

Špela Arhar Holdt, Trojina, zavod za uporabno slovenistikoin Univerza v Ljubljani, Filozofska fakulteta

Polona Gantar, Univerza v Ljubljani, Filozofska fakulteta

Simon Krek, Institut Jožef Stefan in Univerza v Ljubljani, Center za jezikovne vire in tehnologije

V UČILNIČE S SODOBNO SLOVENŠČINO: REZULTATI PROJEKTA *SPORAZUMEVANJE V SLOVENSKEM JEZIKU (SSJ)*

Med rezultati projekta SSJ so trije inovativni jezikovni viri za slovenščino: (1) *Leksikalna baza* kot izhodišče za pripravo novih slovarskih opisov, (2) *Slogovni priročnik* odgovarja na tipične pravopisne zadrege, (3) *Pedagoški slovnični portal* na enostaven način razlaga slovnične pojave, s katerimi imajo učenci in dijaki pri pisanju največ težav.

1 Uvod

Obdobje od začetka tisočletja do danes je bilo in ostaja tehnološko izjemno intenzivno, pri čemer je bil razvoj informacijsko-komunikacijskih tehnologij vedno v prvih vrstah tega procesa. V tem času smo lahko spremljali neslušten razvoj računalništva, interneta, kasneje mobilne telefonije in mobilnega interneta. Svet se je v določenem trenutku zavedel, da stoji na pragu nečesa, kar je dobilo ime »informacijska družba«, v kateri se mnoge dejavnosti, pri katerih smo nekaj stoletij uporabljali pretežno tiskani medij, torej papir, radikalno selijo v digitalno okolje. Skupaj s svetovnim spletom to omogoča povsem nove dimenzije komunikacije in dostopa do informacij. V zadnjem času je kot konica tega razvoja izpostavljeno področje umetne inteligence in glede na trenutno stanje lahko predvidevamo, da bodo naloge, pri katerih si je bilo do sedaj nemogoče predstavljati, da jih opravlja računalnik, sčasoma postale rutinske za prihodnje umetnointeligence sisteme. Ob tem postaja izjemno kritično tudi vprašanje opremljenosti jezikov za prehod v (umetnointeligentno) digitalno dobo. Nekateri avtorji opozarjajo (npr. Kornai 2013), da tega prehoda ne bodo zmogli jeziki, ki bodo predaleč zaostali za možnostmi, ki jih bodo ponujali jezikovnotehnološko boljše opremljeni jeziki.

Tudi izobraževanje ni ostalo izven teh dogajanj. Uvajanje IKT v vse segmente izobraževanja je bilo eden od prepoznavnejših trendov v tem obdobju. Hkrati s tem so se razvijale tudi različne pedagoške in didaktične metode, ki so izkoriščale nove tehnološke možnosti, od obrnjenih učilnic (ang. flipped classroom) do množičnih odprtih spletnih tečajev (ang. massive open online course, MOOC). Toda ob tem se vedno znova pojavlja vprašanje, ali bodo govorke in govorci slovenščine sploh lahko sodelovali v teh novih

izobraževalnih trendih, če bo slovenščina kot tehnološko razmeroma slabo opremljen jezik (prim. Krek 2012a) zaostajala za drugimi jeziki pri prehodu v digitalno okolje. Deloma je to nevarnost omilil projekt, ki ga opisujemo v nadaljevanju.

2 Predstavitev projekta

Projekt *Sporazumevanje v slovenskem jeziku*¹ je bil petletni projekt (2008–2013), namenjen pripravi novih jezikovnih virov za slovenščino. Projekt sta financirala Evropska unija iz Evropskega socialnega sklada ter Ministrstvo za šolstvo in šport.²

Središčna projektna aktivnost je bila izdelava **novih besedilnih korpusov za slovenščino**, med katerimi so ključnega pomena pisni referenčni korpus Gigafida z uravnoteženim podkorpusom Kres (Logar Berginc idr. 2012), korpus govornjene slovenščine GOS (Verdonik in Zwitter Vitez 2012) in korpus šolskih besedil Šolar (Rozman idr. 2012). V povezavi s pripravo korpusov je potekala tudi izboljšava avtomatskega označevanja slovenskih besedil na oblikoskladenjski in skladenjski ravni (Grčar idr. 2012; Dobrovoljc idr. 2012), pa tudi razvoj novega, uporabniško prijaznega korpusnega vmesnika (Arhar Holdt idr. 2012).

Na podlagi novih korpusnih podatkov so bili pripravljene trije vsebinsko in oblikovno inovativni jezikovni viri za slovenščino. *Leksikalna baza za slovenščino* je izhodišče za pripravo na sodobni slovenščini temelječih slovarskih opisov, strukturiranih v obliki, ki jo je mogoče uporabiti tudi za jezikovno-tehnološke namene. *Slogovni priročnik* s kombinacijo napredne obdelave jezikovnih podatkovnih baz in odgovorov strokovnjaka odgovarja na tipične pravopisne, pravorečne, slovnične in slogovne zadrege jezikovnih uporabnikov. *Pedagoški slovnični portal* na enostaven način razlaga slovnične pojave, s katerimi imajo učenci in dijaki pri šolskem pisanju največ težav, obenem pa uči uporabe sodobnih jezikovnih virov. V nadaljevanju po vrsti predstavljamo našete tri portale, pri čemer se deloma opiramo na besedilo, ki je navedeno na projektni spletni strani <http://www.slovenscina.eu>.

3 Leksikalna baza za slovenščino

V delu projekta *Sporazumevanje v slovenskem jeziku*, ki je zajemal pripravo podatkovnih baz,³ je bila izdelana tudi *Leksikalna baza za slovenščino* (LBS), katere namen je bil oblikovati izhodišča za celovit leksikografski opis sodobne slovenščine in hkrati zagotoviti jezikovne podatke v obliki, ki bo služila računalniški obdelavi jezika in razvoju sodobnih jezikovnotehnoloških aplikacij za slovenščino. Tako zasnovan opis naj bi bilo mogoče integrirati v druge leksikalne vire za slovenščino ter ga uporabiti kot izhodišče za izdelavo drugih za slovenščino potrebnih leksikalnih in slovničnih opisov, kot so denimo slovar sodobnega slovenskega jezika (Gorjanc idr. 2015), kolo-kacijski slovar slovenskega jezika (Krek idr. 2016), vezljivostni leksikon slovenskih glagolov ipd.

