži še zgodovino sprememb (recimo sprememb vlog učiteljev). Desna stran zaslona prikazuje graf s podatki o številu oddaj vsakega člana ekipe in graf, ki izrisuje število vseh oddaj po regijah in vrsti šole. Podatki v obeh grafih se sproti osvežujejo.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



Slika 7: Zavihek Ekipa.

Na modri pasici zgoraj desno so gumb za odjavo iz portala (Odjava) in meniji za upravljanje, ki jih predstavljamo v nadaljevanju. Uporabnik z administratorskim uporabniškim računom ima poleg menija Več informacij o sodelovanju še menija Upravljalj z ekipo in Administracijski meni.

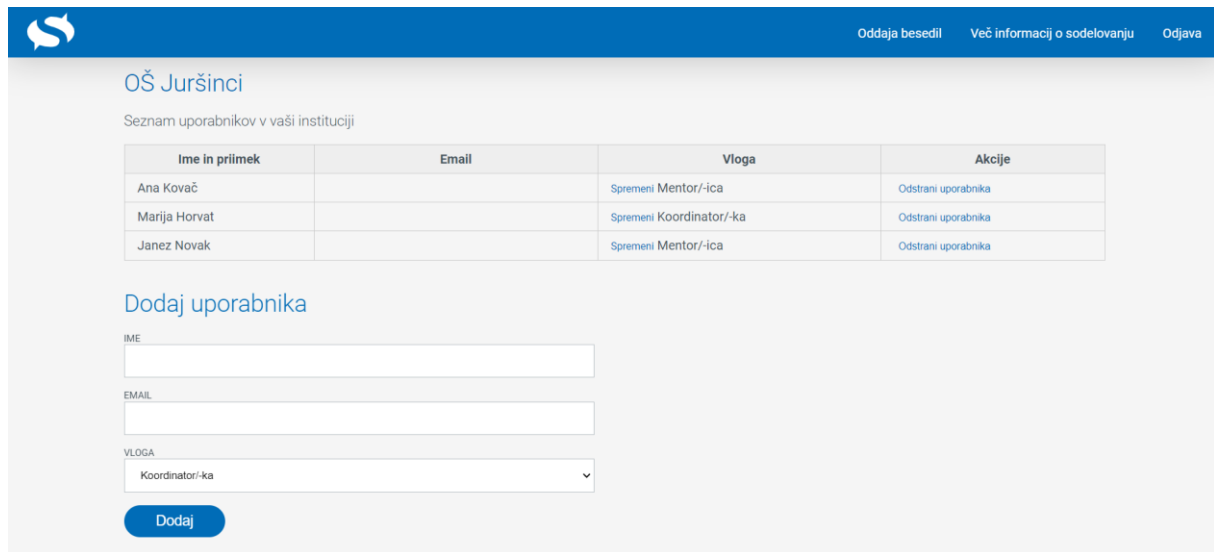


Slika 8: Meniji za upravljanje na modri pasici v administracijskem računu.

**Več informacij o sodelovanju:** Klik na ta gumb nas preusmeri na spletišče <https://rsdo.slovenscina.eu/zbiranje-besedil-za-korpus-solar>, kjer je opisano sodelovanje v projektu (opisan na začetku tega razdelka).

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

**Upravljalj z ekipo:** V tem meniju najdemo pregled nad ekipo v instituciji, mogoče pa je tudi ročno dodajanje članov.



Seznam uporabnikov v vaši instituciji

Ime in priimek	Email	Vloga	Akcije
Ana Kovač		Spremeni Mentor/-ica	Odstrani uporabnika
Marija Horvat		Spremeni Koordinator/-ka	Odstrani uporabnika
Janez Novak		Spremeni Mentor/-ica	Odstrani uporabnika

**Dodaj uporabnika**

IME

EMAIL

VLOGA  
Koordinator/-ka

**Dodaj**

**Slika 9:** Meni Upravljalj z ekipo

**Administracijski meni:** Tu se nahaja seznam vseh aktivnih uporabnikov portala z informacijami o njihovi vlogi in instituciji. Pod tem so funkcionalnosti za dodajanje/odstranitev uporabnika oz. uporabnika glede na institucijo, spreminjanje e-naslava in imena/priimka uporabnika ter spreminjanje vloge uporabniškega računa (navadni uporabnik ali administrator). Nižje na strani najdemo še naprednejše funkcije, recimo združevanje institucij, odstranjevanje vnosa, posodobitev metapodatkov že oddanih vnosov itd.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

UPORABNIKI

Aktivni uporabniki

ID	Ime in priimek	Email	Vloga	ID institucije	Vloga v instituciji
36	Marija Horvat		admin	23	coordinator
38	Janez Novak		user	23	mentor
39	Ana Kovač		user	23	mentor

Dodaj uporabnika

Ime in priimek:

Email:

Geslo:

Spremeni email uporabnika

ID uporabnika:

Nov email:

**Slika 10:** Administracijski meni

## 5 Špela Arhar Holdt, Iztok Kosem, Mojca Stritar Kučuk, Luka Krsnik, Tina Munda: Program Svala in novi format

### 5.1 Program Svala

#### 5.1.1 Izvor programa Svala

Novi program za prikaz in urejanje besedil z jezikovnimi popravki CJVT-Svala predstavlja poleg portala za zbiranje besedil prav tako pomemben napredek pri metodologiji gradnje korpusov usvajanja slovenščine in s tem pri njihovi učinkoviti nadgradnji.

Pred posvojitvijo tega odprtokodnega programa smo preategorizacijo jezikovnih popravkov opravljali v za naše namene prilagojenem orodju Sketch Engine. Ta postopek je bil izjemno časovno obremenjujoč – korpus smo morali najprej pretvoriti v format VERT za uvoz v Sketch Engine; tam smo učiteljske jezikovne popravke označevali z novimi kategorijami (gl. razdelek 3.1), kar je spremljalo sprotno izvažanje korpusnih datotek in njihovo pretvarjanje v XML, da smo lahko nove oznake kategorij, ki se v Sketch Enginu beležijo posebej, vključili v zapis korpusnih datotek lokalno; nato je bilo treba spet opraviti pretvorbo za ponoven uvoz v Sketch Engine. (Arhar Holdt et al. 2022: 27)

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Tak potek dela nas je motiviral, da smo se ozrli po praksah evropske jezikovnotehnološke skupnosti. Švedski kolegi z univerz v Göteborgu in Stockholmu so za gradnjo korpusa švedščine kot drugega ali tujega jezika zasnovali orodje Svala (Wirén 2019), ki smo ga ocenili kot dobro izhodišče za učinkovitejšo doseganje naših projektnih ciljev. Največja prednost orodja je, da združuje več korakov delotoka izgradnje korpusa usvajanja jezika, in sicer psevdomizacijo, normalizacijo in označevanje jezikovnih popravkov v besedilih (Arhar Holdt idr 2022: 27). Poleg tega portal SweLL (Volodina et al. 2019), v katerega je orodje Svala vključeno, skrbi še za vodenje delotokov za zbiranje in urejanje korpusnega gradiva. To orodje smo v projektu RSDO prilagodili za slovenščino in ga nadgradili, da omogoča delo s korpusoma KOST in Šolar. V luči potreb smo v CJVT-Svala prioritarno prenesli funkcionalnosti za transkripcijo, preprosto anonimizacijo in označevanje napak, naprednejše module, tj. avtomatsko podprto anonimizacijo in vodenje označevalnih delotokov pa smo odložili na prihodnje priložnosti za razvoj. Orodje je na voljo na spletnem mestu <https://orodja.cjvt.si/svala/>, v nadaljevanju pa predstavljamo njegovo delovanje in uporabo.

### 5.1.2 Izgled in funkcionalnosti programa CJVT-Svala

Izvirno besedilo uvozimo, prilepimo ali transkribiramo v okence pod napisom 'izvirno besedilo'. S klikom na gumb 'kopiraj v popravljeno' se izvirno besedilo prenese v okence pod napisom 'popravljeno besedilo', kamor vnesemo besedilo z jezikovnimi popravki. V okencu na dnu strani, v t. i. grafu povezav, se eno pod drugim izpišeta izvirno in popravljeno besedilo, njuni deli pa se med seboj avtomatsko povežejo s črticami, ki jim žargonsko rečemo *špageti*. S klikom na vsako črtico je mogoče pripisati kategorijo oznak za popravke, in sicer ločeno za korpus Šolar (več o oznakah gl. razdelek 3.1) in za korpus KOST (več o oznakah gl. razdelek Označevanje napak). Na Sliki 8 je v meniju na levi strani zaslona prikazan sistem oznak za Šolar, pri čemer se do željene oznake lahko preklikamo prek kategorij in podkategorij ali pa v iskalno polje nad seznamom kategorij ('vpišite oznako') vpišemo del opisa popravka in avtomatsko nas bo premestilo na oznako na seznamu, ki se vsaj delno ujema z iskalnim nizom. Z izbiro oznake s seznama se ta obesi na izbran *špaget* v grafu povezav.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

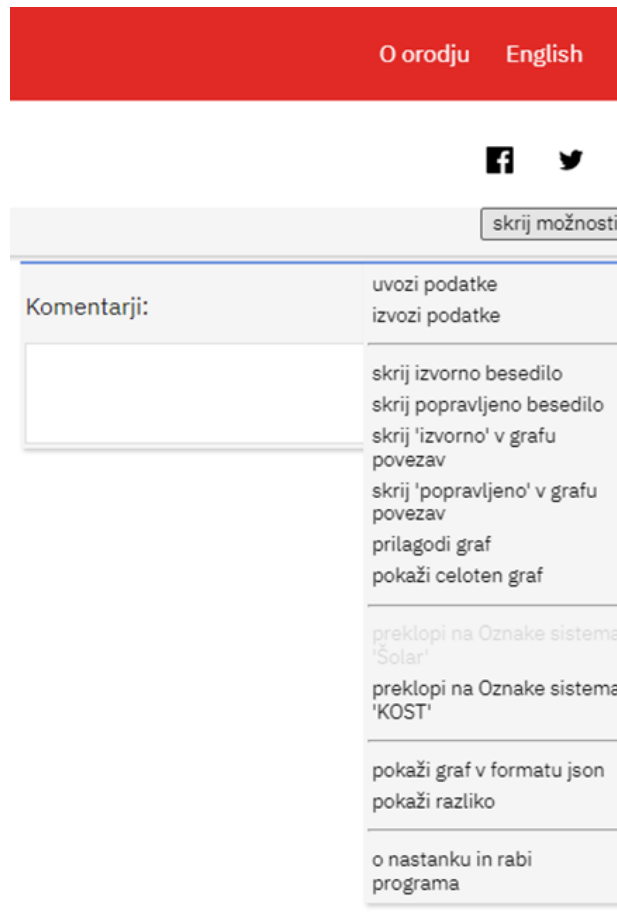
**Slika 8:** Primer izvornega in popravljenega besedila v vmesniku CJVT-Svala s sistemom oznak za Šolar

Poleg načina izpisa obeh verzij besedila in medbesedilnih povezav s pripeto oznako za jezikovno težavo so vizualna pomoč tudi barve – napaka iz izvornega besedila je obarvana rdeče, popravek napake pa zeleno. Nad iskalnim poljem v levem meniju najdemo še praktične gumbе z ukazi za premik na prejšnjo/naslednjo povezavo, prejšnjo/naslednjo spremembo, za ročno povezavo ali razvezavo neustrezno povezanih pojavnic (recimo ko se več besed v izvornem besedilu zamenja z eno v popravljenem ali obratno; gl. Sliko 9).

