

Uporaba jezikovnotehnoloških virov in postopkov pri razvoju učnega e-okolja Slovenščina na dlani

Darinka Verdonik,* Simona Majhenič,* Špela Antloga,* Sandi Majninger,* Marko Ferme,*
Kaja Dobrovoljc,* Simona Pulko,† Mira Krajnc Ivič,† Natalija Ulčnik†

* Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru

Koroška cesta 46, 2000 Maribor

darinka.verdonik@um.si; simona.majhenic@um.si; s.antloga@um.si; sandi.majninger@um.si; marko.ferme@um.si;
kaja.dobrovoljc@um.si

† Filozofska fakulteta, Univerza v Mariboru

Koroška cesta 160, 2000 Maribor

simona.pulko@um.si; mira.krajnc@um.si; natalija.ulcnik@um.si

Povzetek

Prispevek predstavlja, kako so jezikovni viri in jezikovnotehnološki pristopi uporabljeni za razvoj učnega e-okolja Slovenščina na dlani. Jezikovnotehnološka podpora omogoča pomemben korak naprej v razvoju digitalnih učnih e-okolij, saj prinaša podporo prožnim oblikam poučevanja, poučevanju na daljavo, lajša učiteljevo delo, digitalno okolje pa omogoča tudi motiviranje učencev prek igričarskih elementov. V prispevku predstavljamo zasnovo e-okolja, podporne jezikovne vire in izvedbo vaj za štiri osrednje vsebinske sklope (pravopis, slovnico, frazeme in pregovore, besedila) ter razlage k vajam.

Using Language Resources and Tools to Develop the E-Environment for Learning Slovene “Slovene in the Palm of Your Hand”

This paper addresses the issue of using language resources and tools in the development of the e-environment for learning, “Slovene in the Palm of Your Hand”. The use of language resources enables an important step forward in the development of digital e-environments. It supports flexible learning, learning at distance and facilitates teacher’s work, while digital environments hold the potential to motivate learners with gaming elements. The paper represents the conceptual framework of the “Slovene in the Palm of Your Hand” environment, the developed language resources and the designed tasks for four major domains (orthography, grammar, phrasemes and proverbs, and texts) and accompanying explanations.

1. Uvod

Jezikovni viri, jezikovnotehnološka orodja in postopki so postali uveljavljeno področje raziskovanja jezika in njegove rabe; so nujen sestavni del številnih tehnoloških postopkov, zlasti na področjih velikih podatkov in umetne inteligence, ter temelj razvoja končnih jezikovnotehnoloških produktov, kot so prevajalnik, razpoznavalnik in sintetizator; nepogrešljivi so postali pri razvoju slovarjev in drugih jezikovnih priročnikov.

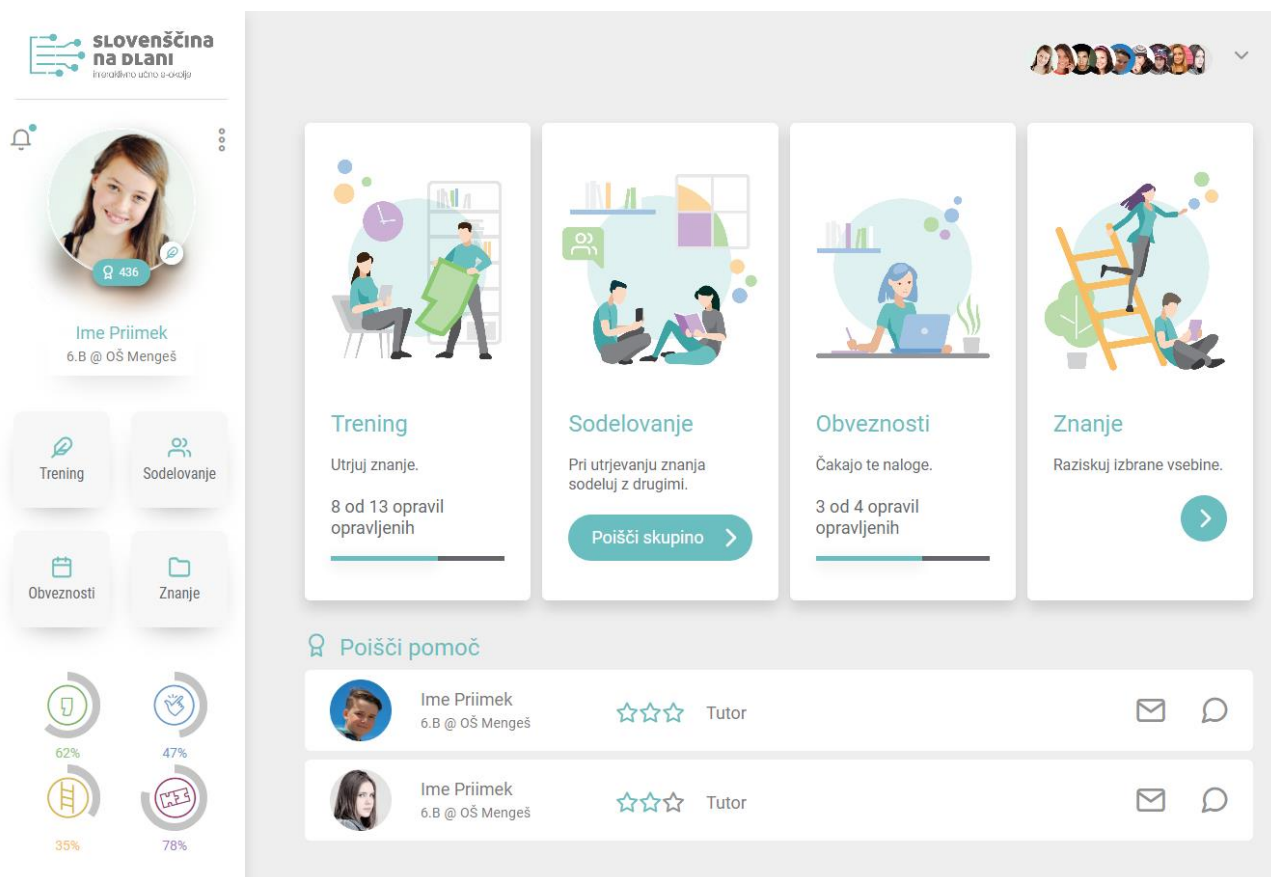
V projektu Slovenščina na dlani¹ smo raziskovali načine, kako jezikovne vire in jezikovnotehnološke pristope uporabiti pri razvoju didaktičnih pripomočkov za učenje jezika (Rozman et al., 2020). Cilj je bil razviti avtomatizirano, interaktivno, z jezikovnimi opisi podprto prosto dostopno učno e-okolje za slovenščino, v okviru katerega so na dlani različne dopolnjujoče se aktivnosti, s svojo dinamično zasnovo bistveno drugačne od obstoječih e-gradiv. Podpora z jezikovnimi viri in jezikovnotehnološkimi pristopi je bila pri tem ključen element, ki omogoča pomemben napredek v primerjavi s podobnimi obstoječimi okolji, zlasti v smislu upoštevanja prožnih in inovativnih oblik učenja, učne diferenciacije in individualizacije, tj. prilagajanja zmoglostim in potrebam vsakega posameznega uporabnika, ter formativnega spremljanja posameznikovega napredka kot učinkovitega vzvoda za prevzemanje odgovornosti za lastno učenje in napredovanje.