Leksikalna baza je bila zasnovana na študiji sorodnih projektov evropskih jezikov (prim. *Nizozemski splošni slovar*,⁴ *Veliki slovar poljskega jezika*⁵ in

¹ Spletna stran: <http://www.slovenscina.eu>.

² Financiranje je potekalo v okviru Operativnega programa razvoja človeških virov za obdobje 2007–2013 (razvojna prioriteta: razvoj človeških virov in vseživljenjskega učenja; prednostna usmeritev: izboljšanje kakovosti in učinkovitosti sistemov izobraževanja in usposabljanja 2007–2013).

³ Poleg *Leksikalne baze* je bil v tem sklopu pripravljen tudi leksikon besednih oblik *Sloleks* (Dobrovoljc idr. 2015), ki je dostopen na povezavi: <http://www.slovenscina.eu/sloleks>.

⁴ *Algemeen Nederlands Woordenboek*: <http://anw.inl.nl/show?page=search1>.

⁵ *Wielki słownik języka polskiego*; demo različica: <http://www.wsjp.pl/>.

druge; Gantar 2015), ki temeljijo na spoznanjih sodobne pomenoslovne teorije (Fillmore idr. 2003; Hanks 2013), inovacijah na področju e-leksikografije (Granger in Paquot 2010) in uporabniških raziskavah (Lew 2015a, 2015b). Dobre prakse sorodnih projektov smo v procesu izdelave leksikalne baze prilagodili specifikam slovenščine in upoštevali možnosti prenosa inovativnih slovarskih praks, zlasti na področju pomenskega opisa in oblikovanja slovarskih razlag, v slovensko okolje. Celostni leksikalni opis, ki ga prinaša LBS, izhaja primarno iz kontekstualne rabe jezika in sporočanjjskih okoliščin ter temelji izključno na realni podobi sodobne slovenščine.

Izbor besed v leksikalni bazi temelji na geslovníku 5.000 najpogostejših besed v korpusih FidaPLUS in Gigafida, upoštevali pa smo tudi izbor besed iz osnovno- in srednješolskih učbenikov, da bi se čim bolj približali realnim problemom šolarjev in dijakov, saj je bil eden od primarnih ciljev LBS izdelati model jezikovnega opisa, ki bi zlasti šolarjem in učencem slovenščine kot tujega jezika ponudil podatke o pomenu besed, njihovem tipičnem okolju, posebnostih rabe, stalnih zvezah in frazeologiji.

Ob zaključku projektne aktivnosti je LBS obsegala 2.500 iztočnic⁶ oziroma 10.946 leksikalnih enot, tj. pomenov, podpomenov, stalnih zvez in frazeoloških enot, poleg tega pa še približno 50.000 kolokacij oz. tipičnih besednih sopojavitvev, ki nimajo samostojnega pomena, kažejo pa tipično in naravno rabo besed oz. njihovih pomenov v skladijskem in besedilnem okolju, ter več kot 7.000 stavčnih vzorcev in več kot 150.000 korpusnih zgledov, ki potrjujejo v leksikalni bazi opisane leksikalne podatke.

Kot prikazuje Slika 1, je slovarska baza organizirana kot mreža medsebojno povezanih leksikalno-gramatičnih podatkov na šestih hierarhičnih nivojih.

Na prvem nivoju je izpostavljena iztočnica v osnovni obliki skupaj s podatkom o besedni vrsti.⁷ Na zadnjem nivoju so obravnavane frazeološke enote, ki predstavljajo samostojni razdelek geselskega članka in so z iztočnico povezane zgolj tehnično. Zanje velja, da so pomensko neprozorne in za razliko od drugih večbesednih enot ekspresivne oz. metaforične.

Na pomenskem nivoju, ki je hierarhično najvišji glede na druge podatke, so zabeleženi pomeni, ki jih obravnavamo kot samostojne leksikalne enote. Pomenski opis je sestavljen iz treh medsebojno povezanih informacij: pomenskega indikatorja, ki je kratek, največkrat enobesedni sinonim, pomenske sheme, ki predstavlja izhodišče za izdelavo stavčne razlage, ter iz oznake. Teh je lahko več vrst, v osnovi pa smo ločili slovnične, področne in komunikacijske. Večbesedne enote, ki nimajo metaforičnega pomena, med katere sodijo stalne besedne zveze in področno specifični izrazi, kot npr. *pametni telefon*, *topla greda*, *jedilna čokolada* ipd., smo obravnavali bodisi kot samostojne razdelke gesla, tj. od pomena neodvisno, ali pa smo jih beležili znotraj posameznega pomena iztočnice. V tem primeru stalne zveze izkazujejo pomensko povezavo s katerim od registriranih pomenov iztočnice, npr. *bazičen* – ‚osnovni; temeljni‘: *bazična znanost*.

Na skladijskem nivoju smo beležili tri tipe podatkov: pri glagolskih iztočnicah smo navajali stavčne vzorce, v katerih se glagoli pojavljajo, sledijo t. i. skladijske zveze kot tipični, zlasti predložni besedni nizi, ki jih v jeziku uporabljamo kot skladijsko in oblikoslovno spremenljivo celoto, npr. *pod*

⁶ Obseg je bil določen na podlagi presoje, da slovarska obdelava 2.500 lem lahko zagotovi zadostno količino različnih leksikalnih in leksikografskih problemov, na podlagi katerih bi bilo mogoče določiti vrste leksikalno-gramatičnih podatkov in način njihove notranje organizacije, ki bi učinkovito delovala tudi na obsežnejšem izboru.