**Slika 9:** Ročno povezovanje pojavnic v CJVT-Svali

Operacija Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Na desni strani zaslona je okno 'Komentarji', kjer lahko označevalec poda dodaten komentar, in spustni seznam z možnostmi, do katerega dostopamo s klikom na 'prikaži možnosti', kot prikazuje Slika 10.



**Slika 10:** Okno za komentar in možnosti v CJVT-Svali

Tukaj so funkcionalnosti za uvoz/izvoz besedila, prikaz v vmesniku (skrij/pokaži izvorno besedilo/popravljeno in nastavitve prikaza grafa povezav), preklop med oznakami sistema KOST in Šolar, možnost prikaza grafa povezav v formatu JSON in povezava na uporabniški priročnik orodja (v angleščini).

V skrajnem zgornjem desnem kotu je možnost menjave jezika v angleščino in polje 'O orodju' z opisom vmesnika.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



### 5.1.3 Evalvacija programa

Orodje smo evalvirali na dveh nalogah: za gradnjo korpusa KOST 1.0, kjer je bil v vmesniku označen del besedil, in nadgradnjo Šolarja na različico 3.0, natančneje popravljanje in usklajevanje oznak iz različice 2.0 glede na posodobljene smernice za označevanje (več v razdelku 3.1). Izkušnja z delom v Svali je bila izredno pozitivna. Da bi preverili enostavnost uporabe tega orodja širše, so ga preizkusili še študenti 3. letnika dodiplomskega študija slovenistike. Pred nalogo smo jim delo v Svali le na grobo predstavili, saj nas je zanimalo, kako se s programom soočijo brez natančnih napotkov. Sodelovalo je 19 študentov, ki so v Svali označili skupno 83 besedil. Označena besedila je pregledala urednica KOST-a in izkazalo se je, da je bilo 35 % oznak za popravke neustreznih, a to ima opraviti s sistemom označevanja (več v razdelku Označevanje napak v poglavju 4). Kar se tiče uporabe orodja, nobeden od označevalcev ni imel večjih težav, kar je spodbudno (še največji izziv na splošno je bilo združevanje besed v skupine). Študenti so pri predstavitvi te naloge v okviru seminarja pri predmetu Slovenščina kot tuji jezik delo v CJVT-Svali označili za pozitivno izkušnjo, med katero so pri zelo dobrodošlem praktičnem delu preverili tudi svoje jezikoslovno znanje. (Arhar Holdt et al. 2022: 28, 29)

### 5.2 Novi korpusni format

Poleg orodja za učinkovito urejanje besedil z jezikovnimi popravki (gl. razdelek 5.1) je še ena pomembna rešitev za tovrstne korpusne vire, zasnovana v tem projektu, standardni format za zapis korpusa z jezikovnimi popravki. Verzija Šolar 2.0 je že odsevala shemo TEI, a še ni bila povsem usklajena s TEI standardom, ki je v rabi za druge korpuse, smo po posvetovanjih v širši projektni ekipi zasnovali popolnoma nov format korpusa z jezikovnimi popravki. Najprej smo izvožene datoteke v formatu JSON, ki je privzeti datotečni format za izvoz iz Svale, pretvorili v XML po shemi TEI<sup>8</sup>. V svojem ustroju TEI XML predvideva element za zapis popravkov v besedilu, a smo glede na množico in vrsto podatkov, ki smo jih želeli zakodirati v korpusno datoteko, morali obstoječo strukturo ponovno premisliti. Nov format korpusa usvajanja jezika je posledica predvsem pristopa, ki ga uvaja Svala. Gre za tri izhodne datoteke JSON (nepopravljena (Šolar - učenci) in popravljena besedila (Šolar - učitelji) ter datoteka s povezavami

---

<sup>8</sup> <https://tei-c.org/>

med vsako pojavnico iz nepopravljene in popravljene verzije datoteke, skupaj z oznakami za jezikovne popravke), ki so nato pretvorjene v XML in združene v eno datoteko XML, ki je skladna s shemo TEI.

Takšno sestavo ima tudi KOST, od Šolarja pa se slednji razlikuje po metapodatkih, ki vsebujejo tudi informacijo o maternem jeziku, spolu in letnici rojstva tvorca, in po tipologiji jezikovnih popravkov.

Korpusa sta na voljo še v bolj robustnih formatih CoNLL-U in JSON, ki sta primernejša za strojno procesiranje, ter VERT, ki ga tipično zahteva uvoz v konkordančnike (npr. noSketch Engine).

### 5.3 Konkordančnik

Korpus Šolar 3.0 smo uvozili v konkordančnika NoSketchEngine in KonText, ki sta del infrastrukture CLARIN.SI. Za vsak korpus smo datoteki z izvornimi in s popravljenimi besedili uvozili ločeno, saj specializiran konkordančnik, ki bi omogočal prikaz nepopravljenih in popravljenih besedil z vsebinskimi oznakami hkrati v našem okolju še ni bil razvit. Za celostno primerjavo besedil je tako treba odpreti obe verziji besedil (v primeru Šolarja: Šolar – učenci in Šolar - učitelji) v ločenih oknih. Postavitev enega okna ob drugem na zaslonu je lahko začasna rešitev za primerjanje podatkov, ne glede na to pa že sama vključitev korpusov v konkordančnike omogoča jasen pregled označenega gradiva in jezikoslovne analize.

Query **Č\_SKLOP\_nj** 193 (101.18 per million)

Page 1 of 10 | Go | Next | Last

KUS-G-slo-1-GO-E-2009-10001.yes	večjih oblikah; najpogostejša sta ljubezen do bližnjih	Č_SKLOP_nj Č_KONZ_izpust Č_SKLOP_nj Č_KONZ_odvce	, s katero se srečujemo vsaki dan ter z vsakim in
KUS-G-slo-1-GO-E-2009-10002.yes	kot [per: Antigona ] . (V) Ismeni po mojem mnenju	manja Č_SKLOP_nj Č_KONZ_izpust	kanček več poguma da bi premagala strah pred
KUS-G-slo-1-GO-E-2009-10003.yes	oz. vrata, kot se temu reče v žargonu, zajela	manja Č_SKLOP_nj Č_KONZ_izpust	nevhta, ki bi bila za naš majhen, komaj šest
KUS-G-slo-2-GO-E-2009-10004.yes	pri materinem jeziku. To je značilno za	razsvetljenstvo Č_SKLOP_nj Č_KONZ_odvce	kjer je tudi bila ustvarjena komedija.
KUS-G-slo-4-GO-E-2009-10012.yes	drugih. Če povežem še z naslovno mislijo. V času	razsvetljenstva Č_SKLOP_nj Č_KONZ_odvce	so se ljudje začeli zavedati razuma. Tako
KUS-G-slo-4-GO-E-2009-10013.yes	se neizmerno veseli, da [per: Polinejka ] ni, spodbuja	zvonjenje Č_SKLOP_nj Č_KONZ_izpust	zvonov, da oznanjajo mir: „Naj še zvone v
KUS-G-slo-4-GO-E-2009-10018.yes	človek, ki je obremenjen z dogajanjem na	takratnem Č_SKLOP_nj Č_KONZ_odvce	tržaškem ozemlju in zato nekoliko umaknjen v
KUS-G-slo-4-GO-E-2009-10020.yes	žene [per: Nadje ] , ki mu je predstavljala edini	življenski Č_SKLOP_nj Č_KONZ_izpust	, intimni smisel v življenju. Po prihodu domov
KUS-G-slo-4-GO-E-2009-10020.yes	Predstavlja partizana, ki pa imata različni	življenski Č_SKLOP_nj Č_KONZ_izpust	zgodbi. Berka bi lahko opisali kot netipičnega
KUS-G-slo-4-GO-E-2009-10030.yes	ljudi. S tem ko mu je umrla žena je izgubil vso	življensko Č_SKLOP_nj Č_KONZ_izpust	voljo in se preselil v svoj domači kraj, kjer je
KUS-G-slo-4-MB-E-2010-10036.yes	je pravkar pokopal svojo ženo na pokopališču na	gorenskem Č_SKLOP_nj Č_KONZ_izpust	, ki jo je zelo ljubil. Izgubil tudi smisel za
KUS-G-slo-4-MB-E-2010-10036.yes	, ter na začetku tudi [per: Bubi ] niso hoteli naučiti	slovenskega Č_SKLOP_nj Č_KONZ_izpust	jezika. To je slovence Šebolj razrezilo in
KUS-G-slo-4-MB-E-2010-10038.yes	, pri katerih je slovensko ime napisano dosti	manjše Č_SKLOP_nj Č_SKLOP_premet	. Nestrpnost se kaže tudi med slovinci in hrvati
KUS-G-slo-3-LJ-E-2009-10047.yes	so se na [loc: Glavni trg ] saj so morali zapustiti	prejšnje Č_SKLOP_nj Č_KONZ_izpust O_KAT_spol	stanovanje zaradi renoviranja. Pričakovanje
KUS-G-slo-3-LJ-E-2009-10059.yes	oblasti. Lik, ki predstavlja oblast in njen	takratni Č_SKLOP_nj Č_KONZ_odvce	odnos do ljudi, je župnik. Je [per: Jermanov ] najhujši
KUS-G-slo-3-LJ-E-2009-10059.yes	tem pa stoji njihova neiskrenost. Iz njegovega	življenskega Č_SKLOP_nj Č_KONZ_izpust	spoznanja o nesmiselnosti vsega sledi posebna
KUS-G-slo-3-LJ-E-2009-10061.yes	spoznanja o nesmiselnosti vsega sledi posebna	življenska Č_SKLOP_nj Č_KONZ_izpust	drža. Zavrača absurd s tem, da noče sodelovati v
KUS-G-slo-3-LJ-E-2009-10061.yes	, saj ga obsodijo, zaprejo in na koncu se njegova	življenska Č_SKLOP_nj Č_KONZ_izpust	pot konča z obglavitvijo. [per: Meursault ] je bil
KUS-G-slo-4-GO-E-2009-10073.yes	razmišljal o življenju, ni imel nobenega	življenskega Č_SKLOP_nj Č_KONZ_izpust	cilja, živel je z danes na jutri. Zato mislim, da
KUS-G-slo-4-GO-E-2009-10073.yes	[per: Kreona ] oz. na kompromisih temelječi vsakdanji	življenski Č_SKLOP_nj Č_KONZ_izpust	praksi. Antigona se v svoji neosebni