Pojem prožne oblike poučevanja (po definiciji Evropske komisije) vsebuje učenje in poučevanje, ki je odzivno na potrebe učečega se in njegove močne strani. Fleksibilno učenje ponuja učečemu se izbiro glede načina, okolja in časa učenja z namenom spodbujanja motivacije in vztrajnosti, lahko tudi v primerih, ko je prisotnost na lokaciji izvajanja učnega procesa otežena. Prožne oblike učenja in poučevanja tako omogočajo fleksibilno obliko dela, saj gre za sistem poučevanja in učenja, v katerem imajo učeči se možnost, da (lahko) del učenja opravijo tudi izven šolskega okolja, omenjene oblike učenja in poučevanja namreč omogočajo prilagajanje izvedbe učnih programov ob upoštevanju zahtev učnega programa ter pričakovanih učnih izidih, opredeljenih z učnimi načrti in katalogi znanj. Ob upoštevanju teh izhodišč smo pri pripravi e-okolja *Slovenščina na dlani* ustrezno prilagodili tudi metode in oblike poučevanja in učenja.

2. Zasnova učnega e-okolja Slovenščina na dlani

Učno e-okolje Slovenščina na dlani ponuja vaje in podporne strokovne vsebine s štirih področij slovenskega jezika: 1) pravopisa, 2) slovnice, 3) frazemov in pregovorov ter 4) besedil. Pri pravopisu in slovnici je poudarek na tistih temah in vsebinah, pri katerih se pomanjkanje znanja pogosto kaže skozi napake v pisanju učečih se. Na področju frazeologije se posvečamo podobi, pomenu in rabi frazemov ter pregovorov, pri čemer prav

¹ Projekt sofinancirata Republika Slovenija in Evropska unija iz Evropskega socialnega sklada..



Slika 1. Uvodna stran e-okolja za učenca

tako skušamo odpraviti primanjkljaje, ki se kažejo pri poznavanju in razumevanju teh leksikalnih enot, s čimer težimo k izboljšanju frazeološke kompetence med uporabniki e-okolja. Pri besedilih uporabniki e-okolja razvijajo (kritično) sporazumevalno jezikovno zmožnost z reševanjem vaj in nalog o zunajjezikovnem in jezikovnem kontekstu, in sicer z avtentičnimi besedili osmih tematskih sklopov. Vsak tematski sklop zajema vsaj tri sklope besedil, slikovito predstavljene s slogani. Vsak slogan pa predstavlja kombinacijo vsaj dveh pogostih in za večino govorcev slovenščine pomembnih besedilnih skupin.

Učno e-okolje loči dva glavna tipa uporabnikov: učitelje in učence. Učitelj v e-okolju upravlja sezname in aktivnosti svojih skupin in njihovih udeležencev; dodeljuje naloge; pregleduje uspešnost skupin in udeležencev; komunicira z njimi; si pripravlja teste in delovne liste; ali pa tudi dodaja in pregleduje nove vsebine. Učenec rešuje vaje, bodisi tiste, ki jih določi učitelj, bodisi jih samostojno izbira; seznanen se lahko z znanji, potrebnimi za izvajanje posameznih vaj, ali pa prosto brska po bazi znanj; pregleduje svojo uspešnost; poišče pomoč učitelja ali tutorjev; pošilja sporočila in se povezuje z drugimi uporabniki; vnaša nove vsebine v e-okolje.

Lastnosti učnega e-okolja Slovenščina na dlani so avtomatiziranost, individualizacija, interaktivnost tako s sistemom kot med uporabniki in igričarski pristop.

Avtomatiziranost pomeni, da učno e-okolje učenca samo vodi skozi vaje in daje povratne informacije o učenčevi uspešnosti. Individualizacija se kaže v

avtomatskem prilagajanju vsebine, količini in zahtevnosti vaj vsakemu posameznemu uporabniku, odvisno od rezultatov o njegovi uspešnosti pri posameznih vsebinah. Interaktivnost med uporabniki je omogočena skozi povezovanje uporabnikov med seboj, funkcionalnost tutorske pomoči ali pomoči učitelja, možnost združevanja v skupine, dodeljevanje nalog uporabnikom s strani učiteljev, pošiljanje sporočil ipd. Interakcija pa poteka tudi med sistemom in uporabnikom, s tem ko sistem uporabniku sproti signalizira uspešnost njegovega dela, mu pomaga z namigi, omogoča prebiranje razlag, po drugi strani pa lahko uporabnik tudi sam prispeva vsebine v učno e-okolje z vnašanjem razlag. Igričarski pristop je izveden tako skozi grafično oblikovanje kot tudi skozi samo zasnovo e-okolja, pri kateri uporabnik učenec zbira točke, osvaja medalje ter tekmuje s samim seboj ali z drugimi.

Oblikovanje uporabniškega vmesnika Slovenščina na dlani je bilo izvedeno v sodelovanju s profesionalnim oblikovalcem.² Slika 1 prikazuje uvodno stran e-okolja za učenca.

3. Vaje iz pravopisa in slovnice

Osnova za vaje iz pravopisa in slovnice je korpus besedil, ki zajema leposlovna, publicistična in spletna besedila, izbrana tako, da so vsebinsko in po zahtevnosti primerna in zanimiva za šolajočo se populacijo, hkrati pa tudi jezikovno pregledana.

² Samo Kramberger (<https://samokramberger.com/>).

3.1. Korpus MAKS

Korpus MAKS (akronim za MIAdinski KorpuS) obsega pribl. 10 mio. besed oz. pribl. 12 mio. pojavnic. Dobro polovico tega sestavljajo besedila iz mladinskega in drugega leposlovja ali priročnikov. Dobrih 40 % besedil je zajetih iz publicistike, manjši delež, dobrih 300.000 besed, pa s spleta. Omejevanje zajetih vsebin s spleta je bilo potrebno predvsem zaradi tega, ker je jezikovni pregled pri teh manj natančen, za korpus MAKS pa je bil to pomemben kriterij, saj je za izvajanje vaj in nalog iz slovnice in pravopisa treba priklicati zgolj slovnično in pravopisno pravilne primere. Besedilodajalci³ so bili večinoma založbe, posamezni leposlovni avtorji, kar nekaj besedil pa je bilo prevzetih tudi iz korpusa Gigafida.

Vsa zajeta besedila so bila ročno pregledana; preverjeno je bilo, ali po vsebini (vsebina ni oglasna, ideološko zaznamovana, nasilna, spolna, zelo strokovna in težko razumljiva ipd.) in jezikovno (rabljen je knjižni jezik) ustrezajo specifičnim potrebam e-okolja Slovenščina na dlani. Pokazalo se je, da so vsebine, ki za učeče se niso primerne vir povedi za vaje iz slovnice in pravopisa, zelo pogoste: v leposlovju, zlasti nemladinskem, so bile pogoste vsebine, povezane z nasiljem, občasno pa smo morali vsebine izločiti tudi zaradi slenga, vulgarizmov, narečnega ali starinskega jezika. V publicistiki so se po drugi strani pojavljale propagandne ali ideološko zaznamovane vsebine, občasno pa tudi jezikovno nezadostno pregledana besedila. Zahtevnim strokovnim vsebinam pa smo se skušali izogniti že pri izboru virov.

