

Predstavitveni portal spletnih jezikovnih virov za slovenščino

Špela Arhar Holdt,* ♦ Kaja Dobrovoljc,* Iztok Kosem* ♦

* Zavod za uporabno slovenistiko Trojina (CUJT),
Trg republike 3, 1000 Ljubljana
spela.arhar@trojina.si, kaja.dobrovoljc@trojina.si, iztok.kosem@trojina.si
♦ Filozofska fakulteta Univerze v Ljubljani,
Aškerčeva 2, 1000 Ljubljana

Povzetek

Prispevek predstavlja Portal jezikovnih virov, rezultat dveh manjših projektov, ki ju je Ministrstvo za kulturo RS sofinanciralo v okviru *Javnega razpisa za sofinanciranje projektov, namenjenih predstavljanju, uveljavljanju in razvoju slovenskega jezika* v letih 2014 in 2015. V prispevku opišemo zasnovo, izvedbo in rezultate projektov. Portal jezikovnih virov je zasnovan kot strukturirana knjižnica videoposnetkov, ki na poljuden način predstavljajo vsebino in strukturo izbranih jezikovnih virov za slovenščino. Za izbrane vire so bila gradiva na portalu v sklopu aktivnosti *Vir meseca* dopolnjena z raznovrstnimi zanimivostmi, ki so v strnjeni obliki na voljo za nadaljnje diseminacijske in izobraževalne namene.

Portal for the Presentation of Language Resources for Slovenian Language

The paper presents the Portal of Language Resources, a result of two small projects funded in 2014 and 2015 by the Ministry of Culture of the Republic of Slovenia under the *Public call for funding projects focused on the presentation, promotion and development of the Slovene language*. The conception and rationale of both projects are discussed, and the results presented. The Portal is conceived as a structured library of videos which in a straightforward and clear manner present the content and structure of different language resources for Slovene. Various interesting aspects of the resources in the Portal were also promoted through the *Resource of the month* activity. All the information coming out of this activity is available for further dissemination and use for educational purposes.

1 Namen portala jezikovnih virov

S prehodom družbe v digitalno dobo se tudi za slovenščino viša število jezikovnih virov, priročnikov in orodij, ki so za rabo prosto dostopni na spletu. Avtorji teh izdelkov vedno uspešneje izrabljajo možnosti novega medija in veliko truda namenjajo zagotavljanju dostopnosti in prijaznosti svojih izdelkov, vendar ti v javnosti pogosto ostajajo premalo opaženi. Za diseminacijo, ki je v projektnih časovnicah tipična zadnja naloga, v praksi namreč pogosto zmanjka časa, izvedba izobraževanja uporabnikov pa običajno presega domet projektov, v katerih se viri pripravljajo. Na drugi strani naraščajoč izziv za uporabnike predstavlja raznolikost razpoložljivih možnosti, tako v smislu namena oz. vsebine kot konkretnih vmesniških rešitev. Tudi v primeru, da uporabniki določen vir na predstavitev ali izobraževanjih dovolj natančno spoznajo, ob neredni uporabi hitro izgubijo veščine, potrebne za uspešno pridobivanje in interpretacijo jezikovnih podatkov.¹

Da bi avtorjem jezikovnih virov, priročnikov in orodij poenostavili diseminacijo projektnih rezultatov, uporabnikom pa na enem mestu omogočili pregledno seznanitev z možnostmi, ki so trenutno na voljo, smo s pomočjo *Javnega razpisa za sofinanciranje projektov, namenjenih predstavljanju, uveljavljanju in razvoju slovenskega jezika* Ministrstva za kulturo RS v letih 2014 in 2015 pripravili predstavitveni portal spletnih jezikovnih virov za slovenščino. Portal, ki trenutno predstavlja 15 virov, je dostopen na spletni strani <http://viri.trojina.si>. Namen prispevka je predstaviti zasnovo, izvedbo in rezultate obeh projektov: *Izdelava spletne strani z opisi jezikovnih virov in orodij za slovenščino ter osnovnimi*

(video)navodili za njihovo uporabo ter Nadgradnja in popularizacija predstavitvenega portala spletnih jezikovnih virov za slovenščino.

2 Poglavlja s predstavitvenimi posnetki

Glavni doprinos Portala jezikovnih virov so izobraževalni posnetki, ki na zgoščen, poljuden način predstavljajo vsebino in strukturo posameznega obravnavanega jezikovnega vira, in posnetki, ki na konkretnih primerih uporabe kažejo, kako lahko v določenem viru najdemo odgovor na specifično jezikovno vprašanje. Posnetki so urejeni v obliki spletnega portfelja, v katerem vsako poglavje prinaša tudi povezavo na obravnavani jezikovni vir in projektno stran ter informacije o projektu in avtorjih.