Page 1 of 10 | Go | Next | Last

Slika 11: Prikaz Šolarja 3.0 – učenci v konkordančniku NoSketchEngine Bonito

Query **Č\_SKLOP\_nj** 193 (100.86 per million)

Page 1 of 10 | Go | Next | Last

KUS-G-slo-1-GO-E-2009-10001.yes	v več oblikah; najpogostejši sta ljubezen do bližnjih	Č_SKLOP_nj Č_KONZ_izpust Č_SKLOP_nj Č_KONZ_odvce	, s katero se srečujemo vsak dan ter z vsakim, in
KUS-G-slo-1-GO-E-2009-10002.yes	kot [per: Antigona ] . (V) Ismeni po mojem mnenju	manjka Č_SKLOP_nj Č_KONZ_izpust	kanček več poguma, da bi premagala strah pred
KUS-G-slo-1-GO-E-2009-10003.yes	oz. vrata, kot se temu reče v žargonu, zajela	manjka Č_SKLOP_nj Č_KONZ_izpust	nevhta, ki bi bila za naš majhen, komaj šest
KUS-G-slo-2-GO-E-2009-10004.yes	pri materinem jeziku. To je značilno za	razsvetljenstvo Č_SKLOP_nj Č_KONZ_odvce	, kjer je tudi bila ustvarjena komedija.
KUS-G-slo-4-GO-E-2009-10012.yes	drugih. Če povežem še z naslovno mislijo. V času	razsvetljenstva Č_SKLOP_nj Č_KONZ_odvce	so se ljudje začeli zavedati razuma. Tako
KUS-G-slo-4-GO-E-2009-10013.yes	se neizmerno veseli, da [per: Polinejka ] ni, spodbuja	zvonjenje Č_SKLOP_nj Č_KONZ_izpust	zvonov, da oznanjajo mir: „Naj še zvone v
KUS-G-slo-4-GO-E-2009-10018.yes	človek, ki je obremenjen z dogajanjem na	takratnem Č_SKLOP_nj Č_KONZ_odvce	tržaškem ozemlju in zato nekoliko umaknjen v
KUS-G-slo-4-GO-E-2009-10020.yes	žene [per: Nadje ] , ki mu je predstavljala edini	življenski Č_SKLOP_nj Č_KONZ_izpust	, intimni smisel v življenju. Po prihodu domov
KUS-G-slo-4-GO-E-2009-10020.yes	Predstavlja partizana, ki pa imata različni	življenski Č_SKLOP_nj Č_KONZ_izpust	zgodbi. Berka bi lahko opisali kot netipičnega
KUS-G-slo-4-GO-E-2009-10030.yes	ljudi. S tem ko mu je umrla žena, je izgubil vso	življensko Č_SKLOP_nj Č_KONZ_izpust	voljo in se preselil v svoj domači kraj, kjer je
KUS-G-slo-4-MB-E-2010-10036.yes	je pravkar pokopal svojo ženo na pokopališču na [loc: Gorenskem	Č_SKLOP_nj Č_KONZ_izpust	, ki jo je zelo ljubil. Izgubil tudi smisel za
KUS-G-slo-4-MB-E-2010-10036.yes	ter na začetku tudi [per: Bubi ] , niso hoteli naučiti	slovenskega Č_SKLOP_nj Č_KONZ_izpust	jezika. To je [per: Slovence ] še bolj razrezilo in
KUS-G-slo-4-MB-E-2010-10038.yes	, pri katerih je slovensko ime napisano dosti	manjše Č_SKLOP_nj Č_SKLOP_premet	. Nestrpnost se kaže tudi med [per: Slovenci ] in [per: Hrvati ]
KUS-G-slo-3-LJ-E-2009-10047.yes	so se na [loc: Glavni trg ] , saj so morali zapustiti	prejšnje Č_SKLOP_nj Č_KONZ_izpust O_KAT_spol	stanovanje zaradi prodaje hiše. Pričakovanje
KUS-G-slo-3-LJ-E-2009-10059.yes	oblasti. Lik, ki predstavlja oblast in njen	takratni Č_SKLOP_nj Č_KONZ_odvce	odnos do ljudi, je župnik. Je [per: Jermanov ] najhujši
KUS-G-slo-3-LJ-E-2009-10059.yes	tem pa stoji njihova neiskrenost. Iz njegovega	življenskega Č_SKLOP_nj Č_KONZ_izpust	spoznanja o nesmiselnosti vsega sledi posebna
KUS-G-slo-3-LJ-E-2009-10061.yes	spoznanja o nesmiselnosti vsega sledi posebna	življenska Č_SKLOP_nj Č_KONZ_izpust	drža. Zavrača absurd s tem, da noče sodelovati v
KUS-G-slo-3-LJ-E-2009-10061.yes	, saj ga obsodijo, zaprejo in na koncu se njegova	življenska Č_SKLOP_nj Č_KONZ_izpust	pot konča z obglavitvijo. [per: Meursault ] je bil
KUS-G-slo-4-GO-E-2009-10073.yes	razmišljal o življenju, ni imel nobenega	življenskega Č_SKLOP_nj Č_KONZ_izpust	cilja, živel je z danes na jutri. Zato mislim, da
KUS-G-slo-4-GO-E-2009-10073.yes	[per: Kreona ] oz. na kompromisih temelječi vsakdanji	življenski Č_SKLOP_nj Č_KONZ_izpust	praksi. Antigona se v svoji neosebni

Page 1 of 10 | Go | Next | Last

Slika 12: Prikaz Šolarja 3.0 – učitelji v konkordančniku NoSketchEngine Bonito

Operacijo Razvoj slovensčine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Korpus je prosto dostopen v konkordančnikih NoSketchEngine, in sicer v starejšem uporabniškem vmesniku (Šolar 3.0 – učenci<sup>9</sup>, Šolar 3.0 – učitelji<sup>10</sup>) in novem (Šolar 3.0 – učenci<sup>11</sup>, Šolar 3.0 – učitelji<sup>12</sup>), ter v konkordančniku KonText (Šolar 3.0 – učenci<sup>13</sup>, Šolar 3.0 – učitelji<sup>14</sup>).

## 6 Špela Arhar Holdt, Iztok Kosem: Prioritete za nadaljnji razvoj

Prva prioriteta za nadaljnji razvoj korpusa je njegova vsebinska nadgradnja. Poskrbeti je treba za **povečanje korpusovega obsega in njegove reprezentativnosti** po regijah, vrsti šole, razredu/letniku avtorja in predmetu, pri katerem je besedilo nastalo. Želja je zagotoviti **redno korpusno posodabljanje**, kar pomeni zbiranje, vzorčenje in transkribiranje vsako tretje šolsko leto. Da bi slednje lahko uspelo, je treba **dvigniti ozaveščenost in spodbujati šole** k rednemu sodelovanju. Zahtevane kadrovske kapacitete za takšen kontinuiran razvoj so 1 FTE, ki si ga na letni ravni delita jezikoslovec, ki skrbi za zbiranje in pripravo gradiva, ter tehnični sodelavec, ki skrbi za korpusni format in dostopnost v vseh želenih orodjih.

Druga prioriteta je izboljšati dostopnost in povečati izrabo korpusnih podatkov. V projektu RSDO je bil razvit format korpusov z jezikovnimi popravki - specializirani XML TEI, ki je omogočil vključevanje v širše rabljene konkordančnike, kot so noSkE Bonito, noSkE Crystal in KonText. Za osnovne korpusne analize je ta vključitev izrednega pomena, vendar obstoječa orodja ne omogočajo polne izrabe bogato označenega gradiva, ki ga prinaša korpus Šolar. V nadaljevanju je treba razviti **specializirani konkordančnik**, ki bo uporaben za vse korpuse z jezikovnimi popravki (poleg korpusa Šolar sta v slovenskem prostoru takšna še korpusa KOST (Stritar Kučuk 2022) in Lektor (Popič 2014)). Novi konkordančnik mora biti po zasnovi primerljiv obstoječim, da se omogoči uporabniški prenos znanja, obenem pa mora imeti dodatne možnosti za izrabo metapodatkov, s pomočjo katerih bi lahko natančneje interpretirali posamezne rezultate iskanj po korpusu. Še bolj nujna je možnost preglednega

---

<sup>9</sup> [https://www.clarin.si/noske/sl.cgi/first\\_form?corpname=solar30\\_orig;align=](https://www.clarin.si/noske/sl.cgi/first_form?corpname=solar30_orig;align=)

<sup>10</sup> [https://www.clarin.si/noske/sl.cgi/first\\_form?corpname=solar30\\_corr;align=](https://www.clarin.si/noske/sl.cgi/first_form?corpname=solar30_corr;align=)

<sup>11</sup> [https://www.clarin.si/ske/#dashboard?corpname=solar30\\_orig](https://www.clarin.si/ske/#dashboard?corpname=solar30_orig)

<sup>12</sup> [https://www.clarin.si/ske/#dashboard?corpname=solar30\\_corr](https://www.clarin.si/ske/#dashboard?corpname=solar30_corr)

<sup>13</sup> [https://www.clarin.si/kontext/query?corpname=solar30\\_orig](https://www.clarin.si/kontext/query?corpname=solar30_orig)

<sup>14</sup> [https://www.clarin.si/kontext/query?corpname=solar30\\_corr](https://www.clarin.si/kontext/query?corpname=solar30_corr)

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

prikazovanja jezikovnih napak skupaj s popravki, zmožljivo iskanje po izvornih in popravljenih oblikah ter klikljive statistike najpogostejših jezikovnih popravkov. Z razvojem specializiranega konkordančnika bo Šolar postal širše uporaben jezikovni vir, zanimiv za pisce učnih gradiv, oblikovalce kurikulov, učitelje ali tiste, ki jih zanima jezik na splošno. Omogočal bo prepoznavo najpogostejših jezikovnih napak, značilnih za govorce določenih prvih jezikov, in s tem pripravo bolj osredotočenih učnih gradiv, pa tudi ustreznejše poudarke v samem pedagoškem procesu. Za najširšo možno rabo je treba zagotoviti tudi **izobraževanja učiteljev** o rabi novega korpusa in o izrabi jezikovnotehnoloških virov pri pouku slovenščine (in širše).