Ob pregledu smo besedilom pripisali vir, leto objave, naslov, avtorja, primernost za osnovno ali srednjo šolo ter teme, vsa besedila pa smo nato oblikoslovno, skladijsko in imenskoentitetno označili.

Za oblikoslovno in skladijsko označevanje smo uporabili Stanford Parser V3 (Qi et al., 2018). Pred označevanjem smo besedilo segmentirali na povedi in besede z orodjem Obeliks4J (Grčar et al., 2012). Na koncu smo z Stanford NER (Finkel et al., 2005) izvedli še označevanje imenskih entitet v besedilu.

Korpus MAKS je dostopen prek Clarinovega konkordančnika NoSke. Kot odprtodostopna zbirka ni dostopen zaradi omejitev pri izhodiščnih avtorskih pravicah.

3.2. Definiranje vaj

Vaje iz slovnice in pravopisa so zasnovane tako, da pokrivajo področja, ki se kažejo skozi napake pri pisanju, pri čemer smo si pomagali z analizo napak iz korpusa Šolar (Kosem et al., 2012, Rozman et al., 2020).

V sklopu pravopisa je največ nalog namenjenih pravilni rabi vejice, in sicer smo pripravili vaje za vejico pri podredju, priredju, pastavkih, pristavkih, vrvkih in pri zahtevnejših primerih, kot je na primer vejica pri večbesednih veznikih. Velik delež nalog predstavljajo tudi zahtevnejši primeri zapisa, pri katerih uporabniki lahko vadijo črkovanje, zapis prevzetih besed in druge zahtevnejše primere, zapis z veliko ali malo začetnico ter (ne)stičen zapis ali zapis z vezajem. V nekoliko manjšem obsegu so zastopane vaje za rabo drugih ločil.

Vsaki vaji smo določili zahtevnostno stopnjo na lestvici od 1 do 3 (osnovna, srednja, zahtevna), zato so vaje

primerne za uporabnike z različnim predznanjem, torej za tiste, ki šele usvajajo pravopisna ali slovnična pravila, ali za tiste, ki želijo svoje znanje zgolj osvežiti. Zlasti za slednjo skupino so primerne zaključne naloge pravopisa (zahtevna raven), pri katerih uporabnik na igriv način rešuje pravopisne probleme vseh poglavij.

Pri slovnici je največ nalog namenjenih samostalnikom (naloge, povezane s sklanjanjem, zanikanim roditeljskim idr.), zaimkom (oziralni zaimki, ujemanje zaimka, zahtevnejši primeri idr.), glagolom (dvojina, opisni deležnik na -l, namenilnik, nedoločnik idr.) in predlogom, nekoliko manj nalog pa obravnava veznike in pridevnike ter besedni red.

Pravopisne in slovnične naloge temeljijo na skupno 16 osnovnih tipih nalog, s katerimi smo želeli zagotoviti ravnovesje med igrivostjo in avtomatizacijo nalog. Na osnovni stopnji uporabnik najprej prepozna napake (npr. z izbiranjem, iskanjem, označevanjem (ne)pravilnega odgovora idr.), vendar je podoba nalog grafično različna, tako da uporabnik odgovor išče npr. v rešetku, ga kot sestavljanke dopolnjuje z danimi odgovori ali, tako kot v igri spomin, išče pare s podobnimi napakami. Pri zahtevnejših nalogah mora uporabnik napake prepoznati in jih tudi odpraviti (npr. s popravljanjem, premikanjem, vstavljanjem ipd.). Najzahtevnejše naloge pa od uporabnika zahtevajo tudi, da svojo odločitev utemelji. Da bi zagotovili čim bolj raznoliko aktivnost uporabnika, e-okolje poleg izbiranja nudi tudi naloge vpisovanja in premetank z odvečnimi črkami v različnih grafičnih oblikah.

3.3. Programiranje vaj

Programiranje vaj obsega več nalog, ki jih mora zagotoviti učno e-okolje. Prva naloga je najti primere povedi, ki ustrezajo področju, zahtevnosti naloge in tipu naloge, za katerega so uporabljeni. Navedeni primeri morajo zato vsebovati ustrezne gradnike, ki jih mora računalnik razpoznati, da lahko sestavi programsko predstavitev problema, ki ga nato ponudi uporabniku. Znotraj učnega e-okolja smo formalno definirali slabih 200 gradnikov, ki so lahko zelo preprosti (npr. vejica) ali pa zelo kompleksni (npr. vrstilni, ločilni in množilni števniki iz vsaj dveh sestavin).

Druga naloga je takšno programsko predstavitev problema prikazati uporabniku, mu hkrati glede na izbrano vajo prikazati navodila in omogočiti interaktivno reševanje omenjenega problema ter ustrezno hranjenje uporabnikovih odgovorov. Tretja naloga je vrednotenje odgovorov uporabnika, ki lahko poteka avtomatsko ali z zunanjim (ročnim) vrednotenjem.

Pri prvi nalogi, iskanju ustreznih primerov, smo z leksikalno analizo in formalnim zapisom pogojev za primere posameznih vaj dobili kandidate za primere posameznih nalog. Navedeni kandidati so bili nato ročno pregledani, da smo lahko izločili vsebinsko neprimerne povedi in tiste povedi, kjer je pri leksikalni analizi prišlo do napak. Končni nabor povedi tako predstavlja množico povedi, iz katerih se v učnem e-okolju tvorijo vaje. Pri vsaki posamezni vaji je izdelan tudi uporabniški vmesnik za urejanje navodil ter pregled in dodajanje primerov, kar omogoča, da je v učno e-okolje mogoče dodajati tako nove

³ Navedeni so na spletni strani projekta <http://projekt.slo-na-dlani.si/>.

primere kot tudi nove vaje. Za posamezno vajo je poleg formalnega zapisa pogojev za iskanje primerov ustvarjen tudi formalni zapis pogojev za gradnike, ki predstavljajo interaktivne elemente (npr. število izbrisanih vejic, gradnik, v katerem smo spremenili velikost začetnice, itd.).

