

Predlog za izdelavo Slovarja sodobnega slovenskega jezika



Simon Krek
Amebis, d.o.o., Kamnik in
Institut "Jožef Stefan"
(Laboratorij za umetno
inteligenco)

Iztok Kosem
Trojina, zavod za uporabno
slovenistiko

Polona Gantar
raziskovalka

Dokument prinaša podroben opis postopkov izdelave spletnega *Slovarja sodobnega slovenskega jezika* na podlagi dobrih leksikografskih praks, jezikovnotehnoloških postopkov in korpusne metodologije. Predlog temelji na konceptu izrabe leksikalno-gramatičnih podatkov, strukturiranih v slovarski bazi, in podaja način izvedbe s tehničnega, vsebinskega, organizacijskega in finančnega vidika.

v1.1 | 22. 5. 2013

AVTORJI:

dr. Simon Krek – Amebis, d.o.o., Kamnik / Institut »Jožef Stefan«, Laboratorij za umetno inteligenco

Doktoriral leta 2011 na Oddelku za slovenistiko, Filozofska fakulteta Univerze v Ljubljani z disertacijo *Pridobivanje jezikovnih podatkov iz besedilnih korpusov za namen izdelave enojezičnih slovarjev in slovníc*. Med leti 1994-2004 je bil zaposlen na založbi DZS kot urednik za slovarje. Od 1995-2006 je bil glavni urednik slovarja Veliki angleško-slovenski slovar Oxford-DZS. V okviru tega projekta je v letih 1997-2000 koordiniral sestavljanje 100-milijonskega besedilnega korpusa FIDA. Med 2005-2007 je bil zaposlen na Filozofski fakulteti Univerze v Ljubljani kot raziskovalec in koordinator gradnje besedilnega korpusa FidaPLUS, 620-milijonskega korpusa slovenskega jezika. Od 2007 je zaposlen na Institutu »Jožef Stefan« kot raziskovalec – v tem času je v projektu Jezikoslovno označevanje slovenskega jezika koordiniral skupino, ki je pripravila strokovne osnove za označevanje učnega korpusa za avtomatsko obdelavo besedil v slovenskem jeziku. Od 2008 zaposlen na podjetju Amebis, d.o.o., Kamnik kot koordinator projekta Evropskega socialnega sklada Sporazumevanje v slovenskem jeziku, ob dodatni raziskovalni zaposlitvi v Laboratoriju za umetno inteligenco na Institutu »Jožef Stefan«, kjer kot strokovnjak za jezikovne tehnologije sodeluje pri evropskih raziskovalnih projektih. Strokovno se ukvarja z leksikografijo in leksikogramatiko, korpusnim jezikoslovjem, računalniško obdelavo naravnih jezikov, jezikovnotehnološko infrastrukturo in računalniško podprtim učenjem in poučevanjem jezikov. Med leti 2007-2010 je bil tajnik Slovenskega društva za jezikovne tehnologije, od l. 2010 je član izvršnega odbora evropskega združenja leksikografov Euralex. Je tudi član organizacijskih odborov mednarodnih konferenc *Electronic lexicography in the 21st century* 2011 na Bledu in 2013 v Tallinnu.

dr. Iztok Kosem - Trojina, zavod za uporabno slovenistiko, Ljubljana

Doktoriral leta 2010 na Univerzi Aston v Veliki Britaniji z disertacijo *Designing a model for a corpus-driven dictionary of academic English (Izdelava modela za slovar akademske angleščine, temelječ na polnem korpusnem pristopu)*. Med leti 2000-2005 je delal kot leksikograf na slovarju Veliki angleško-slovenski slovar Oxford-DZS, v letih 2004-2005 je opravil magisterij iz Jezikoslovja in leksikografije na Univerzi v Birminghamu v Veliki Britaniji z magistrsko nalogo *A model for definitions in a new monolingual dictionary of the Slovenian language (Model za definicije v novem enojezičnem slovarju slovenskega jezika)*. V času doktorskega študija (2006-2010) je bil na Univerzi Aston pomočnik direktorja pri projektu Aston Corpus Network, v okviru katerega so se zbirali in označevali enojezični in vzporedni korpusi angleščine, španščine, francoščine in nemščine. Od leta 2010 je direktor zavoda za uporabno slovenistiko Trojina in pomočnik koordinatorja pri projektu Evropskega socialnega sklada *Sporazumevanje v slovenskem jeziku*. Ukvarja se z leksikografijo in jezikovnimi tehnologijami za podporo leksikografskemu delu, korpusnim jezikoslovjem, poučevanjem jezikov in izdelavo uporabniško prijaznih jezikovnih orodij in vmesnikov. Je vodja organizacijskih in znanstvenih odborov mednarodnih konferenc *eLex (Electronic lexicography in the 21st century)* 2011 na Bledu in 2013 v Talinu, bil je vodja in pomočnik vodje organizacijskih odborov mednarodnega simpozija Aston Corpus Symposium (Aston Corpus Symposium) v letih 2007, 2008 in 2009. Letos je bil povabljen v elitno družbo leksikografov, ki se bo

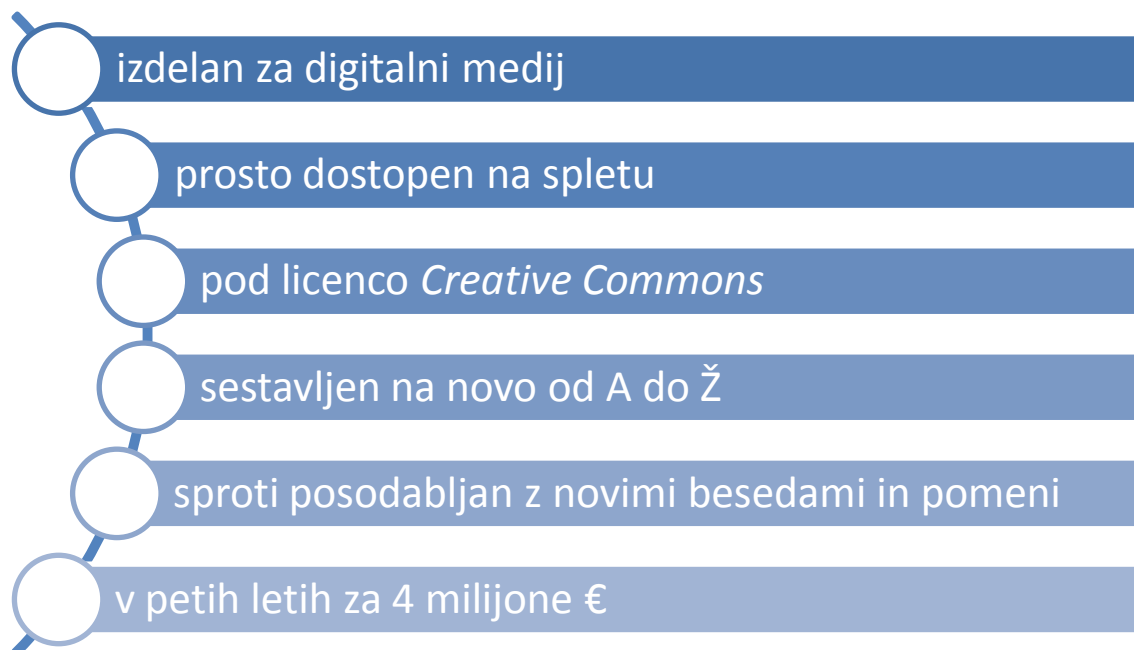
sestala poleti 2013 v Oxfordu in razpravljala o (elektronski) prihodnosti najpomembnejšega slovarja angleškega jezika – Oxford English Dictionary.

dr. Polona Gantar

Doktorirala leta 2005 na Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani z disertacijo *Frazem in njegovo besedilno okolje*. Leta 1997 se je kot mlada raziskovalka zaposlila na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU, kjer se je raziskovalno usposabljala v Sekciji za terminološke slovarje, danes pa je zaposlena kot znanstvena sodelavka Leksikološke sekcije. Po opravljenem doktoratu je med leti 2005–2008 vodila Fakulteto za slovenske študije Stanislava Škrabca Univerze v Novi Gorici, kasneje pa bila koordinatorica študijskih programov. Med leti 2007–2011 je na Fakulteti za humanistične študije Univerze v Novi Gorici predavala leksikalno-gramatične vsebine, korpusno jezikoslovje in jezikovne tehnologije. Med leti 2008 in 2012 je vodila aktivnost »Izdelava leksikalne baze za slovenščino« v okviru projekta Evropskega socialnega sklada Sporazumevanje v slovenskem jeziku in sodelovala pri oblikovanju uporabniškega vmesnika za iskanje po korpusu Gigafida in pri izdelavi pedagoškega slovničnega portala. Je članica organizacijskega in znanstvenega odbora mednarodnih konferenc eLEX 2011 in eLEX 2013: Electronic lexicography in the 21st century in urednica jezikoslovne revije Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave. Poleg korpusne leksikografije, leksikologije in semantike se raziskovalno ukvarja s slovensko frazeologijo in terminologijo.

O SLOVARJU NA ENI STRANI

- ▶ v celoti **izdelan na novo** na podlagi podatkov o sodobni slovenščini od leta 1991 naprej
- ▶ zgleduje se po sodobnih evropskih leksikografskih praksah, ki opuščajo tiskani medij in vsebino slovarjev prilagajajo **digitalnemu okolju**
- ▶ na enem mestu **vsebuje vse vrste podatkov** o slovenskem jeziku, od besednih pomenov, slovničnih informacij, normativnih opozoril, historičnega izvora ter multimedijskih vsebin
- ▶ v času trajanja projekta **redno posodabljan** v skrbno načrtovanem procesu, ki zajema avtomatsko zaznavanje novih besed in pomenov, leksikografsko obdelavo in takojšen prikaz
- ▶ s pomočjo spletnega sistema za izkoriščanje moči množic (*crowdsourcing*) pri izdelavi lahko **sodeluje vsak materni govorec** slovenskega jezika
- ▶ temelji na uporabniku prijaznem razlagalnem jeziku s prikazom vseh podatkov na intuitiven in hitro dostopen način, **učenje uporabe slovarja ni potrebno**
- ▶ zasnovan kot obsežna slovarska baza podatkov, ki **omogoča rabo za različne namene**, denimo za potrebe jezikovnih tehnologij za slovenščino
- ▶ spletna različica in baza podatkov sta dostopni vsem uporabnikom pod **licenco Creative Commons** – Priznanje avtorstva+Deljenje pod enakimi pogoji, podobno kot Wikipedija
- ▶ vsakdo lahko slovarsko bazo kadarkoli **shrani na svojem računalniku** in jo uporablja, spreminja in dopolnjuje v skladu z licenco
- ▶ osnovni podatki o vseh geslih bodo dostopni že **pol leta po začetku** projekta
- ▶ skupna vrednost je nekaj nad 4 milijone evrov, 100.000 gesel, **v petih letih** od začetka projekta (letni strošek približno 800.000€)



O SLOVARJU NA ENI STRANI	3
UVOD	7
1 Stanje na področju slovenskih leksikalnih virov in evropski okvir	9
1.1 Slovenski leksikalni viri in orodja.....	9
I. Baze podatkov.....	10
II. Spletni portali (in podatkovne baze, na katerih temeljijo)	10
III. Elektronske zbirke besedil (korpusi)	11
IV. Jezikovnotehnološka orodja	11
1.2 Jezikovnotehnološka in slovarska podpora v evropskem prostoru	12
1.3 Primeri dobre prakse	13
1.3.1 DANTE (Database of Analysed Text of English)	13
1.3.2 Splošni nizozemski slovar (ANW: Algemeen Nederlands Woordenboek)	14
1.3.3 Leksikalna podatkovna baza za francoščino (BFL: Base lexicale du français).....	16
1.3.4 Veliki slovar poljskega jezika (Wielki słownik języka polskiego).....	17
1.3.5 Primeri dobre prakse - sklep	18
2 Slovar sodobnega slovenskega jezika	20
2.1 Splošne lastnosti	20
2.1.1 Uporabnik.....	20
2.1.2 Medij	22
2.1.3 Obseg	23
2.2 Vsebina	24
2.2.1 Vrsta slovarskih informacij in elementi slovarja	25
Osnovni vmesnik.....	25
Slovarski meni.....	27
Naslovna vrstica	27
Zavihek: Pomen	29
Zavihek: Frazeologija.....	36
Zavihek: Oblike	37
Zavihek: Sinonimi	38
Zavihek: Izvor	39
Zavihek: Govor	40
Zavihek: Norma.....	41
Zavihek: Vizualizacija.....	43
Zavihek: Multimedija	44
Zavihek: Statistika	44
2.2.2 Določitev prioritete zaporedja priprave gesel	45
2.2.3 Detektiranje pomenskih sprememb besedišča	46

2.2.4 Sprotno posodabljanje slovarskih gesel	48
2.2.5 Možnosti tiskanja	49
2.3 Izvedba: izdelava slovarskega gesla v posameznih fazah	52
1. Rdeča faza – avtomatsko izdelano geslo	54
2. Oranžna faza – prečiščeno geslo	55
3. Rumena faza – pomensko razčlenjeno geslo	58
4. Modra faza – geslo z dodanimi »zunajbaznimi« podatki	60
5. Zelena faza – končano geslo	60
3 Leksikalna baza za slovenščino kot vir podatkov za <i>Slovar sodobnega slovenskega jezika</i>	61
3.1 Motiv in ozadje	61
3.2 Namen in uporabnik	61
3.3 Razmerje med leksikalno bazo in slovarjem	62
3.4 Zgradba in organizacija leksikalno-gramatičnih podatkov	62
3.5 Shema DTD	64
3.6 Orodja za pridobivanje in analizo leksikalno-gramatičnih podatkov	68
3.6.1 Analiza konkordančnega niza	69
3.6.2 Leksikalno-gramatični profil leksikalne enote in orodje Sketch Engine	69
3.6.3 Avtomatsko pridobivanje podatkov iz korpusa	71
3.7 Leksikografski postopki – opis in teoretični premisleki	72
3.7.1 Lema – iztočnica	72
3.7.2 Pomen	78
3.7.2.1 Pomenska členitev	78
3.7.2.2 Razvrstitev pomenov	80
3.7.2.3 Pomenski opis	80
3.7.2.3.1 Pomenski meni	81
3.7.2.3.2 Pomenski indikatorji	81
3.7.2.3.3 Pomenske sheme	83
3.7.2.3.4 Defnicijske strategije	86
3.7.2.3.5 Stalne besedne zveze	89
3.7.2.3.6 Frazeološke enote	91
3.7.2.4 Oznake in restrikcije	93
3.7.3 Skladenjski nivo	96
3.7.3.1 Skladenjske strukture	97
3.7.3.2 stavčni vzorci	99
3.7.4 Kolokacije	100
3.7.4.1 Razširjene kolokacije	101
3.7.4.2 Skladenjske zveze	102
3.7.5 Korpusni zgledi	104

4 Organizacija dela.....	105
4.1 Organizacija leksikografskega dela.....	105
4.2 Organizacija dela – drugi človeški viri.....	106
4.3 Organizacija dela – drugo	108
5 Finančni načrt	109
5.1 Razmerja med tipi stroškov	109
6 Kazalniki.....	111
Kazalnik K1.....	113
Kazalnik K2.....	113
Kazalnik K3.....	114
Kazalnik K4.....	114
Kazalnik K5.....	114
Kazalnik K6.....	115
Kazalnik K7.....	115
Kazalnik K8.....	115
Kazalnik K9.....	115
Kazalnik K10.....	116
Kazalnik K11.....	116
Kazalnik K12.....	116
Kazalnik K13.....	116
Kazalnik K14.....	117
Kazalnik K15.....	117
Kazalnik K16.....	117
7 REFERENCE/BIBLIOGRAFIJA	119
Reference	119
Bibliografija.....	122
I. Kritike SSKJ	122
II. Leksikalna baza za slovenščino.....	123
III. Slovarske in semantične podatkovne baze	124
IV. Sodobna slovarska teorija in praksa	129
V. E-leksikografija	134
VI. Korpusi in korpusna leksikografija.....	138
VII. Jezikovne tehnologije	143

UVOD

V slovenski jezikovni skupnosti, tako laični kot strokovni, že od 90. let dalje obstaja kontinuirana in eksplicitna potreba po jezikovnih priročnikih, ki bi odražali sodobno stanje jezika tako glede upoštevanja jezikovnih sprememb kot sodobnih leksikografskih metod pri njihovi izdelavi. V zadnjem času se je temu pridružila še zahteva po njihovi prosti dostopnosti in digitalizirani obliki.¹

Za področje slovarskih opisov, na katere se osredotočamo v konkretnem predlogu, je bil leta 2009 pod okriljem SAZU in Inštituta za slovenski jezik Frana Ramovša ZRC SAZU organiziran Strokovni posvet o novem slovarju slovenskega jezika (2009), ki pa je bil kljub podpori Ministrstva za kulturo RS zaprte narave. Na posvetu so sodelovali povabljeni predstavniki slovenskih akademskih ustanov, pretežno Inštituta za slovenski jezik Frana Ramovša in Filozofske fakultete Univerze v Ljubljani, ter gostji z Inštituta za hrvatski jezik in jezikoslovje ter Jezikoslovnega oddelka Inštituta za slovaški jezik. V celoti je izostala jezikovnotehnološka skupnost in evropska slovarska praksa zadnjih dvajsetih let. Na posvetu so bili med drugim izpostavljeni sklepi, da naj bi novi enojezični razlagalni slovar temeljil na izročilu Slovarja slovenskega knjižnega jezika (npr. prevzete terminološke gradivske osnove; vztrajanje pri opisovanju jezika na podlagi njegovih teoretičnih jezikovnosistemskih lastnosti), zasnovan pa bi bil primarno v knjižni obliki (Strokovni posvet 2009: 12, Snoj 2012: 96). Čas izdelave slovarja je bil predviden na podlagi izkušenj pri izdelavi SSKJ, in sicer: 1 leto za kadrovske, tehnološke, gradivske in konceptualne priprave in 6 let za izdelavo slovarja v obsegu 80.000 gesel, pri katerem bi sodelovalo 10 avtorjev s polno delovno obveznostjo (Strokovni posvet 2009: 20). V času od posveta pa do danes koncept novega slovarja ni bil javno predstavljen, če izvajamo prispevek na Škrabčevih dnevih (Snoj 2012), iz posameznih odzivov ob izidu Slovarja novejšega besedja slovenskega jezika (Bizjak Končar idr. 2013) pa je mogoče razbrati, da se novi slovar kot nadaljevalec vloge SSKJ kot osrednjega slovarskega jezikovnega dela premika v smer digitalnega okolja.

V novem Nacionalnem programu za jezikovno politiko 2012–2016² je poudarjeno, da je »od časa sprejema pretekle resolucije o nacionalnem programu jezikovne politike prepoznavna sprememba na področju jezikovne infrastrukture predvsem pospešena digitalizacija in razvoj informacijskih in komunikacijskih tehnologij. Ta dva procesa postavljata okvir za nadaljnji razvoj infrastrukture, ki bo glede na svetovne in evropske trende ter hitrost sprememb predvsem digitalna, spletna in mobilna.« O obstoječem slovarju (in slovnici) resolucija pravi: »Druga pomanjkljivost obeh temeljnih del je, da sta bili napisani za prikaz v knjižni obliki in da je le temeljni slovar prešel v digitalno okolje, kateremu zaradi prvotne knjižne zasnove ni prilagojen v polni meri.«

Stanje v leksikografiji kot tudi njen nadaljnji razvoj in pogled na slovarsko informacijo v smislu zgornjih navedb določajo spremembe, ki so se v zadnjih 25 letih, predvsem pa v zadnjih nekaj letih, zgodile na področju digitalnih medijev. V ospredju je **uporabnik**, ki je več digitalnih medijev in zna izrabljati njihove prednosti, s tem pa posredno odloča tudi o vrsti, obliki in načinu predstavitve slovarskih in drugih

¹ Pregled recenzij je na voljo na spletni strani Inštituta za slovenski jezik Frana Ramovša: <http://goo.gl/DGLyf>.

² Spletna stran na Ministrstvu za kulturo: <http://goo.gl/a1P06>.

jezikovnih podatkov. Knjižne oblike priročnikov so tako postale manj uporabne, na novo pa se postavljajo tudi vprašanja in izzivi, kaj vse lahko dejansko ponudi novodobni slovar in v kakšni obliki je mogoče najbolj učinkovito predstaviti za jezikovnega uporabnika zanimive informacije. Raziskave sodobnih leksikografskih potreb in trendov³ so pokazale, da so med pričakovanji slovarskih uporabnikov najvišje ovrednotene: zanesljivost vsebine, jasnost informacije, hitra in brezplačna dostopnost, prilagodljivost vsebin, možnost sprotnega posodabljanja, večpredstavne vsebine in povezave na jezikovne vire in enciklopedične podatke. Z vidika snovalcev slovarske informacije na drugi strani velja, da od sodobnega jezikovnega uporabnika ne morejo več pričakovati, da bo svoj jezikovni problem identificiral glede na obstoječo metajezikovno klasifikacijo, tj. glede na to, ali je njegova jezikovna zadrega pomenske, pravopisne ali stilistične narave, pač pa, da je sposoben problem identificirati in ubesediti v obliki in mediju, ki ga navadno uporablja v procesu komuniciranja, besedilne produkcije in interpretacije.

Poleg menjave medija in novih uporabniških pričakovanj je za slovarski predlog pomemben tudi izjemno hiter razvoj **jezikovnih tehnologij**, v našem primeru jezikovnih tehnologij za slovenščino, ki v novih medijih (splet, mobilna telefonija in internet, računalništvo v oblaku itd.) potrebujejo vsa orodja, ki so na voljo za druge (evropske uradne) jezike in so v temelju odvisne od sodobnih jezikovnih virov, kot je denimo predlagani slovar oz. slovarska baza, v kateri je opisana sodobna slovenščina. Teh potreb pri sestavljanju bodočega slovarja ni mogoče spregledati, kajti kot ugotavljajo strokovnjaki s tega področja,⁴ so manjši evropski jeziki izrazito podhranjeni v primerjavi z tistimi, ki imajo več govorcev, pri katerih je komercialni interes nevprašljiv, imajo strokovnjake s tega področja in so v tehnologije in vire vlagali dolga leta. Vsak novi slovar, v katerega bo slovenska govorna skupnost vložila določena sredstva, mora torej služiti dvojnemu namenu – za informiranje govorcev o vseh vidikih njihovega jezika in za (predvsem semantično) podporo tehnološkim aplikacijam, brez katerih noben jezik ne bo preživel v digitalni dobi, ki slovenščine v prihodnjem obdobju v nobenem primeru ne bo obšla. Da bi bil novi slovar čim bolj koristen tudi za jezikovnotehnološke namene, mora biti že od začetka zasnovan kot baza podatkov, ki je tudi računalniško procesljiva, predvsem pa mora biti od začetka načrtovan z licenco, ki omogoča povsem prosto razpolaganje in izkoriščanje te baze tudi za druge, vnaprej nepredvidljive namene. Predlagana licenca za slovar je torej enaka, kot jo ima Wikipedija, trenutno najbolj uporabljena baza podatkov znotraj jezikovnih tehnologij: **Creative Commons – priznanje avtorstva, deljenje pod enakimi pogoji CC BY-SA**.

ZASNOVO NOVEGA SLOVARJA JE SMISELNO OBRAVNAVATI ZNOTRAJ CELOSTNE SLIKE JEZIKOVNE OPREMLJENOSTI JEZIKA, TAKO V SMISLU JEZIKOVNIH PRIROČNIKOV KOT V SMISLU JEZIKOVNIH PODATKOV, KI JIH JE MOGOČE RAČUNALNIŠKO PROCESIRATI, PRI TEM PA VKLJUČITI VSE KOMPETENTNE AKTERJE.

³ Glej razdelek E-LEKSIKOGRAFIJA v poglavju Bibliografija.

⁴ Glej razdelek JEZIKOVNE TEHNOLOGIJE v poglavju Bibliografija.