Tretja prioriteta je nadaljnji razvoj metodologije zbiranja. Velik časovni prihranek bi ponudila dopolnitev delotokov z **optičnim branjem ročno napisanih besedil**, pri čemer bodo potrebne adaptacije za šolsko rabo (kjer so v besedilih prisotne črkovalne napake in druge značilnosti razvojnega pisanja) ter natančno pregledovanje ter popravljanje optično prebranih rokopisov. Druga možnost za pohitritev dela je **strojno podprta identifikacija, vpis in kategorizacija učiteljskih jezikovnih popravkov** v ročno napisanih ali digitalnih besedilih - učiteljski popravki so v določeni meri predvidljivi in ponavljajoči se, kar bi bilo mogoče izkoristiti. Tretja možnost za pohitritev postopka je **vklučitev množičenja z didaktično perspektivo** v proces korpusnega grajenja. Pri tem bi bilo mogoče sodelovati s predavatelji, ki poučujejo jezikovno didaktiko in sorodne predmete na terciarni stopnji in bi v transkribiranje ter označevanje popravkov vključili študente in študentke, ki se pripravljajo na podajanje jezikovne povratne informacije učencem in dijakom. Množičenje je mogoče organizirati tudi za širšo populacijo, pri čemer pa je treba zagotoviti ustrezno kontrolo kvalitete in motivacijo za sodelovanje. V projektu RSDO smo ugotovili, da je v nadaljevanju treba nekoliko bolje urediti **pretvorbo besedil iz formata json**, ki ga uporablja program Svala, ter končnim želenim XML TEI. Izziv je zlasti zapisovanje ločil, ki se za označevanje v Svali ločijo od besednih pojavnic, za končni format pa jih je treba (ustrezno stično oz. nestično) spet urediti v izvorno obliko (ki je v šolskih besedilih lahko tudi neskladna s trenutnimi jezikovnimi pravili). Nenazadnje, evalvirati je treba, v kolikšni meri jezikovne napake v korpusnih besedilih vplivajo na **natančnost strojnega jezikovnega označevanja** na posameznih označevalnih ravneh, in zagotoviti ustrezne metodološke nadgradnje ali opozorila.

Predvsem pri ciljih, ki se vežejo na metodologijo, je treba slediti **mednarodnim iniciativam, rešitvam in dobrim praksam**, ne le na področju razvojnih korpusov, ampak širše na področju digitalne

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

humanistike (npr. za metodologijo optičnega branja, transkribiranja itd.). Stremeti je treba tudi k oblikovanju **mednarodnih standardov** za gradnjo korpusov z jezikovnimi popravki, saj bi to izboljšalo njihovo primerljivost in olajšalo samo uporabo (lažji bi bil tudi prenos znanja in rešitev).

Nenazadnje, zagovoriti je treba projektno financiranje za raziskave, ki bodo zagotovile **sintetične analize in podatke** za pripravo pedagoških učnih gradiv, jezikovnih priročnikov in orodij. Velik potencial predstavlja **primerjava podatkov** iz korpusa Šolar (šolska produkcija) s podatki, ki reprezentirajo šolsko recepcijo (npr. učbeniki, mladinska književnost, uporabniško generirane spletne vsebine), na drugi strani pa primerjava šolskega pisanja v kontekstih, kjer se slovenščino poučuje kot prvi jezik v primerjavi s poučevanjem slovenščine kot drugega/tujega jezika. Potrebne so tudi nadaljnje adaptacije korpusnih podatkov za namene strojnega učenja na področju jezikovnega izobraževanja, npr. dopolnitev z dodatnimi oznakami jezikovnih težav po modelu učnih korpusov. Komplementarno zasnovi korpusa Šolar je treba dodati zbiranje v smer širjenja korpusne vsebine na eni strani proti pisni produkciji **v nižjih razredih** in na drugi proti **študentskemu pisanju** (slednje je bilo že vključeno v raziskovalni projekt ARRS J7-3159, vendar le na ravni razvoja metodologije).

## 7 Priloge

**Priloga 1:** Analiza raziskave: Evalvacija Portala Šolar (RSDO)

**Priloga 2:** Pogodba za zbiranje besedil – šola

**Priloga 3:** Pogodba za zbiranje besedil – učenci

**Priloga 4:** Pogodba za zbiranje besedil – starši

## 8 Literatura

**Arhar Holdt et al. 2022** = Arhar Holdt, Špela, Kosem, Iztok in Stritar Kučuk, Mojca. Metode in orodja za lažjo pripravo korpusov usvajanja jezika. V: Nataša Pirih Svetina, Ina Ferbežar (ur.): *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41*. Ljubljana: Založba Univerze v Ljubljani, 2022. <https://centerslo.si/simpozij-obdobja/zborniki/obdobja-41/>.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



**Arhar Holdt et al. 2017** = Arhar Holdt, Špela, Kosem, Iztok, Gantar, Polona, 2017: Corpus-based resources for L1 teaching: the case of Slovene. V: Ann Marcus-Quinn, Triona Hourigan (ur.): *Handbook on digital learning for K-12 schools*. Cham: Springer. 91–113.

**Granger 2008** = Granger, Sylviane: Learner corpora. V: Anke Ludeling, Merja Kyto (ur.): *Corpus Linguistics. An International Handbook*. Berlin: Mouton de Gruyter, 2008. 259–275.

**Kosem et al. 2016** = Kosem, Iztok, Rozman, Tadeja, Arhar Holdt, Špela, Kocjančič, Polonca in Laskowski, Cyprian Adam. Šolar 2.0 : nadgradnja korpusa šolskih pisnih izdelkov. V: Erjavec, Tomaž (ur.), Fišer, Darja (ur.). *Zbornik konference Jezikovne tehnologije in digitalna humanistika, 29. september - 1. oktober 2016, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana, Slovenija*. 1. izd. V Ljubljani: Znanstvena založba Filozofske fakultete, 2016. str. 95-100, ilustr. ISBN 978-961-237-862-2. [http://www.sdjt.si/wp/wp-content/uploads/2016/09/JTDH-2016\\_Kosem-et-al\\_Solar-2-0-nadgradnja-korpusa-solskih-pisnih-izdelkov.pdf](http://www.sdjt.si/wp/wp-content/uploads/2016/09/JTDH-2016_Kosem-et-al_Solar-2-0-nadgradnja-korpusa-solskih-pisnih-izdelkov.pdf).

**Popič 2014** = Popič, Damjan, 2014. Revising translation revision in Slovenia. V: Tamara Mikolič Južnič, Kaisa Koskinen, Nike Kocijančič Pokorn (ur.): *New horizons in translation research and education 2*. Joensuu: University of Eastern Finland. [http://epublications.uef.fi/pub/urn\\_isbn\\_978-952-61-1657-0/urn\\_isbn\\_978-952-61-1657-0.pdf](http://epublications.uef.fi/pub/urn_isbn_978-952-61-1657-0/urn_isbn_978-952-61-1657-0.pdf).

**Stritar Kučuk 2022** = Mojca Stritar Kučuk: KOST med korpusi usvajanja tujega jezika. V: Nataša Pirih Svetina, Ina Ferbežar (ur.): *Na stičišču svetov: slovenščina kot drugi in tuji jezik*. Obdobja 41. Ljubljana: Založba Univerze v Ljubljani, 2022. 323–334. [https://centerslo.si/wp-content/uploads/2022/11/Stritar-Kucuk\\_Obdobja-41.pdf](https://centerslo.si/wp-content/uploads/2022/11/Stritar-Kucuk_Obdobja-41.pdf).

**Volodina et al. 2019** = Volodina, Elena, Granstedt, Lena, Matsson, Arild, Megyesi, Beáta, Pilán, Ildikó, Prentice, Julia, Rosén, Dan, Rudebeck, Lisa, Schenström, Carl-Johan, Sundberg, Gunlög, Wirén, Mats, 2019: The SwELL Language Learner Corpus: From Design to Annotation. *Northern European Journal of Language Technology* 6. 67–104.

**Wirén et al. 2019** = Wirén, Mats, Matsson, Arild, Rosén, Dan, Volodina, Elena, 2019: SVALA: Annotation of Second-Language Learner Text Based on Mostly Automatic Alignment of Parallel Corpora. Inguna

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

Skadina, Maria Eskevich (ur.): *Selected papers from the CLARIN Annual Conference 2018, Pisa, 8–10 October 2018*. Linköping: Linköping University Electronic Press. 227–239.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



**Priloga 1: Analiza raziskave: Evalvacija Portala Šolar (RSDO)**

# ANALIZA RAZISKAVE: Evalvacija Portala Šolar (RSDO)

## 1. METODOLOGIJA IZVEDBE

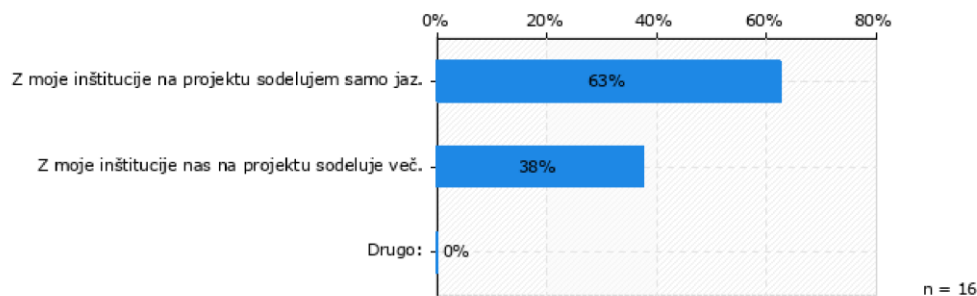
V strokovni evalvaciji je sodelovalo 16 učiteljic in učiteljev z 13 različnih osnovnih in srednjih šol. Evalvacijski obrazec smo pripravili s pomočjo programa 1-ka in ga skupaj z navodili posredovali prek e-pošte. Evalvacija je potekala med 11. 03. 2022 in 11. 06. 2022. V evalvaciji so učiteljice in učitelji sodelovali poimensko, vendar za namen poročila rezultate predstavljamo anonimizirano.

## 2. SAMOSTOJNA ALI EKIPNA IZKUŠNJA

V 10 primerih so bili evalvatorji edini predstavniki svoje inštitucije, v 6 primerih pa je na projektu sodelovalo več učiteljev z inštitucije. To je pomembno, saj smo na tak način lahko preverili tudi funkcionalnosti, ki omogočajo ekipno sodelovanje.