Prikaz vaje in njeno reševanje predstavlja osrednji del programiranja vaj. Programska predstavitev vaje je uporabniku prikazana v obliki spletnega vmesnika, katerega videz močno variira glede na tip naloge. Osnovni elementi interakcije, kot so vstavljanje preko besedilnih vnosnih elementov ali izbira posameznih elementov, ne predstavljajo večjih izzivov, bolj zahteven pa je prikaz interaktivnih elementov, ki jim uporabnik lahko spreminja pozicijo v besedilu (npr. ločila), ali čisto ločeni uporabniški vmesniki, kot so za rešeta ali sestavljanke. V vseh primerih je treba ustrezno hraniti uporabnikovo interakcijo, saj se poleg samega vnosa, kot je besedilo ali izbira, beležita tudi uporabnikovo napačno ravnanje ter čas, ki ga potrebuje za posamezno akcijo. Akcije se sproti beležijo in posredujejo v zaledje učnega sistema, kjer so ovrednotene, saj se na njihovi podlagi po logični shemi prožijo različne akcije. Logična shema se prav tako obnaša dinamično, saj je odvisna od vsebine same vaje, od vrste opravila, ki ga uporabnik opravlja, ter tudi od informacij, ki jih imamo o uporabniku. Glede na vse navedeno je bodisi ponujena pomoč ob neuspešnem reševanju bodisi obvestilo o pravilno ali napačno rešenemu primeru, hkrati pa ima navedeno tudi vpliv na naslednjo ponujeno vajo.

Samo vrednotenje uporabnikovih rešitev vaj poteka v zaledju sistema na dva načina. Prvi način je samodejno vrednotenje, kjer že pri tvorjenju vaje poznamo pravilno rešitev in jo lahko v času reševanja ovrednotimo. Samo vrednotenje je vezano na področje in zahtevnost vaje, je pa lahko v sklopu širšega nabora vaj tudi ustrezno uteženo. Samodejno vrednotenje daje uporabniku ob reševanju neposredno povratno informacijo, hkrati pa tvori njegov profil, v katerega beleži zgodovino uspešnosti reševanja. Drugi način vrednotenja je vezan na naloge, kjer pri generiranju vaje ni jasno določena pravilna rešitev. Takrat mora rešitev ovrednotiti drug uporabnik, običajno v vlogi učitelja ali tutorja. Slednje se dogaja predvsem pri nalogah, kjer mora uporabnik vnesti večjo količino besedila in kjer računalniški algoritmi še niso dovolj zanesljivi, da bi ponudili ustrezno vrednotenje.

4. Vaje s področja frazeologije

Na področju frazeologije smo predvideli vaje, ki so vezane na frazeme kot osnovne frazeološke enote, in na pregovore, ki jih umeščamo k frazeologiji v širšem smislu. Ciljni uporabniki e-okolja Slovenščina na dlani, tj. osnovno- in srednješolci, se z obojimi srečujejo pri jezikovnem pouku slovenščine ter jih z ozirom na obstoječa učbeniška gradiva sistematično obravnavajo, in sicer praviloma v 7. razredu osnovne šole ter v 2. letniku srednje šole. Učitelji slovenščine pri tem opazajo, da učečim se težave povzročajo njihovo razumevanje, zaradi česar so negotovi tudi pri njihovi rabi, ob tem pa poseben problem predstavlja dejstvo, da v slovenskem prostoru ni na voljo priročnika, ki bi se posebej posvečal slovarski obravnavi

idiomatskih enot in bi bil prilagojen šolajoči se populaciji. Izhajajoč iz teh ugotovitev smo naprej pripravili izbor sto frazemov in sto pregovorov, za katere smo v naslednji fazi pripravili vsestranske slovarske opise, na njihovi osnovi pa smo zasnovali tudi vaje s področja frazeologije. Pri tem smo težili k iskanju možnosti za njihovo avtomatizacijo.

4.1. Zbirka frazemov in pogovorov FRIDA

Zasnovali smo zbirko frazemov in pogovorov ter jo poimenovali z akronimom FRIDA (Frazemi in pRegovorI na DIAni).

Najprej smo izdelali ožji seznam enot. Izbrali smo sto frazemov in sto pogovorov (Ulčnik, 2019; Ulčnik in Meterc, 2019) ter pri tem izhajali iz naslednjih kriterijev: 1) aktualnost enot, 2) didaktična relevantnost enot, 3) pokrivanje različnih tematskih skupin. Aktualnost enot je bila določena na osnovi prisotnosti v učbeniških gradivih⁴ ter frekventnosti pojavljanja v korpusih sodobnih pisnih besedil.⁵ Za pregovore smo ob tem lahko upoštevali izdelan paremiološki optimum (Meterc, 2017), tj. seznam tristotih najbolj poznanih in uporabljenih enot, medtem ko za frazeme tovrsten seznam ne obstaja. Pri določanju didaktične relevantnosti smo izhajali iz predpostavke o oteženem razumevanju enot z izkazano večjo stopnjo idiomatičnosti oz. frazeološke pomenske neprozornosti, upoštevali pa smo tudi konkretne predloge učiteljev – ti so izpostavili enote, ki so med učečimi se slabše poznane ali pa je bilo pri pouku opaženo njihovo pomanjkljivo razumevanje (npr. *gordijski voz*, *pri ti z dežja pod kap*). Pokrivanje različnih tematskih skupin je izhajalo iz osnovne zamisli o vključevanju frazemov in pogovorov z različnih tematskih skupin (npr. človek, odnosi, dejavnost in bivanje, predmetnost in pojavnost, čas in prostor, količina, mera, stopnja), s čimer smo želeli opozoriti na raznovrstnost njihove rabe in prisotnost na različnih semantičnih poljih. Vključili smo tudi enote, ki so značilne za slovenski govorni prostor in obenem v rabi izkazujejo zadostno frekventnost (npr. *vleči dreto*).

Izbrane enote smo v naslednji fazi celovito slovarsko opisali, pri čemer imajo opisi dvojno funkcijo – namenjeni so podrobnejšemu seznanjanju s konkretnimi frazemi, ki se pojavljajo tudi znotraj vaj, obenem pa služijo kot zgled za samostojno pripravo slovarskih opisov še neobdelanih enot s širšega seznama. Za vnašanje slovarskih opisov smo izdelali programski vmesnik in v njem predvideli večnivojsko podatkovno shemo. Na prvem nivoju smo predstavili oblikovne in tipološke lastnosti izbranih enot, na drugem nivoju pomensko-pragmatične lastnosti, na tretjem nivoju pa njihove slovnične lastnosti. Formulacije smo vseskozi prilagajali primarnim uporabnikom e-okolja, torej osnovno- in srednješolcem. Četrto nivo je bil namenjen navajanju dodatnih zgledov rabe, ki so uporabni pri pripravi vaj. Pri tem smo bili posebej pozorni na zglede z več frazeološkimi enotami, na pojasnjevanje izvora posameznih enot (npr. na mitološka pojasnila in etimološke razlage), na opazene dobesečne pomene pri enotah, ki omogočajo t. i. dvojno branje (Kržišnik, 2006: 260), na prisotnost uvajalnih sredstev ipd., obenem pa smo pazili tudi na zadostno navajanje besedilnega okolja, iz katerega

⁴ Frazemi in pregovori so bili izpisani iz izbranih gradiv založb Mladinska knjiga in DZS (npr. iz beril od 6. do 9. razreda in od 1. do 4. letnika). Več o gradivu in poteku analize v Ulčnik, 2019; Ulčnik in Meterc, 2019.