1 STANJE NA PODROČJU SLOVENSKIH LEKSIKALNIH VIROV IN EVROPSKI OKVIR

1.1 SLOVENSKI LEKSIKALNI VIRI IN ORODJA

Slovenščina je, upoštevajoč obstoječe jezikovne priročnike ter razvitost in dostopnost jezikovnih virov, v nezavidljivem položaju. Poročilo evropske mreže odličnosti META-NET za slovenščino (Krek 2012a) med drugim ugotavlja, da:

- ▶ manjkajo računalniško procesljivi viri, ki poleg jezikovnotehnoloških orodij predstavljajo temelj za uporabo v različnih digitalnih okoljih, kot sta denimo internet in mobilna telefonija;
- ▶ obstaja izrazita vrzel pri enojezičnih in dvojezičnih leksikalno-skladenjsko-semantičnih virih;
- ▶ ne obstajajo sistematične študije o standardni slovenščini, kar posledično povzroča diskrepanco med realno jezikovno rabo in slovenščino, ki se ohranja pri življenju znotraj obstoječih jezikovnih priročnikov in šolskih učbenikov.

Slovar slovenskega knjižnega jezika (SSKJ) kot trenutno najbogatejši vir leksikalnih informacij o besedišču slovenskega jezika je z vidika časa nastanka glede na današnje potrebe nerelevanten in pomanjkljiv.⁵ Ne vsebuje sodobnih izrazov, pomenov in rabe. Uporablja slovarski metajezik, ki je težko razumljiv in neintuitiven. Ker informacije v njem niso izražene eksplicitno, jih za računalniško procesiranje ni mogoče uporabljati, hkrati pa kot podatkovna zbirka ni dostopen niti za raziskovalne namene, torej ga ni mogoče povezati z drugimi obstoječimi bazami podatkov (npr. sloWNet). Slovar novejšega besedja slovenskega jezika (Bizjak Končar idr. 2013), ki naj bi zapolnil 20-letno vrzel na področju slovarskega opisa slovenske leksike, potreb uporabnika ne more zadovoljevati, saj ni zasnovan kot sodobna (digitalna) slovarska baza podatkov, ki jo je mogoče prikazati na spletu izven izvirnega knjižnega formata, in ne temelji na bazi procesljivih leksikalnogramatičnih podatkov. Čeprav je produkt proračunsko financirane javne znanstvene ustanove, naj bi bil (brezplačno?, za prenos?) na voljo na spletu šele leta 2015. Ker ne predvideva možnosti sprotnega aktualiziranja glede na detektirane leksikalne spremembe v jeziku, je v smislu slovarskega koncepta, ki metodološko, vsebinsko in metajezikovno sledi SSKJ, vedno znova obsojen na odražanje odmikajoče se jezikovne situacije in predstavlja v slovenskem leksikografskem razvoju in praksi prej nazadovanje kot napredek.

OD SODOBNEGA *SLOVARJA SLOVENSKEGA JEZIKA* SE ZATO PRIČAKUJE, DA BO HKRATI ZADOVOLJEVAL DVE TEMELJNI POTREBI: **1. JEZIKOVNIM UPORABNIKOM BO NA VOLJO KOT *SPLETNI SLOVARSKI PRIROČNIK*, V KATEREM BODO PODATKI PREDSTAVLJENI NA RAZUMLJIV IN INTUITIVEN NAČIN, SPROTI POSODABLJANI TER POVEZANI Z DRUGIMI (ŽE OBSTOJEČIMI IN NA NOVO ZGRAJENIMI) BAZAMI JEZIKOVNIH PODATKOV**

⁵ Glej razdelek KRITIKE SSKJ v poglavju Bibliografija.

2. ZA UPORABO V DIGITALNIH OKOLJIH IN ZA RAZVOJ SODOBNIH JEZIKOVNIH APLIKACIJ BO NA VOLJO CELOTNI JEZIKOVNOTEHNOLOŠKI SKUPNOSTI KOT **DIGITALNA IN S TEM RAČUNALNIŠKO PROCESLJIVA **PODATKOVNA BAZA**, V KATERI BODO STRUKTURIRANI LEKSIKALNI, SKLADENJSKI IN SEMANTIČNI PODATKI JEZIKOSLOVNO INTERPRETIRANI IN POVEZLJIVI Z DRUGIMI PODATKOVNIMI BAZAMI.**

Med že izdelanimi viri za slovenščino in jezikovnotehnološkimi orodji, prilagojenimi za slovenski jezik, ki so posredno ali neposredno povezani z izdelavo *Slovarja sodobnega slovenskega jezika*, je treba omeniti:

I. BAZE PODATKOV

- ▶ **Leksikalna baza za slovenščino**,⁶ izdelana pri projektu *Sporazumevanje v slovenskem jeziku* (SSJ),⁷ vsebuje 2.500 iztočnic in zanje podaja izčrpen pomensko-skladenjski opis s pomenskega, skladenjskega, kolokacijskega in frazeološkega vidika. Leksikalna baza je na voljo v formatu XML s pripadajočo formalno shemo, ki jo je mogoče prilagoditi za različne tipe slovarskih priročnikov; [licenca CC BY-NC-SA]
- ▶ slovenski oblikoslovni leksikon **Sloleks**,⁸ izdelan v okviru projekta SSJ, ki vsebuje pregibne oblike za več kot 100.000 slovenskih besed; [licenca CC BY-NC-SA]
- ▶ **sloWNet**,⁹ slovenski semantični leksikon, zgrajen na podlagi angleške baze WordnNt, ki razvršča leksiko v sopomenske nize (sinsete) in ugotavlja medbesedna pomenska razmerja, kot so sopomenskost ter nad- in podpomenskost; [licenca CC BY-NC-SA]

II. SPLETNI PORTALI (IN PODATKOVNE BAZE, NA KATERIH TEMELJIJO)

- ▶ **Slogovni priročnik**,¹⁰ izdelan v okviru projekta SSJ, vsebuje na podlagi analiz z jezikom povezanih vprašanj, ki jih uporabniki zastavljajo na spletnih forumih in v jezikovnih svetovalnicah, oblikovan nabor odgovorov, ki vključujejo vpogled v realno jezikovno stanje v odnosu do obstoječe standardizacije; [licenca CC BY-NC-SA]
- ▶ **Pedagoški slovnici portal**,¹¹ izdelan v okviru projekta SSJ, prinaša na podlagi analize pogostih slovnicih problemov šolske populacije razlage in rešitve, izdelane na podlagi korpusnih analiz, in sicer na uporabniku prijazen in razumljiv način; [licenca CC BY-NC-SA]
- ▶ **Jezikovni viri starejše slovenščine IMP (1584–1919)**,¹² ki vsebujejo digitalno knjižnico, slovar starejšega slovenskega knjižnega jezika in referenčni korpus starejše slovenščine; [licenca CC BY]

⁶ Opis, dostop in prenos: <<http://www.slovenscina.eu/spletni-slovar/leksikalna-baza>>.

⁷ Projekt financira Evropska unija iz Evropskega socialnega sklada ter Ministrstvo za šolstvo in šport. Operacija se izvaja v okviru Operativnega programa razvoja človeških virov za obdobje 2007-2013, razvojne prioritete: razvoj človeških virov in vseživljenjskega učenja; prednostne usmeritve: izboljšanje kakovosti in učinkovitosti sistemov izobraževanja in usposabljanja 2007-2013. Spletna stran: <<http://projekt.slovenscina.eu/Vsebine/SI/Domov/Domov.aspx>>.

⁸ Opis, dostop in prenos: <<http://www.slovenscina.eu/sloleks/opis>>.

⁹ Možnost pregledovanja s pomočjo orodja SloWTool: <<http://nl.ijs.si/slowtool/>>. Dostopnost in prenos: <<http://loje.lugos.si/~darja/slownet.html>>.

¹⁰ Spletni servis: <<http://www.slovenscina.eu/portali/slogovni-prirocnik>>.

¹¹ Opis in demo: <<http://www.slovenscina.eu/portali/pedagoski-slovnici-portal>>.

- ▶ **Wikivir**, spletna digitalna knjižnica; [licenca CC BY-SA / GFDL]
- ▶ **Wikipedija**, prosta večjezična enciklopedija, ki jo soustvarjajo sodelavci z vsega sveta; [licenca CC BY-SA / GFDL]
- ▶ **Slovar slovenskega znakovnega jezika**,¹³ ki vsebuje slovenske besede, opis kretenj, s katerimi se izvaja in videoposnetek. [FRI, Uni-Lj in Zveza društev gluhih in naglušnih Slovenije]

III. ELEKTRONSKE ZBIRKE BESEDIL (KORPUSI)

- ▶ Korpus **Gigafida**, izdelan pri projektu SSJ,¹⁴ vsebuje 1,2 milijarde slovenskih besed, zbranih na podlagi slovenskih besedil (časopisi, revije, internet, stvarna besedila, leposlovje in drugo) med leti 1990 in 2011; [licenca CC BY-NC-SA za korpus ccGigafida, sicer pogodba o nekormercialni uporabi]
- ▶ Korpus **Kres**,¹⁵ 100-milijonski uravnoteženi podkorpus korpusa Gigafida, vsebuje po 20 % časopisnih, revijalnih in internetnih besedil, 18 % stvarnih besedil, 17 % leposlovja in 5 % drugih besedil; [licenca CC BY-NC-SA za korpus ccKres, sicer pogodba o nekormercialni uporabi]
- ▶ Korpus **Gos**,¹⁶ referenčni korpus slovenskega govornega jezika, izdelan v projektu SSJ, vsebuje milijon besed, transkribiranih na glasoslovni, leksemski, oblikoslovni in skladenjski ravni; [licenca CC BY-NC-SA za transkripcije]
- ▶ druge prosto dostopne **elektronske besedilne zbirke**, npr.
 - korpus Šolar, besedila, ki so jih učenci slovenskih osnovnih in srednjih šol samostojno tvorili pri pouku¹⁷
 - korpus slovenskih spletnih besedil slWaC
 - KoRP – korpus besedil odnosov z javnostmi
 - korpus slovenskih besedil s področja informatike in nekaterih drugih terminoloških področij
 - korpusi besedil iz družabnih omrežij (npr. korpus slovenskih tvitov) itd.¹⁸

IV. JEZIKOVNOTEHNOLOŠKA ORODJA

- ▶ Orodje **Sketch Engine**, prilagojeno za slovenščino, s pomočjo katerega je mogoče pridobivati relevantne leksikalne, slovnične in pemske podatke iz korpusa z uporabo aplikacij, kot so:
 - konkordančnik za analizo in urejanje konkordančnega niza za določeno besedo, besedno obliko, zvezo ipd.
 - Besedne skice – avtomatsko profiliranje leksikalnega obnašanja besed in besednih zvez na podlagi oblikoskladenjsko označenega korpusa
 - Primerjalne skice – avtomatska primerjava besedilnega okolja za različne besede, besedilne žanre ipd.

¹² Dostopno na: <<http://nl.ijs.si/imp/>>.

¹³ Dostopno na: <<http://sszj.fri.uni-lj.si/>> in <<http://www.zveza-gns.si/slovarji/>>.

¹⁴ Korpus Gigafida: <<http://www.gigafida.net/>>.

¹⁵ Korpus Kres: <<http://www.korpus-kres.net/>>.

¹⁶ Korpus Gos: <<http://www.korpus-gos.net/>>.

¹⁷ Korpus Šolar: <<http://www.korpus-solar.net/>>.

¹⁸ Zadnjih pet in drugi korpusi so dostopni na: <<http://nl.ijs.si:3003/cuwi/>> ali <<http://nl.ijs.si/bonito/>>.

- Tezaver – izdelava seznamov pomensko sorodnih besed idr.
- ▶ **Slovnica besednih skic za slovenščino:** izdelana za namene avtomatskega luščenja leksikalnih podatkov iz korpusa ter slovenščini prilagojena **aplikacija za avtomatski izbor dobrih zgledov** GDEX v orodju Sketch Engine;
- ▶ **Skladenjski razčlenjevalnik MSTParser:** za namen avtomatskega luščenja skladijsko-leksikalnih podatkov iz korpusa. Skladijski označevalnik se poleg orodja Sketch Engine uporablja predvsem za luščenje vezljivostnih podatkov ter stalnih besednih zvez (*multi-word units*);
- ▶ **Spletni pajek** (Web crawler): program, ki na sistematičen način avtomatsko »brska« po spletu, izlušči povezave (tj. vsebovane spletne naslove) in po čiščenju ohrani naslov, povzetek in glavno besedilo;
- ▶ **Posebno programsko orodje za luščenje normativnih podatkov iz korpusa:** v okviru projekta *Sporazumevanje v slovenskem jeziku* je bilo razvito orodje, s pomočjo katerega je iz korpusa mogoče izluščiti variantne oblike in pridobivati druge podatke, ki so relevantni za podajanje normativnih informacij v slovarju (Krek 2012b);
- ▶ **Prepoznavalnik imenskih entitet:** razpoznavanje in klasificiranje pojavnih oblik imen in številskih izrazov (*entity extraction, named entity recognition, entity identification*) je pomembna naloga pri izločanju informacij iz besedil in pripomore k izločanju leksikografsko relevantnih podatkov iz korpusa (dobri zgledi, semantični tipi itd.).

1.2 JEZIKOVNOTEHNOLOŠKA IN SLOVARSKA PODPORA V EVROPSKEM PROSTORU

Pojem **jezikovnotehnološka leksikalna podatkovna baza**, ki se osredotoča na leksikalne podatke, uporabne primarno za **računalniško obdelavo**, je v evropskem jezikoslovnem kontekstu znan že več kot 20 let, dejavnosti v povezavi s tem so se pri angleščini začele že v 60-ih letih prejšnjega stoletja, zelo pa so se intenzivirale z evropskimi jezikovnotehnološkimi projekti v 90-ih letih, ko so takrat še Evropska gospodarska skupnost in posamezne močnejše evropske države finančno in raziskovalno podprle jezikovne tehnologije in vire za svoje jezike (Krek 2009: 99). Namen gradnje tovrstnih baz je bil vzpostaviti mrežo jezikovnih virov, namenjenih za zagotavljanje jezikovnih potreb (večjih) jezikovnih skupnosti znotraj EU.

Na tej podlagi so se za številne evropske jezike, primarno pa za angleščino, nemščino, francoščino, španščino, italijanščino, nizozemščino ter večino skandinavskih jezikov oblikovale bodisi **jezikovnotehnološke leksikalne baze**, ki so namenjene zlasti računalniški obdelavi naravnega jezika in razvoju jezikovnih tehnologij¹⁹ in t. i. **slovarske leksikalne baze**, ki predstavljajo vir jezikovnih opisov, namenjenih človeškemu uporabniku. Kljub temeljnemu ločevanju na jezikovnotehnološke in jezikovnopriročniške leksikalne baze pa se v sodobni praksi meja med njimi zabrisuje.²⁰ Možnosti, ki jih z

¹⁹ npr. [STO](#) – leksikalna podatkovna baza za danščino, [CLIPS](#) – italijanska leksikalna podatkovna baza, [SenSem](#) in [ADESSE](#) – špansko-katalonski jezikovnotehnološki podatkovni bazi, [Vallex](#) – vezljivostni slovar čeških glagolov itd.

²⁰ Glej razdelek E-LEKSIKOGRAFIJA v poglavju Bibliografija.

novimi mediji ponuja splet, namreč omogočajo uporabo strukturiranih slovarskih podatkov tudi kot jezikovnotehnoloških virov in obratno.

1.3 PRIMERI DOBRE PRAKSE²¹

V evropskem prostoru smo v zadnjih letih priča nastanku vrste **hibridnih jezikovnotehnoloških in slovarskih podatkovnih baz**, ki so prosto dostopne na spletu. Tovrstni spletni slovarji oz. jezikovnoinformacijski portali zagotavljajo jezikovne informacije za različne profile končnih uporabnikov, v različnih stopnjah zahtevnosti in različnih načinih vizualizacije. Zagotavljajo pa tudi možnosti pridobivanja podatkov v surovi obliki za nadaljnje procesiranje, kar je zanimivo predvsem za računalniško skupnost.

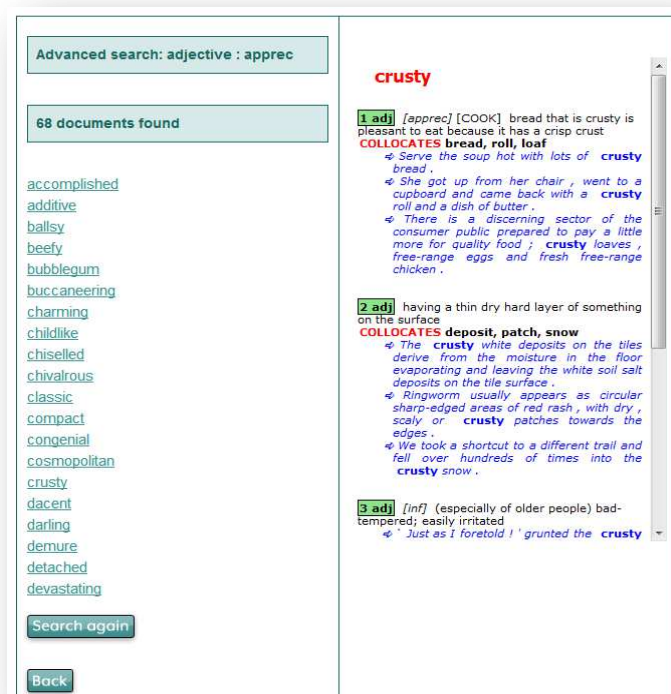
Med primeri dobre prakse, po katerih se je mogoče zgledovati, lahko omenimo: DANTE – *Database of Analysed Text of English*, Splošni spletni nizozemski slovar: *Algemeen Nederlands Woordenboek* (ANW), Leksikalno podatkovno bazo za francoščino: *Base lexicale du français* (BFL), in sodobni Poljski spletni slovar: *Wielki słownik języka polskiego*.

1.3.1 DANTE (DATABASE OF ANALYSED TEXT OF ENGLISH)

Podatkovna baza DANTE prinaša sistematični opis pomenov, slovničnega in kolokacijskega obnašanja ter besedilne značilnosti za približno 50.000 najpogostejših angleških besed. Osnovno leksikografsko orodje pri projektu je bil Sketch Engine, vsaka beseda je bila analizirana v 1,7 milijardnem korpusu v dveh fazah. Prva faza je zajemala identifikacijo posameznih rab ali pomenov, v drugi fazi pa so za vsak pomen ugotovljena relevantna jezikovna dejstva. Posamezni pomen predstavlja ključno enoto podatkovne baze in tako za razliko od številnih slovarjev postavlja v središče semantični in ne sintaktični (besednovrstni) princip organiziranja podatkov. Pri določanju besednih pomenov so se avtorji osredotočali na sobesedilo. Na podlagi analize sobesedila so registrirali tipične pomenske in skladijske lastnosti pomena, njegovo kolokacijsko obnašanje in besedilnotipske preference, kot so področje rabe, register, stil ipd. Drugi temeljni izziv, ki ga uresničuje DANTE, je beleženje vseh ugotovljenih dejstev v obliki formaliziranih kod, ki jih je mogoče računalniško procesirati. V bazi, ki je prosto dostopna na spletu,²² lahko iščemo po vseh jezikovnih podatkih, ki so v leksikalni bazi strukturirani v obliki 40 različnih podatkovnih tipov, kot npr. besedna vrsta, inherentne slovnične lastnosti, kot so npr. števnost in neštevnost pri samostalnikih, glagolske oblike pri glagolih, vezljivostne lastnosti pri glagolih, samostalnikih in pridevniki, skladijski kontekst pri posameznih besednih vrstah in številni tipi slovničnih, stilnih, področnih in drugih oznak. Posamezne podatkovne tipe je mogoče med seboj tudi kombinirati.

²¹ Več o opisanih projektih v razdelku PRIMERI DOBRE PRAKSE v poglavju Bibliografija.

²² DANTE: <<http://www.webdante.net/index.html>>.



SLIKA 1: PRIKAZ SLOVARSKIH PODATKOV V LEKSIKALNI PODATKOVNI BAZI ZA ANGLEŠČINO DANTE

Podatki, ki jih vsebuje DANTE, so namenjeni založnikom, ki izdajajo različne tipe, zlasti elektronskih slovarjev, jezikovnim tehnologom za avtomatično razdvajanje pomenov, izdelavo črkovalnikov, ekstrakcijo informacij ter za razvoj prevajalnih sistemov. Poleg tega pa tudi jezikoslovcem, raziskovalcem in učiteljem, ki se ukvarjajo z besediščem angleškega jezika.

1.3.2 SPLOŠNI NIZOZEMSKI SLOVAR (ANW: ALGEMEEN NEDERLANDS WOORDENBOEK)

Elektronski slovar sodobnega nizozemskega jezika, ki pokriva področje Nizozemske in Flandrije, je projekt Inštituta za nizozemsko leksikologijo in bo zajemal obdobje od leta 1970 do 2019. Gre za korpusni leksikografski projekt, katerega cilj je prikazati jezikovno-pomenski razvoj posamezne besede (z vključitvijo pomenskih in skladijskih opisov konkretne besede v posameznih obdobjih, vse od leta 500 do danes) ter opisati sodobno stanje nizozemskega, zlasti pisnega jezika, v najširšem pomenu leksikalne informacije. Prednost digitalnega slovarja je v tem, da se od leta 2009, ko je bila objavljena prva poskusna verzija, slovar na spletu posodablja vsaka dva meseca, pri čemer so upoštevani tako odzivi uporabnikov kot sodobne pridobitve leksikografske teorije in prakse.

Poleg pomenskega opisa besede ali zveze predvideva ANW tudi prikaz pomenskih razmerij (npr. sopomenke in nadpomenke), izgovornih posebnosti ter pregibnih oblik, semagramov,²³ primere rabe, tipične skladijske zveze, stalne besedne zveze, pregovore in reke in prikaz besedne družine (zlasti zloženke in izpeljanke). Celotna slovarska baza je interaktivna; podatki o pomenskih lastnostih besed in

²³ O tem glej: Moerdijk 2008.

medpomenskih razmerjih so dopolnjeni s fotografijami, videoposnetki, zvoki ter notranjimi in zunanjimi povezavami (npr. na Wikipedijo). Uporabniški vmesnik²⁴ omogoča iskanje od besede k pomenu, lahko pa tudi v obratni smeri, od posamezne pomenske lastnosti, ki jo uporabnik zapiše v obliki kratkega pomenskega opisa (npr. tipično rusko glasbilo; žival, ki daje mleko itd.), ali v obliki ontoloških semantičnih razredov (npr. glasbilo; žival itd.) h konkretni besedi ali pomenski skupini besed. Mogoče je določiti tudi zahtevnejše iskalne pogoje, ki vključujejo iskanje po kraticah, sinonimih, besednih vrstah, zgodovinskih oblikah besed, slogovnih posebnostih in glede na različne tipe prenosnikov.

The screenshot shows the ANW (Algemeen Nederlands Woordenboek) website. The main header features the ANW logo and the text 'Algemeen Nederlands Woordenboek'. Below the header, there are navigation tabs: 'Word — Meaning', 'Meaning — Word', 'Advanced search', 'Find examples', 'Neologisms', and 'Help · About the ANW'. The main content area displays the entry for 'bier'. On the left, there is a sidebar with a search bar and a list of results for 'bier': '1.0: alcoholische drank', '1.1: biersoort', and '1.2: glas bier'. The main entry for 'bier' includes the following information:

- Verkleiningsvorm:** biertje (biertje)
- Betekenisbetrekking**
 - Algemeen:** metonymie
 - Specifiek:** stof : soort stof
 - Betrokken betekenissen:** 1.0 : 1.1
- Combinations met adjectief ervoor**
 - een alcoholvrij bier; een Belgisch bier; een Vlaams bier; alcoholvrije bieren; Belgische bieren; Vlaamse bieren
- Example**
 - Alcoholvrije bieren mogen minder dan 0,5% alcohol bevatten.
 - <http://www.bierengezondheid.be/>, 2001
- Combinations met adjectivisch tegenwoordig deelwoord**
 - een bovengistend bier; een ondergistend bier; bovengistende bieren; ondergistende bieren
- Example**
 - Door dezelfde wort te fabriceren met gedemineraliseerd water en deze te vergisten met een pilsgist bij lage temperatuur hoop ik een beetje inzicht te verwerven in de verschillen tussen boven- en ondergistende bieren.
 - <http://home.wxs.nl/~triple-w/frames2.htm>
- Combinations met substantief ervoor**
 - een assortiment bieren
- Example**
 - Het predikaat 'Bij Koninklijke Beschikking Hofleverancier' vormt de kroon op het werk van alle ALFA-medewerkers, die er in geslaagd zijn een assortiment bieren te ontwikkelen, die als uniek worden ervaren op de Nederlandse markt.
 - <http://www.alfa-bier.nl/>

At the bottom left, there is a 'Show:' section with links to 'Whole article', 'Semagrams', 'Word family', 'Relations to other words', 'Proverbs', 'Combinations', 'Fixed expressions', and 'Examples'.