⁷ Pri vključevanju slovničnih podatkov smo se izogibali uporabi krajšav, s čimer smo se želeli izogniti morebitnim napačnim interpretacijam (prim. Rozman 2010), hkrati pa smo želeli razbremeniti uporabnike s seznanjanjem slovarskega metajezika, saj ta zunaj slovarja nima informativne vrednosti. Neokrajšane opise omogoča tudi elektronski medij, za katerega je bila LBS zasnovana, saj prostorske omejitve, ki so značilne za tiskani slovar, zanj ne obstajajo.

okriljem česa, v zameno za koga/kaj ipd., navadno pa ne potrebujejo pomenskega opisa. Tretji podatek – skladijske strukture – prikazuje življenje besede v danem pomenu znotraj njene besednozvezne skladnje, kar je ključno za avtomatsko pridobivanje kolokacijskih podatkov iz korpusa.

Na kolokacijskem nivoju smo beležili kolokatorje, tj. besede, ki se v določenem pomenu tipično povezujejo z besedo, ki je obravnavana kot iztočnica. Take povezave – kolokacije – so v slovarski bazi beležene glede na že omenjene skladijske strukture.

Samostojni nivo je bil namenjen beleženju korpusnih zgledov, ki potrjujejo vse registrirane podatke v slovarski bazi.

I. LEMA	• iztočnica • besedna vrsta	grmeti <i>glagol</i>	
II. POMEN	• indikator	1 oddajati glasen zvok 1.1 o nevihti	2 glasno govoriti
	• pomenska shema	kadar grmi, se sliši glasen doneč zvok, navadno ko dežuje ali se približuje nevihta	če ČLOVEKE grmi, govori zelo glasno in odločno, navadno, če se jezi ali če se s kom ali s čim ne strinja
	• oznaka	samo v 3. osebi	
III. SKLADNJA	• vzorec	----- -----	▪ kdo grmi ▪ kdo grmi s česa ▪ kdo grmi nad kom/čim
	• sklad. zveza	grmeti in se bliškati	
	• struktura	Glag + Inf-GLAG prisl + GLAG	prisl + GLAG
IV. KOLOKACIJE	• kolokacija	[začeti] grmeti [zunaj, močno] grmi	[glasno] grmeti grmeti z [odra]
V. ZGLEDI	• zgled	Bili so ravno v restavraciji, ko je začelo grmeti.	Grmel je z govorniškega odra, dokler mikrofoni neneadoma umolknil.
	• stalna zveza	-----	-----
VI. FRAZEOLOGIJA	• fraz. enota • indikator	Če malega travna grmi, slane se kmet več ne boji. pregovor	

Slika 1: Zgradba Leksikalne baze za slovenščino

Največ pozornosti smo pri izdelavi LBS namenili dejstvu, da pomenskega opisa ni mogoče strogo ločiti od skladijskega okolja besede. Tu smo sprejeli za slovenščino največ novih leksikografskih rešitev. Oddaljili smo se od jezikovnosistemskega koncepta, ki definira besedni pomen na podlagi prepoznavanja kategorialnih slovničnih in pomenskih sestavin pomena (Vidovič Muha 2000) in išče niz zadostnih ali potrebnih pogojev za njegovo realizacijo. Izhajali smo namreč iz spoznanja, da se pomeni besed oblikujejo v konkretni sporočanjski situaciji in so odvisni od vsakokratnega konteksta, iz česar sledi, da se govorci pomenov besed učimo iz kontekstov, v katerih jih tipično slišimo oz. uporabljamo. Na podlagi tega smo sporočanjško izhodišče uporabili pri vseh tistih pomenskih opisih, pri katerih je opis interakcije med udeleženci ključen za razbiranje pomena, zlasti torej pri pomenski razlagi glagolov, frazeoloških enot ter pri razlagi metaforičnih pomenov in posebnosti rabe (vrednotenje, stil, register, okoliščine). Na ta način so podatki o pomenu besed, njihovem tipičnem besedilnem okolju, posebnostih rabe, stalnih zvezah in frazeologiji v LBS ubesedeni na način, ki ga uporabnik pozna iz vsakodnevne komunikacije.

Z vidika potencialnih uporabnikov so splošnemu in šolskemu uporabniku namenjeni predvsem pomenski opisi besed v obliki kratkih pomenskih indikatorjev, ki tvorijo pomenske menije za lažjo navigacijo po večpomenskem geslu, ter pomenske sheme, ki vsebujejo stavčne razlage. Koloka-

cije in korpusni zgledi prikazujejo besede v njihovem realnem in najbolj tipičnem besedilnem okolju, zato predstavljajo neposredno informacijo o življenjskem okolju besed, ki je zlasti pomembna za učenje slovenščine kot tujega jezika. Jezikoslovec bo sposoben prepoznati temeljne vezljivostne vzorce v stavčno oblikovanih razlagah in jih povezati z različnimi možnimi skladejskimi realizacijami, ki jih govorci slovenščine najpogosteje uporabljamo v pisni komunikaciji. Primarno jezikovnim tehnologom pa so na voljo kodirane skladejske strukture in stavčni vzorci za vsak registrirani pomen besede, kar je predpogoj za izboljšanje avtomatskega označevanja slovenskih besedil na oblikoskladejski, skladejski in pomenski ravni ter za razvoj jezikovnotehnoloških aplikacij za slovenščino. Slika 2 prikazuje del slovarskega gesla za iztočnico *globalen* v LBS.

globalen *pridevnik*

1 svetovni
1.1 o okoljskih spremembah
1.2 splošno veljaven
2 ki zadeva celoto

1 svetovni
 globalni PROCESI zlasti EKONOMSKI in KOMUNIKACIJSKI potekajo po vsem svetu
PRIDEVNIKO + SAMOSTALNIKO

globalna [ekonomija, konkurenčnost, komunikacija]
globalni [kapitalizem, trg, problem]

- Angleški jezik postaja jezik *globalne* komunikacije.
- Epidemija virusa HIV in boleznai aids dosega *globalne* razsežnosti in je v zelo dinamičnem porastu.

◆ **globalna vas**
 če človek poimenuje svet *globalna vas*, ga vidi kot majhno okolje, v katerem je med seboj vse povezano in soodvisno

1.1 o okoljskih spremembah
ekologija
 globalne PODNEBNE SPREMEMBE vplivajo na celotni zemeljski planet

◆ **globalno segrevanje/ogrevanje**
globalno segrevanje je postopno zviševanje povprečne temperature v Zemljinem ozračju in oceanih

1.2 splošno veljaven; razširjen
 če postanejo neke DEJAVNOSTI ali LASTNOSTI globalne, jih upošteva vedno več ljudi po svetu
POMOŽNI GLAGOL + PRIDEVNIKO

[postajati, postati] **globalen**

- Merila, kakšna ženska je lepa, postajajo vse bolj *globalna*.