⁵ Izhodišče za preverjanje rabe je predstavljal korpus Gigafida v. 2.0 Dedup, upoštevali pa smo tudi KAS (akademska besedila), slWac (Slovene Web) in Janes (družbena omrežja).

je možno razpoznati pomen predstavljenih frazeoloških enot. Pripravljeni opisi, ki že v osnovi sledijo jasnemu, nazornemu in razumljivemu podajanju podatkov, so v e-okolju prikazani v prilagojeni zaslonski sliki, pri čemer sledimo uporabnikovemu interesu in omogočamo večnivojski prikaz z možnostjo selektivnega izbiranja podatkov v razponu med osnovnim in razširjenim oz. celostnim prikazom (prim. tudi Jesenšek in Ulčnik, 2014: 286). Za frazeme so tako sprva prikazani le izhodiščna enota, parafraza in zgled, šele v naslednji fazi pa pomenski opis, pragmatična pojasnila, morebitne so- in protipomenske enote ipd. Podatke, zbrane na vmesniku, smo skušali v čim večji meri izkoristiti tudi pri zasnovi in pripravi vaj.

4.2. Definiranje vaj

Pri definiranju vaj smo izhajali iz osnovne delitve na: 1) vaje, ki smo jih zasnovali z ozirom na podatke, predstavljene v slovarskih opisih, in 2) dodatne vaje, ki predpostavljajo uporabnikovo ustvarjalnost. Na podlagi spoznanj teoretične frazeologije in skladno z izsledki frazeodidaktike smo vaje iz prve skupine vezali na tri podsklope: 1) podoba, 2) pomen, 3) raba. Osnovni namen prvega podsklopa je doseči, da bodo uporabniki e-okolja frazeme in pregovore prepoznali v besedilu, da se bodo zavedali njihove oblikovne oz. sestavinske podobe (večbesednost, ustaljenost) in da bodo znali ločevati med frazemi ter pregovori. V drugem podsklopu preverjamo razumevanje pomena frazeoloških enot, (pre)poznavanje medfrazemskih razmerij (so- in protipomenskost), zmožnost njihovega pomenskega pojasnjevanja ter sposobnost njihovega nadomeščanja s slogovno nezaznamovanimi besedami. V tretjem podsklopu se osredinjamo na ustrezno frazeološko rabo ter sposobnost nadomeščanja nezaznamovanega izražanja s frazeološkim. Pri pregovorih smo posebej pozorni tudi na uveljavljeno rabo s t. i. uvajalnimi sredstvi (npr. *kot pravi znan slovenski pregovor*).

Pri vseh treh podsklopih je bilo v ospredju iskanje možnosti za čim večjo avtomatizacijo pri preverjanju odgovorov in posledično pri točkovanju opravljenih nalog, kar v končni fazi uporabniku zagotavlja takojšnjo povratno informacijo. Druga skupina vaj temu kriteriju ni mogla zadostiti, saj je vezana na kreativno rabo in izhaja iz problemskega pristopa, ki se nujno povezuje z doseganjem višjih taksonomskih ravni (analiza, sinteza, vrednotenje), te pa je zaradi neoznačenih in nepredvidljivih odgovorov (skoraj) nemogoče avtomatično ocenjevati. Posledično smo za tovrstne vaje predvideli samovrednotenje oz. samorefleksijo na osnovi vnaprej pripravljenih možnih rešitev. Dva različna principa pri zasnovi vaj pa sta zahtevala tudi različne pristope z vidika njihovega programiranja.

4.3. Programiranje vaj

Podobno kot pri vajah iz pravopisa in slovnice je tudi pri frazeologiji programiranje izvedeno v več korakih. Prvi korak je izbrati frazeološke enote, ki ustrezajo zahtevnosti in tipu naloge. Naslednji korak je tvorjenje prikaza naloge in uporabniku omogočiti interaktivno reševanje. Na koncu

je potrebno še vrednotenje pravilnosti uporabnikovega odgovora.

Pri izbiri ustrezne frazeološke enote je pomembno, da so v zbirki za izbrano enoto izpolnjeni vsi podatki, potrebni za tvorjenje naloge. Prav tako moramo upoštevati, da je izbrana frazeološka enota primerna za zahtevnostno stopnjo, ki jo določa naloga.

Tvorjenje prikaza naloge uporabniku vključuje izpis navodila za reševanje, izbor ustreznih namigov, ki so uporabniku v pomoč pri reševanju, ter izpis same naloge v spletnem uporabniškem vmesniku. Izpis je odvisen od tipa naloge ter od naprave, ki jo uporablja uporabnik. Nekateri tipi nalog potrebujejo še smiselne napačne odgovore. Pokazalo se je, da za določene naloge s področja frazeologije računalnik ne zna avtomatsko poiskati ustreznih alternativnih odgovorov, zato zahtevnost nalog ne bi bila primerna.⁶ To težavo smo rešili tako, da smo zbirko frazeoloških enot ročno dopolnili s smiselnimi napačnimi odgovori, ki jih pri posamezni nalogi lahko izbere računalnik. Na tak način smo povečali zahtevnost tovrstnih nalog in njihovo kvaliteto.

Med reševanjem se uporabnikove akcije sproti beležijo v zalednem delu e-okolja. Pri vrednotenju končnega uporabnikovega odgovora se lahko upoštevajo še vmesni koraki, napačni poskusi in porabljen čas. Način vrednotenja je odvisen od naloge. Pri tem dopuščamo tudi nekatere nenamerne napake uporabnika, kot so npr. odvečni presledki pri vnosu besedila. E-okolje vsebuje tudi naloge, ki jih ni mogoče natančno avtomatsko ovrednotiti. Za take naloge uporabljamo računalniške algoritme, ki podajo približno oceno, lahko pa te naloge ovrednoti tudi drug uporabnik, običajno v vlogi učitelja ali tutorja.

5. Vaje iz besediloslovja

Delo z besedili v učno-vzgojnem procesu predstavlja svojevrsten izziv. Pogosto se zgodi, da sicer zanimivemu besedilu sledijo tri ali štiri naloge, ki od učečega se zahtevajo poznavanje kraja in časa nastanka besedila, udeležencev, njihovega družbenega razmerja, ki je običajno povsem zadostno opisano s terminoma (ne)formalno in (ne)javno, sledita še vprašnji o temi in sporočevalnem namenu. Za natančno razumevanje besedil pa zadostuje določitev prevladujočega slogovnega postopka in nekaj koreferenc k eni ali dvema referencama. Ostale naloge se nanašajo na slovnica in/ali leksikalna vprašanja. Temu načinu dela se želimo izogniti, in sicer tako, da smo v ta namen oblikovali zbirko avtentičnih besedil in da se vaje in naloge nanašajo izključno na besedilne značilnosti posameznega besedila, s čimer želimo uporabniku pomagati pri razvijanju bralne in funkcionalne pismenosti.