SLIKA 2: PRIKAZ DELA INTERAKTIVNEGA SLOVARSKEGA GESLA V SPLETNEM SPLOŠNEM NIZOZEMSKEM SLOVARJU

Spletni slovar sodobnega nizozemskega jezika je dober primer sproti aktualiziranega slovarskega priročnika, hkrati pa predstavlja tudi vir računalniško procesljivih podatkov, ki jih je mogoče uporabiti za pripravo ali izboljšavo različnih jezikovnotehnoloških aplikacij za nizozemščino.

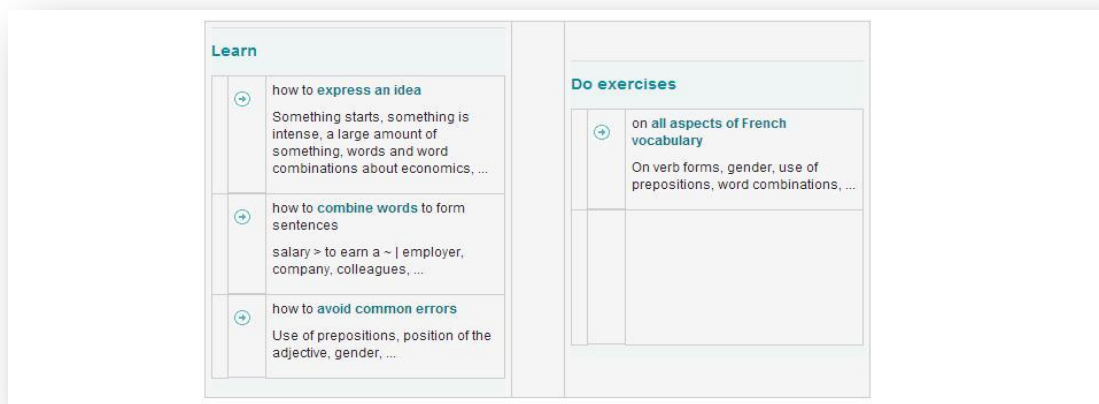
²⁴ ANW:< <http://anw.inl.nl/search>>. V beta verziji je mogoče uporabljati tudi angleško verzijo vmesnika.

1.3.3 LEKSIKALNA PODATKOVNA BAZA ZA FRANCOŠČINO (BFL: BASE LEXICALE DU FRANÇAIS)

Cilj oblikovalcev Leksikalne podatkovne zbirke za francoščino,²⁵ ki je nastala na Jezikovnem inštitutu Univerze v Leuvenu, je združiti informacije v obstoječih elektronskih francoskih slovarjih, zlasti enojezičnega slovarja za tujce, portala za učenje francoščine kot tujega jezika (ALFALEX: *Actieve leeromgeving Frans voor anderstaligen – lexicon*) in elektronskega slovarja poslovne francoščine (DAFA: *Dictionnaire d'apprentissage du français des affaires*), jih dopolniti s korpusnimi, leksikonskimi in pravopisnimi podatki ter uporabniku omogočiti različne možnosti za čim bolj učinkovit dostop do njih. Cilj projekta je torej ustvariti celostno, prosto dostopno spletno orodje za učenje in rabo sodobne francoščine, pri čemer je poudarek na interakciji slovarskih podatkov, učnega okolja ter računalniških orodij za analizo korpusov.

The screenshot shows the website for the Lexical Database for French (BFL). The header includes the KU Leuven logo and navigation links. The main title is 'Lexical Database for French (Base lexicale du français - BLF) - new site'. Below the title, there is a search bar and a description: '(Almost) everything you always wanted to know about... French words'. The main content area is divided into four sections: 'Find information on', 'Find the translation of', 'Check', and 'Help me'. Each section contains a list of interactive options with icons. The 'Find information on' section includes 'a word' and 'a word combination/phrase'. The 'Find the translation of' section includes 'a word to French' and 'a word combination/phrase to French'. The 'Check' section includes 'the use of a (sequence of) word(s)' and 'whether a translation is correct'. The 'Help me' section includes 'understand a text', 'translate a text', and 'write a text'. A sidebar on the right contains 'INTERFACE' options (in English, Coming soon, en français, in het Nederlands) and 'ACCENTED LETTERS AND SPECIAL CHARACTERS'.

²⁵ BFL: <http://ilt.kuleuven.be/blf/>. Baza je dostopna tudi prek angleške verzije vmesnika: Interactive language Toolbox: <https://ilt.kuleuven.be/inlato/>.



SLIKA 3: SPLETNI VMESNIK ZA DOSTOP DO RAZLIČNIH SKLOPOV PODATKOV V LEKSIKALNI BAZI ZA FRANCOŠČINO

Spletni vmesnik, ki je na voljo v francoščini, nizozemščini in angleščini, vključuje šest med seboj povezanih podatkovnih baz ali modulov, med katerimi lahko uporabnik izbira pri iskanju odgovora na svoj jezikovni problem. V slovenščini bi posamezne dele jezikovnega portala, ki so prikazani na Sliki 3, lahko poimenovali takole: (1) poišči podatke o besedi ali besedni zvezi, (2) poišči prevod besede ali besedne zveze, (3) preveri, npr. tipično rabo besede ali besedne zveze, pravilnost prevoda ipd., (4) pomagaj mi, npr. razumeti, prevesti ali napisati besedilo, (5) nauči se npr. izraziti določeno misel, idejo, oblikovati besede v stavke, odpraviti najpogostejše napake in (6) izboljšaj svojo francoščino z vajami za rabo različnih besednih oblik, spolov, predlogov besednih zvez itd.

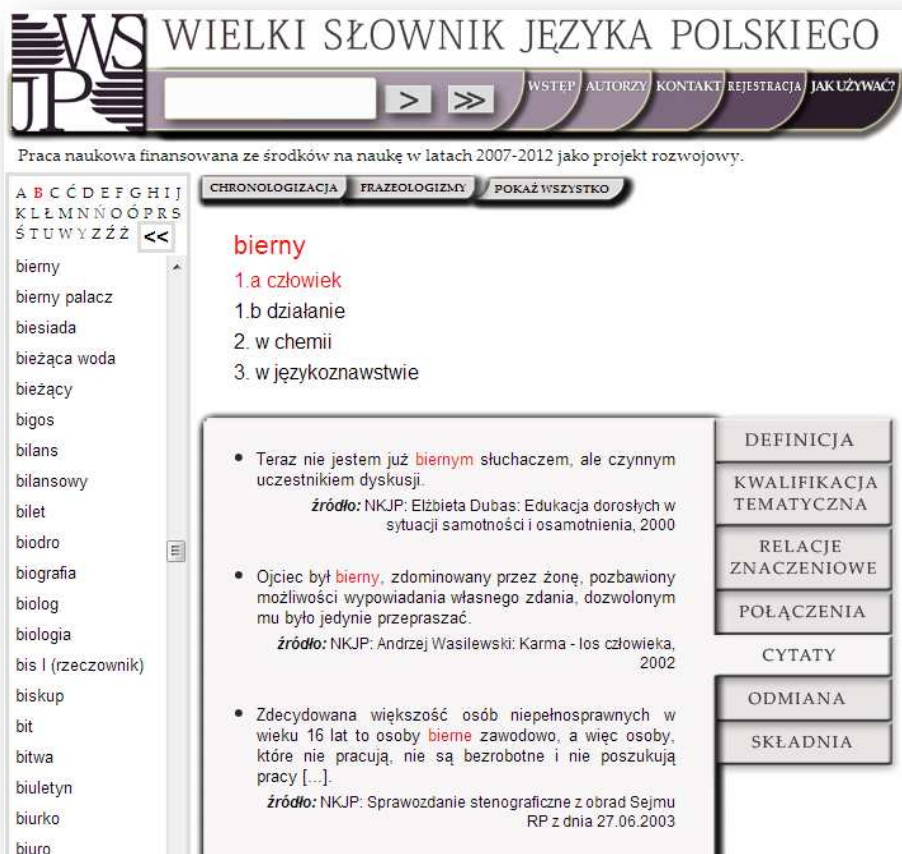
1.3.4 VELIKI SLOVAR POLJSKEGA JEZIKA (WIELKI SŁOWNIK JĘZYKA POLSKIEGO)²⁶

Ob koncu leta 2004 se je t. i. Odbor jezikovnih znanosti Poljske akademije znanosti obrnil na skupino jezikoslovcev s pozivom za oblikovanje predloga za izdelavo splošnega velikega slovarja poljskega jezika, ki bi kot glavni vir védenja o poljskem besedišču zamenjal že močno zastarel Slovar Poljske akademije znanosti in umetnosti.²⁷ Rezultat teh prizadevanj je Veliki slovar poljskega jezika, ki je financiran iz proračuna kot razvojni projekt v letih od 2007 do 2012. V projekt je trenutno vključenih več deset strokovnjakov iz različnih znanstvenih centrov na Poljskem, organizacijsko pa jih koordinira inštitut Poljske akademije znanosti in umetnosti. Temeljni namen slovarja je različnim jezikovnim uporabnikom zagotoviti izčrpen opis poljskega besedišča, in sicer v skladu s sodobnimi leksikografskimi standardi in znanjem poljskih jezikoslovcev. Načrtovani slovar bo uporabnikom na voljo le v elektronski obliki oz. bodo do njega lahko dostopali brezplačno na spletu. Trenutno je na spletu na voljo testna različica slovarja,²⁸ in sicer približno 15.000 gesel, iz katerih je razviden osnovni koncept slovarja, ki pa se skozi leksikografski opis še posodablja in dopolnjuje.

²⁶ O slovarju tudi: <<http://goo.gl/i3pgi>>.

²⁷ Słownik języka polskiego v 11 zvezkih je izhajal med leti 1958-1969 pod uredniškim vodstvom Witolda Doroszewskega. Po njem je mogoče iskati tudi na spletu: <<http://doroszewski.pwn.pl/>>.

²⁸ Veliki slovar poljskega jezika, testna verzija: <<http://www.wsjp.pl/>> .



SLIKA 4: SLOVARSKI VMESNIK ZA PREGLEDOVANJE VELIKEGA SLOVARJA POLJSKEGA JEZIKA

Preizkusna različica slovarja ponudi seznam trenutno razpoložljivih gesel. Ob kliku na izbrano geslo se odpre pomenska zgradba gesla. Posamezni deli celotnega gesla so za vsak posamezni geselski pomen na voljo z izbiro zavihkov: »definicija«, »tematska klasifikacija« (pomenske povezave v obliki ontologije), »medpomenska razmerja« (npr. so- in protipomenke), kolokacije oz. »tipično besedilno okolje«, »primeri rabe« ter »leksikon besednih oblik«. Nekatera gesla imajo poleg tega opisana tipična skladenjska razmerja v zavihku »skladnja« in etimologijo oz. izvor. V zgornjem delu zaslona so zavihki z distribucijsko oz. hierarhično neodvisnimi jezikovnimi podatki, kjer je mogoče izbrati pogled gesla v celoti z možnostjo tiskanja, njegovo predstavitev v posameznih slovarjih skozi celotno časovno obdobje pojavljanja (kronologizacija), frazeološke enote, ki vključujejo obravnavano besedo in podatke po posameznih zavihkih (definicija, primeri rabe itd.), ter različne besedilne variante, zlasti pri večbesednih leksikalnih enotah.

1.3.5 PRIMERI DOBRE PRAKSE - SKLEP

Pregled spletnih slovarskih portalov za evropske jezike pokaže, da se je predstava o klasičnem tipu slovarskega priročnika korenito spremenila. V ospredju je na eni strani **jezikovni uporabnik**, predvsem v procesu izobraževanja (učenec maternega jezika in učenec drugega/tujega jezika), profesionalni

uporabnik jezika (pisec, novinar, prevajalec, učitelj, lektor, oglaševalec, predstavnik stikov z javnostmi, raziskovalec itd.) ali splošni uporabnik z različnimi (pravopisnimi, slogovnimi, pomenskimi, slovničnimi) jezikovnimi problemi, ki ima na drugi strani na voljo **prosto dostopni jezikovno-informacijski spletni portal**, namenjen celotni jezikovni skupnosti določenega jezika ali specifičnim vrstam uporabnikov. Skupni imenovalac dobrih jezikovnih praks na tem področju je mogoče strniti v nekaj točk:

- ▶ Informacije so **hitro dostopne**: odgovor na jezikovno vprašanje je mogoče dobiti razmeroma hitro in brez vnaprejšnjega jezikoslovnega znanja ali poznavanja strukturiranosti podatkov;
- ▶ informacije so **učinkovite**: uporabnik dejansko najde rešitev za svoj jezikovni problem, saj so spletni slovarji izdelani na podlagi predhodnih analiz potreb jezikovne skupnosti oz. analiz potreb potencialnih skupin jezikovnih uporabnikov;
- ▶ jezikovni podatki so **redno posodabljeni**: zagotovljeno je posodabljanje in dopolnjevanje glede na dejanske spremembe v jeziku;
- ▶ informacije so **zanesljive**, tj. vzpostavljajo razmerje med veljavno normo in aktualnim stanjem v jeziku ter v skladu s tem ustrezno vrednotene;
- ▶ informacije so **brezplačne**: jezikovni spletni portali so navadno izdelani s pomočjo proračunskih in projektih sredstev in v skladu s potrjenim jezikovnopolitičnim načrtom za določeno obdobje ter tako javno in prosto dostopni celotni jezikovni skupnosti.

V nadaljevanju predstavljamo predlog slovarja, ki se zgleduje po naštetih praksah in upošteva vsa nova dejstva, ki izhajajo iz razvoja jezikoslovnih spoznanj, leksikografije, jezikovnih tehnologij, informacijsko-komunikacijskih tehnologij v obdobju od objave zadnje knjige SSKJ.

2 SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

Spletni *Slovar sodobnega slovenskega jezika*, kot ga predlagamo v pričujočem dokumentu, je spletni jezikovni priročnik, ki vsebuje celotno paleto slovarskih in drugih z leksiko povezanih jezikovnih informacij, ki so zanimive za najširši krog uporabnikov. Njegov namen je nadomestiti *Slovar slovenskega knjižnega jezika* v smislu temeljnega in vsestranskega slovarskega priročnika za slovenščino v digitalni dobi.

Novi slovar slovenskega jezika, če želi nadaljevati vlogo SSKJ kot osrednje slovarske institucije/avtoritete, se mora nujno odmakniti od njegove (knjižne) tradicije. Temeljiti mora na korpusni metodologiji pri pridobivanju jezikovnih podatkov ter uporabi orodij za njihovo analizo; leksikografske postopke je potrebno racionalizirati in pospešiti, kar omogoča avtomatizacija urejanja slovarskega gradiva v leksikalni/slovarski bazi in izkoriščanje moči množic (ang. *crowdsourcing*). Leksikalne podatke je potrebno predstaviti tako, da bodo uporabniku lahko dostopni in razumljivi ter posledično uporabni pri razreševanju konkretne jezikovne zadrege. Pri tem je mogoče s pridom izkoristiti spletni medij oz. digitalno/mobilno okolje, ki omogoča nove načine predstavljanja slovarskih informacij. Skladno z nastajanjem slovarja je treba zagotoviti dinamično rast referenčnega pisnega korpusa ter korpusa govorjenega jezika, kar omogoča sprotno posodabljanje na korpusnih podatkih temelječega slovarja. Podatke je v teku projekta in po njem treba pripravljati, hraniti in distribuirati v obliki, ki zadovoljuje tudi potrebe jezikovnotehnološke skupnosti.

2.1 SPLOŠNE LASTNOSTI

2.1.1 UPORABNIK

Določitev zahtev različnih vrst uporabnikov je za izdelavo slovarja ključnega pomena in mora biti izhodišče tako pri izbiri medija kot sestavnih delov slovarske vsebine. Kot je bilo omenjeno v razdelku 2.3, je *Slovar sodobnega slovenskega jezika* zasnovan za dva profila »človeških« uporabnikov: šolskega in splošnega tj. odraslega uporabnika. Ob dejstvu, da je v letu 2010 imelo srednješolsko izobrazbo skoraj 90 % mladih v Sloveniji in da je bila v študijskem letu 2009/10 v terciarno izobraževanje vključena polovica mladih, kar nas med državami EU-27 uvršča na prvo mesto,²⁹ gre pri obeh skupinah za uporabnike, ki se bodisi v izobraževalnem procesu ali pri svojem poklicnem ali vsakdanjem delu ukvarjajo s tvorjenjem in razumevanjem besedil in pri tem potrebujejo ustrezno lahko dostopno podporo glede vseh vrst jezikovnih informacij. Ob tem je treba upoštevati, da se je besedilna produkcija, dostopna širokemu krogu uporabnikov, v času širitve spleta in družabnih omrežij, in samo število piscev, ki pri tem sodelujejo, potisočeno, s čimer se je spremenil tudi temeljni družbeni model publiciranja, ki je zamenjal prejšnjega, temelječega na tisku in založniški industriji. Slovar torej mora svojim ciljnim uporabnikom uspešno služiti v novem digitalnem in spletnem modelu komuniciranja.

²⁹ Vir podatkov: <http://www.stat.si/doc/pub/izobrazevanje.pdf> (Do leta 2020 naj bi se delež delovnih mest v EU-27, za katera se bo zahtevala visoka izobrazba, zvišal s 25,1 % na 31,3 % vseh delovnih mest, število delovnih mest, za katera se zahteva srednješolska izobrazba, pa se bo zvišalo z 48,3 % na 50,1 %).

Za **splošnega uporabnika** Slovarja sodobnega slovenskega jezika se torej predvideva, da je materni govorec slovenščine, ki doživlja jezik in z njim povezane probleme skozi sporazumevalne situacije, kar pomeni, da je **njegov cilj rešitev določene sporazumevalne zagate, manj pa raziskovanje jezika oziroma njegovih prvin**. Informacija, ki jo pričakuje, je aktualna pomenska opredelitev besed in zvez, s poudarkom na tistih, ki so v slovenskem besedišču nove in še neuveljavljene, ter besed, ki so sicer specifične za določena specializirana področja, ki pa pod vplivom družbenopolitičnih, gospodarskih in kulturnih okoliščin zlasti prek informativnih medijev prehajajo v splošni jezik. Poleg pomenskega opisa splošnega uporabnika zanimajo tudi slovnično-pomenske lastnosti besed in besednih zvez, njihova izgovorjava, različne oblike – tudi glede na obstoječo normo – pa tudi umestitev besed in zvez v sobesedilo (kolokacije, področne in žanrske oznake), pomen in raba stalnih zvez in frazeoloških enot, sinonimija, izvor besed ipd. Uporabniki od sodobnega slovarja še pričakujejo, da je hitro in enostavno dostopen, npr. prek različnih digitalnih, tudi mobilnih naprav, ki so jih pri izobraževanju, v svojem poklicnem ali vsakdanjem delu vajeni uporabljati, in sicer na preprost in intuitiven način. To pomeni, da posebno izobraževanje ne sme biti potrebno oz. da uporabniki ne bodo potrebovali posebnih navodil za uporabo slovarja.

Šolski uporabnik in učenec slovenščine kot tujega jezika zahtevata specifično besedišče, zato je smiselno že pri snovanju slovarske baze prioriteto upoštevati besedišče aktualnih srednješolskih učbenikov, hkrati pa je tej populaciji smiselno prilagoditi tudi pomenske opise in zahtevnejše skladenjske podatke. Na podlagi ugotovitev, da eksplicitno navajanje slovničnih podatkov v slovarju pogosto zaradi prezahtevnega metajezika ne doseže uporabnikov oz. da jih uporabniki največkrat ignorirajo, je dobro slediti praksi, po kateri so skladenjsko-pomenski podatki vključeni v stavčne razlage, oz. jih je mogoče priklicati po potrebi, glede na namen in stopnjo zahtevnosti. V bazo vključene kolokacije (statistično izstopajoče in s tem tipične sopojavitve dveh ali več besed) in korpusni zgledi prikazujejo besede v njihovem realnem in tipičnem besedilnem okolju, zato predstavljajo neposredno informacijo o dejanskem življenju besed, kar je zlasti pomembna informacija za učence slovenščine kot tujega jezika, poleg šolskega in tudi splošnega uporabnika slovarja.

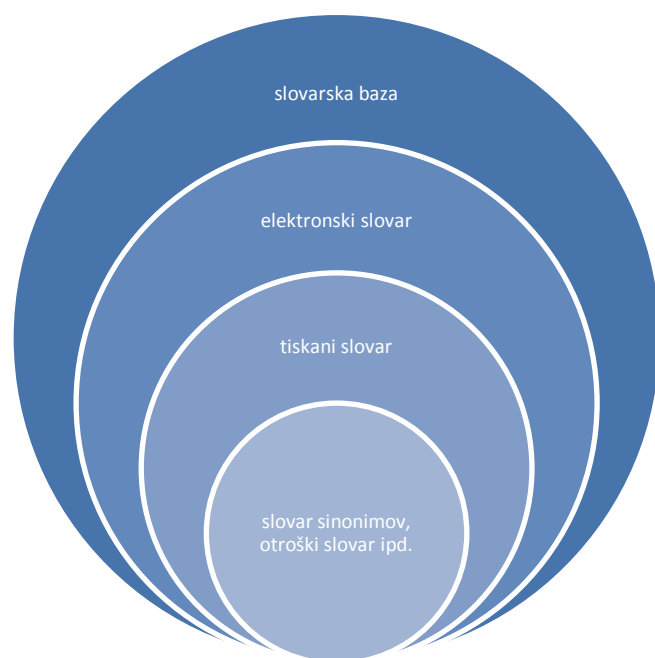
Organizacija pomenskih, skladenjskih, slovničnih in kolokacijskih podatkov v različnih nivojih omogoča tudi specializiranemu uporabniku, kot je npr. **jezikoslovec**, učitelj, lektor, prevajalec itd., pridobivanje in kombiniranje različnih tipov podatkov, ki so nujno potrebni za podrobne semantične in skladenjske analize. Ker je slovarska baza zgrajena na podlagi korpusnih podatkov, odražajo pridobljeni podatki realno stanje sodobne pisne (in tudi govorjene) slovenščine, in so osnova za pripravo učnih gradiv kot tudi za aktualiziranje in prevrednotenje obstoječih učnih pripomočkov.

Za jezikoslovce bo zanimiva tudi slovarska baza, ki bo dostopna v obliki podatkovne zbirke v formatu XML (pod podobno licenco kot Wikipedija). Na podlagi teh podatkov bo mogoče izvajati različne raziskave o pomenskih, skladenjskih, vezljivostnih, kolokacijskih, normativnih in drugih vidikih slovenskega jezika.

2.1.2 MEDIJ

Formati, v katerih uporabnikom lahko prikažemo slovarsko vsebino, so različni, vendar dejansko govorimo o dveh med seboj zelo različnih medijih: tiskanem in digitalnem. Slovarji v digitalnem oz. elektronskem mediju imajo več oblik (npr. spletni slovar, slovar na mobilnih napravah, slovar na CD-ju), vendar v zadnjih letih skoraj v celoti prevladuje spletna oblika (ter s tem povezane aplikacije v mobilnih napravah) in tak razvoj lahko pričakujemo tudi v prihodnosti.³⁰

Dejstvo je, da mora biti slovar na voljo v čim več različnih formatih, če hoče doseči čim večje število uporabnikov, k čemur v digitalnem okolju lahko najbolj pripomore prosta dostopnost slovarske baze in fleksibilna organiziranost podatkov. Konceptualizacija slovarja zato ne sme biti omejena zgolj na en format oz. medij, temveč mora vključevati izdelavo digitalne **slovarske baze**, ki vključuje informacije, na podlagi katerih je mogoče izdelati slovarje z različnimi kombinacijami informacij v različnih formatih (gl. Sliko 5), s čimer lahko posledično zadovoljimo potrebe različnih tipov uporabnikov.