2 ki zadeva celoto; celostni
 globalni PROCES se nanaša na pojav v celoti

- S pomočjo plakata postanejo učencem misli bolj jasne in končno odkrijejo *globalni* pomen besedila.

◆ **globalna spremenljivka**
računalništvo

◆ **globalno potrdilo o lastništvu**
bančništvo

↳ Slika 2: Del gesla⁸ za iztočnico *globalen* v LBS

Ob koncu projekta smo podatke iz LBS prikazali v demo verziji *Spletnega slovarja slovenskega jezika*: <http://www.slovenscina.eu/spletni-slovar>, celotna slovarska baza, vključno z definicijo zgradbe (DTD; Document type Definition), pa je v formatu XML jezikovni skupnosti na voljo pod licenco Creative Commons Priznanje avtorstva-Nekomercialno-Deljenje pod enakimi pogoji 2.5 Slovenija.⁹

4 Slogovni priročnik

Slogovni priročnik je spletni portal, ki pomaga reševati tipične pravopisne, slovnične in slogovne težave govorcev slovenščine. Namen portala je zago-

⁸ Razvoj različnih leksikalnih, zlasti elektronskih virov, zahteva tudi prilagoditev izrazja: v prispevku uporabljamo izraz geslo kot sinonim za geselski članek (prim. Gantar 2015). Za razliko od izraza slovarski članek, ki implicira zgolj geslo v slovarju, imamo v našem primeru v mislih tudi opise v slovarskih bazah in drugih neslovarskih leksikalnih virih.

⁹ Leksikalno bazo je v stisnjeni datoteki mogoče prenesti s spletne strani projekta: <http://www.slovenscina.eu/spletni-slovar/prenos>.

točiti, da vsak uporabnik spleta – tudi brez specialističnega jezikoslovnega znanja – lahko pridobi informacije o jezikovni zadregi, pred katero se je znašel v takšni ali drugačni konkretni jezikovni situaciji (Krek 2012b). Slogovni priročnik je brezplačno dostopen na naslovu slogovni.slovenscina.eu, kjer je zasnovano in delovanje portala mogoče preizkusiti na 15 demonstracijskih zgledih.

Portal ponuja iskalno okence, v katerega uporabnik vtipka svoje jezikovno vprašanje. Odgovori so podani v obliki sestavkov, ki so oblikovno razdeljeni v tri rubrike (Slika 3).

Slogovni priročnik [Oglejte si kazalo vsebine](#)

www.slovenscina.eu
sporazumevanje v slovenskem jeziku

klemna ali klemena

Všeč mi je Tweet Priporoči to v Googlu

Sklanjanje moških samostalnikov z neobstojnimi samoglasnikom: slovenska lastna imena

KRATKO IN JEDRNATO

Na grafu si lahko ogledate podatke o rabi oblik lastnega imena **Klemen** v korpusu Gigafida. Pogostost je prikazana za niz oblik po variantah, npr. **Klemen, Klemna, Klemnu** itd. in **Klemen, Klemena, Klemenu** itd. Obe varianti sta skladni s trenutnim pravopisnim standardom.

Oblika	Pogostost
Klemna	3.541
Klemena	1.893

NA DOLGO IN ŠIROKO

Samostalnike moškega spola, ki se končujejo na polglasnik, zapisan s črko **e**, in soglasnik (npr. **pisec, boben**) praviloma sklanjamo tako, da polglasnik izpustimo: npr. **pozabil sem ime pisca; luknje v bobnu**. Polglasnik izpade pri tistih samostalnikih, katerih osnova se pred neničto končnico končuje na sklop nezvočnika in zvočnika (**posel – posla**), isto velja za samostalnike s priponskim obrazilom -c ali -k (**starec, dedek – starca, dedka**). Pri nekaterih samostalnikih polglasnik sicer ohranjamo, pogosto tam, kjer bi izpust polglasnika povzročil težji ali popačen izgovor besede: npr. **jezdec – dvaindvajset jezdecev** (ne ***jezdcev**).

Enako pravilo velja tudi za slovenska lastna imena (npr. **Kolšek – pogovor s Petrom Kolškom**), tudi med njimi pa so izjeme, kjer polglasnik ohranjamo: npr. **Kocbek – razprava je posvečena Kocbeku** (ne ***Kocbku**).

Obstajajo tudi posamezna imena, pri katerih se poleg standardne oblike pogosto uporabljajo tudi nestandardne: npr. **Klemen – za Klemna/Klemena**. V nekaterih imenih so neobstojni tudi samoglasniki **a**, **i** in **o**: npr. **Jakac, Sajovic, Vovčok – Jakca, Sajovca, Vovčka** –, obstaja pa tudi skupina imen, kjer raba omahuje: npr. **Lajovic – koncert je bil posvečen skladatelju Aleksandru Lajovcu/Lajovcu**. Ne glede na predpis vedno prevlada tudi morebitna osebna preferenčna oblika imenovane osebe. Če nismo prepričani, katera oblika se uporablja pogosteje, lahko podatke o tem dobimo v korpusu.

ZA NAVDUŠENCE

Slovenski pravopis – pravila (2001):
 Stran 82 - Besede in oblike z neobstojnimi samoglasnikom
 Stran 90 - Posebnosti 1. moške (o-jevske) sklanjatve – krajšanje osnove

Preverite tudi, kaj o vašem iskalnem pogoju pravijo digitalizirani slovenski pravopisi in starejše slovnice, ki so izšli v obdobju od 1899 do 2001.

▼ Slika 3: Sestavek, ki predstavlja sklanjanje imena *Klemen*

- Rubrika *Kratko in jedrnato* podaja grafični prikaz distribucije jezikovnih variant, kot se izkazujejo v referenčnem korpusu, npr. primerjava pogostosti oblike *Klemna* v primerjavi z obliko *Klemena*.