5.1. Zbirka besedil BERTA

Izhodišče za tvorjenje vaj in nalog področja besedil učnega e-okolja *Slovenščina na dlani* predstavlja zbirka BEsedil pRakTičnega sporazumevanjA BERTA. Zbirka ima dva podkorpusa, in sicer govornega (G-BERTA), ki zajema 59 besedil, in pisnega (P-BERTA), ki zajema 216 besedil, skupno 275 besedil. V zbirko vključena besedila so skrbno izbrana. Pred začetkom zbiranja smo namreč

⁶ Npr. pri vaji, kjer je izločena ena sestavina frazema in mora uporabnik med štirimi ponujenimi odgovori izbrati ustreznega (primer: *narediti iz muhe ...* – ob ustreznih

sestavini *slona* z avtomatskim iskanjem težko dobimo relevantne alternativne odgovore, ki zagotavljajo tudi ustrezno zahtevnost naloge).

pregledali učne načrte in učbeniško gradivo ter popisali besedilne skupine, ki so del učne snovi od šestega razreda osnovne šole dalje. Med popisanimi besedilnimi skupinami smo izbrali npr. novico, vremensko napoved, telefonski pogovor, razgovor, govorni nastop, prijavnico, pritožbo, življenjepis, opis postopka, opis življenja osebe, pripoved o življenju osebe, poljudnoznanstveni prispevek. Izbiri besedilnih skupin je sledil izbor tematike oziroma tematskih sklopov. Izhajali smo iz sporazumevalnih tem za poučevanje slovenščine ali katerega koli drugega jezika kot drugega jezika (SEJO, 2010), npr. osebna identiteta, zdravje, odnosi z drugim, izobraževanje, storitve. Skladno s tem smo si zamislili slogane kot načine dela v e-okolju in zbirali avtentična besedila. Prednost smo dajali besedilom, ki so prototipični predstavniki izbranih besedilnih skupin in ki nastajajo v za večino govorcev slovenščine pomembnih ali priljubljenih, v vsakem primeru pa živih in aktualnih komunikacijskih položajih. Z vsebinskega vidika smo zajeli besedila tistih sporazumevalnih tem, ki so po načinu obravnave teme blizu šolajoči se mladini od 11. do 19. leta starosti.

Besedil torej nismo združevali le po besedilnih skupinah – nabor besedilnih skupin smo omejili na 26, kar ni veliko, a je za prvi tovrstni način dela dovolj –, temveč tudi po tematiki. Podobne tematike, npr. *agrumi*, *češnja*, *čokolada*, *sivka*, *paradižnik*, smo povezali v enotne sklope, naslovljene s slikovitimi slogani, npr. *Lačen kot Lakotnik*. Te pa smo združili po dveh merilih, in sicer glede na tematski sklop (*narava*) in glede na razred/letnik, za katerega predvidevamo, da je besedilo primerno za obravnavo. Združevanje besedil hkrati po kriteriju besedilne skupine in tematike učečemu se omogoča, da razlike med besedili iste tematike doživi in dojame kot posledice različnih kontekstov tvorjenja obravnavanega besedila. Učeči se namreč znotraj istega slogana rešuje vaje in naloge za isto tematiko, npr. tematika *agrumi* je predstavljena kot novica (*Veste, zakaj se pomaranče vedno prodajajo v rdečih mrežastih vrečkah?*) ali kot opis postopka (*Solata z agrumi*). To izhodišče je zbiranje gradiva precej počasnilo, saj je bilo treba tako rekoč za vsako besedilo pridobiti soglasje avtorja. Če soglasja nismo uspeli pridobiti, je to pomenilo spremembo koncepta včasih tudi celotnega slogana, npr. glede na zastavljena merila smo našli ustrezno besedilo k nekemu sloganu, vendar za to besedilo nismo uspeli pridobiti soglasja avtorja ali več avtorjev znotraj sklopa. V takih primerih smo bili primorani preoblikovati slogan, kdaj neuporabiti tudi besedila, za katera smo sicer soglasja pridobili.

Da bi bila obravnava besedila kar najbolj podobna interpretaciji besedil, torej spontanemu razbiranju bistvenih in izločanju manj pomembnih podatkov, da lahko iz besedila dojamemo njegov smisel in namen, je bilo treba besedila prikazati v realnem kontekstu, torej ustrezno večpredstavnostno (ali multimodalno). To omogoča in zagotavlja le celotni prikaz besedila (npr. pdf-posnetek besedila), pri čemer so vidni tudi neverbalni elementi besedila (oblikovanje, grafični elementi, video/avdio itd.). Poleg primernosti in zanimivosti tematike ter besedilne skupine za določeno starostno skupino uporabnikov e-okolja je bilo treba upoštevati tudi dolžino besedila. Ta so obsegovno zelo različna: od nekajvrstičnih šal in tudi do

petnajststranskih poljudnoznanstvenih prispevkov. Tako obsežna besedila pri vnašanju nismo delili na odlomke, saj uporabnik sicer ne bi dobil informacij o obsegu besedila kot eni od lastnosti besedilne skupine.

Pri definiranju natančnejšega seznama besedilnih skupin je bilo treba najprej opredeliti različne termine za poimenovanje in pojmovanje množice besedil. V pregledanem osnovno- in srednješolskem gradivu se namreč za poimenovanje množice besedil uporablja termin besedilna vrsta. Besedilne vrste, kot jih razumemo tukaj, se, upoštevaje večrazsežnostni model opredelitve besedila, »konstituirajo prek prototipičnih ena z drugo povezanih kontekstualnih in strukturnih značilnosti. Oblikujejo okvir za prototipične na osnovi konvencij jezikovnih uporabnikov temelječih jezikovnih/govornih vzorcev s karakterističnimi funkcijskimi, medijsko-položajnimi in tematskimi značilnostmi kot tudi glede na te značilnosti ustrezno obliko strukture besedila« (Gansel, Jürgens, 2007). Besedilne vrste so produkti konvencionalnih jezikovnih dejanj znotraj posameznega komunikacijskega področja (Heinemann, 2000). Tovrstna opredelitev besedilne vrste je pokazala, da v učnih procesih učeči se pravzaprav ne spoznavajo značilnosti besedilnih vrst ali le redko, temveč v veliki večini primerov spoznavajo značilnosti besedilnih tipov, npr. prošnja, zahvala, voščilo. Besedilni tip je skupina besedil s skupnimi ali podobnimi le jezikovnimi značilnostmi in je vir za oblikovanje besedilnih vrst. Da bi se izognili nejasnostim in zmedi, smo bili prisiljeni vpeljati termin, ki vključuje tako besedilne tipe kot vrste, to je termin besedilna skupina.

5.2. Definiranje vaj

Ker niti dve besedili nista enaki, četudi obravnavata isto tematiko, je skladno s tem treba oblikovati tudi vaje in naloge. Zato smo najprej popisali tipe nalog, ki so primerni za učno e-okolje in omogočajo tvorjenje več različnih nalog s pomočjo enega vnosa.

Prvi sklop nalog je vezan na zunajjezikovne kontekstualne spremenljivke, to so podatki o udeležencih, njihovem medsebojnem družbenem razmerju, kraju in času tvorjenja besedila, naslovu, mestu objave, dostopnosti besedila za širši krog naslovnikov in kanalu, ki so skupaj s tematskim sklopom, besedilno skupino, temo in jezikovno zvrstjo bili že popisani pri vnašanju besedila v vmesnik BERTA.

Temu sledi vnos dodatnih informacij, ki se uporabijo pri tvorjenju nalog, vezanih na vsebino vsakega posameznega besedila, kamor sodijo tudi naloge o sporočevalnem namenu in slogovnem postopku, in nalog, vezanih na značilno leksiko ali druge besedilotvorne značilnosti besedila.