SLIKA 5: RAZMERJE MED SLOVARSKO BAZO IN RAZLIČNIMI OBLIKAMI SLOVARJEV

Problematiko napačnega konceptualnega pristopa lahko ponazorimo s še vedno dokaj razširjeno prakso izhodiščne izdelave tiskanega slovarja in kasnejšo izdajo njegove elektronske oz. spletne verzije, kar je bilo morda smiselno v obdobju prehajanja s tiskanega medija v digitalno okolje, ne pa ob načrtovanju izdelave slovarja za obdobje, ki prihaja. Pri tiskanih slovarjih je omejitev predstavljal prostor oz. število znakov, zato je šlo veliko leksikografskega razmisleka v rešitve, kako te omejitve omiliti, npr. izbor končnega števila gesel, okrajšave, skrajšani zgledi, gnezdenje podatkov, izogibanje krožnosti definicij ipd. Če takšen »tiskani« slovar neposredno prenesemo v elektronsko obliko, nam umanjajo številne informacije, ki so jih uporabniki novih elektronskih referenčnih virov že vajeni (npr.

pokritost skoraj celotnega besedišča, celostavčni zgledi, različne iskalne možnosti, hiperpovezave itd.). Naknadno dopolnjevanje izvirno tiskanega slovarja z novimi informacijami je zamuden, drag in

³⁰ SURS – Uporaba informacijsko-komunikacijske tehnologije v gospodinjstvih in pri posameznikih, podrobni podatki, Slovenija, 2012 (http://www.stat.si/novica_prikazi.aspx?id=5179): "Rednih uporabnikov informacijsko-komunikacijskih tehnologij (IKT) je bilo več med mlajšimi kot med starejšimi osebami. Med osebami, starimi 16–24 let, je bilo rednih uporabnikov računalnika kar 98 %, rednih uporabnikov mobilnih telefonov 99 %, rednih uporabnikov interneta 97 % in rednih uporabnikov mobilnih naprav za dostop do interneta 74 %; med osebami, starimi 65–74 let, pa je bilo rednih uporabnikov računalnika 20 %, rednih uporabnikov mobilnih telefonov 72 %, rednih uporabnikov interneta 17 % in rednih uporabnikov mobilnih naprav za dostop do interneta le 2 %."

pravzaprav nesmislen postopek, saj se je potrebno vrniti k ponovnem luščenju podatkov, njihovi analizi in vnašanju v slovarsko bazo.

V nadaljevanju ne opisujemo slovarske baze v formatu XML (razen deloma v povezavi z Leksikalno bazo v poglavju 3), poudarek je na opisu spletne oblike slovarja, predvsem zaradi lažje predstave, kako bo večina končnih uporabnikov dostopala do podatkov. V spletni obliki bodo omogočene številne funkcije, ki jih potrebujejo uporabniki v digitalnem okolju, npr. takojšnja in sprotna dostopnost, redno in sprotno posodabljanje, pa tudi različne oblike kompleksnejših iskalnih mehanizmov ter jasna razdelitev in stopenjskost prikaza informacij. S tem se uporabniška izkušnja spremeni iz »klasične«, kjer je v ploski in zgoščeni strukturi (tj. na slovarski strani v knjigi) treba najti relevantno informacijo, v »globinsko«, kjer uporabnik strukturirano bazo raziskuje v globino, pri čemer prehaja z informacije na informacijo tako horizontalno kot vertikalno, vedno pa ima na ekranu na voljo le obvladljivo količino informacij. O tem podrobneje v poglavju 2.2.1 o elementih slovarja, predvsem v poglavju Pomeni in podpomeni.

Pomembne prednosti spletne oblike predlaganega slovarja so še:

- a) brezplačen dostop
- b) sprotni odziv – predvideno je dopolnjevanje/posodabljanje slovarskih informacij na podlagi spremljanja (uspešnih in neuspešnih) poizvedb uporabnikov³¹
- c) preprosta raba – ker bo slovar uporabljal rešitve, kot jih najdemo v ostalih spletnih virih, s katerimi se uporabniki redno srečujejo, ne bo potrebe po obsežnem učenju uporabe slovarja
- d) vključitev multimedijskih informacij (npr. avdio posnetkov, izgovorjave, slik, videoposnetkov ipd.)
- e) enostavna prilagodljivost za različne skupine uporabnikov (npr. slabovidni)
- f) prilagodljivost za različne platforme, npr. mobilnike in tablice

2.1.3 OBSEG

Cilj projekta je v petih letih izdelati približno **100.000 gesel**, kar izhodiščni, v okviru projekta izdelani Slovar sodobnega slovenskega jezika uvršča med slovarje večjega srednjega ali »namiznega« obsega po klasičnem pojmovanju³² (za primerjavo, SSKJ vsebuje nekaj manj kot 107.000 gesel in podgesel). Po tem obdobju je predvideno sprotno spremljanje sprememb v besedišču in hitra reakcija nanje z dodajanjem novih besed oz. pomenov. Izhodiščni geslovnik bo zajemal občnoimensko besedišče polnopomenskih besed (samostalniki, glagoli, pridevniki in prislovi), in sicer tako enobesedne kot večbesedne iztočnice, poleg tega lastnoimenske iztočnice, kolikor so zanimive bodisi zato, ker izkazujejo besedilno okolje, ki je zanimivo s stališča leksikalnega opisa (npr. imena prebivalcev, nekatera zemljepisna imena), ali s stališča norme (npr. variantne oblike Shakespearjev-Shakespearov). Stalne besedne zveze in frazeološke enote bodo v slovarski strukturi ohlapno povezane s slovarskimi sestavki iztočnic oz. besed, ki jih sestavljajo, in jih po tradicionalnem pojmovanju lahko obravnavamo kot samostojna gesla. Njihovo gnezdenje pod

³¹ Spletna analitika (ang. *web analytics*) je postalo pomembno področje, ki je postalo zelo relevantno tudi v okviru e-leksikografije. Glej npr. Lorentzen in Theilgaard 2010.

³² Tiskani *college dictionaries* v angloameriškem prostoru imajo tipično 180.000-200.000 gesel, *desk* (ali *desktop*) *dictionaries* med 60.000-80.000 gesel (Landau 2001: 36-37).

relevantnimi gesli (npr. *osnovna šola* pod *šola*) je namreč smiselno le v slovarjih, kjer so gesla razvrščena po abecednem vrstnem redu in ni drugih iskalnih mehanizmov. Funkcijske in slovnične besede, kot so vezniki, predlogi itd., bodo obravnavane drugače od ostalih gesel, poudarek bo na njihovi vlogi v besedilu oz. slovničnem opisu.

Po obsegu bo torej predlagani slovar primerljiv z obstoječim slovarjem slovenskega jezika, v katerem pa je že zdaj precejšen del besedišča zanimiv zgolj s historičnega vidika.³³ V primeru, da bi bilo mogoče povezati bazo SSKJ in novi slovar, bi bilo ob uspešni integraciji (ki sicer ni del predlaganega projekta) število gesel lahko še bistveno večje. Na ravni obsega – in v povezavi z načrtom dela – so bolj ključne soodvisne novosti, ki odražajo sodobne leksikografske trende, in sicer:

- | |
|--|
| <ul style="list-style-type: none">a) določitev prioritetenega zaporedja priprave geselb) detektiranje pomenskih in leksikalnih sprememb besedišča inc) sprotno ažuriranje slovarskih gesel v posameznih fazah. |
|--|

Izhodišče za prioritetno obravnavo besedišča predstavlja frekvenčna lista korpusa Gigafida, v kombinaciji z natančno in razmeroma kompleksno statistično obravnavo podatkov iz korpusa Kres, Gos in drugih baz, ki so na voljo (Pravopis, Sloleks, Slogovni priročnik itd.). Kompleksna obravnava obsega primerjalne statistike glede na pojavljanje iztočnic po žanrih, obdobjih, prisotnost/odsotnost v različnih bazah, pojavljanje v prioritetnih podkorpusih, kot je denimo učbeniški podkorpus, in podobno. Izhodiščni geslovnik bo obsegal približno četrtno več gesel od predvidenega končnega števila iztočnic. Prvi spisek, izdelan v šestih mesecih po začetku projekta, bo obsegal okrog 100.000 korpusnih lem, kar po izračunih generira približno 130.000 slovarskih iztočnic.

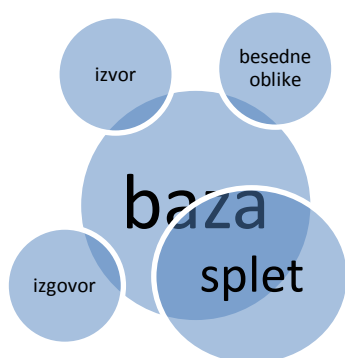
Kot kaže leksikografska praksa, je neizogibno, in tudi potrebno, da se geslovnik v teku izvajanja projekta prilagaja, ker je treba sproti spremljati nastajanje novih leksemov oz. dodajati pomene obstoječim. Zato je treba v okviru projekta predvideti tako redno vzdrževanje korpusa kot tudi vzpostaviti mehanizem za avtomatično opozarjanje na potencialne nove iztočnice ali pomene (s pomočjo kontekstualnih podatkov).

2.2 VSEBINA

Pri slovarju je treba ločiti med več vrstami vsebin: (1) vsebino slovarskih gesel v leksikalni oz. slovarski bazi kot podatkovni zbirki, ki je obsežnejša in vsebuje tudi podatke, ki niso namenjeni človeškemu uporabniku, zlasti npr. podatke za jezikovnotehnološke namene,³⁴ (2) vsebino slovarskih gesel, do katere dostopa uporabnik slovarja prek spleta in so torej ustrezno vizualizirana, in (3) vsebine, ki so vizualizirane, vendar niso del slovarske baze, temveč so organizirane kot samostojne baze, s slovarsko bazo pa so povezane preko identifikatorjev na ustreznih mestih. Vsebine so med seboj povezane, saj so

³³ Približno desetina gesel/pomenov je v SSKJ označenih s kvalifikatorji, ki signalizirajo odsotnost realne rabe že v času izhajanja slovarja (1970-1991), npr. 1819 označenih s "starinsko" ali "zastarelo" (docirati, dojmiti, fronki, golomraz, kamižola, pojata, pozoj itd.), 1596 z "redko" (bandera, barusa, čestitkar, hučati, krcelj, polšina, prižmuriti itd.), 5008 s "knjižno neobčevalno" (agnoscirati, akribija, bedeker, bleskot, cisterca, defraudant, donebnik, gladež, lornjon, pasjica, prednočje itd.), če omenimo samo nekatere oznake, ki opozarjajo na zastarelost ali redko rabo besed.

³⁴ O tem glej Poglavlje 3.



SLIKA 6: VRSTE VSEBIN V SLOVARSKI BAZI

neobdelani avtomatsko izluščeni korpusni podatki – prek izbire dobrih možnosti s pomočjo izkoriščanja moči množic in v leksikografskem prepoznavanju leksikalnih enot in njihovih pomenskih tendenc – oblikujejo v celovito jezikovno informacijo, vključno z enciklopedičnim znanjem in izrabo multimedijskih možnosti. Nekoliko bolj tehnični opis nekaterih delov baze, ki ne bodo neposredno vidni na portalu, vsebuje poglavje 3 o Leksikalni bazi.

2.2.1 VRSTA SLOVARSKIH INFORMACIJ IN ELEMENTI SLOVARJA

Na uvodni strani načrtujemo dokaj standardne dele slovarskega vmesnika, kot so: geslovnik, pomoč pri iskanju informacij v slovarju, opis organizacije in vrste podatkov, predstavitev koncepta in metodologije, sodelujoči partnerji ipd.

Poleg omenjenega na prvi strani predvidevamo še prikaz drugih v zvezi s slovenskim besediščem zanimivih podatkov o aktualni jezikovni rabi, kot je denimo beseda dneva, najbolj iskana ali pogosta beseda, najnovejša beseda ipd. Zasnova spletnega portala omogoča tudi sodelovanje uporabnikov prek spletnega foruma, socialnih omrežij, implementiranih besednih iger ipd., sprejemanje predlogov uporabnikov za nove besede v slovarju itd.

Dostop do podatkov v slovarju je torej v osnovi dveh vrst: preko iskalnika, tudi z naprednimi možnostmi, kot je iskanje po posameznih delih slovarske baze, ki so dostopni preko filtrov, ter preko geslovnika, ki je poljubno vizualiziran, bodisi kot tabela, 2D ali 3D mreža.

OSNOVNI VMESNIK

Podatki, vezani na iskano besedo ali zvezo, se v spletnem slovarju prikazujejo v naslednjih segmentih:

- naslovna vrstica
- osrednje prikazno okno
- levi stolpec s slovarskim menijem
- desni stolpec z okvirčki

Okvirčki v desnem stolpcu vsebujejo predvsem možnost filtriranja in dostopanja do podrobnejših informacij znotraj posameznega segmenta geselskega članka. Podatki, ki jih izberemo levo v slovarskem meniju, se prikazujejo v osrednjem delu zaslona in so v statičnem smislu odvisni od tega, v kateri redakcijski fazi se geslo nahaja, kar označuje semafor barvnih oznak in datum zadnje posodobitve gesla.³⁵ V dinamičnem smislu se podatki spreminjajo glede na to, kateri del gesla (pomen ali podpomen) oz. tip leksikalne enote (stalne zveze, frazeološke enote) nas natančneje zanima. V desnem stolpcu so podatki organizirani v več okvirčkih, njihova vsebina pa se prilagaja vsebini »odprtega« osrednjega dela gesla oz. posameznega (pod)pomena ali stalne zveze. Prvi okvirj je namenjen seznamu stalnih zvez, ki deluje kot klikljiv meni, drugi pa »uredniškemu« izboru kolokacij za dano besedo v obliki besednega oblaka.

SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

globalen pridevnik /globálen/ P 3000

naslovna vrstica

1. svetovni; mednarodni

1.1 splošno veljaven; razširjen

2. zemeljski; planetarni

3. ki zadeva celoto; celostni

1. svetovni; mednarodni

globalni procesi, zlasti gospodarski in politični, zajemajo ves svet

»V New Yorku naj bi državniki in podjetniki razpravljali o **globalni** varnosti.

Tudi največje svetovne firme, ki danes obvladujejo **globalni** trg, so se razvile iz malih podjetij in obratov.

Motorola je zaradi **globalne** recesije v visokotehnoloških gospodarskih panogah lani odpustila 48.400 zaposlenih.

1.1 splošno veljaven; razširjen

če postanejo neke dejavnosti ali lastnosti globalne, jih upošteva vedno več ljudi ali držav po svetu

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

2. zemeljski; planetarni ekologija

globalne spremembe v okolju vplivajo na celoten zemeljski planet

Eden najbolj preprostih in praktično izvedljivih načinov za zmanjšanje **globalnega** segrevanja je sajenje novih dreves.

Krčenje ledenikov in spremembe v pokrajini sta zelo očitni in lahko razumljivi posledici **globalnega** ogrevanja.

Vodik bo tudi občutno skrčil emisije ogljikovega dioksida in učinke **globalne** otoplitve.

3. ki zadeva celoto; celostni

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

Ta študija zlasti opozarja na vrsto vprašanj, ki se zastavljajo glede oblikovanja **globalne** nacionalne kulturne politike.

Gre za **globalni** pristop do človeka, ki mu pomaga, da vzame življenje v svoje roke in ga znova progamira, kot si ga je sam zamislil.

stalne zveze

- globalno segrevanje/ogrevanje
- globalna vas
- globalni nomad
- globalna spremenljivka
- globalno potrdilo o lastništvu

ekonomija gospodarstvo izziv

kapitalizem komunikacija

konkurenčnost korporacija kriza

lokalen mreža

ogrevanje otoplitev

politika problem raven

razsežnost recesija

regionalen segrevanje

sklad sprememba

temperatura trend trg

vojna

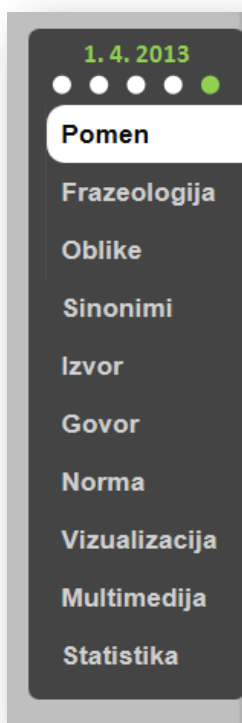
osrednje okno

© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 7: OSNOVNI VMESNIK

³⁵ Izdelava gesel je predvidena v petih stopnjah: rdeči – avtomatski, oranžni – prečiščeni, rumeni – pomensko razdelani, modri – zunajbazni oz. enciklopedijski in zeleni – zaključni fazi. Glej poglavje 2.3.

SLOVARSKI MENI



SLIKA 8: SLOVARSKI MENI

Prek slovarskega menija, ki vsebuje zavihke: *Pomen*, *Frazeologija*, *Oblike*, *Sinonimi*, *Izvor*, *Govor*, *Norma*, *Vizualizacija*, *Multimedija* in *Statistika*, uporabnik izbira podatke, ki ga v zvezi z iztočnico podrobneje zanimajo. V zavihku *Pomen* so mu na voljo podatki o posameznem pomenu ali podpomenu iztočnice in stalnih zvez, podatki o besednozveznih in stavčnih vzorcih, v katerih se iztočnica pojavlja, podatki o kolokacijah ter zgledi, ki vse omenjeno ponazarjajo v realni rabi. V zavihku *Frazeologija* je naveden seznam frazeoloških enot, vezanih na iztočnico, njihov pomen, oblike rabe in zgledi. Do različnih oblik, v katerih se pojavlja iztočnica glede na svojo sklonsko ali spregatveno paradigmo, dostopa uporabnik v zavihku *Oblike*, medtem ko je zavihek *Sinonimi* namenjen prikazovanju podatkov o pomenskih razmerjih med besedami. Spletni slovar tu črpa podatke iz obstoječega semantičnega leksikona, podatke iz korpusa, generirane s pomočjo orodja Sketch Engine (funkcija Primerjalne skice), in podatke iz Leksikalne baze. Zavihek *Izvor* prinaša dva tipa etimoloških podatkov. Prvi je vezan na izvor besede, drugi pa na starinske oblike, kot se pojavljajo v posameznih časovnih obdobjih. Podatki o etimologiji so besedi dodani v predzadnji, modri fazi, medtem ko je podatke o starinskih oblikah mogoče pridobiti avtomatsko že na ravni iztočnice, kjer črpamo iz korpusa in Slovarja starejšega slovenskega besedja. Zavihek *Govor* je namenjen

prikazovanju podatkov, ki jih vsebuje korpus govornenega jezika Gos. Uporabniku je na jasen in razumljiv način, v obliki slik in grafov, predstavljena porazdelitev govornih variant v slovenskem prostoru. Do podatkov, ki opisujejo obstoječo jezikovno normo in opozarjajo na standardizacijske specifike iztočnice, uporabnik dostopa prek zavihka *Norma*. Zavihek *Vizualizacija* je namenjen predstavitvi vseh podatkov v zvezi z iztočnico, ki so strukturirani v slovarski bazi, v obliki hiperboličnega grafa. V posameznih fazah se vizualizacija spreminja glede na pomensko razčlenjenost besede in druge v leksikografskem procesu pripisane podatke. Spletni slovar po vzoru dobrih evropskih praks predvideva tudi vključitev multimedijskih podatkov, kot je dodajanje slikovnega materiala in videoposnetkov – temu je namenjen zavihek *Multimedija*. V zavihku *Statistika* pa so vsebovani podatki o statističnem obnašanju besede v različnih obdobjih, besedilnih tipih in žanrih.

NASLOVNA VRSTICA

Naslovna vrstica v spletnem slovarju vsebuje podatke, vezane neposredno na iztočnico, in sicer besedno vrsto, izgovor (studijski posnetek in zapis), ikono za znakovni jezik in podatek o pogostosti besede v korpusu Gigafida. Vsi podatki, vezani na izrazno podobo iztočnice, vključno s sklonskimi in spregatvenimi oblikami v zavihku *Oblike* (z izjemo zapisa in posnetka izgovora, ki sta iztočnici dodana v naslednjih fazah geselske izdelave) so uporabniku na voljo že ob avtomatski izdelavi gesla, tj. v rdeči fazi (glej poglavje 2.3).



SLIKA 9: NASLOVNA VRSTICA

IZTOČNICA

Iztočnica je navedena v osnovni obliki, ki je skladna z osnovno obliko (lema), kakršna je besedi pripisana skladno z oblikoskladenjskimi pravili JOS³⁶ in uporabljena tudi v korpusih Gigafida, GOS in drugih. Kot naslovna enota geselskega članka je zasnovana v skladu z možnostmi optimalne izrabe elektronskega medija ter z upoštevanjem uporabniških potreb pri iskanju informacij slovarskega tipa. V elektronski verziji ali spletni postavitvi dostop do slovarskih podatkov in njihova vsebina namreč nista povezana z uporabnikovim poznavanjem makro- in mikrostrukturnih razmerij, ki jih uporablja tiskani slovar (npr. sistem abecedne razporeditve iztočnic, kazalčni sistem, sistem podgesel in gnezdenja ter drugih rešitev, povezanih s konkretno organizacijo slovarske strukture), pač pa je pomembna predvsem opredelitev leksikalne enote oz. segmenta besed, ki predstavlja za uporabnika potencialno iskalno možnost oz. slovarski problem, ali z drugimi besedami: za uspešno iskanje je pomembno, kateri črkovni niz se pojavlja kot del specifičnega strukturnega elementa v slovarski bazi. Zato ima iztočnica v celotni geselski zgradbi spletnega slovarja v prvi vrsti namen povezati posamezne leksikalne enote, tj. pomene, podpomene, stalne zveze in frazeologijo, pod isto streho, ob tem, da je do podatkov v zvezi z njimi mogoče priti tudi neodvisno od iztočnice.

V zvezi z iztočnico, ki predstavlja naslovno enoto gesla,³⁷ je skladno s specifikami spletnega medija obravnavana tudi homonimija (tj. enakopisne in enakoizrazne besede znotraj iste besedne vrste, npr. *bor* – samostalnik: 1. drevo, 2. kemijski element) in besednovrstno povezane iztočnice, kot so npr. posamostaljeni pridevniki, samostalniki v pridevniški rabi, deležniki ipd. V tiskanih slovarjih je obravnavala te problematike zaradi omejenih možnosti iskanja po slovarju zahtevala vrsto odločitev in posledično navodil, ki jih je bil uporabnik prisiljen obvladati, medtem ko je v spletnem slovarju pomembna predvsem konsistentnost ločevanja med tehničnim ali formalnim izrazom, ki je v našem primeru iztočnica, in njeno vsebino, ki jo obravnavamo kot leksikalno enoto. V spletnem slovarju je torej problem homonimije in besednovrstno povezanih besed rešen z upoštevanjem tehničnih možnosti, kamor sodi formalno prepoznavanje oblike – tj. iztočnice kot povezovalne enote za potrebe iskanja – od vsebine, ki je vezana na različne leksikalne nize, ki so z iztočnico povezani in vsebujejo različne leksikalno-gramatične podatke. Konkretno to pomeni, da pomenskih vsebin ne obravnavamo na ravni iztočnice (npr. ne predvidevamo dveh različnih iztočnic za samostalnik *bor*), pač pa na ravni njenih pomenskih lastnosti: samostalnik *bor* ima torej dva pomena in eno samo formalno podobo.

³⁶ Glej spletno stran: <<http://nl.ijs.si/jos/>>.

³⁷ Iztočnica je lahko beseda ali besedna zveza, kakor je opredeljeno v nadaljevanju. Kjer je v besedilu uporabljena »beseda« v iztočnici, je vedno mišljena tudi »besedna zveza«.