**Vprašanje: Ali z vaše inštitucije na projektu sodelujete samo vi ali vas sodeluje več?**

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Z moje inštitucije na projektu sodelujem samo jaz.	10	63%	63%	63%
	2 - Z moje inštitucije nas na projektu sodeluje več.	6	38%	38%	100%
	3 - Drugo	0	0%	0%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		



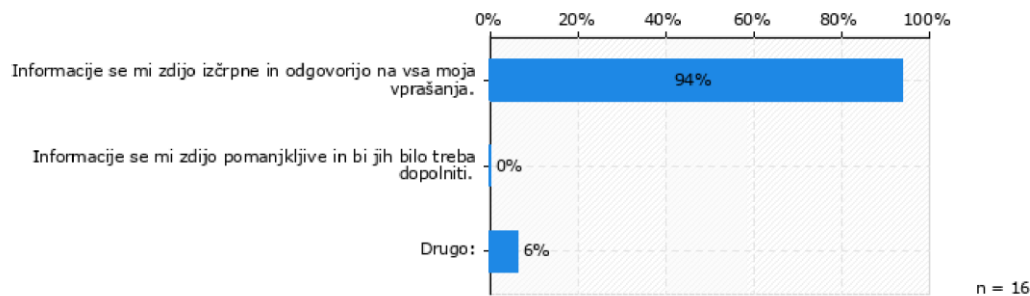
## 3. INFORMACIJE O SODELOVANJU

Evalvatorji so ocenili, da so informacije, ki so o zbiranju besedil na voljo na spletni strani projekta, izčrpne in odgovorijo na vsa vprašanja. Dilemo, ki jo je izpostavila ena od učiteljic v rubriki Drugo, smo na spletni strani dodatno pojasnili.

**Vprašanje: Poleg povezave na portal ste prejeli splošne informacije o zbiranju šolskih besedil. Kako ocenjujete te informacije?**

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Informacije se mi zdijo izčrpane in odgovorijo na vsa moja vprašanja.	15	94%	94%	94%
	2 - Informacije se mi zdijo pomanjkljive in bi jih bilo treba dopolniti.	0	0%	0%	94%
	3 – Drugo »Predlog sem zapisala spodaj.«	1	6%	6%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		



**Vprašanje: Prosimo, opišite predlog za izboljšave.**

- Informacije so izčrpane in odgovorijo na večino mojih vprašanj. Vprašanje, na katerega ne odgovarjajo, je, da pogodba, podpisana s strani staršev velja za 2 leti oz. čas sodelovanja v projektu.

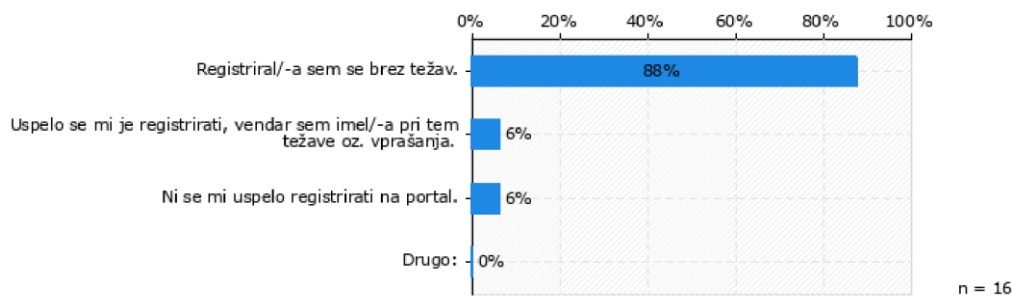
#### 4. REGISTRACIJA

V večini primerov je registracija potekala brez težav, pri čemer je bila v določenih primerih potrebna pomoč. Pomemben je komentar ene od učiteljic, ki se ponovi tudi pri naslednjem vprašanju, da traja zelo dolgo, da se spletna stran naloži oz. odpre.

**Vprašanje: Kakšna je bila vaša izkušnja z registracijo na portal?**

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Registriral/-a sem se brez težav.	14	88%	88%	88%
	2 - Uspelo se mi je registrirati, vendar sem imel/-a pri tem težave oz. vprašanja.	1	6%	6%	94%
	3 - Ni se mi uspelo registrirati na portal.	1	6%	6%	100%
	4 - Drugo	0	0%	0%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



**Vprašanje: Prosimo, opišite izkušnjo oz. predlog za izboljšave.**

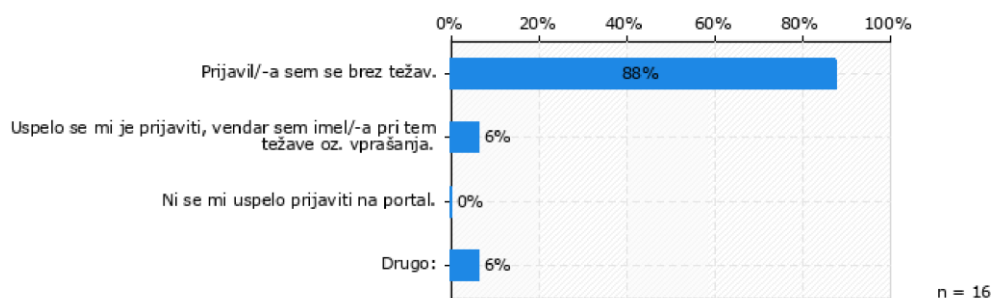
- *Imela sem težavo zaradi osebne površnosti pri registraciji v portal, zato mi je vse zablokiralo. Prijazna Eva mi je pri tem potrpežljivo in zelo nazorno pomagala. Hvala. Izboljšave niso potrebne, ker mi je po nazornem vodenju šlo vse gladko.*
- *Z registracijo sem imela težave. Kontaktirala sem koordinatorko, ki pa je težavo zelo hitro odpravila in nato sem se lahko uspešno prijavila. Zelo dolgo pa traja, da se začetna stran portala sploh odpre.*

**5. PRIJAVA**

Tudi ob prijavi na portal je bil identificiran problem s počasnostjo portala. Glede na učiteljske komentarje gre za dva različna problema: nalaganje oz. odpiranje spletne strani ter čakanje na potrditev registracije. Medtem ko gre v prvem primeru za tehnični problem, ki ga je nujno nasloviti pred lansiranjem portala, pa je drugi problem rešljiv z opozorilom, da registracija potrebuje potrditev oz. čaka na potrditev. Opozorilo je mogoče vključiti tako v avtomatsko sporočilo o registraciji kot na sam portal, kot je predlagala ena od evalvatorok.

**Vprašanje: Kakšna je bila vaša izkušnja s prijavo na portal?**

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Prijavil/-a sem se brez težav.	14	88%	88%	88%
	2 - Uspelo se mi je prijaviti, vendar sem imel/-a pri tem težave oz. vprašanja.	1	6%	6%	94%
	3 - Ni se mi uspelo prijaviti na portal.	0	0%	0%	94%
	4 – Drugo »Kar nekaj časa traja, da se začetna stran portala sploh odpre.«	1	6%	6%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		



**Vprašanje: Prosimo, opišite izkušnjo oz. predlog za izboljšave.**

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

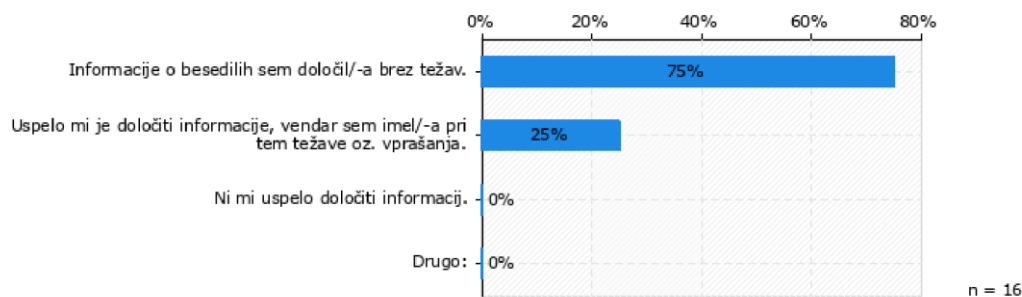
- Morda se nazorno navedejo posamezni koraki v registracijo in vstop v portal Šolar.
- Mogoče bi lahko že na začetno stran namesto besedila \ "Zbiranje besedil za korpus šolar poteka po naslednjem postopku, ki prinaša tudi točke za napredovanje v pedagoški naziv. \ " napisali, da je po vpisu podatkov potrebno počakati nekaj časa, da se portal odpre.

## 6. PODATKI O BESEDILIH

Vsem sodelujočim je uspelo določiti metainformacije o besedilih za oddajo, vendar so imeli nekateri pri tem vprašanja ali težave. Izkazalo se je, da je moteče oz. zavajajoče poimenovanje rubrike »Letnik« namesto ustrežnejšega »Letnik/razred«, ki bi že na prvi pogled naslavljal tudi osnovne šole. Popravke potrebuje tudi meni »Vrsta besedila«, pri katerem ni jasno, na kaj se kategorije nanašajo. Večje težave so se pojavile pri poskusu večkratnega zaporednega oddajanja besedil, kar je resnejši problem, ki zahteva prioritarno reševanje.

**Vprašanje: Kakšna je bila vaša izkušnja z dodajanjem informacij o šolskih besedilih (regija šole, predmet, stopnja itd.)?**

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Informacije o besedilih sem določil/-a brez težav.	12	75%	75%	75%
	2 - Uspelo mi je določiti informacije, vendar sem imel/-a pri tem težave oz. vprašanja.	4	25%	25%	100%
	3 - Ni mi uspelo določiti informacij.	0	0%	0%	100%
	4 - Drugo	0	0%	0%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		



**Vprašanje: Prosimo, opišite izkušnjo oz. predlog za izboljšave.**

- V osnovni šoli imamo razrede, zato me je beseda letnik zmotila. Letnik mi pove zaporedno šolanje dijaka (SŠ).
- Razdelek letnik na začetku uporabe vzbuja dvom o tem, kaj izbrati za osnovnošolce.
- Vrsta besedil - tam je kar nekaj ločeno z vejico in ni jasno, ali morajo veljati vsi kriteriji za eno besedilo ali ne.
- Pri tretjem vnosu ni bilo na voljo orodne vrstice besedilna vrsta, zato sem naložila brez tega. Naknadno sem pri zgodovini sodelovanja videla, da je avtomatično izpisalo kot prejšnji vnos. Pri 4. vnosu mi je avtomatično napačno zapisalo mesto, ker nisem bila več pozorna na to, je sedaj v zgodovini sodelovanja mesto zapisano napačno.