Poseben sklop nalog je vezan na značilnosti besedila kot predstavnika besedilne skupine. Večina navedenih nalog bo vsaj deloma avtomatizirana.⁷ Za avtomatsko generiranje besediloslovnih nalog besedilom pripisujemo že zgoraj navedene informacije, npr. informacije o udeležencih, temi, sporočevalnem namenu, slogovnem postopku, primernosti besedila za uporabnika določenega razreda oz. letnika, in druge informacije, npr. pravilne in napačne trditve o vsebini konkretnega posameznega besedila.

⁷ Priprava vaj iz besediloslovja je aktivnost, ki je v teku in se zaključuje leta 2021.

Še en prav tako poseben sklop pa predstavljajo naloge tvorjenja besedil, pri katerih želimo, da uporabnik povzame, obnovi izhodiščno besedilo ali tvori besedilo iste ali podobne besedilne skupine.

5.3. Programiranje vaj

Tudi pri vajah iz besediloslovja sistem izvaja tri naloge: izbor ustreznih besedil, tvorjenje naloge ter vrednotenje uporabnikovega odgovora.

Pri izboru ustreznega besedila upoštevamo primernost besedila glede na starost uporabnika, ustreznost besedila za izbrano nalogo ter vsebnost vseh potrebnih informacij, ki jih potrebujemo za tvorjenje naloge.

Vaje iz besediloslovja se pogosto nanašajo na pomen obravnavanega besedila. Računalniški algoritmi imajo težave z razumevanjem pomena besedila, zato takšnih nalog ne moremo povsem avtomatsko tvoriti. Zbirko besedil smo zato opremili z dodatnimi informacijami, ki smo jih vnesli ob pregledu besedil. Računalnik na podlagi teh informacij lahko tvori različne tipe nalog. Če npr. pri vsakem besedilu zapišemo nekaj vprašanj in pravih odgovorov, lahko potem računalnik tvori naloge tipa 'Odgovori na vprašanje', 'Izberi pravi odgovor' ali celo uporabi več vprašanj in tvori nalogo 'Poveži vprašanja s pripadajočimi odgovori'. S takim polavtomatskim pristopom smo prihranili čas, ki bi ga sicer potrebovali za ročno tvorjenje nalog in hkrati povečali nabor vaj, ki se nanašajo na pomen zapisanega ali govorenega besedila.

Pri izpisu nalog v spletnem uporabniškem vmesniku smo poskrbeli za preglednost naloge in obravnavanega besedila. Uporabniku je v vsakem trenutku reševanja omogočen vpogled v izhodiščno besedilo. Hkrati smo naloge opremili z namigi in s kljuki na bazo znanja, ki so uporabniku v pomoč pri reševanju.

Vrednotenje pravih odgovorov poteka enako kot pri ostalih sklopih vaj. Pri nalogah, ki predvidevajo daljši vnos besedila, upoštevamo dva scenarija. Če smo za nalogo vnesli primer pravilne rešitve, lahko z računalniškimi algoritmi izračunamo približno oceno uporabnikovega odgovora. Poleg tega mu lahko ponudimo možnost, da se sam ovrednoti na način, da mu izpišemo predviden pravilni odgovor. Če predvidenega pravih odgovora nismo vnesli, lahko uporabnikov odgovor ovrednoti drug uporabnik, običajno v vlogi učitelja ali tutorja. Izbiro načina ovrednotenja v takšnih primerih prepuščamo uporabniku, to je običajno učitelj, ki dodeli naloge v reševanje učencem.

6. Vključevanje razlag

Ob izbranih vajah vseh štirih vsebinskih sklopov bodo uporabnikom na voljo krajše sprotne razlage in dostop do baze znanja, kjer bodo obširneje predstavljene izbrane jezikovne teme.⁸

Osnovni cilj vključevanja razlag je takojšnja podpora pri reševanju vaj in nalog ter s tem povezano doseganje nemotene uporabe uporabnikovega napredovanja. Sprotne razlage ob reševanju nalog bodo namreč uporabnikom omogočale, da se zaradi težav pri reševanju naloge ali nepoznavanja konkretnega pravopisnega, slovničnega,

frazeološko-paremiološkega ali besedilnega problema seznanijo z vsebino, ki bo nazorno in razumljivo pojasnila izpostavljene jezikovne teme. Če bo uporabnik želel ponujene vsebine še dodatno raziskati, bo lahko po njih brskal tudi prek posebnega zavihka v e-okolju, ki bo opravljal funkcijo jezikovnega priročnika.

Razlage⁹ bodo prilagojene zmožnostim oz. predznanju učečih se in bodo smiselno vključene k pripravljenim vajam in nalogam. Da bi se približali celotni ciljni publiki, tj. učencem med 11. in 19. letom, bodo razlage razumljive, preproste in opremljene z verodostojnimi in nazornimi primeri. Uporabnik bo lahko priklical razlago kadarkoli med reševanjem nalog s klikom na ustrezno ikono. Sprotna razlaga ob reševanju nalog in vaj se bo prikazala najprej kot povzetek oziroma kratka razlaga v povedi ali dveh. To bo uporabniku pomoč, na kaj mora biti pri nalogi najbolj pazljiv ali da si na hitro razjasni največjo zagato pri reševanju. Če mu to ne bo zadostovalo za uspešno reševanje naloge, se mu bo s klikom na gumb *Preberi več* odprla daljša razlaga, ki bo podrobnejša, obširnejša, po potrebi opremljena s slikovnimi elementi in tabelami ter primeri. Ta razlaga bo zadostovala za uspešno rešeno nalogo in seznanitev z najpomembnejšimi jezikovnimi vsebinami pri nalogi. Če bo uporabnik o ponujeni temi želel izvedeti še več ali ga bodo zanimale povezane jezikovne vsebine, bo s klikom na gumb *Preberi še* preusmerjen k poglavju, kjer bodo povezane izbrane vsebine predstavljene še bolj celostno, ali s klikom na dodane povezave k drugim relevantnim spletnim jezikovnim priročnikom in virom.

Za vnos razlag smo pripravili poseben vmesnik, v katerega bomo vnesli več kot 200 razlag, ločenih v sprotne razlage za pomoč pri izvajanju vaj ter združene v poglavja baze znanj Orehek. Poleg razlag, ki jih bomo pripravili avtorji e-okolja Slovenščina na dlani, pa bodo razlage lahko vnašali tudi uporabniki sami. Teme teh razlag bodo definirane vnaprej in povezane s posameznimi vajami in nalogami, njihovo vsebino pa bo izpolnil uporabnik učenc in pregledal in popravil učitelj in/ali član projektne skupine Slovenščina na dlani, preden bo vidna ostalim uporabnikom.