BESEDNA VRSTA

Glede na izpostavljeno formalno vlogo iztočnice kot povezovalca pripadajočih leksikalnih enot, tudi besednovrstna opredelitev v spletnem slovarju sledi njeni formalni in ne vsebinski podobi. Besedno vrsto torej napoveduje osnovna oblika besede, kot je določena skladno s tabelo oznak JOS. Na primer, beseda *odvisen* je na ravni iztočnice opredeljena kot pridevnik, znotraj gesla pa združuje tudi druge skladijske rabe, npr. samostalniško: *‘odvisni od drog’ – ‘odvisniki’*, in druge leksikalne enote, npr. stalne zveze in frazeološke enote, ki so ugotovljene v postopku pomenskega členjenja: *odvisni stavek*, *odvisna spremenljivka*. Ti podatki, podobno kot pri obravnavi homonimije, niso pripeti na dve različni iztočnici, pridevniško in samostalniško, kot je praksa tiskanih slovarjev, pač pa je iztočnica ena sama, v izhodišču opredeljena kot pridevnik, znotraj nje, torej na ravni (pod)pomenov, pa je mogoče dobiti podatke o njenih različnih skladijskih vlogah in s tem povezanih različnih pomenskih vsebinah. Posledično tudi uporabnikom za dostop do zelenega podatka ni potrebno obvladati teorije besedne konverznosti, ker je to naloga jezikoslovnega strokovnjaka in leksikografa. Ta mora namreč svoje znanje uporabiti tako, da podatke organizira na način, da bo za uporabnika dostop do njih enostaven, intuitiven in učinkovit. Poudariti je treba, da je prednost besednovrstne opredelitve osnovnih kategorij, kot so samostalnik, glagol, pridevnik in prislov, in podkategorij, kot so deležniki, deležja, glagolniki, ki je usklajena z oblikoskladijskimi oznakami JOS, tudi v tem, da omogoča optimalno iskanje in procesiranje v slovarski bazi strukturiranih podatkov.

IZGOVOR



Podatek o tem, kako se beseda izgovori, je za uporabnika nadvse koristen. Upoštevati je namreč treba, da je večina slovarskih poizvedb na receptivni in produktivni ravni opravljena pri branju/tvorjenju pisnih besedil. Če pa želi uporabnik kasneje uporabiti besedo tudi v govoru, potrebuje podatek o njeni normativni izgovorjavi. Elektronski medij nam omogoča, da se uporabniku predvaja posnetek izgovora besede, hkrati pa tudi zapis z naglasnimi znamenji. Podatki o izgovoru, če v zvezi z obravnavano besedo v iztočnici že obstajajo, so na iztočnico pripeti že v avtomatski fazi, sicer pa so izdelani v predzadnji, modri fazi.

ZNAKOVNI JEZIK



V spletnem slovarju je na ravni iztočnice predvidena tudi povezava na Slovar slovenskega znakovnega jezika, ki je namenjen njegovemu razumevanju in učenju. V njem je za posamezne besede in besedne zveze prikazan način kretanja in opisane kretnje, ki so uporabljene pri posamezni besedi.

ZAVIHEK: POMEN

Zavihek Pomen je del spletnega slovarja, ki vsebuje večino informacij pomenske, skladijske in kolokacijske narave, poleg tega še informacije o stalnih besednih zvezah ter skladijsko in pomensko fiksni večbesedni enotah, ki jih imenujemo »besedni nizi«. Ker vsebuje številne slovarske novosti, ki so vezane na spletni/digitalni medij, mu pri opisu namenimo nekoliko več prostora.

POMENSKI MENI

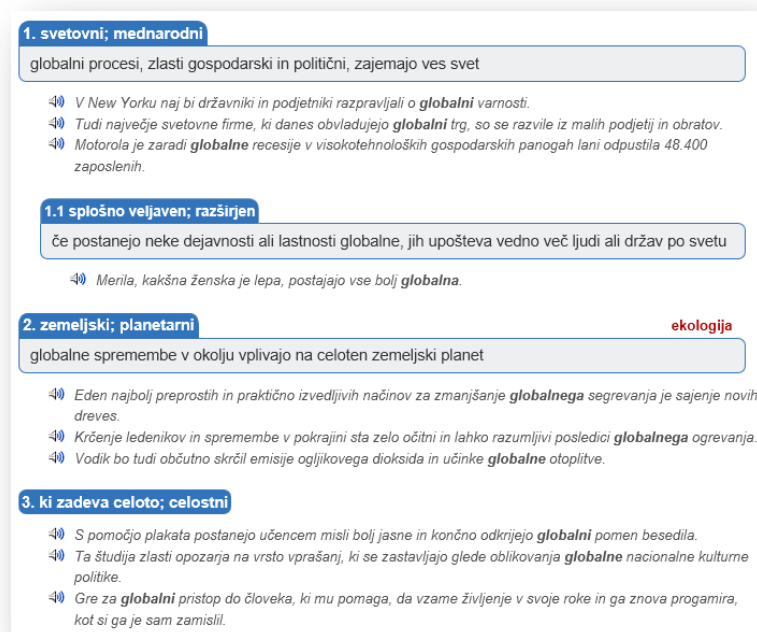
Pomenski meni je izpostavljen del gesla, ki s kratkimi opisi pomenov in podpomenov omogoča lažjo in hitrejšo navigacijo po njem. Gre za element, ki so ga uvedli slovarji za tuje govorce angleščine, in sicer že v tiskanih verzijah. V elektronski obliki je vloga pomenskega menija postala še pomembnejša, saj je s pomočjo hiperpovezav premikanje po geslu precej enostavnejše. Ključnega pomena za učinkovit meni je način ubeseditve pomenskih opisov. Ti morajo biti kratki in jasni, pa tudi dovolj razločevalni, da lahko uporabnik že na ravni menija jasno loči med posameznimi (pod)pomeni. Pomenska zgradba gesla, kot je prikazana v pomenskem meniju, je prepletena z drugimi podatki znotraj geselske strukture in neposredno povezana z možnostmi njihovega prikazovanja na spletu. Zato je oblikovanje pomenskega menija predvideno že na ravni slovarske baze, saj je koristen tako za leksikografsko urejanje pomenskih podatkov kot za slovarske uporabnike.



SLIKA 10: POMENSKI MENI

POMENI IN PODPOMENI

Slovar sodobnega slovenskega jezika je zasnovan tako, da v svojem osrednjem delu predvideva prikaz podatkov o pomenskih lastnostih besede. Sem sodijo stopnja pomenske členitve (pomeni in podpomeni), opis pomenov in podpomenov v obliki stavčne definicije ter skladenjski (vzorci, besedni nizi) in pomenski (kolokacije) prikaz sobesedila, v katerem se beseda ali besedna zveza tipično uporablja.



SLIKA 11: PRIKAZ POMENOV IN PODPOMENOV

Na podlagi pomenske členitve, ki jo leksikograf zabeleži v slovarski bazi, je vsak pomen in podpomen v spletnem slovarju opredeljen s kratkim pomenskim indikatorjem, lahko pa ima še pomensko razlago oz. definicijo. V zavihku pomen so na prvi ravni prikazane kratke verzije (pod)pomenov, kjer z indikatorji, razlagami in zgledi skušamo uporabniku čim hitreje pomagati, da najde (pod)pomen, ki ga zanima. Kot namreč kažejo raziskave, je največ neuspešnih iskanj povezanih z izbiro napačnega pomena oz. nezmožnostjo odkritja pravega pomena med veliko količino podatkov. Definicije so v

spletnem slovarju oblikovane na podlagi podatkov, ki so v slovarski bazi zabeleženi v obliki t. i. pomenskih shem (po vzoru baze FrameNet). Pomenske sheme pojasnjujejo pomen besede tako, da ga vključijo v stavčni vzorec, v katerem se tipično pojavlja. Posamezna mesta, ki so v stavčnem vzorcu za konkretni pomen predvidljiva, so v slovarski bazi posebej označena s t. i. semantični tipi (z velikimi črkami), ki zastopajo pomensko skupino besed na teh mestih, npr. za enega od podpomenov glagola *grmeti*: če neka **NAPRAVA** ali **VOZILO** *grmi*, pri svojem delovanju oddaja *glasen in bobneč zvok* ali če **KJE** *grmi*, se *sliši glasen in bobneč zvok*, ker se *približuje nevihta*. Na tej podlagi oblikovane definicije prinašajo v slovenski leksikografski prostor novost v pomenskem opisu besed. Temeljijo namreč na prepričanju, da se različne pomenske tendence besed uveljavljajo v konkretnih bolj ali manj predvidljivih jezikovnih situacijah oz. rabah, ki jih je mogoče rekonstruirati v obliki stavčne razlage, ki te situacije najbolje povzema. Registracijo prototipičnega in alternativnih stavčnih vzorcev, v katerih se beseda pojavlja, opravi leksikograf v procesu pomenske členitve, saj so prav spremembe v številu, obliki in pomenskih skupinah udeležencev eden od ključnih pokazateljev spremembe pomenskih tendenc obravnavane besede. Ker pa je podatek o semantičnih tipih in alternacijah osnovnega vzorca zanimiv predvsem za računalniško procesiranje, ostaja na ravni slovarske baze, v spletnem slovarju pa je prikazana na tej podlagi oblikovana stavčna razlaga.

KOLOKACIJE, RAZŠIRJENE KOLOKACIJE, VZORCI IN BESEDNI NIZI

Kolokacije so obvezen element slovarjev, ki prikazujejo rabo besede v njenem statistično in skladenjsko prepoznavnem okolju. Kolokacijski podatki so zato nepogrešljivi pri prepoznavanju pomenov besed, določanju večbesednih leksikalnih enot in pri prepoznavanju njihovega skladenjskega okolja (vzorcev). Besednozvezni in stavčni vzorci, registrirani v slovarski bazi, obenem služijo tudi za pridobivanje podatkov iz korpusa, nadaljnje procesiranje ter za jezikoslovne analize.



SLIKA 12: IZBOR KOLOKACIJ

V spletnem slovarju so kolokacije prikazane na dveh nivojih oz. v dveh stopnjah podrobnosti. Na ravni iztočnice (oz. vseh pomenov) je v okvirčku v desnem stolpcu v obliki besednega oblaka prikazan »uredniški« izbor kolokacij, vezanih na besedo v iztočnici, in sicer ne glede na njeno pomensko razčlenjenost. Gre za izbor kolokacij, ki so za prikazano besedo najpogostejše tako v smislu besednozveznih vzorcev, ki jim pripadajo (npr. **PRIDEVNIK + SAMOSTALNIK**, kjer večja pisava nakazuje, da gre za iztočnico), kot v smislu pogostosti kolokatorja znotraj takega vzorca. Na ta način dobi uporabnik osnovno sliko o besedilnem okolju besede ne glede na njeno pomensko razčlenitev.

Podrobnejšo sliko o tem, kako se razvrščajo kolokacije za posamezni pomen ali podpomen besede, dobi uporabnik na drugem nivoju, tj. s klikom na posamezni (pod)pomen.

1. 4. 2013

Pomen

Oblike

Sinonimi

Izvor

Govor

Vizualizacija

Multimedija

Statistika

SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

globalen pridevnik
 /globálen/
P 3000

1. svetovni; mednarodni

1.1 splošno veljaven; razširjen

2. zemeljski; planetarni

3. ki zadeva celoto; celostni

KLIK NA 1. POMEN

VZORCI IN BESEDNI NIZI PRI 1. POMENU

1. svetovni; mednarodni

globalni procesi, zlasti gospodarski in politični, zajemajo ves svet

« V New Yorku naj bi državniki in podjetniki razpravljali o **globalni** varnosti.

« Tudi največje svetovne firme, ki danes obvladujejo **globalni** trg, so se razvile iz malih podjetij in obratov.

« Motorola je zaradi **globalne** recesije v visokotehnoloških gospodarskih panogah lani odpustila 48.400 zaposlenih.

stalne zveze

- globalna vas
- globalni nomad

vzorci

PRIDEVNIK + SAMOSTALNIK

VEZNI GLAGOL + PRIDEVNIK₁

besedni nizi

globalen in lokalni

globalen in regionalni

na globalni ravni

KLIK NA "RECESIJA" (GLEJ NASLEDNJO SLIKO)

globalna [ekonomija, kriza, konkurenčnost, **recesija**, korporacija]

globalna [razsežnost, komunikacija, pravičnost]

globalni [kapitalizem, trend, trg, problem, fenomen, izziv, terorizem]

[postajati, postati] globalen

KOLOKACIJE PO VZORCIH

© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 13: PRIKAZ PODATKOV PRI POSAMEZNEM POMENU

1. 4. 2013

Pomen

Oblike

Sinonimi

Izvor

Govor

Vizualizacija

Multimedija

Statistika

SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

globalen pridevnik
 /globálen/
P 3000

1. svetovni; mednarodni

1.1 splošno veljaven; razširjen

2. zemeljski; planetarni

3. ki zadeva celoto; celostni

SKLADENJSKI VZOREC IZBRANE KOLOKACIJE

1. svetovni; mednarodni

globalni procesi, zlasti gospodarski in politični, zajemajo ves svet

« V New Yorku naj bi državniki in podjetniki razpravljali o **globalni** varnosti.

« Tudi največje svetovne firme, ki danes obvladujejo **globalni** trg, so se razvile iz malih podjetij in obratov.

« Motorola je zaradi **globalne** recesije v visokotehnoloških gospodarskih panogah lani odpustila 48.400 zaposlenih.

stalne zveze

- globalna vas
- globalni nomad

vzorci

PRIDEVNIK + SAMOSTALNIK

Vsi vzorci

besedni nizi

globalen in lokalni

globalen in regionalni

na globalni ravni

IZBRANA KOLOKACIJA (GLEJ PREJŠNJO SLIKO)

globalna recesija

nazaj...

- [grozeča, prihajajoča, zdajšnja, **globoka**, nova, velika] globalna recesija
- globalna [gospodarska] recesija
- [grožnja, posledica, vpliv, razmera] globalne recesije
- [čas, obdobje, začetek, konec] globalne recesije
- globalna recesija [prizadene, povzroči]

"RAZŠIRJENE" KOLOKACIJE

« Kljub grožnji **globalne** gospodarske **recesije** McDonalds po svetu odpira nove in nove izpostave.

« Po številnih ocenah aktualna oblast ni storila dovolj, bolje povedano ni storila vsega, kar bi lahko, da bi omilila posledice **globalne** recesije.

ZGLEDI PRI IZBRANI KOLOKACIJI

© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 14: PRIKAZ PODATKOV PRI IZBIRI POSAMEZNE KOLOKACIJE

Na tem nivoju uporabnik izbor kolokacij lahko filtrira glede na vzorec, v katerem se beseda pojavlja, in širi sliko glede na stopnjo trdnosti obsežnejših leksikalnih nizov, kot so t. i. razširjene kolokacije, npr. gospodarska recesija – [svetovna, huda, globalna] gospodarska recesija, in besedni nizi. Besedni nizi prikazujejo obsežnejše prepoznavne dele jezika, ki imajo celostno skladenjsko vlogo, predvidljive sestavne elemente in so pomensko omejene, npr. v *letih od – do*, iz *generacije v generacijo*, *upati si povedati naglas*, *angažirati se na področju česa* ipd. Kot prikazujeta Sliki 13 in 14, je podatke mogoče filtrirati prek seznama vzorcev in besednih nizov, vezanih na (pod)pomen v okvirčkih v desnem stolpcu ekrana.

STALNE ZVEZE

Stalne zveze so večbesedne leksikalne enote, ki imajo lasten pomen, pogosto specifično besedilno okolje in se lahko uporabljajo kot del splošnega ali specializiranega besedišča, npr. *globalna vas*; *globalna spremenljivka*. Podatki o stalnih zvezah so strojno izluščeni iz korpusa Gigafida, vendar v rdeči fazi prikazani na ravni kolokacij, saj se strukturno ne razlikujejo od prostih besednih zvez in kolokacij. Identificiramo in pomensko opišemo jih v procesu pomenske členitve, zato so informacije o stalnih zvezah dodane v rumeni fazi ter terminološko pregledane in dopolnjenje v izpopolnjeni – modri fazi gesla.

The screenshot displays the web interface of the 'SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA'. On the left is a dark sidebar with navigation links: '1. 4. 2013', 'Pomen', 'Oblike', 'Sinonimi', 'Govor', 'Vizualizacija', 'Multimedija', and 'Statistika'. The main content area is titled 'državni zbor' with a small icon and the phonetic transcription '/državni zbor/'. Below the title, a blue box labeled '1. parlament' contains the definition: 'državni zbor je osrednje zakonodajno telo v Sloveniji, ki ga sestavlja 90 poslank in poslancev'. Two example sentences follow, each preceded by a speaker icon. Below these are several phrases in brackets, such as '[slovenski] državni zbor', '[ratificira, uzakoni, sprejme, potrdi, zavrne]', '[razpustiti, obveščati, obvestiti, seznaniti] državni zbor', '[predlagati, predložiti, posredovati, poročati, predsedovati] državnemu zboru', '[poslati, vložiti, posredovati, priti] v državni zbor', '[zasedanje, seja, delo, poslanec, poslanka] državnega zbora', '[volitve, izvolitev, vstop] v državni zbor', and '[poslanec, poslanka, obravnava, glasovanje, odločanje, branje] v državnem zboru'. On the right side, there are two yellow boxes: the top one lists 'državen (3)' and 'zbor (1.1)', and the bottom one is titled 'vzorci' and lists various grammatical combinations like 'PRIDEVNIK+ ZVEZA', 'ZVEZA₁ + GLAGOL', etc. At the bottom right, a section titled 'besedni nizi' shows 'državni zbor in vlada'. The footer of the interface reads '© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA'.

SLIKA 15: PRIKAZ PODATKOV PRI STALNIH ZVEZAH

Pri pomenskem členjenju leksikograf v slovarski bazi identificira tri kategorije stalnih zvez: (1) stalne zveze, ki podobno kot enobesedne iztočnice vsebujejo celotno paleto na posamezni (pod)pomen vezanih podatkov; (2) stalne zveze, ki ne predstavljajo očitne in nedvoumne pomenske povezave s katerim od (pod)pomenov besede v iztočnici, in (3) stalne zveze, ki izkazujejo pomensko povezavo s katerim od registriranih pomenov. V spletnem slovarju je, za razliko od slovarske baze, vsaka stalna

zveza, ne glede na pomensko povezanost s katerim od registriranih pomenov besede, ki je njen sestavni del, obravnavana kot iztočnica. Na ta način poleg pomenskih vsebujejo stalne zveze vedno tudi podatke, ki so neposredno vezani na besednozvezno iztočnico, in sicer: posnetek in zapis izgovora, znakovni jezik ter pripadnost konkretni frekvenčni skupini v korpusu.

Uporabnik do vseh oz. podrobnejših podatkov v zvezi s posamezno stalno zvezo dostopa bodisi prek seznama stalnih zvez v okvirčku v desnem stolpcu ekrana pri konkretni iztočnici, ki se pojavlja kot njen sestavni del, ali prek iskalnega okenca, v katerem išče stalno zvezo kot celoto. Klik na posamezno stalno zvezo v okvirčku znotraj konkretnega (pod)pomena besede v iztočnici vodi uporabnika do podatkov o tem, kaj stalna zveza pomeni (indikator in/ali definicija), v katero strokovno področje sodi, kakšno je njeno tipično besedilno okolje, če obstaja, in kako se rabi v realni situaciji (zgledi in oznake). Poleg tega lahko uporabnik poizvedovanje o stalni zvezi osredotoči na posamezni vzorec ali besedni niz, ki ga izbere v seznamu vzorcev v okvirčku v desnem stolpcu ekrana.

OZNAKE

Pri beleženju oznak za namene slovarske baze in za namene spletnega slovarja obstajajo določene razlike. V podatkovni bazi je zabeleženo predvsem obnašanje besed glede na tipično področje rabe, diskurza, stila oz. konotacijo, hkrati pa tudi glede na slovnične lastnosti, omejitve in druge relevantne informacije, ki so povezane z besedo ali njenim pomenom. Pri izdelavi slovarske baze nam taki opisi v elementih geselske strukture pomagajo pri sistematizaciji in pri avtomatski ekstrakciji besedišča, ki je tako ali drugače označeno, na primer besedišča, ki se rabi pretežno v neformalnih situacijah ali z določenimi konotacijskimi preferencami, besedišča, ki se tipično rabi na določenem specializiranem področju, kontekstu ipd.

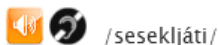
SLIKA 16: PRIKAZ SLOVNIČNIH IN PODROČNIH OZNAK

V spletnem slovarju so ti opisi glede na slovarsko bazo selekcionirani in v obliki kratkih pojasnil pripeti bodisi na definicije kot njihovo dopolnilo bodisi se nanašajo neposredno na iztočnico ali njen posamezni (pod)pomen. Kot je razvidno iz prikaza gesla v predvidenem

v spletnem slovarju, dobi uporabnik prek oznak osnovno informacijo o tem, v kakšnem kontekstu se beseda običajno rabi in kako učinkuje v realni komunikaciji.

S pomočjo oznak uporabnik lahko tudi hitreje išče informacije, ki ga zanimajo. Denimo če pri enem od pomenov pri večpomenski besedi vidi oznako računalništvo, ki ga ne zanima, lahko ta pomen avtomatično preskoči.

sesekljati glagol



1. narezati na koščke
2. onemogočiti; uničiti
3. zlahka premagati

1. narezati na koščke

z nožem ali napravo narezati na majhne koščke

- ☞ Za nadev **sesekljamo** čebulo, 2 paradižnika, 100 g šampinjonov, 1 žličko kaper in 4 rezine šunke.
- ☞ Na kuhinjski deski z nožem **sesekljamo** čebulo, zeleno in peteršilj.

2. onemogočiti; uničiti

neformalno

SLOGOVNA OZNAKA

zlasti v politiki onemogočiti, navadno z ostro kritiko

- ☞ Gre za zakonski predlog finančnega ministra Mitja Gasparija, ki ga je sekretar za zakonodajo državnega zbora Milan Baškovič z desetimi stranmi pripomb dobesedno **sesekljal**.
- ☞ Francoski premier je po vrsti **sesekljal** evropske vizije, kot so jih v minulem letu predstavili nemški zunanji minister, predsednik države in na koncu še kancelar.

KONTEKSTUALNA OZNAKA

3. zlahka premagati

neformalno

v športu premagati z veliko prednostjo ali razliko

- ☞ Našemu hokejskemu strategu Matjažu Seklju od srca želimo, da bi s svojimi fanti na vročem ledu **sesekljal** naše severne sosesde.

SLIKA 17: PRIKAZ KONTEKSTUALNIH IN SLOGOVNIH OZNAK

ZGLEDI RABE

Zgledi rabe so pomemben element slovarskega gesla, saj je njihova vloga večplastna: prikažejo iztočnico, kot se v jeziku uporablja, tj. v njenem naravnem okolju, služijo kot opora razlagi, do neke mere pa tudi uporabniku pomagajo pri navigaciji po geslu, kajti uporabnik lahko lažje odkrije zanj relevanten pomen oz. podpomen, če v slovarju najde podoben primer, kot ga je srečal oz. ga tvori.

Najpogosteje omenjeni kriteriji za dober slovarski zgled so:

- Avtentičnost. Zgledi rabe se črpajo iz korpusov. Uporaba izmišljenih zgledov, kot je bila to pogosta praksa v obdobju pred korpusi, je v sodobni leksikografski praksi nesprejemljiva.
- Tipičnost in naravnost. Zgled mora ponazoriti tipično rabo besede v določenem pomenu, podpomenu ipd. z vidika konteksta, skladnje, kolokacij ipd.

- Informativnost. Zgled mora dati geslu neko dodano vrednost. Ta kriterij se nanaša predvsem na izogibanje kopičenju zgledov, predvsem takšnih, ki ponazarjajo enako ali zelo podobno rabo besede.
- Razumljivost. Zgled ne sme vsebovati preveč zahtevnega ali redkega besedišča, da se lahko uporabnik osredotoči na rabo iztočnice in mu ni treba iskati razlag za ostale besede v zgledu. Seveda je to zlasti pri terminih in področno omejeni rabi zaradi specifičnosti konteksta težko vedno doseči, je pa pomembno, da se omenjenega kriterija skušamo čim dosledneje držati.
- Celostavčna oblika. Zgledi morajo biti podani kot celi stavki, saj je podajanje zgolj kratkih fraz, ki vsebujejo iztočnico, kot je to še vedno praksa v nekaterih slovarjih (tudi v SSKJ) abstraktno, nenaravno in tudi zavajajoče.
- Ustrezna dolžina. Zgledi morajo biti ravno prave dolžine, ne smejo biti ne predolgi, kar vpliva na berljivost in preglednost gesla, in ne prekratki, ker so tako lahko kontekstno podhranjeni in dejansko ne dosežejo svojega namena.