Prioriteta pri pripravi posnetkov je bila zagotoviti optimalno uporabniško izkušnjo na različnih vrstah naprav, od računalnikov z velikimi zasloni do mobilnih telefonov z majhnimi. Proces editiranja posnetkov je bil v večji meri pogojen s prikazom na telefonih, za katerega je bilo treba zagotoviti ustrezno približevanje obravnavanim delom ekrana, vendar na način, da uporabnik pri gledanju posnetka ne izgubi občutka za vmesnik kot celoto. Na drugi strani smo širšo dostopnost vsebin za različne skupine uporabnikov in v različnih situacijah rabe skušali omogočiti tudi s pripravo slovenskih in angleških podnapisov.

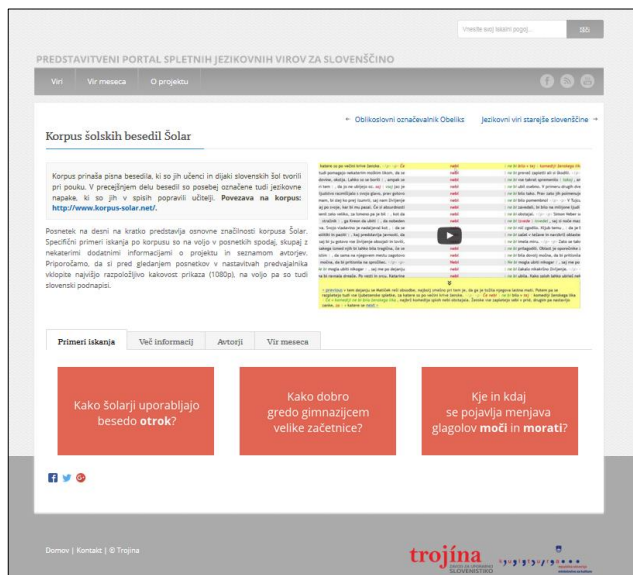
Posnetke smo pripravljali po naslednjem postopku: (I) pregled obravnavanega vira, preizkus iskalnih možnosti, rezultatov itd., (II) izbira reprezentativnih primerov in pisanje prve različice scenarija, (III) testno snemanje in popravljanje scenarija, (IV) snemanje posnetka, (V) editiranje posnetka, npr. krajšanje, približevanje, dodajanje

¹ V tem smislu indikativne so povratne informacije udeležencev jezikovnotehnološkega izpopolnjevanja učiteljev, ki je potekalo po slovenskih šolah v letih 2013 in 2014 (<http://ucitelji.sdjt.si/>).

Učitelji, ki so nove možnosti po večini ocenjevali zelo pozitivno, so kasneje poročali, da je novih vsebin preveč in so preveč raznolike, da bi imeli nad njimi celovit pregled.

napisov,² (VI) izvoz in objava na kanalu *YouTube*, (VII) dodajanje podnapisov in (VIII) vgradnja novega posnetka v ustrezno poglavje na strani portala. Pred objavo posnetka na portalu smo zbrali in upoštevali tudi povratne informacije predstavnikov avtorjev obravnavanih virov.

Slika 1 predstavlja poglavje, posvečeno korpusu šolskih besedil Šolar. Desno zgoraj je na voljo posnetek, ki pregledno predstavi glavne značilnosti korpusnega vmesnika in možnosti za izvedbo različnih vrst iskanj. Spodaj so na voljo povezave do treh krajših posnetkov, ki kažejo, kako poiskati in interpretirati korpusne podatke, da dobimo odgovor na specifična jezikovna vprašanja (npr. »Kje in kdaj se pojavlja menjava glagolov *moči* in *morati*?«). S klikom na zavihke med gornjim in spodnjim delom ekrana dostopamo do osnovnih informacij o jezikovnem viru in projektu, podatkov o avtorjih projekta, pri nekaterih poglavjih pa so na voljo tudi dodatne zanimivosti, ki so bile pripravljene v okviru akcije *Vir meseca* (več o tem v razdelku 4).



Slika 1: Poglavje na Portalu jezikovnih virov.