7. Zaključek

V projektu Slovenščina na dlani s pomočjo jezikovnih virov in avtomatiziranih postopkov izdelujemo interaktivno učno e-okolje za podporo učenju slovenščine v osnovnih šolah od 6. razreda naprej in v srednjih šolah. Za potrebe tega e-okolja smo razvili manjši korpus besedil, primernih za mladino (MAKS), večmodalno zbirko besedil praktičnega sporazumevanja (BERTA), slovar frazemov in pregovorov (FRIDA) in bazo razlag (Orehek). Izdelani viri bodo na voljo tudi kot jezikovni vir pod licencami CC BY, razen v primerih, kjer to ni mogoče zaradi omejitev pri izvornih avtorskih pravicah besedil oz. vsebin. Dostopni so oz. bodo prek repozitorija CLARIN.SI oz. prek Clarinovih konkordančnikov (<https://www.clarin.si/noske/>).

Do oktobra 2020 je pripravljenih okrog 1000 različnih vaj s področja pravopisa, slovnic in frazeologije ter

⁸ V slovenskem prostoru je podoben princip že uporabljen v Pedagoškem slovničnem portalu, ki je nastal v okviru

projekta Sporazumevanje v slovenskem jeziku (prim. <http://slovnica.slovenscina.eu/>).

⁹ Aktivnost priprave razlag še ni zaključena.

besedil, izdelane pa bodo tudi vsebinske razlage za več kot 200 različnih tematik. Za vsako vajo je na voljo večje število različnih primerov, pri pravopisu in slovnici tudi do 500 za eno vajo.

V šolskem letu 2020/21 nadaljujemo z zadnjo fazo izgradnje vaj, tj. s priklicem vaj iz zalednih baz v spletni vmesnik e-okolja, dodajanjem besedilnih nalog in pisanjem razlag, izdelavo različnih funkcionalnosti, kot so prikazovanje doseženih rezultatov, prikazovanje razlag ob vajah ipd., ter začnemo testiranje e-okolja in analizo vedenja uporabnikov. S šolskim letom 2021/22 bo e-okolje odprto dostopno zainteresirani javnosti.

Avtomatizacija s podporo jezikovnih tehnologij in digitalno okolje imata nekaj prednosti, ki jih papirni medij ne omogoča: veliko količino primerov in vaj, prilagajanje zahtevnosti vaj znanju uporabnika, avtomatsko ocenjevanje in usmerjanje med vajami, enostavno priklicljivo pomoč v obliki razlag, prilagojenih vsaki posamezni nalogi, podporo pri reševanju z namigi ali sprotno komunikacijo z drugimi uporabniki sistema, motiviranje z grafičnimi elementi, načini točkovanja ali sodelovanjem v skupinah. Ob raziskovanju možnosti za motivacijo in spodbujanje ustvarjalnosti pa smo ugotavljali tudi omejitve v primerjavi z osebnim stikom pri učenju. Ob tem da je učno e-okolje visoko avtomatizirano in individualizirano, namreč omogoča veliko manj ustvarjalnih in interaktivnih nalog. Avtorji e-okolja Slovenščina na dlani zato že od same zasnove okolja naprej sledimo načelu, da je digitalno učno e-okolje koristno dopolnilo in popestritev pouka, nikakor pa ne nadomestilo za tradicionalne oblike poučevanja.

8. Literatura

- Christina Gansel in Frank Jürgens. 2007. *Textlinguistik und Textgrammatik. Eine Einführung*. 2. Auflage. Vandenhoeck & Ruprecht, Göttingen.
- Miha Grčar, Simon Krek in Kaja Dobrovoljc. 2012. Obeliks: statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik. V: *Zbornik Osme konference Jezikovne tehnologije*, str. 82–87. (<http://nl.ijs.si/ijst12/JezikovneTehnologije2012.pdf>)
- Wolfgang Heinemann. 2000. Textsorte – Textmuster – Texttyp. V: K. Brinker, G. Antos, W. Heinemann, S. F. Sager, ur., *Text- und Gesprächslinguistik: ein internationales Handbuch zeitgenössischer Forschung*, str. 507–523, Handbücher zur Sprach- und Kommunikationswissenschaft, Zv. 16. Walter de Gruyter, Berlin, New York.
- Vida Jesenšek in Natalija Ulčnik. 2014. Spletni frazeološko-paremiološki portal: redakcijska vprašanja ob slovenskem jezikovnem gradivu. V: V. Jesenšek, S. Babič, ur., *Več glav več ve: Frazeologija in paremiologija v slovarju in vsakdanji rabi*, str. 276–292, Oddelek za germanistiko, Filozofska Fakulteta Univerze v Mariboru in ZRC SAZU Ljubljana, Inštitut za slovensko narodopisje.
- Iztok Kosem, Mojca Stritar, Sara Može, Ana Zwitter Vitez, Špela Arhar Holdt, Tadeja Rozman. 2012. *Analiza jezikovnih težav učencev: korpusni pristop*. Trojina, zavod za uporabno slovenistiko, Ljubljana.
- Mira Krajnc Ivič. 2020. Obravnava besedil: merila za razlikovanje med besedilno vrsto in besedilnim tipom. *Slavistična revija*, 68(1):55–71. (<https://srl.si/ojs/srl/article/view/2020-1-1-4>)
- Erika Kržišnik. 2006. Izraba semantične potence frazemov. *Slavistična revija*, 56(1):259–279.
- Matej Meterc. 2017. *Paremiološki optimum: Najbolj poznani in pogosti pregovori ter sorodne paremije v slovenščini*. Založba ZRC, ZRC SAZU, Ljubljana.
- Pedagoški slovnični portal*. Sporazumevanje v slovenskem jeziku (<http://slovnica.slovenscina.eu/>).
- Jenny Rose Finkel, Trond Grenager in Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. V: *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, str. 363–370. (<http://nlp.stanford.edu/~manning/papers/gibbscrf3.pdf>)
- Tadeja Rozman, Irena Krapš Vodopivec, Mojca Stritar, Izток Kosem. 2020. *Empirični pogled na pouk slovenskega jezika*. Znanstvena založba FF UL, Ljubljana. (<https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/227/327/5303-1>)
- SEJO = Skupni evropski jezikovni okvir: učenje, poučevanje, ocenjevanje*. Irena Kovačič (pr). El. knjiga. Ljubljana: Ministrstvo RS za šolstvo in šport, Urad za razvoj šolstva, 2011. (<https://centerslo.si/wp-content/uploads/2015/10/SEJO-komplet-za-splet.pdf>)
- Peng Qi, Timothy Dozat, Yuhao Zhang in Christopher D. Manning. 2018. Universal Dependency Parsing from Scratch. V: *Proceedings of the CoNLL 2018 UD Shared Task*, str. 160–170. (<https://www.aclweb.org/anthology/K18-2000.pdf>)
- Natalija Ulčnik. 2019. Izbor frazemov za bazo FRIDA. V: N. Ulčnik, ur., *Slovenščina na dlani 2*, str. 37–45, Univerzitetna založba Univerze v Mariboru. (<https://press.um.si/index.php/ump/catalog/book/447>)
- Natalija Ulčnik in Matej Meterc. 2019. Izbor pregovorov za bazo FRIDA. V: N. Ulčnik, ur., *Slovenščina na dlani 2*, str. 47–55, Univerzitetna založba Univerze v Mariboru. (<https://press.um.si/index.php/ump/catalog/book/447>)