Leksikografska praksa kaže, da je zelo težko v korpusu za posamezno iztočnico najti več zgledov, ki ustrezajo vsem naštetim kriterijem. Težava se pojavlja zlasti pri zadnjih dveh kriterijih: celostavčni zgledi so pogosto predolgi, imajo preveč kompleksno strukturo in/ali besedišče, kar vpliva na njihovo razumljivost. Posledično se pri izbiri zgledov za slovar v primeru pomanjkanja ustreznih korpusnih zgledov dovoli tudi uporaba daljših in mogoče manj primernih zgledov iz slovarske baze, ki se jih za namene slovarja lahko skrajša, pri čemer morajo obdržati celostavčno obliko in ne smejo izgubiti sporočilnega bistva. Najpogostejši primer omenjenega skrajševanja je odstranitev »pogrešljivih« odvisnikov in lastnih imen.

Zaradi navidezne neomejenosti prostora, s katerim razpolaga elektronski medij, številni leksikografi že dolgo pozivajo k vključitvi velikega števila zgledov v slovarska gesla. To se zdi sicer do neke mere smiselno, vendar pa je treba opozoriti, da je tudi prostor na računalniškem ekranu omejen in da uporabniki niso preveč navdušeni nad drsenjem po strani v neskončnost. Rešitev je prikaz zgledov v dveh stopnjah – najprej se predstavi privzeto število zgledov, ki se zdi še obvladljivo za večino uporabnikov, uporabnikom pa se ponudi dostop do dodatnih zgledov.

V spletnem slovarju smo predvideli tudi možnost poslušanja izbranih zgledov s pomočjo sintetizatorja govora, na kar opozarja ikona z zvočnikom na začetku vsakega zgleda.

ZAVIHEK: FRAZELOGIJA

V fazi pomenskega členjenja leksikograf identificira tudi frazeološke enote. Za razliko od stalnih zvez frazeološke enote v podatkovni bazi niso registrirane v povezavi s katerim od pomenov besede v iztočnici, pač pa so pomensko neodvisne, do podatkov o njih pa uporabnik spletnega slovarja dostopa prek samostojnega zavihka *Frazeologija* v slovarskem meniju. V zvezi z vsako frazeološko enoto prikazuje spletni slovar podatke o njenih tipičnih variantah in oblikah rabe, pomenih in tipičnem besedilnem okolju. Frazeološke enote so posebej opredeljene tudi glede na svoje konotacijske in pragmatične lastnosti. Kot pri vseh leksikalnih enotah je tudi v primeru frazeoloških njihova dejanska raba ponazorjena s korpusnimi zgledi.

1. 4. 2013

Pomen

Frazeologija

Oblike

Sinonimi

Izvor

Govor

Vizualizacija

Multimedija

Statistika

SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

dihati

glagol

 /dihati/

P 5000

1	dihati komu za ovrtnik/vrat	dihati komu za ovrtnik / dihati komu za vrat
2	ne pustiti (niti) dihati komu ne dati dihati komu	1. v športu ali politiki če pri tekmovanju rečemo, da kdo diha komu za ovrtnik, želimo povedati, da se močno približuje vodilnemu
3	nehati dihati	40 Kolesar iz Zimbabveja je v gorskih etapah vseskozi dihal za ovrtnik Perezu Cuapiu.
4	(skoraj) pozabiti dihati	40 Spet so se izkazali Ribničani, ki v Velenju sicer niso zmagali, so pa večji del tekme vodili in vse do zadnjega dihali za ovrtnik favoriziranemu Gorenju.
5	dihati na škrge	40 Tony Blair vodi, Michael Howard mu dih a za ovrtnik, oba pa računata na podporo v parlamentu še vedno dokaj močnih Kennedyjevih liberalnih demokratov.
6	dihati s polnimi pljuči	2. izvajati pritisk na koga; nadzorovati
7	komaj/težko dihati	40 Evropska inšpekcija nam dih a za ovrtnik.
8	komaj (si) upati dihati skoraj (si) ne upati dihati	40 Busheva administracija je našla novo malopridno državo, ki ji dih a za ovrtnik.
9	lažje dihati	40 Finančna kriza obalnega prvoligaša, ki mu za vrat dih a tudi davkarja, vnaša med igralce precej nezadovoljstva.
10	dihati isti zrak	3. pri kaznivih dejanjih če policisti rečejo, da dihajo storilcu za ovrtnik, menijo, da ga bodo kmalu izsledili ali aretirali
11	dihati s kom isti zrak	40 Kdo je umor naročil, je še danes glavno in nepojasnjeno vprašanje, čeprav odgovorni v policiji že ves čas zatrjujejo, da storilec dih ajo za ovrtnik.
12	(za vedno) nehati dihati	

© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 18: PRIKAZ PODATKOV O FRAZELOGIJI

ZAVIHEK: OBLIKE

Neposredno z iztočnico so povezani podatki o njenih različnih oblikah v skloni in spregatveni paradigmi in njihovih oblikoslovnih lastnostih. Gre za podatke o občno- oz. lastnoimenskosti ter o oblikovni povezanosti in variantnosti besed, pa tudi za podatke o pogostosti oblike ter o njeni opredeljenosti v normativnih virih. Te podatke za spletni slovar pridobivamo iz Leksikona besednih oblik Sloleks, uporabnik pa do njih dostopa prek zavihka *Oblike*. Obstoječim podatkom v Sloleksu bodo v procesu izdelave slovarja dodani tudi podatki o naglasu posameznih pregibnih oblik.

SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

globalen pridevnik /globálen/ P 3000

moški ženski srednji

EDNINA			DVOJINA			MNOŽINA		
OSNOVNIK			OSNOVNIK			OSNOVNIK		
oblika	sklon	pog.	oblika	sklon	pog.	oblika	sklon	pog.
globalen	imenovalnik	513	globalna	imenovalnik	13	globalni	imenovalnik	1.185
nedoločno			globalnih	rodilnik	2	globalnih	rodilnik	1.452
globalni	imenovalnik	4.308	globalnima	dajalnik	1	globalnim	dajalnik	114
določno			globalna	tožilnik	7	globalne	tožilnik	935
globalnega	rodilnik	2.772	globalnih	mestnik	0	globalnih	mestnik	632
globalnemu	dajalnik	363	globalnima	orodnik	4	globalnimi	orodnik	385
globalen	tožilnik	207						
nedoločno								
globalni	tožilnik	2.529						
določno								
globalnega	tožilnik	120						
globalnem	mestnik	2.618						
globalnim	orodnik	598						

POVEZANE OBLIKE:
globalno prislav
globalnost samostalnik

© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 19: PRIKAZ PODATKOV O BESEDNIH OBLIKAH

ZAVIHEK: SINONIMI

Prek zavihka *Sinonimi* v slovarskem meniju uporabnik dostopa do avtomatsko generiranih sinonimov za besedo, ki je obravnavana v iztočnici. Sinonimi so pridobljeni iz slovenske različice semantičnega leksikona sloWNet, zgrajene po vzoru baze WordNet, kjer vsaka beseda spada v enega ali več sinonimnih nizov oz. sinsetov, povezanih v ontološko mrežo. Sinonimi se iztočnici v slovarju pripisujejo v prvi fazi ne glede na pomen, v nadaljnjih fazah, oz. po končani pomenski členitvi, pa je mogoče upoštevati postopek prekrivanja posameznih pomenov v slovarski bazi s pomensko členitvijo v SloWNetu, za kar je bila že narejena pilotska študija z obetavnimi rezultati (Fišer idr. 2012).

Sinonimi za dano iztočnico so v spletnem slovarju prikazani skupaj z besedami, ki se s sinonimom najpogosteje sopojavljajo (t. i. kolokatorji), in sicer na 5-stopenjski lestvici, ki uporabniku pomaga pri odločitvi, kakšno besedilno okolje je za določen sinonim v razmerju do iztočnice najbolj sprejemljivo. Prva stopnja prikazuje kolokatorje, ki se vežejo samo z iztočnico, druga stopnja kolokatorje, ki se vežejo bolj na iztočnico kot na njen sinonim, tretja stopnja izkazuje kolokatorje, ki se enakovredno vežejo z iztočnico in njenim sinonimom, četrta stopnja kolokatorje, ki se vežejo bolj na sinonim kot na iztočnico in peta stopnja kolokatorje, ki se vežejo zgolj s sinonimom. Kot je razvidno iz Slike 20, so posamezne stopnje prikazane v stolpcih, kjer je tako z odtenki sive barve kot z velikostjo kolokatorja v odnosu do

iztočnice mogoče hitro ugotoviti, katere povezave so za slovenščino bolj in katere manj običajne. Ob kliku na puščico ob posameznem kolokatorju se za konkretno kombinacijo sinonima in iztočnice v posebnem okvirčku odpre seznam korpusnih zgledov, ki dano kombinacijo prikazujejo v realnem okolju.

Podatke o različnih kolokatorjih glede na sinonim iztočnice pridobivamo iz korpusa Gigafida s pomočjo orodja Sketch Engine oz. funkcije Sketch Difference.

SLIKA 20: PRIKAZ PODATKOV O SINONIMIH

ZAVIHEK: IZVOR

Pod zavihkom *Izvor* najde uporabnik dva tipa historičnih podatkov o iztočnici oz. njenem pomenu v primeru homonimije, ki jo v predlaganem slovarju upoštevamo znotraj večpomenskosti in ne dveh ali več samostojnih gesel (gl. poglavje 2.2.1, Iztočnica), in sicer: (1) podatek o etimološkem izvoru besede in njenih sorodnih oblikah v različnih jezikih ter (2) podatek o starinskih oblikah ali zapisih besede v slovenskem jeziku glede na časovno umeščenost besedila, v katerem se je konkretna oblika pojavila. Klik na letnico na časovnem traku ponudi v ustreznem besedilu tudi zglede, ki prikazujejo starejšo obliko v konkretnem besedilu.

Podatke o starejših oblikah za besedo v iztočnici je mogoče dobiti avtomatsko na podlagi Jezikovnih virov starejše slovenščine, ki vsebujejo Slovar starejšega slovenskega knjižnega jezika in starejše oblike, oboje izdelano v okviru projekta IMP. Temu se nato v modri fazi izdelave gesla dodajo še opombe o izvoru besede in njeni etimologiji, ki jih iztočnici pripiše etimolog.



SLIKA 21: PRIKAZ PODATKOV O IZVORU

ZAVIHEK: GOVOR

V zavihku *Govor* so uporabniku na voljo podatki o različnih regijskih govornih različicah besede v iztočnici, kot jih je mogoče avtomatsko pridobiti iz korpusa govorne slovenščine Gos, zato so uporabnikom na voljo že v prvi – rdeči fazi izdelave gesla. Do podatkov o regijskih variantah uporabnik dostopa tako, da s klikom na posamezno regijo prikličemo tabelo, v kateri so navedene za izbrano regijo izpričane oblike v svoji sklonski ali spregatveni paradigmi. Posamezne oblike v tabeli so, kot prikazuje Slika 22, klikljive in vodijo uporabnika do posnetkov izjav, v katerih je oblika izpričana v avtentičnem okolju. Uporabnik prek posamezne oblike lahko dostopa tudi do drugih podatkov, ki so v zvezi s posamezno izjavo zabeleženi že v korpusu Gos, kot denimo: vrsta diskurza, tip in opis dogodka, število udeležencev, vir ter podatki o govorcu (npr. starost, spol, regija...). Pri iztočnicah, za katere podatki v korpusu Gos niso na voljo, zavihke *Govor* ne bo prikazan.

1. 4. 2013

Pomen

Frazeologija

Oblike

Sinonimi

Izvor

Govor

Norma

Vizualizacija

Multimedija

Statistika

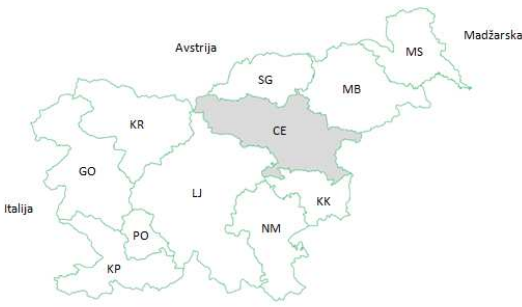
SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

reči

glagol

 /rěči/

P 1000



SEDANJIK

oseba	oblika
EDNINA	
prva	rečem
druga	rečeš
tretja	rječe, reče
DVOJINA	
prva	rečeva
druga	/
tretja	/
MNOŽINA	
prva	rečemo
druga	/
tretja	rječejo

DELEŽNIK

število	oblika
MOŠKI SPOL	
ednina	rik, riku, rek, reku, reko
dvojina	/
množina	rekl, rekli
ŽENSKI SPOL	
ednina	reka, rekla, rjekla
dvojina	/
množina	/
SREDNJI SPOL	
ednina	rekl
dvojina	/
množina	/

NEDOLOČNIK IN NAMENILNIK

glag. oblika	oblika
nedoločnik	reč, rečt
namenilnik	reč

VELELNIK

število	oblika
PRVA OSEBA	
dvojina	/
množina	recimo, rečma
DRUGA OSEBA	
ednina	/
dvojina	/
množina	/

40 eem ja reku je da čakl danes eee prvega avgusta začne že delat

40 pa se mu ful ni dal pol sem mu pa jz reku naj gre na pir vsaj ne

40 ko pa vem da reku hujš to je ze vse znamka veš ko je tut volvo bil

© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 22: PRIKAZ PODATKOV IZ KORPUSA GOS

ZAVIHEK: NORMA

Zavihek *Norma* je znotraj spletnega slovarja namenjen podatkom o normativnih in stilističnih posebnostih besede, ki je obravnavana v iztočnici, saj predvidevamo, da potrebuje uporabnik celovito informacijo o besedišču slovenskega jezika na enem mestu. O tem nas ne prepričuje le sodobna leksikografska literatura, pač pa tudi dobri primeri zlasti evropske leksikografske prakse.

Podatki o tem, kako se beseda ustrezno zapisuje (*ponavadi* ali *po navadi*; *glodalec* ali *glodavec*), izgovarja ali naglašuje (*vreme* ali *ureme*; *pétnajst* ali *petnájst*), sklanja (z *Georgeem/Georgeom/Georgom*; *pri otrokih* ali *pri otrocih*), tvori (*Voltairejev* ali *Voltaireov*) ipd., čeprav niso neposredno povezani s posameznimi pomeni, so iztočnici dodani v bodisi v začetni fazi izdelave gesla, če so na voljo v

Slogovnem priročniku, bodisi v kasnejši modri fazi izdelave gesla, ko na iztočnico pripenjamo tudi t. i. »zunajbazne podatke«, ki jih zagotavljajo jezikovni strokovnjaki, kot so terminologi, etimologi in strokovnjaki za normo oz. standardizacijo jezika.

SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

glodavec samostalnik /glodávec/

Pri mnogih samostalniki, ki se končajo s **prিপিস্কিম obrazilom -lec/-vec**, se v rabi pojavljajo razhajajoči zapisi, kar je posledica pogostih sprememb in nedoločnih pravil v jezikovnem standardu skozi zgodovino. Polemika o zapisovanju obrazil -lec ali -vec je znana tudi kot **problem bralca** ali polemika **bralec – bravec**, zaradi svoje razsežnosti in dolgotrajnosti pa so jo označevali tudi kot novo **črkarsko pravdo**. Zapuščina tega perečega pravopisnega vprašanja je še vedno vidna v jeziku, saj se raba pri nekaterih oblikah vztrajno odmika od jezikovnega standarda, za slednjega pa se zdi, da deluje zoper jezikovno intuicijo, ki se glede na zgodovinski razvoj nagiba k obrazilu -lec. Pri težavah z zapisom posameznih oblik si lahko pomagamo s **korpusom**, ravnamo pa se lahko po naslednjih pravilih:

1. Priponsko obrazilo **-lec** dodajamo:
 - a) glagolski osnovi na **samoglasnik: brati – bralec, maščevati – maščevalec**;
 - b) korenu na samoglasnik: **greti – grelec, vreti – vrelec**.

Priponsko obrazilo **-lec** se lahko zapisuje tudi kot **-vec**, če je na koncu korena **-l-** ali **-lj-**: **voliti – volivec/volilec, ponavljati – ponavljavec/ponavljalec**. Nekateri primeri imajo podobne vzporednice, tvorjene iz pridevnika: **blebetati – blebetalec** proti **blebetav – blebetavec**.

2. Priponsko obrazilo **-vec** dodajamo:
 - a) korenu na samoglasnik: **klati – klavec, briti – brivec, peti – pevec** (največkrat beseda izraža (živega) vršilca nekega dejanja, izjema je **števec**, kadar označuje napravo);
 - b) osnovi na samoglasnik, če je pred glagolsko pripono črka **l** ali sklop **lj**: **voliti – volivec, ponavljati – ponavljavec**.

Kljub temu da obrazilo **-vec** pri nekaterih oblikah pri življenju ohranja predvsem pravopisni standard in se jezikovna raba ne ozira na to, ali je pred glagolsko pripono črka **l** ali sklop **lj**, je nujno pri visoki pojavnosti oblik na **-vec** omeniti tudi vlogo črkovalnikov v urejevalnikih besedil, denimo v Microsoftovi **programski zbirki Office**, kamor je slovenski črkovalnik vključen od leta 1994. Podobne korekcijske vzorce lahko opazimo tudi pri črkovalnikih za druge sisteme oz. programske zbirke. Korpusni podatki denimo kažejo, da je bila oblika **volilec** ob obdobju državnoborskih volitev v Sloveniji leta 1996 veliko pogostejša, neke ob vstopu v novo tisočletje jo je oblika **volivec** dohitela, od leta 2005 naprej pa je oblika **volivec** prepričljivo močnejša od konkurenčne. Prav tako ob takšni dinamiki oblik ne gre zanemariti vpliva izida **Slovenskega pravopisa 2001** – zadnjega normativnega pravopisnega priročnika za slovenski jezik. Pred tem so bila v veljavi pravila, določena leta 1950, od takrat pa se je razmerje med rabo in standardom pri tovrstnih oblikah dodatno nagnilo v prid oblik na **-lec** – za ponovno vzpostavitev razmerja med obraziloma je poskrbela prav nova izdaja pravopisa.

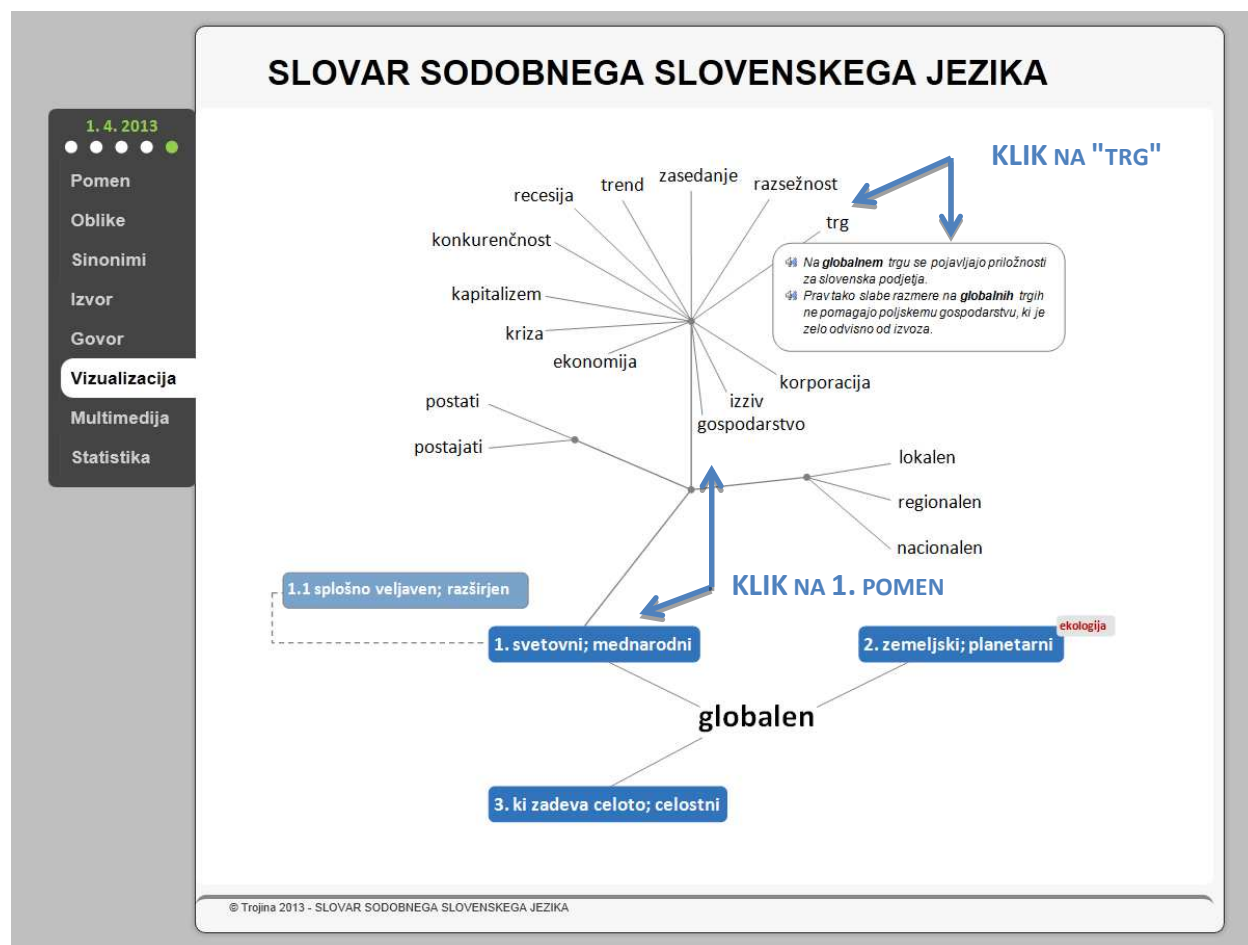
© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 23: PRIKAZ PODATKOV IZ SLOGOVNEGA PRIROČNIKA

V slovarski bazi normativni podatki niso zabeleženi, pač pa jih za namene spletnega slovarja pridobivamo iz Slogovnega priročnika, ki je bil izdelan pri projektu Sporazumevanje v slovenskem jeziku. Odločitev za ločevanje normativnih od izhodiščno leksikalnih podatkov, kamor sodi zlasti pomenska členitev in prikaz tipičnega okolja besede v obliki kolokacij, je povezana s tem, da so denimo posamezne oblike besede lahko tudi glede na obstoječi standard nesprejemljive, hkrati pa v slovenskem besedišču opazno prisotne in zato zastopane tudi v slovarju. Pri tem je pomembno predvsem to, da določene besede ali njene oblike (npr. zajedalec, piersing ipd.) kljub svoji nestandardnosti lahko razvijejo samostojne pomene, oblikujejo stalne zveze in imajo glede na standardne zapise svoje specifično besedilno okolje. Pristop, ki nestandardnih oblik ne izključuje iz obravnave, uporabniku zagotavlja, da nestandardnost sama na sebi ne pomeni hkrati tudi odsotnosti informacije same oz. še več, uporabniku pokaže, kaj in zakaj je v določenih okoliščinah bolj ali manj tipično in/ali bolj ali manj sprejemljivo.

ZAVIHEK: VIZUALIZACIJA

Zavihek *Vizualizacija* vsebuje po vzoru grafičnih prikazov³⁸ semantičnih leksikonov, kot je npr. Wordnet, medsebojno prepletenost in soodvisnost vseh podatkov v zvezi z izbrano besedo, ki so predhodno zabeleženi v slovarski bazi. V tem zavihku so vsi pomeni in podpomeni besede ter z njimi povezani podatki, kot so vzorci, kolokacije in opozorila o posebnostih rabe (oznake), predstavljeni v obliki nevronske mreže, ki prikazuje podatke po vzoru delovanja človeških možganov. Taka predstavitev uporabniku omogoča celostno predstavo o pomenski in kolokacijski razvejanosti besede ter o pomensko sorodnih notranjih povezavah. Uporabnik lahko s klikom na posamezni (pod)pomen širi podobo besede vse do ravni vzorcev in z njimi povezanih kolokacij. Za razliko od »klasične« slovarske postavitve gesla na spletni strani, kjer si pomeni in podpomeni sledijo zaporedno, prikazuje mreža s pomočjo posameznih vozlov oz. mrežnih jeder/točk lastnosti, ki so besedi oz. njenemu (pod)pomenu skupne: na primer vse kolokacije za dani pomen 'svetovni; mednarodni' besede *globalen*, ki ustrezajo vzorcu **PRIDEVNIK + SAMOSTALNIK**. Kot prikazuje spodnja slika/video, se ob kolokacijah na klik v okvirčku odpirajo tudi zgledi, ki predstavljajo konkretno kolokacijo v realnem besedilu.