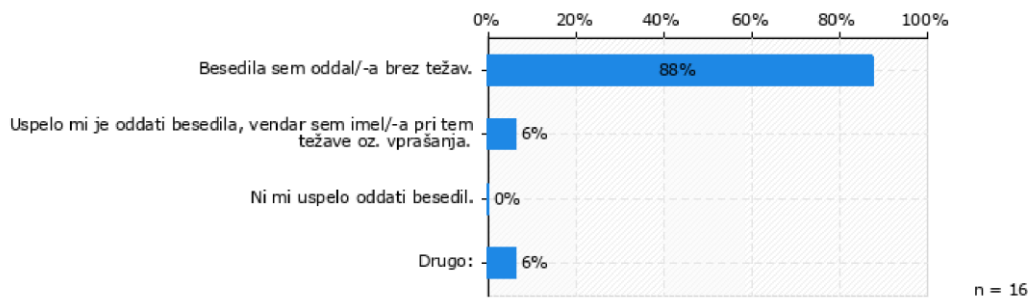
## 7. ODDAJANJE BESEDIL

Pri oddaji besedil večina ni imela težav, relevanten predlog za nadgradnjo pa je, da bi bilo mogoče datoteko, ko je naložena, pred oddajo ponovno odpreti in preveriti, če je prava.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

**Vprašanje: Kakšna je bila vaša izkušnja z oddajanjem šolskih besedil?**

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Besedila sem oddal/-a brez težav.	14	88%	88%	88%
	2 - Uspelo mi je oddati besedila, vendar sem imel/-a pri tem težave oz. vprašanja.	1	6%	6%	94%
	3 - Ni mi uspelo oddati besedil.	0	0%	0%	94%
	4 – Drugo »Besedila oddaja kolegica [Ime Priimek].«	1	6%	6%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		



**Vprašanje: Prosimo, opišite izkušnjo oz. predlog za izboljšave.**

- Zdi se mi smiselno, da bi lahko datoteko, ki si jo naložil, odprl in preveril, če je res prava (preden dokončno oddaš).

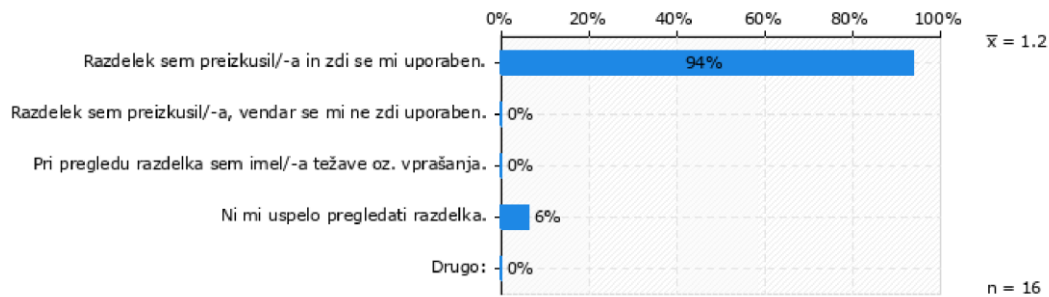
**8. ZGODOVINA SODELOVANJA**

**Vprašanje: Kakšna je bila vaša izkušnja z razdelkom Zgodovina sodelovanja?**

Sodelujoči so, razen ene osebe, razdelek Zgodovina sodelovanja pregledali in zdi se jim uporaben.

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Razdelek sem preizkusil/-a in zdi se mi uporaben.	15	94%	94%	94%
	2 - Razdelek sem preizkusil/-a, vendar se mi ne zdi uporaben.	0	0%	0%	94%
	3 - Pri pregledu razdelka sem imel/-a težave oz. vprašanja.	0	0%	0%	94%
	4 - Ni mi uspelo pregledati razdelka.	1	6%	6%	100%
	5 - Drugo	0	0%	0%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



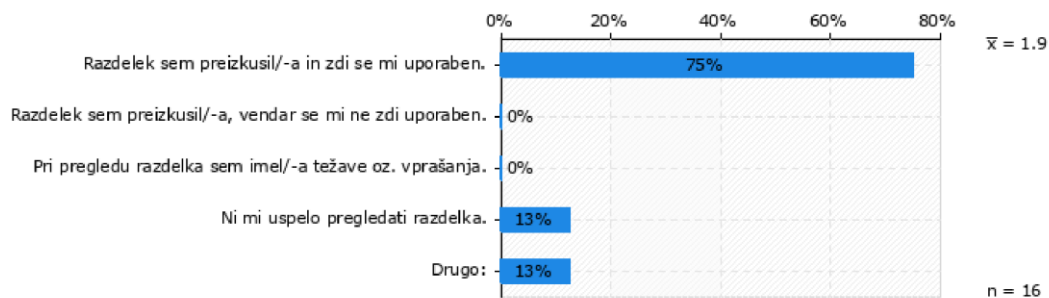
Vprašanje: Prosimo, opišite izkušnjo oz. predlog za izboljšave. (ni bilo odgovorov)

## 9. EKIPA

Vprašanje: Kakšna je bila vaša izkušnja z razdelkom Ekipa?

Trem četrtinam sodelujočih se zdi razdelek uporaben, dvema razdelka ni uspelo pregledati, ena sodelujoča je omenila, da za samostojno delo ni uporaben. Za nadgradnjo je najbolj relevanten komentar, da namen razdelka ni jasan, saj v času evalvacije podatki v njem še niso bili celoviti: želje so kazati statistiko oddanih besedil v primerjavi z drugimi šolami, po vrsti šol in regijah.

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Razdelek sem preizkusil/-a in zdi se mi uporaben.	12	75%	75%	75%
	2 - Razdelek sem preizkusil/-a, vendar se mi ne zdi uporaben.	0	0%	0%	75%
	3 - Pri pregledu razdelka sem imel/-a težave oz. vprašanja.	0	0%	0%	75%
	4 - Ni mi uspelo pregledati razdelka.	2	13%	13%	88%
	5 - Drugo »Ker sodelujem sama na šoli, zavme ne bo uporaben. Za druge pa bo.« »Ne vem točno, kakšen je namen tega razdelka.«	2	13%	13%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		



Vprašanje: Prosimo, opišite izkušnjo oz. predlog za izboljšave.

- Razdelka nisem pregledala.

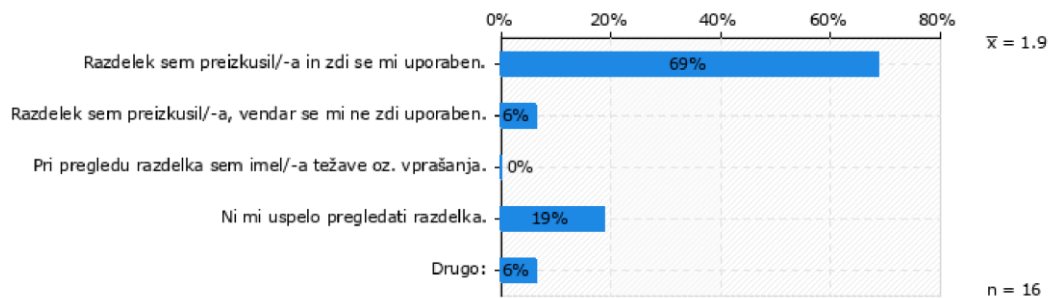
Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

## 10. UPRAVLJANJE Z INŠTITUCIJO

69 % sodelujočih je razdelek pregledalo in se jim zdi uporaben, dvema sodelujočima se ne zdi uporaben, trem pa razdelka ni uspelo pregledati, predvidoma, ker ga niso opazili. Slednje je za dano anketo relativno visoka številka. Administratorje posamezne šole – ki bi do razdelka dejansko dostopali – bi bilo torej smiselno posebej opozoriti na njegov obstoj in delovanje.

**Vprašanje: Na levi zgoraj je možnost upravljanja s člani vaše inštitucije. Kakšna je bila vaša izkušnja s tem razdelkom?**

	Odgovori	Frekvenca	Odstotek	Veljavni	Kumulativa
	1 - Razdelek sem preizkusil/-a in zdi se mi uporaben.	11	69%	69%	69%
	2 - Razdelek sem preizkusil/-a, vendar se mi ne zdi uporaben.	1	6%	6%	75%
	3 - Pri pregledu razdelka sem imel/-a težave oz. vprašanja.	0	0%	0%	75%
	4 - Ni mi uspelo pregledati razdelka.	3	19%	19%	94%
	5 - Drugo: »Ker sodelujem edina na šoli, zame ne bo uporaben.«	1	6%	6%	100%
Veljavni	Skupaj	16	100%	100%	
Manjkajoči	Skupaj		0%		
	SKUPAJ	16	100%		



**Vprašanje: Prosimo, opišite izkušnjo oz. predlog za izboljšave.**

- Nisem pregledala razdelka.
- Nimam komentarja

## 11. SPLOŠNA OCENA

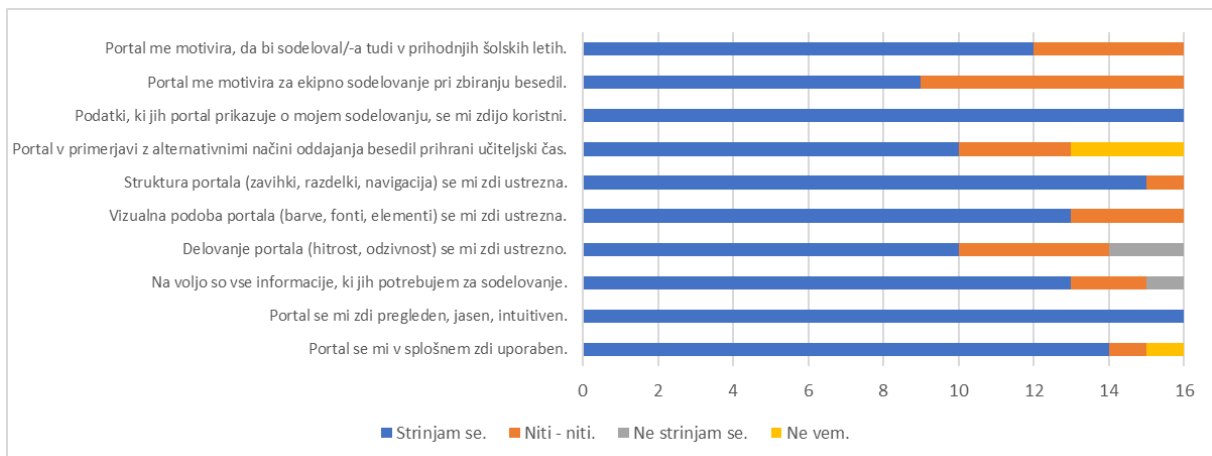
**Vprašanje: Prosimo, opredelite, koliko se strinjate s spodnjimi trditvami.**