- Rubrika *Na dolgo in široko* ponuja daljšo razlago za tiste uporabnike, ki jih zanimajo jezikovna pravila oz. zakaj v primerih, povezanih z njihovim vprašanjem, prihaja do normativnih zadreg. Razlaga je napisana v poljudnem jeziku, vsi uporabljeni strokovni izrazi pa so pojasnjeni v dvižnih menijih.
- Rubrika *Za navdušence* poveže uporabnike z jezikoslovnimi viri, ki obravnavajo izbrano tematiko. Zbirko jezikoslovnih virov sestavljajo digitalizirana strokovna besedila, ki so bila prvotno objavljena v monografijah, strokovnih revijah ali dnevnem časopisju.

Priprava *Slogovnega priročnika* je potekala v več sklopih. Na programerski oz. jezikovnotehnoški strani je bil pomemben korak usmerjen v razvoj sistema, ki avtomatsko črpa jezikovne informacije iz besedilnega korpusa in leksikona besednih oblik in jih prikaže uporabniku na pregleden način. Ta sistem omogoča dinamično prikazovanje podatkov, ki se lahko nenehno prilagajajo novemu stanju v jeziku. Ko se posodobi in poveča besedilni korpus, se statistični podatki iz korpusa prelijejo v leksikon besednih oblik, na portalu pa so ti podatki uporabljeni za vizualizacijo brez potrebe po poseganju v sistem (Krek 2012b). Ta sistem je uporabljen pri izdelavi grafov v rubriki *Kratko in jedrnato*. Programersko zahtevnejši korak je bil tudi razvoj sistema za iskanje in prikaz ustreznega odgovora, ki omogoča, da uporabnik svoje vprašanje lahko vnese v različnih oblikah, tudi brez uporabe jezikoslovne terminologije.

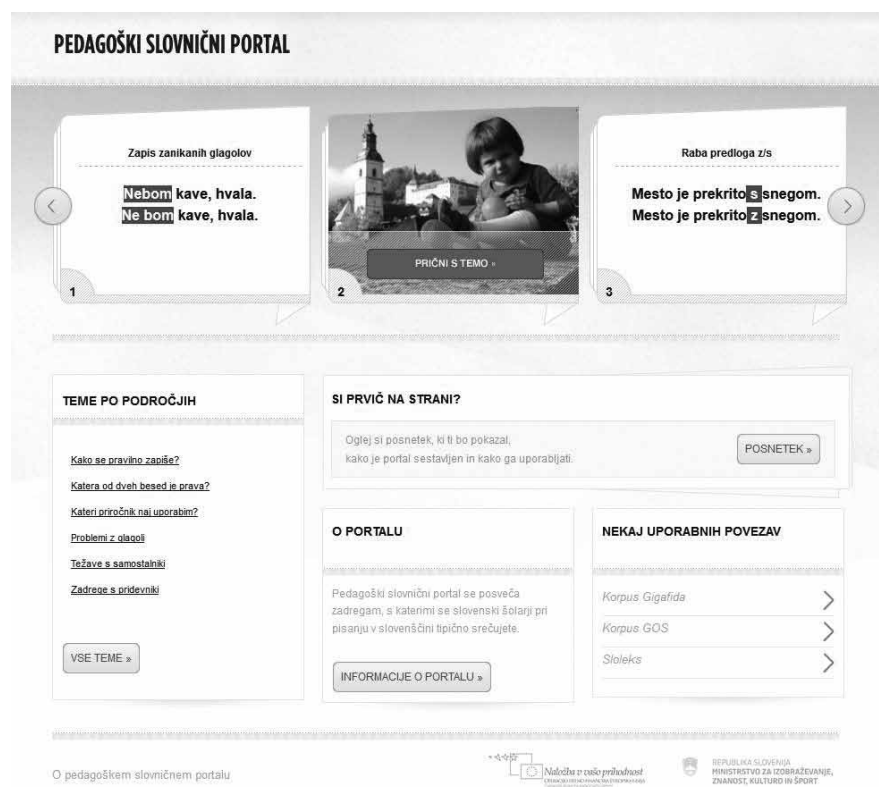
Vzporedno so potekale jezikoslovne aktivnosti, od katerih je najpomembnejša analiza tipičnih jezikovnih vprašanj govorcev in govork slovenščine. S spletnim pajkanjem smo zbrali vprašanja, ki so jih uporabniki glede jezikovne rabe zastavili na različnih spletnih forumih in v spletnih jezikovnih svetovalnicah. Rezultat je približno 700 vprašanj, ki smo jih analizirali in hierarhično uredili po jezikovnih ravniinah v kategorije: pravopis, pravořečje, oblikoslovje, besedotvorje, besedišče, skladnja, besedilo in razno. Nastala ontologija predstavlja kazalo *Slogovnega priročnika*, po katerem lahko uporabnik torej s klikanjem najde vsebino, ki ga zanima (primer s Slike 2 spada denimo v *Oblikoslovje* → *Samostalniki* → *Sklanjanje: Moške sklanjatve* → *Samostalniki z neobstojnim samoglasnikom* → *Sklanjanje moških samostalnikov z neobstojnim samoglasnikom: slovenska lastna imena*). Zadnja projektna aktivnost, ki jo je treba omeniti, pa je bila priprava petnajstih predstavitevnihih gesel, za katera so bile opravljene specializirane kvantitativne in kvalitativne korpusne analize in diahroni ter sinhroni pregled s problemom povezanihih jezikovnih pravil.

Značilno za *Slogovni priročnik*, kot tudi za druge vire *Sporazumevanja v slovenskem jeziku*, je sopostavljanje podatkov o realni jezikovni rabi in trenutno obstoječih jezikovnih pravilih. Ta pogled, ki pred pojavom besedilnih korpusov ni bil mogoč, ni dragocen le za jezikovne uporabnike, ki dobijo možnost na enem mestu videti, kako se določeno jezikovno pravilo udejanja v jezikovni praksi, ampak tudi za jezikoslovce. Izredno dobro namreč nakazuje težavna mesta oz. nerešene naloge tako za normativistiko, slovaropisje in slovničarstvo kot tudi za pripravljavce učnih gradiv za pouk slovenščine (Arhar Holdt idr. 2013).