³⁸ Med najbolj znanimi lahko omenimo: Visual Thesaurus: <http://www.visualthesaurus.com/>, Graph Words: <http://graphwords.com/>, Visuwords: <http://www.visuwords.com/>, Snappy Words: <http://www.snappywords.com/> in vizualizacijsko orodje za slovenski wordnet, SloWTool: <http://nl.ijs.si/slowtool/>.

ZAVIHEK: MULTIMEDIJA

V zavihku *Multimedija* so za iztočnico ali njene izpričane pomene prikazane na spletu prosto dostopne slike in video posnetki, ki pripomorejo k bolj plastični predstavi pomenov in pomenskih tendenc besede. Dodajanje tovrstnih podatkov je v spletnih slovarjih, ki to omogočajo v veliko večji meri ter na preprostejši in predvsem cenovno sprejemljivejši način kot tiskani slovarji, splošna in z vidika uporabnika zaželena praksa. Glede na to, da se slikovno in video gradivo pripenja na posamezne (pod)pomene besede v iztočnici, je na slovarsko geslo pripet v pomenskorazčlenitveni (rumeni) fazi geselske izdelave. V slovarju bomo uporabili samo gradiva, ki so dostopna pod enako licenco kot sam slovar, torej Creative Commons – priznanje avtorstva, deljenje pod enakimi pogoji. Taka gradiva je mogoče najti na različnih platformah za izmenjavo slik ali videoposnetkov, kot so YouTube, Flickr itd.

The screenshot displays the 'SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA' interface. On the left is a sidebar with navigation options: '1. 4. 2013', 'Pomen', 'Oblike', 'Sinonimi', 'Izvor', 'Govor', 'Vizualizacija', 'Multimedija' (highlighted), and 'Statistika'. The main content area for the word 'dihati' (glagol, P 5000) features a video player with the title 'Minute za zdravje - Pravilno dihanje' showing a woman demonstrating breathing exercises. Below the video are several images: two diagrams of the human respiratory system, a woman in a blue athletic top, a man in a grey shirt, a pair of sneakers, and a blue long-sleeved shirt. A small anatomical diagram of the respiratory system is also present on the right. At the bottom, a copyright notice reads '© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA'.

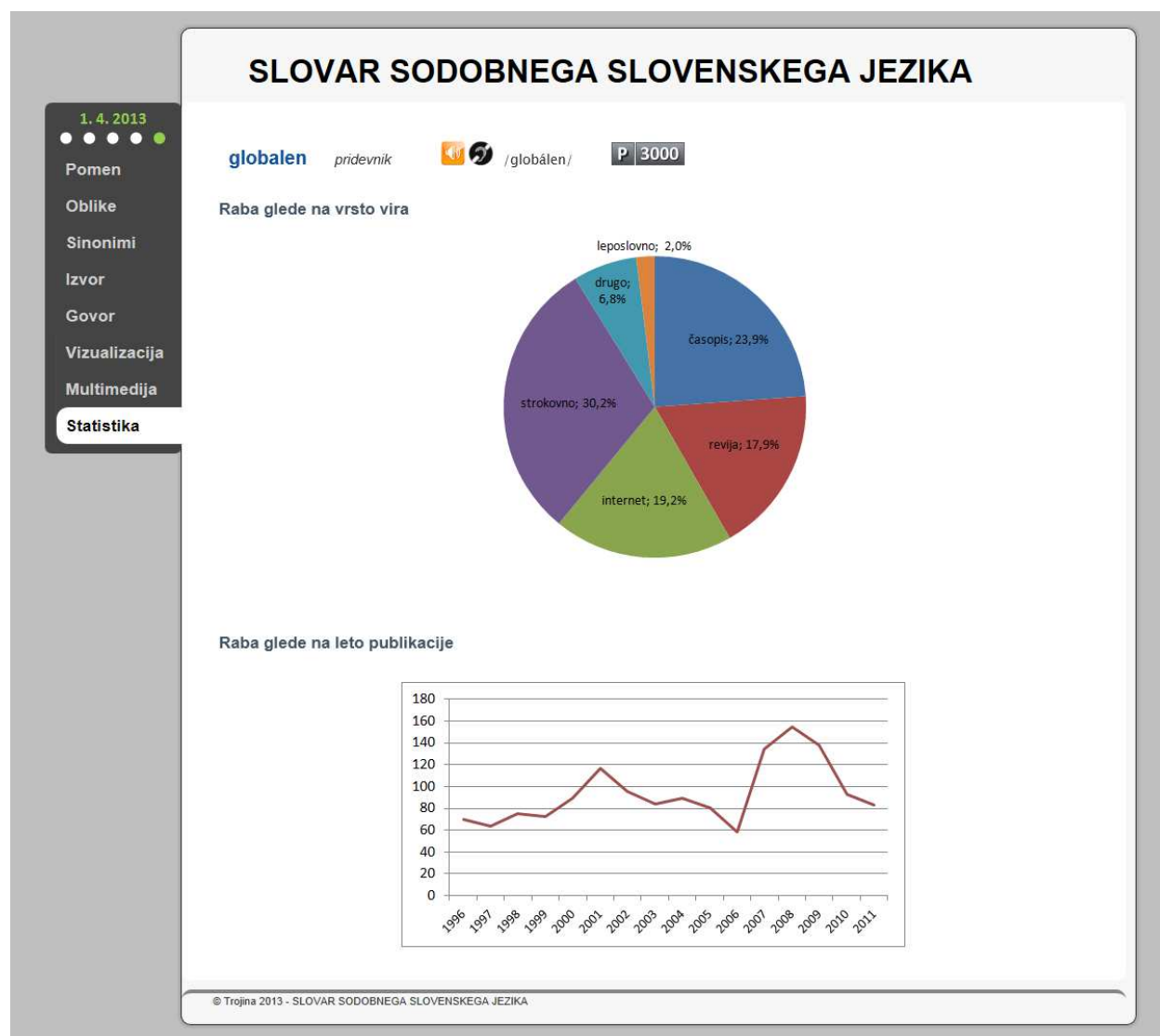
SLIKA 24: PRIKAZ MULTIMEDIJSKIH VSEBIN

ZAVIHEK: STATISTIKA

Pri frekvenčnih podatkih je ključno, da se uporabniku ponudijo na način, ki je razumljiv in tudi primerljiv z drugimi besedami. Pogostost iztočnice v nekem korpusu, absolutna ali relativna (npr. X pojavitev na milijon besed), ki je predstavljena kot številčni podatek, uporabniku ne ponudi jasne slike. Zato se je v na korpusih temelječih slovarjih uveljavil pristop, ki iztočnice umešča znotraj določene frekvenčne skupine, npr. če je beseda med 1000, 2000 ali 3000 najpogostejšimi v jeziku. To prakso bi uporabili tudi v

predlaganem Slovarju, vendar bi veljalo dodati še podatek o frekvenci tudi pri redkejših besedah, ki bi bil za uporabnike koristen zlasti pri produktivni rabi, npr. pri izbiri med besedami, ki so si po pomenu ali drugih lastnostih podobne.

Druga vrsta frekvenčne informacije, ki je za uporabnike zanimiva, je podatek o gibanju (relativne) frekvence besede (ali besedne zveze) skozi čas. Tovrstna informacija je pri novejšem besedju (od l. 1994, ko je v korpusu na voljo dovolj podatkov) lahko dopolnilo etimološkim podatkom (gl. Sliko 21), saj pokaže, kdaj se je beseda pojavila in kdaj uveljavila v jeziku.



SLIKA 25: PRIKAZ STATISTIČNIH PODATKOV IZ KORPUSOV

2.2.2 DOLOČITEV PRIORITETNEGA ZAPOREDJA PRIPRAVE GESEL

Avtomatsko izdelana gesla na podlagi luščenja podatkov iz korpusa Gigafida – skupaj 100.000 izbranih lem – bodo na voljo uporabnikom po šestih mesecih po začetku izdelave slovarja, končna gesla se bodo izdelovala in prikazovala na spletu sproti po predvidenih postopkih. Ker pa za uporabnike vsa gesla niso

pomembna v enaki meri, bomo določili kriterije za njihovo prednostno obravnavo. Glavna kriterija sta dva: relevantnost besede za usvajanje splošnega besedišča in aktualnost besede. Kriterij relevantnosti besede za usvajanje splošnega besedišča je povezan z delitvijo besedišča v štiri skupine (Coxhead 2000, Nation in Waring 1997), in sicer v splošno besedišče, akademsko oz. šolsko besedišče, terminologijo in zelo redke besede. Ker predlagani slovar cilja predvsem na šolske uporabnike ter odrasle uporabnike, ki bodo slovar uporabljali v okviru izobraževanja ter zaradi sporazumevalnih potreb, bo pomembno čim hitreje ponuditi končna gesla, ki bodo pokrivala splošno besedišče, zlasti tisto, ki se uporablja pri izobraževanju v osnovnih in srednjih šolah, deloma tudi na univerzitetni ravni, in sicer nepodročno specifično oz. neterminološko.

Drugi kriterij, aktualnost besede, je povezan z dejstvom, da so zaradi političnih, športnih in podobnih dogodkov nekatere besede oz. besedne zveze, zlasti tiste z določenega področja, v določenih tednih ali mesecih uporabljene precej bolj pogosto kot ponavadi. V zadnjih letih so bili to izrazi, kot so npr. *slaba banka*, *interpelacija*, *doping*, *fiskalno pravilo*, *projektna vlada*. Takšne besede ali zveze zaradi (vsaj začasnega) povečanja rabe postanejo del splošnega besedišča in pričakuje se, da je razlago njihovega pomena mogoče najti hitro in sproti, ko je raba še na vrhuncu. Ker je pričakovano, da bodo uporabniki iskali oz. potrebovali razlago ravno takšnih izrazov (kar sproti detektiramo tudi z beleženjem spletnih poizvedb), se jih zdi smiselno čim prej ponuditi v obliki dokončanih gesel. V ta namen je v okviru projekta predvidena uporaba spletnega pajka (ang. *web crawler*), ki smo ga uporabili že pri izdelavi spletnega dela korpusa Gigafida. Z orodjem bomo redno, npr. enkrat tedensko ali celo dnevno, shranjevali spletne strani glavnih novičarskih portalov in časopisov ter avtomatsko luščili besedišče, kateremu bo raba v primerjavi s preteklimi obdobji statistično pomembno narasla. Za to besedišče se bodo po predvidenem tehničnem protokolu avtomatsko izdelala gesla (če jih še ne bo v geslovniku), strojno izluščile definicije iz besedil (če bodo obstajale v besedilih, shranjenih s spleta), leksikografi pa se bodo takoj lotili izdelave končnih gesel, pri čemer bo glavni poudarek na pisanju oz. dodelavi razlag. V idealnem scenariju bi imel tako uporabnik, ko bi se srečal z eno takšnih besed in besednih zvez, sprva na voljo avtomatsko izdelano geslo, že naslednji dan pa v celoti izdelano geslo z leksikografsko razlago.

Informacije o tem, katere besede oz. besedne zveze so za uporabnike aktualne, bomo poleg spremljanja spletnih objav s statističnimi orodji pridobivali tudi iz dnevnikov uporabniških poizvedb. Spremljanje uporabniških aktivnosti na spletni strani bo koristno tudi za druge namene, npr. ugotavljanja, katere dele gesel uporabniki najpogosteje uporabljajo, katere povezave največkrat kliknejo, katere besede največkrat napačno vtipekajo ipd.

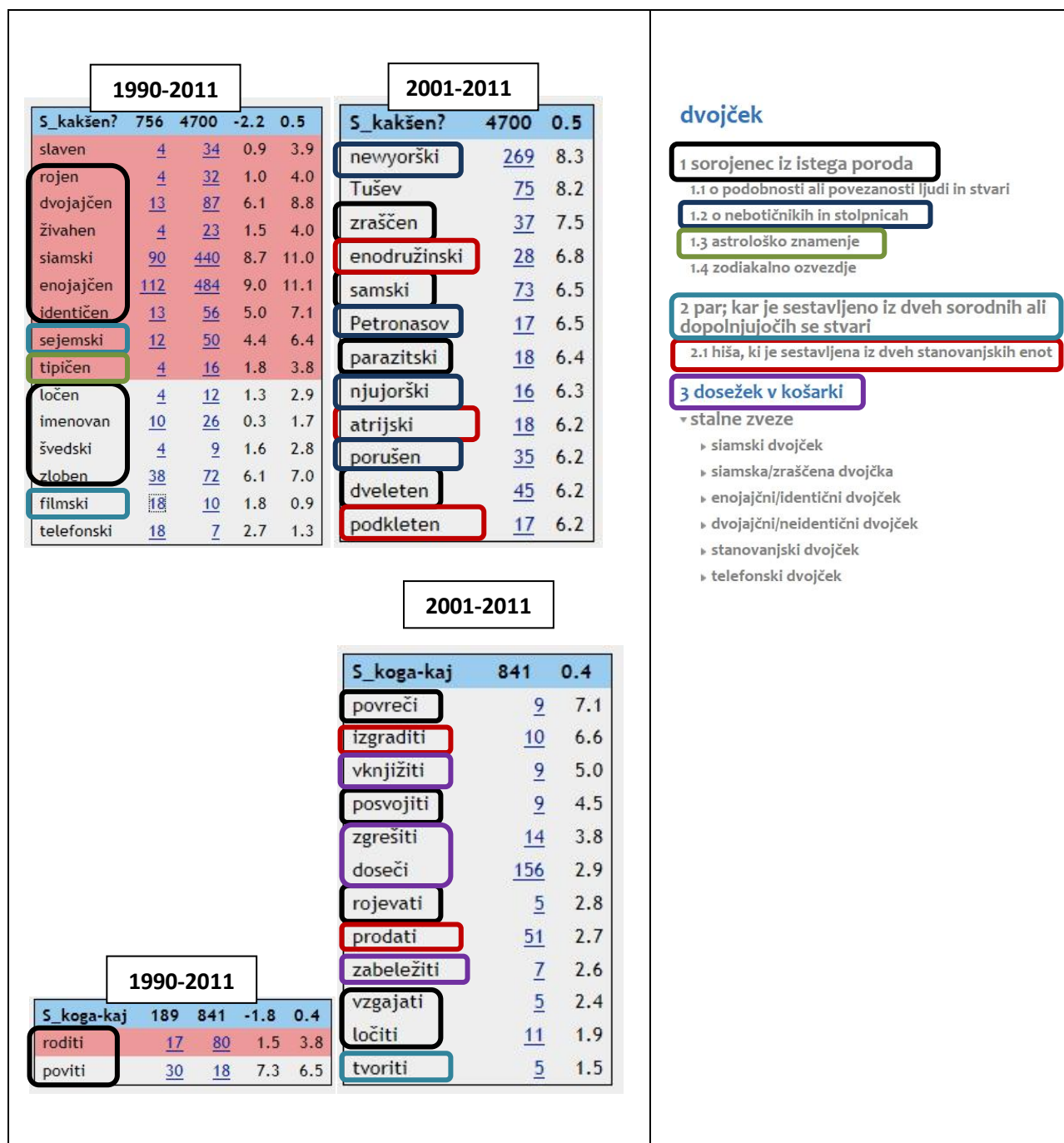
2.2.3 DETEKTIRANJE POMENSKIH SPREMENB BESEDIŠČA

Postopek, ki je ločen od avtomatske detekcije novih besed oz. neologizmov, je prepoznavanje novih pomenov ali pomenskih odtenkov (prek kolokacij). Za namene tega postopka bomo izdelali posebno orodje, ki bo leksikografe avtomatično opozarjalo na spremembe v rabi besed. Orodje bi se opiralo na funkcijo primerjalnih skic v orodju Sketch Engine, ki primerja besedne skice besed v različnih časovnih obdobjih, pri čemer je predpogoj redno posodabljanje korpusnih podatkov. Vzemimo kot primer samostalnik *dvojček* in njegovo pojavljanje v podkorpusu besedil iz obdobja 1990-2000 ter iz obdobja

2001-2011.³⁹ V primerjalni besedni skici so za detekcijo novih kolokacij (in pomenov) relevantne zlasti relacije oz. skladijska vzorca, kot sta PRIDEVNIK+SAMOSTALNIK, ki vsebuje pridevniške kolokatorje, in GLAGOL+SAMOSTALNIK₄, ki vsebuje glagole, ob katerih se *dvojček* pojavlja kot stavčni predmet (Slika 26). Za referenco podajamo poskusno geslo iz SSSJ, izdelano na podlagi teh podatkov. Kot lahko vidimo, se v obeh obdobjih pojavljajo prvi pomen, podpomen 1.3 in drugi pomen, medtem ko se v obdobju 2001-2011 pojavijo še podpomena 1.2 in 2.1 ter tretji pomen, povezan z rabo besede v športu. Poleg tega se v tem obdobju pojavijo tudi kombinacije, povezane s prvim pomenom, npr. *samski dvojček* in *parazitski dvojček*.

Poleg detekcije novih besed, pomenov itd. bi postopek lahko izkoristili tudi za posodabljanje razvrstitve pomenov in podpomenov. V geslu naj bi namreč težili k temu, da pogostejše rabe, za katere obstaja večja verjetnost, da se bodo uporabniki z njimi srečali in jih iskali v slovarju, ponudimo čim prej oz. čim višje v razporeditvi pomenov in podpomenov. Razporeditev torej sledi načelu pogostosti rabe, ne pa historičnemu načelu, čemur so pravzaprav sledili že avtorji SSKJ (Suhadolnik 1963, Suhadolnik 1997). Posledično bo redkejši pomen pomaknjen proti dnu gesla oz. podpomen proti dnu pomena. Pri besedi *dvojček* to npr. velja za stalno zvezo *telefonski dvojček*, saj podatki kažejo, da njena raba upada. V primeru, da določena raba besede izgine, se pomen/podpomen opredeli z ustrezno oznako, kar pomeni, da se ga sčasoma lahko odstrani iz eksplicitne vizualizacije v spletnem slovarju (ne pa tudi iz slovarske baze).

³⁹ Za ponazoritvene namene primerjamo večji obdobji, medtem ko bi dejanski postopek verjetno vključeval primerjavo po posameznih letih oz. primerjavo celotnega korpusa z besedili, dodanimi v zadnjem letu.



SLIKA 26: RELEVANTNI RELACIJE IZ PRIMERJALNE SKICE PODKORPUSOV 1990-2000 IN 2001-2011 ZA DVOJČEK, TER POSKUSNO GESLO IZ SSSJ

2.2.4 SPROTNO POSODABLJANJE SLOVARSKIH GESEL

Sprotno posodabljanje slovarskih gesel pomeni, da slovar nastaja »v živo« in je javno dostopen v vseh fazah izdelave, torej uporabnikom slovarja ni treba čakati, da so pred objavo izdelana vsa gesla. Danes so avtomatske metode obdelave podatkov iz korpusov napredovale do mere, da lahko že brez leksikografske intervencije ponudijo dokaj zadovoljlivo sliko o leksikogramatičnem vedenju besede, zato

uporabnikom na rešitev jezikovne zagate ni treba čakati 5 ali več let, tj. do izida slovarja. Hkrati je tudi škoda, da gesla, ki jih leksikografi izdelujejo v času trajanja projekta, niso sproti in takoj na voljo uporabnikom, in sicer v stanju, v katerem se trenutno nahajajo – z jasnim opozorilom, da geslo še ni do konca obdelano oz. z informacijo, v kateri fazi obdelave se trenutno nahaja. Za razliko od knjižnega formata, kjer je bila dokončnost obdelave gesel nujnost in je bila osnovna logika objavljanja slovarjev tipično vezana na letnico nove izdaje določene knjižne slovarske linije, je v digitalnem mediju prevladala logika sprotnega posodabljanja posameznih gesel, čemur so se prilagodila tudi pričakovanja uporabnikov, predvsem pri prosto dostopnih virih (npr. Wikipedija).

Predlagani postopek sprotnega posodabljanja, ki vsebuje različne faze geselske izdelave (gl. poglavje 2.3 v nadaljevanju) je rešitev obeh omenjenih kritik.

2.2.5 MOŽNOSTI TISKANJA

Za posebne situacije je predvidena tudi možnost tiskanja gesel, npr. za rabo v šolah, kjer ni sprotnega dostopa do spleta in podobno. Trenutno sta predvideni dve vizualizaciji za tisk: »šolska« in »knjižna«. Pri šolski vizualizaciji se predvideva, da uporabnik potrebuje eno geslo ali omejeno število gesel, ki pa vsebujejo nekoliko več informacij iz slovarske baze (slovarske zglede, kombinacijo semantičnih indikatorjev in definicij, vse izpisane oblike iz leksikona besednih oblik in podobno).

"KNJIŽNI NATIS"

globálen -lna -o prid. 1. *svetovni; mednarodni*: globalna ekonomija, globalna komunikacija; v *New Yorku naj bi državniki in podjetniki razpravljali o globalni varnosti* // *splošno veljaven; razširjen*: merila, kakšna ženska je lepa, postajajo vse bolj globalna 2. *zemeljski; planetarni* ekol.: globalna otoplitev, globalna temperatura 3. *ki zadeva celoto; celostni*: s pomočjo plakata postanejo učencem misli bolj jasne in končno odkrijejo globalni pomen besedila ◇ ekol. globalno segrevanje/ogrevanje postopno zviševanje povprečne temperature v Zemljinem ozračju in oceanih; globalna vas svet kot celota, v kateri je vse med seboj povezano in soodvisno; globalni nomad kdor nenehno potuje po svetu in to šteje kot svoj način življenja

"ŠOLSKI NATIS"

globalen pridevnik

/globálen/

P: 3.000

1. svetovni; mednarodni

globalni procesi, zlasti gospodarski in politični, zajemajo ves svet

- V New Yorku naj bi državniki in podjetniki razpravljali o **globalni** varnosti.
- Tudi največje svetovne firme, ki danes obvladujejo **globalni** trg, so se razvile iz malih podjetij in obratov.
- Motorola je zaradi **globalne** recesije v visokotehnoloških gospodarskih panogah lani odpustila 48.400 zaposlenih.
- globalna [ekonomija, konkurenčnost, recesija, korporacija]
- globalna [komunikacija]
- [postajati, postati] globalen

1.1 splošno veljaven; razširjen

če postanejo neke dejavnosti ali lastnosti globalne, jih upošteva vedno več ljudi ali držav po svetu

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

2. zemeljski; planetarni ekologija

globalne spremembe v okolju vplivajo na celoten zemeljski planet

- globalno [segrevanje, ogrevanje]
- globalna [temperatura, otopletev]

3. ki zadeva celoto; celostni

- S pomočjo plakata postanejo učencem misli bolj jasne in končno odkrijejo **globalni** pomen besedila.
- Ta študija zlasti opozorja na vrsto vprašanj, ki se zastavljajo glede oblikovanja **globalne** nacionalne kulturne politike.
- Gre za **globalni** pristop do človeka, ki mu pomaga, da vzame življenje v svoje roke in ga znova progamira, kot si ga je sam zamislil.

STALNE ZVEZE

globalno segrevanje/ogrevanje ekologija

globalno segrevanje je postopno zviševanje povprečne temperature v Zemljinem ozračju in oceanih

- Sivka zdrži z malo vode in naj bi bila odporna proti daljši suši - idealna rastlina za klímo **globalnega** segrevanja.
- Taljenje ledu bi bilo lahko posledica **globalnega** segrevanja, čeprav se znanstveniki še niso zedinili o vzroku.

globalna vas

če človek poimenuje svet globalna vas, ga vidi kot celoto, v kateri je vse med seboj povezano in soddvisno

- Na internet sicer večina gleda pozitivno, saj nudi nove priložnosti, vendar obenem tržišče spreminja v **globalno** vas.