	Podvprašanja	Odgovori					Veljavni	Št. enot	Povprečje	Std. odklon
		Strinjam se.	Niti - niti.	Ne strinjam se.	Ne vem.	Skupaj				
Q21a	Portal se mi v splošnem zdi uporaben.	14 (88%)	1 (6%)	0 (0%)	1 (6%)	16 (100%)	16	16	1,3	0,8
Q21b	Portal se mi zdi pregleden, jasen, intuitiven.	16 (100%)	0 (0%)	0 (0%)	0 (0%)	16 (100%)	16	16	1,0	0,0

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



Q21c	Na voljo so vse informacije, ki jih potrebujem za sodelovanje.	13 (81%)	2 (13%)	1 (6%)	0 (0%)	16 (100%)	16	16	1,3	0,6
Q21d	Delovanje portala (hitrost, odzivnost) se mi zdi ustrezno.	10 (63%)	4 (25%)	2 (13%)	0 (0%)	16 (100%)	16	16	1,5	0,7
Q21e	Vizualna podoba portala (barve, fonti, elementi) se mi zdi ustrezna.	13 (81%)	3 (19%)	0 (0%)	0 (0%)	16 (100%)	16	16	1,2	0,4
Q21f	Struktura portala (zavihki, razdelki, navigacija) se mi zdi ustrezna.	15 (94%)	1 (6%)	0 (0%)	0 (0%)	16 (100%)	16	16	1,1	0,3
Q21g	Portal v primerjavi z alternativnimi načini oddajanja besedil prihrani učiteljski čas.	10 (63%)	3 (19%)	0 (0%)	3 (19%)	16 (100%)	16	16	1,8	1,2
Q21h	Podatki, ki jih portal prikazuje o mojem sodelovanju, se mi zdijo koristni.	16 (100%)	0 (0%)	0 (0%)	0 (0%)	16 (100%)	16	16	1,0	0,0
Q21i	Portal me motivira za ekipno sodelovanje pri zbiranju besedil.	9 (56%)	7 (44%)	0 (0%)	0 (0%)	16 (100%)	16	16	1,4	0,5
Q21j	Portal me motivira, da bi sodeloval/-a tudi v prihodnjih šolskih letih.	12 (75%)	4 (25%)	0 (0%)	0 (0%)	16 (100%)	16	16	1,3	0,4



Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

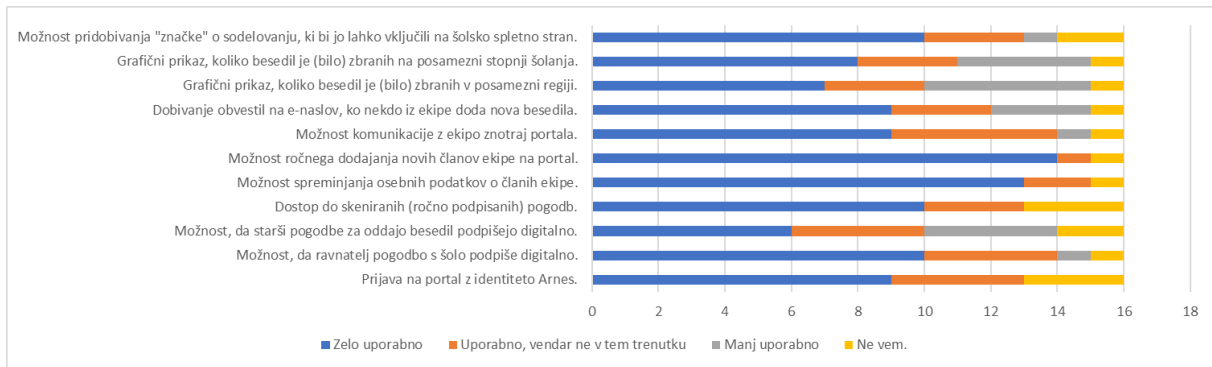


## 12. DODATNE FUNKCIONALNOSTI

Vprašanje: Prosimo, opredelite, koliko bi se vam zdele uporabne spodnje funkcionalnosti, če bi jih dodali na portal.

	Podvprašanja	Odgovori					Veljavni	Št. enot	Povprečje	Std. odklon
		Zelo uporabno	Uporabno, vendar ne v tem trenutku	Manj uporabno	Ne vem.	Skupaj				
Q22a	Prijava na portal z identiteto Arnes.	9 (56%)	4 (25%)	0 (0%)	3 (19%)	16 (100%)	16	16	1,8	1,2
Q22b	Možnost, da ravnatelj pogodbo s šolo podpiše digitalno.	10 (63%)	4 (25%)	1 (6%)	1 (6%)	16 (100%)	16	16	1,6	0,9
Q22c	Možnost, da starši pogodbe za oddajo besedil podpišejo digitalno.	6 (38%)	4 (25%)	4 (25%)	2 (13%)	16 (100%)	16	16	2,1	1,1
Q22d	Dostop do skeniranih (ročno podpisanih) pogodb.	10 (63%)	3 (19%)	0 (0%)	3 (19%)	16 (100%)	16	16	1,8	1,2
Q22e	Možnost spreminjanja osebnih podatkov o članih ekipe.	13 (81%)	2 (13%)	0 (0%)	1 (6%)	16 (100%)	16	16	1,3	0,8
Q22f	Možnost ročnega dodajanja novih članov ekipe na portal.	14 (88%)	1 (6%)	0 (0%)	1 (6%)	16 (100%)	16	16	1,3	0,8
Q22g	Možnost komunikacije z ekipo znotraj portala.	9 (56%)	5 (31%)	1 (6%)	1 (6%)	16 (100%)	16	16	1,6	0,9
Q22h	Dobivanje obvestil na e-naslov, ko nekdo iz ekipe doda nova besedila.	9 (56%)	3 (19%)	3 (19%)	1 (6%)	16 (100%)	16	16	1,8	1,0
Q22i	Grafični prikaz, koliko besedil je (bilo) zbranih v posamezni regiji.	7 (44%)	3 (19%)	5 (31%)	1 (6%)	16 (100%)	16	16	2,0	1,0
Q22j	Grafični prikaz, koliko besedil je (bilo) zbranih na posamezni stopnji šolanja.	8 (50%)	3 (19%)	4 (25%)	1 (6%)	16 (100%)	16	16	1,9	1,0
Q22k	Možnost pridobivanja "značke" o sodelovanju, ki bi jo lahko vključili na šolsko spletno stran.	10 (63%)	3 (19%)	1 (6%)	2 (13%)	16 (100%)	16	16	1,7	1,1

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.



### 13. BODOČE IZBOLJŠAVE

**Vprašanje:** Opišite, kako bi lahko izboljšali portal in celoten postopek zbiranja gradiva. Kako bi lahko proces olajšali, kaj bi vas motiviralo k sodelovanju? *Vsa mnenja, ideje in zadržki nam bodo prišli prav. Hvala za sodelovanje!*

- *Nalaganje oz. zbiranje besedil je enostavno, zato naj tako tudi ostane. zdi se mi enostaven in razumljiv portal, kar je nujno, da bi v prihodnje sodelovalo čim več profesorjev.*
- *Po prvotni izkušnji menim, da ne bo večjih težav.*
- *Ne zdi se mi primerno, da se označi \ "letnik\ " za razvrščanje osnovnošolskih besedil - izbira programa bi lahko omogočila, da se znotraj tega določi letnih za srednješolski program) ter razred za osnovnošolski. Nalaganje več besedil hkrati je smiselno le, če ima učitelj že predhodno skupaj shranjene enake vrste besedila, drugače je potrebno za ...*
- *Portal bi morda potreboval vizualno izboljšavo, drugače pa se mi zdi zelo uporaben in praktičen. zanima me, ali bodo vsi naloženi vnosi vidni vsem.*
- *Postopek se mi zdi ustrezen. veselim se morebitnih nadaljnjih aktivnosti.*
- *Za objektivno oceno je enoletno sodelovanje premalo. Portal bo moral zaživeti.*
- *Nmam pripomb.*
- *Morda bi pri določitvi min. št. oddanih izdelkov morali bolj upoštevati njihovo dolžino. Esej, ki sem ga prispevala, vsebuje več kot 800 besed. Lahko vam zaupam še to, da sem izdelke učencev prenesla v e-obliko s pomočjo telefonske aplikacije za pretvarjanje govora v besedilo. To mi je prihranilo nekaj časa.*
- *Najlažja bi bil oddaja na ustrezen e-naslov, kjer bi prejela vašo povratno informacijo, da sem vse naredila, kot je zaželeno.*
- *Trenutna verzija se mi zdi zelo dobra.*
- *Menim, da je postopek zbiranja besedil precej enostaven, saj omogoča veliko možnosti oz. izbiro formata za posredovanje (od wordovega dokumenta do skeniranih dokumentov). Oddala sem 10 primerov enake besedilne vrste (basen) v enem wordovem dokumentu. Proces ste učiteljem že olajšali, saj so lahko izbrali, v kakšni obliki bodo posredovali besedila učencev. Nabor čim večjega ...*
- *Predlagam, da bi se ta portal nekje bolj obelodanil, saj sem prepričana, da ga ogromno učiteljev ne pozna. morda bi bilo dobro, da se predstavi na posameznih in ustreznih študijskih skupinah in se pojasni, čemu je namenjen, kaj je njegov doprinos k učiteljevemu delu.*
- *Čas med registracijo in prijavo je predolg: 2 uri je trajalo, da sem dobila potrditveno e-pošt; šele nato sem se lahko prijavila. Ostalo deluje brezhibno.*
- *Predlagam jezikovni pregled spletne strani. Portal trenutno dobro deluje. morda bi lahko dodali zasavsko statistično regijo. Lep pozdrav*
- *Nimam predlogov.*

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

## Priloga 2: Pogodba za zbiranje besedil – šola

Univerza v Ljubljani, Kongresni trg 12, 1000 Ljubljana, ki jo zastopa rektor dr. Gregor Majdič; matična številka: 5085063000, davčna številka: 54162513 (v nadaljnjem besedilu: izvajalec)

in

[Ime šole], [ulica in št.], [poštna številka in kraj], ki ga/jo zastopa: [ime priimek], [naziv], v nadaljevanju: šola,

sklepata naslednjo

### POGODBO O SODELOVANJU

Šola si bo prizadevala, da bo izvajalcu omogočila uporabo besedil in bo v ta namen pomagala z zbiranjem in posredovanjem besedil učencev, ki jih bodo učenci napisali v sklopu šolanja (v nadaljevanju besedila), v namene projekta izgradnje **Korpusa šolskih pisnih izdelkov Šolar** (v nadaljevanju Projekt), ki ga izvaja izvajalec v okviru raziskovalnih programov **Slovenski jezik – bazične, kontrastivne in aplikativne raziskave** (P6-0215) in **Jezikovni viri in tehnologije za slovenski jezik** (P6-0411) na Univerzi v Ljubljani.