3 Predstavljeni jezikovni viri

Vire za predstavitev na portalu smo izbrali ob upoštevanju izkušenj z izvedbo jezikovnotehnološkega izpopolnjevanja učiteljev (Stritar in Dobrovoljc 2013), rezultatov spletnega vprašalnika (<http://viri.trojina.si/>) drugi-viri) ter zanimanja avtorjev jezikovnih virov za sodelovanje pri projektu. V nadaljevanju naštevamo trenutno predstavljene vire, skupaj z referenčno literaturo in povezavo na spletno mesto:

- referenčni pisni korpus Gigafida (Logar et al., 2012; www.gigafida.net),
- korpus govorne slovenščine GOS (Verdonik in Zwitter Vitez, 2011; www.korpus-gos.net),
- korpus šolskih besedil Šolar (Kosem et al., 2012; www.korpus-solar.net),

² Pri prvem od projektov smo za snemanje uporabili program *Community Clips*, pri drugem *Debut Video Capture Software*, ki je za razliko od prejšnjega plačljiv, vendar omogoča globlje zajemanje barv na zaslonu in nekatere dodatne funkcije, kot npr. barvno označevanje premika miške ali glasovno okrepljeno klikanje. Za editiranje posnetkov smo uporabili *iMovie*.

³ Akcija je potekala v sodelovanju s predstavniki avtorjev virov, ki so posredovali statistične podatke in zanimivosti v zvezi z

- slovarski portal Fran (Ahačič et al., 2015; www.fran.si),
- slovarski portal Termania (Romih in Krek, 2012; www.termania.net),
- rezultati projekta Viri starejše slovenščine IMP (Erjavec, 2015; <http://nl.ijs.si/imp/>),
- rezultati projekta Signor (Vintar et al., 2012; <http://lojze.lugos.si/signor/>),
- kolaborativni slovar Razvezani jezik (Dolar, 2014; <http://razvezanijezik.org/>),
- Jezikovna svetovalnica ISJFR ZRC SAZU (Dobrovoljc in Bizjak Končar, 2015; <http://isjfr.zrc-sazu.si/svetovalnica#v>),
- Terminologišče ISJFR ZRC SAZU (Žagar Karer, 2015; <http://isjfr.zrc-sazu.si/terminologisce#v>),
- jezikovnodidaktični Pedagoški slovnici portal (Arhar Holdt et al., 2016; www.slovnica.slovenscina.eu),
- leksikon besednih oblik Sloleks (Dobrovoljc et al., 2015; <http://www.slovenscina.eu/sloleks>),
- oblikoslovni označevalnik Obeliks (Grčar et al., 2012; www.slovenscina.eu/tehnologije/oznacevalnik),
- digitalizirani starejši pravopisi (<http://www.trojina.org/pravopisi>),
- in digitalizirane starejše slovnice (<http://www.amebis.si/slovnice>).

Čeprav gornji seznam seveda ni popoln in zaključen, v trenutnem obsegu predstavlja solidno izhodišče za rabo portala v (samo)izobraževalne oz. diseminacijske namene. V nadaljevanju si želimo portal dopolnjevati z novimi poglavji, izziv pa predstavlja tudi posodabljanje vsebin pri virih, ki se razvijajo, k čemur se vračamo v razdelku 5.

4 Diseminacijska aktivnost Vir meseca

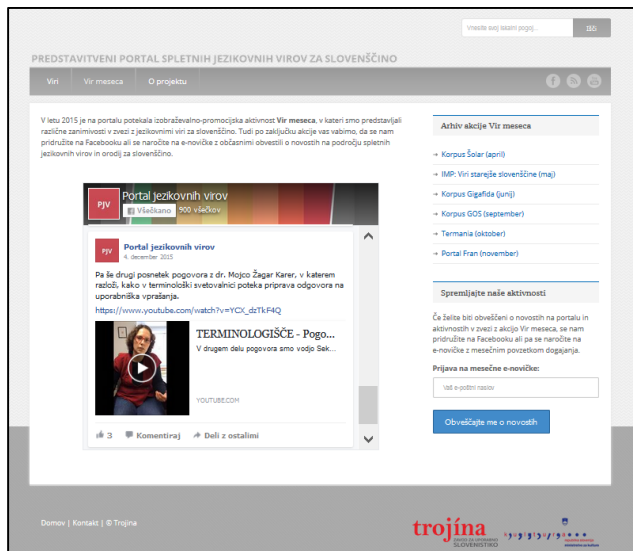
V letu 2015 je potekala v časovnem smislu najbolj obsežna projektna aktivnost, ki smo jo imenovali *Vir meseca*. Namen akcije je bil kontinuirano (v trajanju enega meseca) opozarjati širšo javnost na izbrani jezikovni vir in s pomočjo zanimivosti izobraževati uporabnike ter jih motivirati, da vire v praksi preizkusijo.