5 Pedagoški slovnčni portal

Eden od ciljev projekta *Sporazumevanje v slovenskem jeziku* je bil empirično ugotoviti, na katerih mestih imajo učenci in dijaki pri pisanju v slovenščini probleme, in jim ponuditi razlago in rešitev teh problemov v zanje zanimivi in razumljivi obliki. Rezultat je *Pedagoški slovnčni portal*, ki je prosto dostopen na spletnem naslovu slovnica.slovenscina.eu. Kot velja za portal *Slogovni priročnik*, je tudi *Pedagoški slovnčni portal* pripravljen kot pilotni projekt. V trenutni različici obsega 18 slovnčnih poglavij¹⁰ in 6 poglavij, ki na kratko in poenostavljeno predstavljajo prosto dostopne jezikovne vire za slovenščino.

V skladu s sodobnimi didaktičnimi trendi in splošnim tehnološkim razvojem družbe smo poglavja na portalu pripravili v digitalni obliki, kar omogoča interaktivnost in povezljivost, npr. vključitev posnetkov govora iz korpusa GOS v razlago slovnčnih vsebin. Na vsebinski ravni portal določa predvsem dejstvo, da so poglavja v celoti pripravljena na osnovi podatkov iz besedilnih korpusov. Pri pripravi poglavij so bili uporabljeni korpusi Šolar (Rozman idr. 2012), Gigafida (Logar Berginc idr. 2012) in GOS (Verdonik in Zwitter Vitez 2011), kot je na primeru priprave poglavja o menjavi glagolov *moči* in *morati* natančneje razloženo v Kosem idr. (2012: 108–122).



Slika 4: Vstopna stran Pedagoškega slovnčnega portala

¹⁰ Nekaj primerov poglavij: Ko smo že pri otrocih, z otroki je križ: Sklanjanje besede otrok; Kdo zna rešiti ta problem: Nedoločnik in kratki nedoločnik; S škafcem in krpo gre lažje kot z metlo: Raba predloga z/s; S stricem Sašem sva tekla za avtobusom: Preglas; Če Eva ne more po jabolko, mora pa Adam: Glagola moči in morati.

Glavni namen portala je omogočiti učinkovitejšo izrabo časa, ki ga imajo učitelji za izboljševanje pisne kompetence šolajoče se populacije. Vsako od poglavij se osredotoča na posamezen (povsem konkreten) jezikovni problem. Obenem vsako poglavje predstavlja zaključeno celoto, za razumevanje katere ni potrebno poznavanje drugih slovnčnih vsebin. Velika pozornost je

namenjena razumljivosti razlage: jezik razlage je preprost, vsi zahtevnejši pojmi so dodatno razloženi, razlaga je opremljena z mnemotehnikami in vizualnimi elementi, kot so denimo besedni oblaki. Na koncu vsakega poglavja so na voljo slovnične vaje, ki so – ponovno – pripravljene izključno na osnovi avtentičnega jezikovnega gradiva. Pri vsakem od obravnavanih slovničnih problemov so na voljo vaje s po 150 do 200 primeri, ki se prikazujejo v naključnih sklopih po 10, odgovori se samodejno popravijo in ovrednotijo na način, da lahko učenec spremlja svoj napredek. Vsebine so torej zasnovane tako, da jih je mogoče uporabljati tako pri pouku kot tudi samostojno.¹¹ S tem je omogočena individualizacija v razvoju pisne zmožnosti učencev: namesto da vsi učenci rešujejo nabor raznovrstnih slovničnih nalog, je mogoče učenca usmeriti neposredno k poglavju, ki se osredotoča na njegov dejanski jezikovni problem.

Dejstvo, da jezikovne vaje temeljijo na realnem in ne izmišljenem jeziku, pri čemer so pogosto zasnovane na primerih iz šolske produkcije (npr. esejev), podpira prenos teoretičnega znanja (tj. o določenem jezikovnem pravilu) v prakso pisanja, saj se poviša možnost, da bodo učenci podobne jezikovne strukture, kot jih srečajo v vajah, tudi sami tvorili pri pisanju. Kot težavo tega pristopa je mogoče omeniti predvsem časovno zahtevnost priprave tovrstnega gradiva, saj je iz množice primerov treba izbrati tiste, ki so obenem tudi didaktično ustrezni. Primeri morajo biti mdr. dovolj kratki, da s kompleksno strukturo ne odvrnejo pozornosti učenca od obravnavanega jezikovnega problema, pa obenem ne pomensko presplošni oz. prazni; razen z obravnavanega vidika ne smejo biti jezikovno problematični; prav tako ne smejo biti sporni z vsebinskega vidika (npr. vsebovati prepoznavna lastna imena ali za šolski kontekst potencialno kočljive vsebine). Pri pripravi portala smo upoštevali vse naštetje pogoje in pri tem minimalno posegali v izvorna besedila: primere smo po potrebi skrajšali, ne pa tudi drugače spreminjali. Z adaptacijo primerljivih leksikografskih postopkov (Kosem idr. 2011) smo opravili tudi prve korake na poti k avtomatizaciji izbire ustreznih primerov za jezikovnodidaktične namene, kar se je izkazalo kot dobrodošel in obetaven doprinos k obstoječi metodologiji.

Pedagoški slovnični portal je zasnovan z namenom, da učence in dijake poleg slovničnih pravil uči, kateri jezikovni viri so za slovenščino na voljo ter kdaj in kako jih uporabiti – predvsem, kako pridobiti in interpretirati podatke, ki so v posameznem viru na voljo, manj pozornosti je posvečene razlagi sestave posameznega vira. Ker se jezikovni viri za slovenščino hitro posodablajo in spreminjajo, *Pedagoški slovnični portal* pa je ostal v obliki kot ob izteku financiranja, so nekatera od poglavij s predstavitvenimi posnetki virov v času pisanja tega prispevka že zastarela. Obenem so bila poglavja v vmesnem času v določenih pogledih vsebinsko in oblikovno presežena z izobraževalnim gradivom, pripravljenim v sklopu projekta *Portal jezikovnih virov*.¹² V nadaljevanju razvoja bo torej treba poskrbeti, da se poglavja s posnetki ustrezno posodobijo in še natančneje usmerijo k vsebinam, ki so za učence in dijake v praksi problematične.