Šola se zavezuje, da bo izvajalcu za namene Projekta posredovala le besedila tistih učencev, ki bodo sami ali za njih zakoniti zastopniki podpisali pogodbo, ki jo bo predhodno pripravil izvajalec in bo vsebovala določbe o prenosu ustreznih avtorskih pravic oziroma dovoljenje za uporabo besedil v Projektu ter dovoljenje za uporabo besedil skladno z licenco Creative Commons (CC) BY-SA 4.0. Osnutek obeh pogodb je priloga in sestavni del te pogodbe. Izvajalec se zavezuje, da bo v Projektu uporabil le besedila tistih učencev, ki jim jih bo posredovala šola skladno s prejšnjim odstavkom.

Izvajalec se zavezuje, da originalnih besedil za potrebe Projekta ne bo odnašal iz šole.

Izvajalec se zavezuje, šola pa se strinja, **da bo vse osebne podatke**, ki se morebiti pojavljajo v besedilih, **anonimiziral in varoval** v skladu z Uredbo (EU) 2016/679 Evropskega parlamenta in Sveta z dne 27. aprila 2016 o varstvu posameznikov pri obdelavi osebnih podatkov in o prostem pretoku takih podatkov ter o razveljavitvi Direktive 95/46/ES (Splošna uredba o varstvu podatkov), Zakonom o varstvu osebnih podatkov (ZVOP-1, UL RS 86/4), Zakonom o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA, UL RS št. 30/2006), njenimi podzakonskimi akti in drugo veljavno zakonodajo.

Za razmerja v zvezi s to pogodbo se uporabljajo pravni predpisi Republike Slovenije.

Spore iz te pogodbe bosta stranki reševali po mirni poti. V primeru, da mirna rešitev ne bo mogoča, je za vse spore v zvezi s to pogodbo pristojno sodišče v Ljubljani.

Ta pogodba nadomešča vsa predhodna pogajanja, ponudbe in druge dogovore med strankama.

Ta pogodba je sestavljena v [dveh] istovetnih izvodih, od katerih prejme vsaka stranka po enega.

Kraj, datum:.....

Podpis zastopnika šole (ravnatelja/-ice):

.....

Podpis zastopnika izvajalca

(Univerza v Ljubljani):

.....

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

### Priloga 3: Pogodba za zbiranje besedil – učenci

#### POGODBA O PRENOSU PRAVIC in DOVOLJENJE ZA UPORABO PRAVIC

Podpisani/-a (ime in priimek) ....., stalno bivališče .....

(v nadaljevanju: avtor), skladno s to pogodbo odstopam izvajalcu Univerza v Ljubljani, Kongresni trg 12, 1000 Ljubljana, ki jo zastopa rektor dr. Gregor Majdič; matična številka: 5085063000, davčna številka: 54162513 (v nadaljnjem besedilu: izvajalec) vse avtorske pravice, avtorski sorodne pravice in ostale pravice avtorja v skladu z Zakonom o avtorski in sorodnih pravicah (ZASP) na besedilih, ki sem jih v sklopu šolanja napisal/-a pri pouku v (naziv šole) ..... (v nadaljevanju: besedila), v namene izgradnje **Korpusa šolskih pisnih izdelkov Šolar**, ki ga izvaja izvajalec v okviru raziskovalnih programov **Slovenski jezik – bazične, kontrastivne in aplikativne raziskave** (P6-0215) in **Jezikovni viri in tehnologije za slovenski jezik** (P6-0411) na Univerzi v Ljubljani.

**S to pogodbo** brezplačno, časovno in krajevno neomejeno ter nepreklicno prenašam na izvajalca avtorske pravice reprodukcije, distribucije, dajanja v najem, priobčitve javnosti in predelave besedil ter dajem dovoljenje, da se besedila uporabijo za projekt za namene izgradnje **Korpusa šolskih pisnih izdelkov Šolar**.

S to izjavo avtor izjavljam, da sem imetnik vseh avtorskih, avtorski sorodnih pravic oziroma drugih pravic avtorja v skladu z ZASP na avtorskih delih, ki so predmet te izjave.

Izvajalec se zavezuje, avtor pa se strinja, **da bo vse osebne podatke**, ki se morebiti pojavljajo v besedilih, **anonimiziral in varoval** v skladu z Uredbo (EU) 2016/679 Evropskega parlamenta in Sveta z dne 27. aprila 2016 o varstvu posameznikov pri obdelavi osebnih podatkov in o prostem pretoku takih podatkov ter o razveljavitvi Direktive 95/46/ES (Splošna uredba o varstvu podatkov), Zakonom o varstvu osebnih podatkov (ZVOP-1, UL RS 86/4), Zakonom o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA, UL RS št. 30/2006), njenimi podzakonskimi akti in drugo veljavno zakonodajo.

Izrecno soglašam, da navajanje avtorja na besedilih ni potrebno.

**S to pogodbo** avtor soglašam z nadaljnjim prenosom materialnih pravic, navedenih v tej izjavi, na tretje osebe. Avtor sem seznanjen, da bo Korpus šolskih pisnih izdelkov Šolar na voljo javnosti pod pogoji licence Creative Commons 4.0 "priznanje avtorstva" in s tem izrecno soglašam. Ta licenca omogoča uporabnikom avtorsko delo in njegove predelave reproducirati, distribuirati, dajati v najem, priobčiti javnosti in predelovati samo pod pogojem, da navedejo avtorja tega dela, oz. v konkretnem primeru, ko avtor ne bo naveden, navedejo kot vir korpus Šolar.

Za razmerja v zvezi s to pogodbo se uporabljajo pravni predpisi Republike Slovenije.

Spore iz te pogodbe bosta stranki reševali po mirni poti. V primeru, da mirna rešitev ne bo mogoča, je za vse spore v zvezi s to pogodbo pristojno sodišče v Ljubljani.

Ta pogodba nadomešča vsa predhodna pogajanja, ponudbe in druge dogovore med strankama.

Ta pogodba je sestavljena v [dveh] istovetnih izvodih, od katerih prejme vsaka stranka po enega.

Kraj, datum:.....

Podpis avtorja (dijaka/-inje oz. študenta/-ke):

Podpis zastopnika izvajalca

(Univerza v Ljubljani):

.....

.....

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

**Priloga 4:** Pogodba za zbiranje besedil – starši

**POGODBA O PRENOSU PRAVIC in DOVOLJENJE ZA UPORABO PRAVIC**

Učenec/-ka ....., stalno bivališče  
..... (v nadaljevanju:  
učenc/-ka), ki ga zastopa/-ta spodaj podpisani/-a starš/-a oz. zakoniti/-a zastopnik/-a  
....., s stalnim bivališčem:  
....., skladno s  
to pogodbo odstopam izvajalcu Univerza v Ljubljani, Kongresni trg 12, 1000 Ljubljana, ki jo zastopa rektor dr. Gregor  
Majdič; matična številka: 5085063000, davčna številka: 54162513 (v nadaljnjem besedilu: izvajalec) vse avtorske  
pravice, avtorski sorodne pravice in ostale pravice avtorja v skladu z Zakonom o avtorski in sorodnih pravicah  
(ZASP) na besedilih, ki sem jih v sklopu šolanja napisal/-a pri pouku v (naziv šole)  
..... (v nadaljevanju: besedila), v  
namene izgradnje **Korpusa šolskih pisnih izdelkov Šolar**, ki ga izvaja izvajalec v okviru raziskovalnih programov  
**Slovenski jezik – bazične, kontrastivne in aplikativne raziskave (P6-0215)** in **Jezikovni viri in tehnologije za  
slovenski jezik (P6-0411)** na Univerzi v Ljubljani.

**S to pogodbo** brezplačno, časovno in krajevno neomejeno ter nepreklicno prenašam na izvajalca avtorske pravice  
zlasti reprodukcije, distribucije, dajanja v najem, priobčitev javnosti in predelave besedil ter dajem dovoljenje, da se  
besedila uporabijo za namene izgradnje **Korpusa šolskih pisnih izdelkov Šolar**.

S to izjavo učenec/-ka izjavljam, da sem imetnik vseh avtorskih, avtorski sorodnih pravic oziroma drugih pravic  
avtorja v skladu z ZASP na avtorskih delih, ki so predmet te izjave.

Izvajalec se zavezuje, učenec/-ka pa se strinja, **da bo vse osebne podatke**, ki se morebiti pojavljajo v besedilih,  
**anonimiziral in varoval** v skladu z Uredbo (EU) 2016/679 Evropskega parlamenta in Sveta z dne 27. aprila 2016  
o varstvu posameznikov pri obdelavi osebnih podatkov in o prostem pretoku takih podatkov ter o razveljavitvi  
Direktive 95/46/ES (Splošna uredba o varstvu podatkov), Zakonom o varstvu osebnih podatkov (ZVOP-1, UL RS  
86/4), Zakonom o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA, UL RS št. 30/2006), njunimi  
podzakonskimi akti in drugo veljavno zakonodajo.

Učenec/-ka izrecno soglašam, da navajanje avtorstva na besedilih ni potrebno.

**S to pogodbo** učenec/-ka nadalje izrecno soglašam z nadaljnjim prenosom materialnih pravic, navedenih v tej  
izjavi, na tretje osebe. Učenec/-ka sem seznanjen, da bo Korpus šolskih pisnih izdelkov Šolar na voljo javnosti pod  
pogoji licence Creative Commons 4.0 "priznanje avtorstva" in s tem izrecno soglašam. Ta licenca omogoča  
uporabnikom avtorsko delo in njegove predelave reproducirati, distribuirati, dajati v najem, priobčiti javnosti in  
predelovati samo pod pogojem, da navedejo avtorja tega dela, oz. v konkretnem primeru, ko avtor ne bo naveden,  
navedejo kot vir korpus Šolar.

Za razmerja v zvezi s to pogodbo se uporabljajo pravni predpisi Republike Slovenije.

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega  
sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske  
politike v obdobju 2014-2020.

Spore iz te pogodbe bosta stranki reševali po mirni poti. V primeru, da mirna rešitev ne bo mogoča, je za vse spore v zvezi s to pogodbo pristojno sodišče v Ljubljani.

Ta pogodba nadomešča vsa predhodna pogajanja, ponudbe in druge dogovore med strankama.

Ta pogodba je sestavljena v [dveh] istovetnih izvodih, od katerih prejme vsaka stranka po enega.

Kraj, datum:.....

Podpis starša/zakonitega zastopnika učenca/-ke:

.....

Podpis zastopnika izvajalca  
(Univerza v Ljubljani):

.....

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.