Za promocijo so bili izbrani: korpus Šolar (april), Viri starejše slovenščine IMP (maj), korpus Gigafida (junij), korpus GOS (september), portal Termania (oktober) in slovarski iskalnik Fran z Jezikovno svetovalnico in Terminologiščem ISJFR ZRC SAZU (november). Promocija je potekala po treh glavnih kanalih: prek predstavitevne strani na omrežju Facebook (<https://www.facebook.com/jezikovniviri>), v obliki mesečnih e-novičk in na poštnem seznamu SlovLit (<https://mailman.ijs.si/mailman/listinfo/slovlit>).

Tekom akcije smo vsakega od zgoraj navedenih virov promovirali z rednimi (tipično tremi na teden) objavami na omrežju Facebook. Objave so vsebovale zanimivosti glede priprave vira, informacije o gradivu (npr. različne statistike o jeziku, besedne oblake in druge vrste slikovnega gradiva), predstavitevne posnetke, ideje za uporabo vira v didaktične namene, kratke posnetke pogovora z avtorji in podobno.³

gradnjo, svoje projektne izkušnje in vizijo za prihodnji razvoj. Sodelovali so: dr. Tadeja Rozman (korpus Šolar), dr. Tomaž Erjavec in Katja Zupan (viri starejše slovenščine), dr. Nataša Logar (Gigafida), dr. Ana Zwitter Vitez (GOS), Miro Romih (Termania), dr. Helena Dobrovoljc (Jezikovna svetovalnica) in dr. Mojca Žagar Karer (Terminologišče). Ostali avtorji (po večini gre za projekte z visokim številom sodelujočih) so naštetih v ustrezajočih poglavjih na spletni strani portala.

Za izdelavo mesečnih novičk smo izbrali program *MailChimp*, ki ob kreiranju poštnega sporočila ustvari spletno mesto, na katerem ostane vsebina sporočila trajno dostopna. Z uporabo te možnosti smo promocijske vsebine vključili na spletno stran portala, kjer so prosto na voljo za nadaljnje promocijsko-diseminacijske in izobraževalne aktivnosti.



Slika 7: Arhiv *Vir meseca* na Portalu jezikovnih virov.

Za konec navajamo Tabela 1, v kateri so podatki o številu ogledov pripravljenih posnetkov do 17. maja 2016. Podatki so pridobljeni s kanala Youtube in združujejo ogled osnovnih predstavitvenih posnetkov, kratkih posnetkov na temo izbranih jezikovnih vprašanj in intervjujev, ki so bili pripravljene v akciji *Vir meseca*. Podatki posredno razkrivajo, za spoznavanje katerih virov je med uporabniki največ interesa, in nakazujejo mesta, ki bi se jim bilo v prihodnje smiselno dodatno posvetiti.

Število ogledov	Osnovna predstavitev	Kratki posnetki	Intervju z avtorjem	Skupaj - vir
Termania	79		69	148
Signor	44			44
Jezikovna svetovalnica	188		94	282
Pedagoški slovnčni portal	125			125
Razvezani jezik	107			107
Terminologišče	94		74	168
Gos	189	129		318
Gigafida	208	171	173	552
Šolar	251	200	211	662
Sloleks	248	134		382
IMP	105	149		254
Fran	237			237
Obeliks	128			128
Starejše slovnice	61			61
Starejši pravopisi	78			78
Skupaj - portal				3546

Tabela 1: Število ogledov posnetkov na kanalu Youtube.

5 Zaključek in nadaljnje delo

Po zaključku projektov je Portal jezikovnih virov dosegel velikost, ko ga je mogoče uporabljati kot samostojen vir za (samo)izobraževalne namene. Ker so vsebine pripravljene za širšo javnost in optimalno uporabniško izkušnjo na

različnih vrstah naprav, je vrednost predstavitvenih posnetkov velika. Potencial za nadaljnjo rabo imajo tudi zanimivosti, zbrane za akcijo *Vir meseca*, ki bi jih bilo mogoče preoblikovati, da bi bile neposredno uporabne za izobraževalne namene.

Glavna naloga za naprej je dopolniti portal z novimi poglavji in z obstojem vsebin seznaniti čim širši nabor potencialnih uporabnikov. Projektno financiranje razvoja portala se je sicer zaključilo, zato bi bilo v nadaljevanju k pripravi vsebin smiselno aktivneje vključiti zainteresirane avtorje virov, ki bi prek obstoječe platforme lahko ponudili lastne diseminacijske vsebine. Kontinuirano sodelovanje bi bilo dobro vzpostaviti tudi z avtorji virov, ki se nadgrajujejo ali pogosteje spreminjajo oz. dopolnjujejo. Za slednje bi bilo mogoče v predstavitvena poglavja vnesti posebno rubriko, kjer bi bile sproti predstavljene nadgradnje oz. posodobitve. Tako bi zagotovili ažurnosti in uporabnost portala tudi v prihodnje.