Pedagoški slovnični portal je inovativen ne samo v slovenskem prostoru, ampak tudi širše. Med najbolj izpostavljenimi novostmi so: prioretizacija učnih vsebin glede na empirične podatke o prisotnosti jezikovnih napak v

¹¹ Ciljni uporabniki portala so učenci zadnje triade OŠ in dijaki raznovrstnih SŠ. Razlage so na osnovni ravni usmerjene k osnovnošolskim uporabnikom, pri čemer so za dijake skozi celotno poglavje na klik na voljo tudi zahtevnejše dodatne vsebine. Vaje trenutno niso razvrščene po zahtevnosti – zahtevnost primarno določa obravnavani jezikovni problem. Na osnovi testiranj v razredu bi bilo mogoče v nadaljevanju določiti, za katero stopnjo šolanja so določene vsebine najbolj primerne, pa tudi ugotoviti parametre za označevanje zahtevnosti jezikovnih zgledov in vaj.

¹² <http://viri.trojina.si>

celotni populaciji; osnovanost razlag iz posamezne jezikovne napake, ne iz slovničnega sistema; uporaba izključno avtentičnega jezikovnega gradiva; vključitev velike količine specifičnih vaj za avtomatizacijo transfera znanja; upoštevanje različnih jezikovnih zvrsti, npr. pisni in govorni jezik, standardna¹³ in nestandardna slovenščina (Arhar Holdt idr. 2016). Z vsemi naštetimi novostmi se seveda odpira vprašanje, ali so slovenske učilnice pripravljene na spremembe v odnosu do jezikovne rabe, ki jih prinaša korpusni pristop (o teh več v Krek in Arhar Holdt 2010). Prvi odzivi učiteljev na predstavljeni rezultat so pozitivni. Od 168 učiteljev, ki so se udeležili jezikovnotehnološkega izpopolnjevanja v letu 2014,¹⁴ jih je 88 % ocenilo portal kot relevanten, 73 % pa kot uporaben vir za svoje delo. Med glavnimi težavami je pilotna narava portala, konceptualne rešitve pa bo v prihodnjem koraku treba evalvirati tudi pri dejanski rabi v razredu.

6 Sklep

V prispevku smo opisali rezultate projekta *Sporazumevanje v slovenskem jeziku*, ki vključuje jezikovne vire (pisne, govorne korpusne), baze podatkov (slovenski oblikoslovni leksikon *Sloleks*, *Leksikalna baza za slovenščino*), jezikovnotehnološke aplikacije (oblikoskladenjski označevalnik, skladenjski razčlenjevalnik) in inovativne spletne priročniške portale. Podrobneje smo opisali predvsem *Leksikalno bazo* kot izhodišče za pripravo slovarskih opisov, *Slogovni priročnik*, ki odgovarja na tipične jezikovne zadrege, in *Pedagoški slovnični portal*, ki na enostaven način razlaga slovnične pojave, s katerimi imajo učenci in dijaki pri šolskem pisanju največ težav. Namen projekta je bil omogočiti, da bodo govorke in govorci slovenščine lahko sodelovali v novih izobraževalnih trendih, česar ni mogoče doseči brez dobre (jezikovno)tehnološke opremljenosti slovenskega jezika.

Kot je razvidno iz prispevka, je kot glavno pomanjkljivost predstavljenih portalov mogoče izpostaviti njihovo vsebinsko nedokončanost, zaradi katere so trenutno manj uporabni za konkretno delo v razredu. Ključen korak za naprej ostajajo tudi uporabniške evalvacije (tako učiteljev kot učencev), ki bodo omogočile poenotenje terminologije in nadaljnje vsebinsko ter oblikovno nadgrajevanje. Kot pravo vrednost predstavljenih rezultatov je torej treba razumeti nove postopke pridobivanja, organizacije in vizualizacije jezikovnih podatkov, ki korenito spreminjajo jezikoslovno delo in omogočajo pogled na jezikovno rabo, ki v preteklosti ni bil mogoč. Nova metodologija odpira možnosti za hiter razvoj raznovrstnih, uporabniškimi potrebam prilagojenih izdelkov, dokončno oblikovanje katerih pa mora potekati v tesnem sodelovanju z zainteresiranimi uporabniškimi skupinami. V tem smislu je za učitelje slovenščine pomembno predvsem, da se imajo s pomočjo pripravljenih treh portalov možnost seznaniti z določenimi od teh novih možnosti in se v nadaljevanju vključiti v proces kot strokovni evalvatorji in razvijalci novih vsebin.

¹³ Za natančnejše pojasnilo glede rabe izrazov standardna in nestandardna slovenščina v primerjavi s knjižno zvrstjo gl. Krek 2015 in Krek, Gorjanc in Popič 2015.

¹⁴ <http://ucitelji.sdt.si/>


 POVZETEK

Rezultati projekta *Sporazumevanje v slovenskem jeziku* vključujejo obsežne besedilne korpuse s pisnimi in transkribiranimi govorjenimi besedili, ki so bili zgrajeni z namenom, da bi ugotovili, kako v resnici pišemo in govorimo v sodobni slovenščini. Poleg tega so bile izdelane baze podatkov, ki vsebujejo informacije o slovenski leksiki. To sta predvsem *Leksikalna baza* s pomenskimi, skladijskimi, kolokacijskimi in drugimi podatki ter slovenski oblikoslovni leksikon *Sloleks*, ki vsebuje pregibne oblike za več kot 100.000 slovenskih besed. Razvite so bile tudi nekatere temeljne računalniške tehnologije in baze za potrebe strojne obdelave slovenskega jezika. V prispevku opisujemo predvsem inovativne spletne priročniške portale, na katerih so na uporabniško prijazen način predstavljeni odgovori na tipične pravopisne zadrege ter razlage slovničnih pojavov, s katerimi imajo učenci in dijaki pri šolskem pisanju največ težav.

Ključne besede: slovenščina, bralna pismenost, jezikovni viri, jezikovne tehnologije, inovativna učna okolja


 Viri in literatura

- Arhar Holdt, Špela, Kosem, Iztok in Logar, Nataša, 2012: Izdelava korpusa Gigafida in njegovega spletnega vmesnika. Erjavec, Tomaž in Žganec Gros, Jerneja (ur.): *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut Jožef Stefan. 16–21.
- Arhar Holdt, Špela, Dobrovoljc, Kaja in Popič, Damjan, 2013: Reprezentacija standardnega in nestandardnega v virih SSJ. Žele, Andreja (ur.): *Družbena funkcijskost jezika: vidiki, merila, opredelitve*. Ljubljana: Znanstvena založba Filozofske fakultete. 19–27.
- Arhar Holdt, Špela, Kosem, Iztok in Gantar, Polona, 2016: Corpus-Based Resources for L1 Teaching: The Case of Slovene. Marcus-Quinn, Ann in Hourigan, Triona (ur.): *Handbook on Digital Learning for K-12 Schools*. Springer, v tisku.
- Dobrovoljc, Kaja, Krek, Simon in Rupnik, Jan, 2012: Skladijski razčlenjevalnik za slovenščino. Erjavec, Tomaž in Žganec Gros, Jerneja (ur.): *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut Jožef Stefan. 42–47.
- Dobrovoljc, Kaja, Krek, Simon in Erjavec, Tomaž, 2015: Leksikon besednih oblik Sloleks in smernice njegovega razvoja. Gorjanc, Vojko, Gantar, Polona, Kosem, Iztok in Krek, Simon (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete. 80–105.
- Fillmore, Charles, J., Johnson, Christopher, R. in Petruck, Miriam, R. L., 2003: Background to Framenet. *International Journal of Lexicography* 16/3. 235–250.
- Gantar, Polona, 2015: *Leksikografski opis slovenščine v digitalnem okolju*. Ljubljana: Znanstvena založba Filozofske fakultete UL. <http://www.ff.uni-lj.si/Portals/0/Dokumenti/ZnanstvenaZalozba/e-knjige/Leksikografski.pdf>.
- Gorjanc, Vojko, Gantar, Polona, Kosem, Iztok in Krek, Simon (ur.), 2015: *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete.

- Granger, Sylviane in Paquot, Magali (ur.), 2010: *eLexicography in the 21st century: New challenges, new applications. Proceedings of ELEX2009*. Cahiers du CENTAL. Louvain-la-Neuve: Presses universitaires de Louvain.
- Grčar, Miha, Krek, Simon in Dobrovoljc, Kaja, 2012: Obeliks: statistični obliko-skladenjski označevalnik in lematizator za slovenski jezik. Erjavec, Tomaž in Žganec Gros, Jerneja (ur.): *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut Jožef Stefan. 89–94.
- Hanks, Patrick, 2013: *Lexical Analysis: Norms and Exploitations*. Cambridge: MIT Press.
- Kornai, Andras, 2013: Digital Language Death. *PLoS ONE* 8(10): e77056. doi:10.1371/journal.pone.0077056.
- Kosem, Iztok, Husak, Milos in McCarthy, Diana, 2011: GDEX for Slovene. Kosem, Iztok in Kosem, Karmen (ur.). *Electronic lexicography in the 21st century: new applications for new users : proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ljubljana: Trojina, Institute for Applied Slovene Studies. 150–159.
- Kosem, Iztok, Stritar Kučuk, Mojca, Može, Sara, Zwitter Vitez, Ana, Arhar Holdt, Špela in Rozman, Tadeja, 2012: *Analiza jezikovnih težav učencev: korpusni pristop*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Krek, Simon in Arhar Holdt, Špela, 2010: Slovenski besedilni korpusi: kako v razred?. *Sodobna pedagogika* 61/1. 224–241.
- Krek, Simon, 2012a: *Slovenski jezik v digitalni dobi. The Slovene language in the digital age, (White paper)*. Heidelberg: Springer.
- Krek, Simon, 2012b: Spletni portal Slogovni priročnik. Krakar Vogel, Boža (ur.): *Slavistika v regijah – Koper*. Ljubljana: Zveza društev Slavistično društvo Slovenije: Znanstvena založba Filozofske Fakultete. 225–231.
- Krek, Simon, 2015: Standardni in knjižni jezik – drugi poskus. Smolej, Mojca (ur.): *Obdobja 34: Slovnica in slovar – aktualni jezikovni opis (1. del)*. Ljubljana: Znanstvena založba Filozofske Fakultete. 401–407.
- Krek, Simon, Gorjanc, Vojko, Popič, Damjan, 2015: Med ideologijo knjižnega in standardnega jezika. Gorjanc, Vojko, Gantar, Polona, Kosem, Iztok in Krek, Simon (ur.), 2015: *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete. 32–48.
- Krek, Simon, Gantar, Polona, Kosem, Iztok, Gorjanc, Vojko in Laskowski, Cyprian, 2016: Baza kolokacijskega slovarja slovenskega jezika. *Zbornik konference Jezikovne tehnologije in digitalna humanistika*. V tisku.
- Lew, Robert, 2015a: Dictionaries and their users'. Hanks, Patrick in De Schryver, Gilles-Maurice (ur.): *International Handbook of Modern Lexis and Lexicography*. Berlin/Heidelberg: Springer. 1–9.
- Lew, Robert, 2015b: Research into the use of online dictionaries. *International Journal of Lexicography* 28/2. 232–253.
- Logar Berginc, Nataša, Grčar, Miha, Brakus, Marko, Erjavec, Tomaž, Arhar Holdt, Špela in Krek, Simon, 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.
- Rozman, Tadeja, Krapš Vodopivec, Irena, Stritar Kučuk, Mojca in Kosem, Iztok, 2012: *Empirični pogled na pouk slovenskega jezika*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Rozman, Tadeja, 2010: *Vloga enojezičnega razlagalnega slovarja slovenščine pri razvoju jezikovne zmožnosti*. Doktorska disertacija. Ljubljana: Filozofska fakulteta UL.
- Verdonik, Darinka in Zwitter Vitez, Ana, 2011: *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Vidovič Muha, Ada, 2000: *Slovensko leksikalno pomenoslovje – Govorica slovarja*. Ljubljana: Znanstveni inštitut Filozofske fakultete.