6 Zahvala

Projekta *Izdelava spletne strani z opisi jezikovnih virov in orodij za slovenščino ter osnovnimi (video)navodili za njihovo uporabo in Nadgradnja in popularizacija predstavitvenega portala spletnih jezikovnih virov za slovenščino* je sofinanciralo Ministrstvo za kulturo Republike Slovenije v sklopu *Javnega razpisa za sofinanciranje projektov, namenjenih predstavljanju, uveljavljanju in razvoju slovenskega jezika* v letih 2014 in 2015 (JPR-UPRS-2014 in JPR-UPRS-2015). Posebej se zahvaljujemo avtorjem jezikovnih virov, ki so sodelovali pri pripravi promocijskega gradiva: Helena Dobrovoljc, Tomaž Erjavec, Nataša Logar, Miro Romih, Tadeja Rozman, Katja Zupan, Ana Zwitter Vitez in Mojca Žagar Karer, ter anonimnima recenzentoma prispevka za koristna dopolnila.

7 Literatura

- Kozma Ahačič, Nina Ledinek in Andrej Perdih. 2015. Portal Fran – nastanek in trenutno stanje. V: M. Smolej, ur., *Slovnica in slovar – aktualni jezikovni opis, Obdobja 34*, str. 57–66. Znanstvena založba Filozofske fakultete, Ljubljana.
- Špela Arhar Holdt, Iztok Kosem in Polona Gantar. V pripravi. Corpus-based resources for L1 teaching: The case of Slovene. V: A. Marcus-Quinn, ur.: *Handbook on Digital Learning for K-12 Schools*. Springer.
- Helena Dobrovoljc in Aleksandra Bizjak Končar. 2015. Pravopisno slovaropisje na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU. *Slavia Centralis*, 8(1): 34–50.
- Kaja Dobrovoljc, Simon Krek in Tomaž Erjavec. 2015. Leksikon besednih oblik Sloleks in smernice njegovega razvoja. V: V. Gorjanc, P. Gantar, I. Kosem in S. Krek, ur., *Slovar sodobne slovenščine: problemi in rešitve*, str. 80–105. Znanstvena založba Filozofske fakultete, Ljubljana.
- Kaja Dolar. 2014. Kolaborativni slovar Razvezani jezik. *Slavistična revija* 62(2): 235–252.
- Tomaž Erjavec. 2015. The IMP historical Slovene language resources. *Language resources and evaluation* 49/3, str. 753–775.
- Miha Grčar, Simon Krek in Kaja Dobrovoljc. 2012. Obeliks: statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik V: T. Erjavec, J. Žganec

- Gros, ur., *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012: zbornik 15. mednarodne multikonference. Informacijska družba – IS 2012, zvezek C*, str. 89–94. Institut Jožef Stefan, Ljubljana.
- Iztok Kosem, Mojca Stritar Kučuk, Sara Može, Ana Zwitter Vitez, Špela Arhar Holdt in Tadeja Rozman. 2012. *Analiza jezikovnih težav učencev: korpusni pristop*. Zavod za uporabno slovenistiko Trojina, Ljubljana.
- Nataša Logar, Miha Grčar, Marko Brakus, Tomaž Erjavec, Špela Arhar Holdt in Simon Krek. 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Zavod za uporabno slovenistiko Trojina; Fakulteta za družbene vede, Ljubljana.
- Miro Romih in Simon Krek. 2012. Termania – prosto dostopni spletni slovarski portal V: T. Erjavec, J. Žganec Gros, ur., *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012: zbornik 15. mednarodne multikonference Informacijska družba - IS 2012, zvezek C*, str. 163–166. Institut Jožef Stefan, Ljubljana.
- Mojca Stritar in Kaja Dobrovoljc. 2013. Korpusi na poti v šole: jezikovnotehnološko izpopolnjevanje učiteljev. *Slovenščina 2.0*, 1(1): 181–194.
- Darinka Verdonik in Ana Zwitter Vitez. 2011. *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Špela Vintar, Boštjan Jerko in Marjetka Kulovec. 2012. Compiling the Slovene sign language corpus. V: *8th International Conference on Language Resources and Evaluation, 21–27 May 2012, Istanbul, Turkey. LREC 2012: proceedings*, str. 159–162. ELRA, Istanbul.
- Mojca Žagar Karer. 2015. Terminologiče – kraj, kjer terminolog išče. *Slavia Centralis*, 8(1): 22–33.