- Svet bo do leta 2010 postal **globalna** komunikacijska vas.

globalni nomad

kdor nenehno potuje po svetu in to šteje kot svoj način življenja

- Roubini sam zase pravi, da je »**globalni** nomad«, ker držav ni mogoče spoznavati prek računalnika, ampak je treba iti tja, da jih začutiš.
- Poigraval sem se z mislijo, da bi bil to: **globalni** nomad, ki se v nedogled seli za poletjem.

OBLIKE

• moški spol

	ednina	dvojina	množina
imenovalnik	globalen <i>nedoločno</i> globalni <i>določno</i>	globalna	globalni
rodilnik	globalnega	globalnih	globalnih
dajalnik	globalnemu	globalnima	globalnim
tožilnik	globalen <i>nedoločno</i> globalni <i>določno</i> globalnega <i>živo</i>	globalna	globalne
mestnik	globalnem	globalnih	globalnih
orodnik	globalnim	globalnima	globalnimi

• ženski spol

	ednina	dvojina	množina
imenovalnik	globalna	globalni	globalne
rodilnik	globalne	globalnih	globalnih
dajalnik	globalni	globalnima	globalnim
tožilnik	globalno	globalni	globalne
mestnik	globalni	globalnih	globalnih
orodnik	globalno	globalnima	globalnimi

• srednji spol

	ednina	dvojina	množina
imenovalnik	globalno	globalni	globalna
rodilnik	globalnega	globalnih	globalnih
dajalnik	globalnemu	globalnima	globalnim
tožilnik	globalno	globalni	globalna
mestnik	globalnem	globalnih	globalnih
orodnik	globalnim	globalnima	globalnimi

»Knjižni« natis je še bolj strnjen in je pravzaprav namenjen klasičnemu tisku, torej natisu in vezavi slovarja v knjigo, če se bo morda pri določenih uporabnikih kazala takšna potreba. V tem natisu so ohranjene samo nekatere informacije iz gesla, po vsebini in izgledu pa je zelo podoben standardnim knjižnim slovarjem, tudi SSKJ in SNB. Transformacija gesel iz slovarske baze za natis v katerikoli od omenjenih variant je avtomatska.

2.3 IZVEDBA: IZDELAVA SLOVARSKEGA GESLA V POSAMEZNIH FAZAH

V nadaljevanju prikažemo organizacijo slovarskega dela in vključevanje podatkov, ki jih spletni slovar predvideva v posameznih fazah, kot jih prikazuje spodnja slika in jih predvideva organizacijski načrt, ki je prikazan v nadaljevanju.



SLIKA 27: FAZE IZDELAVE SLOVARSKEGA GESLA

Izdelava gesel po posameznih fazah omogoča, da so podatki uporabniku na voljo že v času izdelovanja slovarja in ne šele po zaključku slovarskega dela – ko so podatki, ki so bili v slovarsko bazo vneseni na začetku projekta, stari vsaj toliko časa, kot je minilo od začetka izdelave. Prednosti, ki jih ima tak postopek, odtehtajo večjo zapletenost procesa, v katerem morajo biti posamezne faze zelo natančno določene, opravila leksikografov in drugih sodelavcev pa usklajena in vnaprej predvidena. Večstopenjskost izdelave gesel omogoča tudi učinkovitejšo in bolj ekonomično delitev dela. V prvi fazi tako večino dela opravi računalnik, človeško delo pa je vključeno šele v kasnejših fazah, kjer se znanje leksikografov izrablja le za specifične leksikografske postopke, ki zahtevajo izkušnje in strokovno usposobljenost. Predhodna delitev nabora iztočnic v težavnostne stopnje omogoča tudi postopno seznanjanje manj izkušenih leksikografov s postopki pomenskega členjenja in opisovanja. Ker je za določena rutinska dela, kot je npr. odstranjevanje neustreznih in nerelevantnih podatkov ter razvrščanje kolokacij in zgledov pod ustrezne pomena predvidena izraba moči množic, je mogoče racionalizirati stroške človeških virov, predvsem pa pospešiti čas izdelave gesel.

Dodajanje podatkov v posamezni fazi prikazuje Tabela 1:

	rdeča	oranžna	rumena	modra	zelena
POMEN	iztočnica	-	-	-	končna redakcija na vseh nivojih
	besedna vrsta	-	-	-	
	pogostost	-	-	-	
	vzorci (strukture)	urejeno	-	-	
	kolokacije	urejeno	-	-	
	zgledi	urejeno	-	-	
	znakovni jezik	-	-	-	
			indikatorji + pomenski meni	-	
			pomenske sheme/definicije	-	
STALNE ZVEZE			stalne zveze	-	
FRAZEOLOGIJA			frazeologija	-	
			oznake	-	
	izgovor	-		izgovor (zapis in avdio)	
				terminološki pregled	
VIZUALIZACIJA	vizualizacija	-	-	-	
STATISTIKA	statistika	-	-	-	
OBLIKE	oblike	-	-	oblike	
SINONIMI	sinonimi	-	pomensko razčlenjeni sinonimi	-	
IZVOR	starinske oblike	-	-	-	
				etimologija	
GOVOR	govorjene oblike po regijah				
NORMA	normativni podatki	-	-	dopolnitev norme	
MULTIMEDIJA			slike in video	-	

TABELA 1: DODAJANJE PODATKOV PO FAZAH

Legenda: siva polja predstavljajo obvezno, bela pa neobvezno prisotnost podatkov; krepki tisk nakazuje, katere faze so ključne za posamezne postopke.

Pri postopku izdelave slovarja v posameznih fazah je zelo pomembno, da uporabnik takoj in jasno prepozna, v kateri fazi se nahaja geslo – in s tem podatki, do katerih dostopa. V ta namen smo predvideli podatek z datumom stanja gesla, ki se generira skladno s spremembami, ki potekajo v celotnem procesu geselske izdelave. Na ta način zagotovimo dvoje: prvič, uporabnik ima možnost navajanja reference na slovarsko geslo, ki velja za določeno stanje tega gesla, kar je zlasti pomembno pri citiranju podatkov v raziskovalne in izobraževalne namene, in drugič, s tem uporabniku nakažemo, kaj lahko od slovarskih podatkov pričakuje tako v smislu njihove količine, urejenosti in stopnje zanesljivosti.

1. RDEČA FAZA – AVTOMATSKO IZDELANO GESLO

V prvi t. i. avtomatski – rdeči fazi izdelave slovarskega gesla so za celotno besedišče, za katerega predvidevamo slovarski opis v spletnem slovarju, tj. za 100.000 korpusnih lem, ki jih določimo glede na prag pogostosti v korpusu Gigafida in glede na predvidene parametre, določene v Kazalniku 2 (gl. poglavje 6), avtomatsko izločeni naslednji podatki: iztočnica v osnovni obliki, besedna vrsta, podatek o pogostosti v korpusu, določena slovnična opozorila, ki se v slovarskem geslu generirajo kot oznake, npr. *pogosto zanikano*, *pogosto v 3. os. ednine*, *pogosto z lastnim imenom* ipd., poleg tega tudi slovnične relacije, ki se v geslu prepišejo v vzorce, ter pripadajoče kolokacije in njihovi zgledi. Postopek avtomatizacije je bil za slovenščino že preizkušen pri izdelavi Leksikalne baze (Kosem idr. 2012), kjer smo s pomočjo v ta namen izdelane slovnice besednih skic, ki deluje v orodju Sketch Engine, s pomočjo posebej prilagojene API skripte, ki vsebuje opise vseh relevantnih slovničnih relacij za luščenje kolokacij, in s pomočjo t. i. GDEX konfiguracije, ki opredeli lastnosti dobrih zgledov, iz korpusa Gigafida izluščili zgoraj navedene podatke in jih avtomatsko prenesli v slovarsko bazo, kjer so bili pripravljeni za nadaljnjo obdelavo.⁴⁰ V prvih šestih mesecih, ko poteka postavitve računalniške infrastrukture, nameravamo izkoristiti tudi uporabo funkcije gručenja (*clustering*) kolokacij na podlagi informacij iz tezavra v orodju Sketch Engine ter implementirati funkcijo povezav na večbesedne leksikalne enote (*MW links*), kar bo omogočilo avtomatično luščenje tudi t. i. razširjenih kolokacij tipa: *[delovno] mesto* → *[prosto, novo] delovno mesto* neposredno v slovarsko bazo in prek nje v spletni slovar.

Poleg avtomatsko izluščenih podatkov iz korpusa, ki so prek slovarske baze preneseni v spletni slovar, so v prvi fazi iztočnici pripisani tudi drugi podatki, ki jih je mogoče avtomatsko pridobiti iz obstoječih baz podatkov. Sem sodijo podatki o izgovoru (če so na voljo, sicer jih dopolnimo v modri fazi), statistični podatki, ki jih pridobimo s pomočjo aplikacij v orodju Sketch Engine, podatki o besednih oblikah iz leksikona besednih oblik Sloleks, sinonimih, ki jih pridobimo iz baze SloWNet, starinskih oblikah iz Slovarja starejšega slovenskega knjižnega jezika in referenčnega korpusa starejše slovenščine, podatki o različnih govornih variantah, značilnih za slovenske pokrajine, ki se zgenerirajo na podlagi podatkov v korpusu govorne slovenščine Gos, ter normativni podatki, ki so vsebovani v Slogovnem priročniku. Vsi podatki, ki so na voljo v prvi fazi geselske izdelave, so prikazani tudi v obliki nevronske mreže v zavihku *Vizualizacija*.

⁴⁰ Za izdelavo spletnega slovarja bomo postopek nekoliko prilagodili, v nadgradnjo sodi predvsem prilagoditev sheme XML vrsti podatkov v slovarski bazi in kasneje slovarju, avtomatsko odstranjevanje kolokatorjev, ki ponudijo same enake zgleds in postavitve leme in/ali kolokatorja pri izpisu v slovarsko bazo v ustrezen sklon, spol in število.

1. 4. 2013

Pomen

Oblike

Sinonimi

Govor

Vizualizacija

Statistika

SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

globalen pridevnik

P 3000

PRIDEVNIK + SAMOSTALNIK

globalno [segrevanje, zasedanje, ogrevanje]

globalna [ekonomija, kriza, delniska, konkurenčnost, recesija, raven]

globalna [razsežnost, korporacija, otoplitev, spremenljivka, vas]

globalna [pravičnost]

globalni [kapitalizem, vodič, nomad, trend, klepet]

globalni [trg, fenomen, izziv, terorizem, motorist]

☞ Danes ni več politika ali politične stranke, ki v svojih nastopih in programih ne bi omenjala tudi klimatskih sprememb in zaskrbljenosti zaradi **globalnega** segrevanja ozračja, skrbi za pitno vodo in zdravo okolje.

☞ Za podjetja je postala zaradi soodvisnosti in povezanosti **globalne** ekonomije nujna za preživetje.

☞ Spremljamo dogajanje na ljubljanski borzi, ki je zaradi **globalne** finančne krize leden začela v rdečih številkah.

☞ Brez podkupnine, ne mešajte s provizijo, v **globalnem** kapitalizmu ni več resnega posla.

☞ Zaradi pospeševanja **globalne** konkurenčnosti slovenskega gospodarstva pa bi bilo treba strateško usmeritev uresničiti brez dodatnih fiskalnih bremen.

PRIDEVNIK + in + PRIDEVNIK

globalen in [transnacionalen, lokalni, regionalni, univerzalni, pavšalni]

globalen in [meddržaven, nacionalni, celovit, radikalen, operativni]

globalen in [regijski, dolgoročni, evropski, konkurenčni, trajni]

globalen in [krajveni, domači, mednarodni, splošni]

☞ Medtem ko lepota stalno tekmuje z novimi trendi zahodne oblačilne mode, je oblačeni vedno vpeto v zapletena pogajanja med lokalnimi, transnacionalnimi in **globalnimi** normami lepote, spola in spolnosti.

☞ Nič drugega kot vrnitev domov, združitve lokalnega in **globalnega**, avstralsko nacionalnega in svetovljansko filmskega.

☞ Njih se tičejo "nacionalna, regionalna in **globalna** varnost", ki jo zagotavlja vojaški stroj.

☞ Proces izobraževanja in usposabljanja zdravnikov postaja vedno bolj celovit, univerzalen in **globalen**.

☞ Poleg te » **globalne** in pavšalne odškodnine« je bila jugoslovanska stran dolžna Italiji s »seznama za prosto razpolaganje« vrniti še 179 nepremičnin, nekoč odvzetih italijanskim upravičencem.

delniska ekonomija

fenomen izziv

kapitalizem klepet

konkurenčnost

korporacija kriza

lokalni

nomad ogrevanje

otoplitev pravičnost raven

razsežnost recesija

segrevanje

spremenljivka

transnacionalni trend

trg vas vodič zasedanje

vzorci

PRIDEVNIK + SAMOSTALNIK

PRIDEVNIK + in + PRIDEVNIK

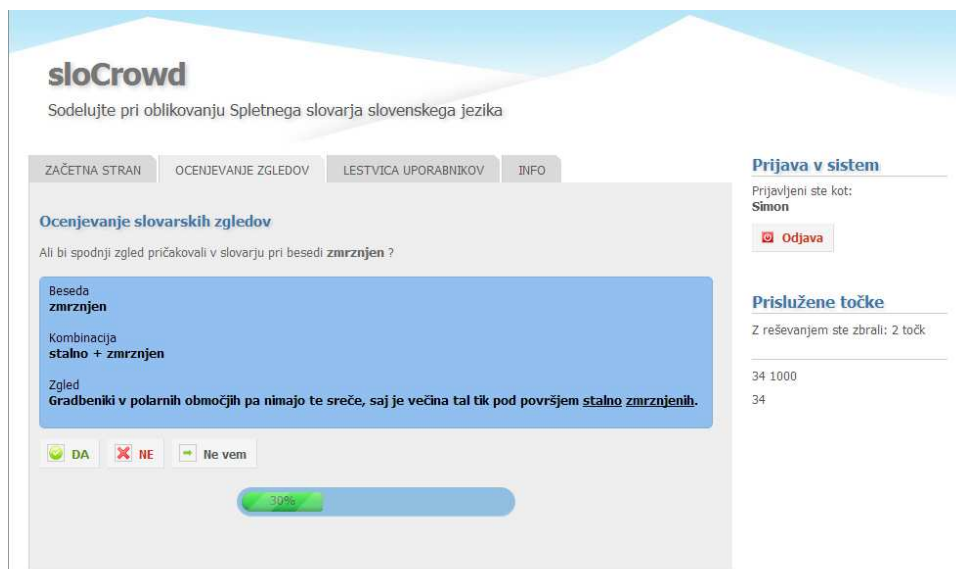
NEUSTREZNE KOLOKACIJE IN ZGLEDI BODO ODSTRANJENI V NASLEDNJI (ORANŽNI) FAZI

© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 28: RDEČA FAZA – AVTOMATSKO IZDELANO GESLO

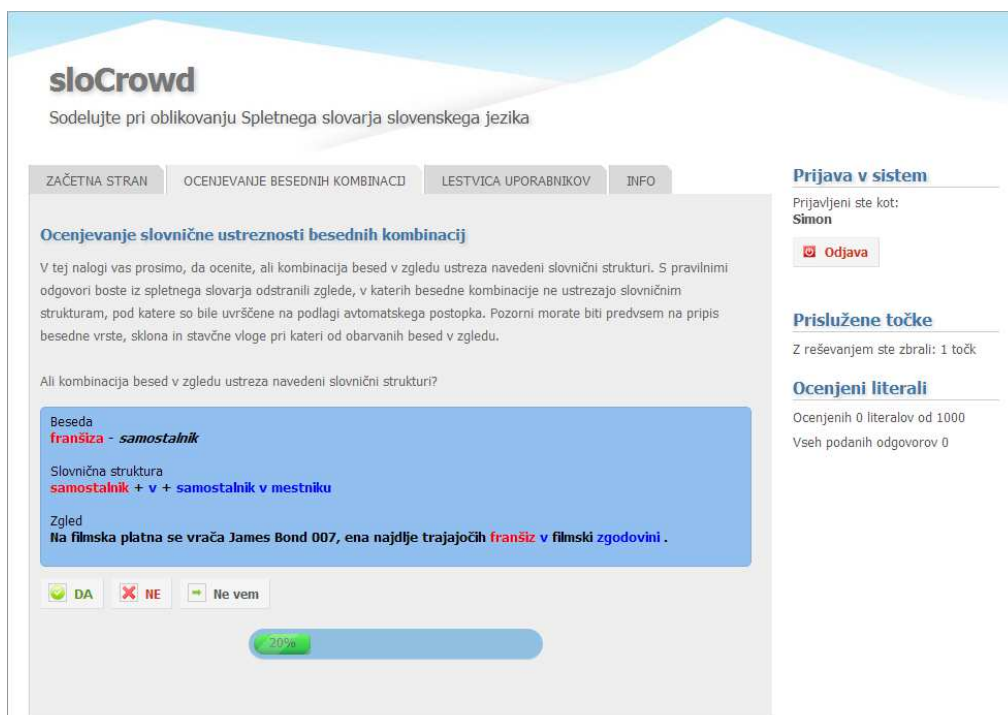
2. ORANŽNA FAZA – PREČIŠČENO GESLO

V drugi, oranžni fazi izdelave slovarskega gesla poteka izločanje nerelevantnih informacij, ki so bile v zvezi s posamezno korpusno lemo, tj. iztočnico v Slovarju, izluščene v postopku avtomatizacije. Ta faza je v celoti predvidena za izrabo moči množic, ki postopek izdelave slovarskega gesla pocení in hkrati časovno precej skrajša, saj gre za opravila, ki ne zahtevajo tehtnejših leksikografskih premislekov, zato z njimi ni smiselno obremenjevati dragocenega leksikografskega časa. Izraba moči množic je predvidena za ekipo zunanjih sodelavcev, ki lahko v proces sodelovanja pri izdelavi slovarja poljubno vstopajo in izstopajo, ker daljše učenje za to delo ni potrebno, njihovo delo pa se plačuje glede na posamezno odločitev, ki jo posredujejo prek za to izdelanega spletnega servisa (gl. poglavje 4.1 – Moč množic in kazalnik K1 v poglavju 6). V izhodišču gre pri izrabi moči množic za dve vrsti opravil:



SLIKA 29: IZKORIŠČANJE MOČI MNOŽIC – SLOVARSKI ZGLEDI

navodil za proces izkoriščanja množic, ki med drugim predvideva navedbo primerov, ki predstavljajo tipične primere in pričakovane odločitve. Postopek je bil preverjen na manjšem številu prostovoljcev v okviru spletnega servisa, ki se v izhodišču uporablja za izboljšavo slovenskega WordNeta.⁴¹ Primer naloge in vprašanja prikazuje Slika 29.



SLIKA 30: IZKORIŠČANJE MOČI MNOŽIC – KOLOKACIJE

(a) Ocenjevanje slovarskih zgledov, kjer prostovoljci ocenjujejo, ali je v zgledu navedena kombinacija, ki predstavlja avtomatsko izluščeno kolokacijo, po njihovem mnenju primerna za vključitev v slovarsko geslo. V procesu izdelave spletnega slovarja je sicer predvidena izdelava

(b) Ocenjevanje vzorcev, kjer prostovoljci ocenjujejo ustreznost avtomatsko izluščenih skladenjskih struktur, ki predstavljajo osnovo za kolokacijo, prikazano v zgledu. Na podlagi pravih odgovorov se iz slovarja odstranijo zgledi,

⁴¹ Dostopno na: <http://nl.ijs.si/slowcrowd/index.php?project=slowcrowdmain>.

v katerih kolokacije ne ustrezajo slovničnim strukturam, pod katere so bile uvrščene na podlagi avtomatskega postopka. Nalogo, ki je bila poskusno preverjena na manjšem številu prostovoljcev, prikazuje Slika 30.

Zaključena oranžna faza slovarskega gesla predvideva avtomatsko geslo, ki smo mu s pomočjo izrabe moči množic odstranili nerelevantne ali napačne informacije na ravni vzorcev, pripadajočih kolokacij ter zgledov. Tako prečiščeno geslo je pripravljeno za nadaljnjo tj. leksikografsko obdelavo, hkrati pa se samodejno posodobijo tudi podatki v zavihku *Vizualizacija*.

The screenshot displays the 'SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA' interface. On the left is a sidebar with navigation options: 'Pomen', 'Oblike', 'Sinonimi', 'Govor', 'Vizualizacija', and 'Statistika'. The main content area shows the entry for 'globalen' (pridevnik) with a frequency of 3000. It lists various forms of the word (globalno, globalna, globalni) and provides example sentences. A 'Vzorci' (Patterns) box on the right lists related terms like 'ekonomija', 'kapitalizem', 'konkurenčnost', etc. The bottom of the page includes a copyright notice: '© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA'.

SLIKA 31: ORANŽNA FAZA – PREČIŠČENO GESLO

Program za izkoriščanje moči množic uporabimo v procesu izdelave gesel še enkrat, in sicer po tem, ko leksikograf besedo pomensko razčleni. Skupina leksikografsko že izvežbanih sodelavcev, ki je sodelovala pri izrabi moči množic, razvrsti posamezne zgleds in kolokacije pod ustrezne pomene. Temu namenu služi naloga, ki jo prikazuje Slika 31:

V kateri pomen besede *aranžma* spada navedeni zgled?

Beseda:

aranžma

Kombinacija:

aranžma + z + banka

Zgled:

Vlada je naložila Srdu, da v okviru aranžmajev z bankami glede na sprostitev hipotek zagotovi sredstva za financiranje proizvodnje v postopku prisilne poravnave.

- ☐ Pomen 1 – turistična ponudba
- ☐ Pomen 2 – ureditev cvetja
- ☐ Pomen 3 – glasbeno delo
- ☒ Pomen 4 – dogovor, sporazum
- ☐ Nič od tega

SLIKA 32: IZKORIŠČANJE MOČI MNOŽIC - POMENI

3. RUMENA FAZA – POMENSKO RAZČLENJENO GESLO

Tretja – rumena faza – je v procesu stopenjske izdelave slovarskega gesla z vidika količine in zanesljivosti slovarskih podatkov najpomembnejša, saj predvideva poseg profesionalnega leksikografa, ki opravi postopek ključne leksikografske redakcije gesla, kamor sodi pomenska členitev besede v iztočnici, oblikovanje pomenskega menija ter identifikacijo stalnih zvez in frazeoloških enot. Poleg tega oblikuje razlage za posamezne (pod)pomene,⁴² stalne zveze in frazeološke enote, pripiše različne tipe oznak: slovnične, diskurzne, stilistične oz. konotacijske in kontekstualne ter dodaja pragmatične komentarje. Pripis področne oznake v zvezi s posameznim pomenom besede ali stalne zveze predstavlja signal za terminološki pregled v naslednji fazi leksikografske obdelave gesla.

Z vidika izvedbe posameznih faz in organizacije slovarskega dela je pomembna predhodna razdelitev gesel v težavnostne stopnje, kjer predvidemo, da bodo pri eno- ali dvopomenskih geslih leksikografovi posegi omejeni na pripis kratkih pomenskih indikatorjev (tj. za en sam pomen, morda še za podpomen), na identifikacijo stalnih zvez in pripis oznak za ugotovljeno tipično področje njene rabe. Pri nekoliko zahtevnejših geslih predvidevamo, da bo leksikograf namenil več pozornosti oblikovanju pomenskega menija ter identifikaciji stalnih zvez in potencialnih frazeoloških enot. Gesla, ki jih predhodno uvrstimo v

⁴² Predhodno oz. vzporedno tudi pomenskih shem, ki vsebujejo računalniško procesljive podatke za nadaljnje analize slovenskega besedišča, ki pa se v slovar ne prenašajo.

najvišjo zahtevnostno stopnjo,⁴³ kamor na podlagi izkušenj pri izdelavi Leksikalne baze sodijo zlasti glagoli, predvidevajo obsežnejši in zahtevnejši angažma leksikografa, vključno z dodatno analizo korpusnih podatkov, prepoznavanjem kompleksnejših slovničnih in skladenjskih vzorcev, podrobno analizo pomenskih lastnosti besede, stalnih zvez in frazeoloških enot ter njihovega besedilnega okolja.

V rumeni fazi izdelave gesla se v skladu s pomensko redakcijo dopolnijo oz. posodobijo tudi podatki o sinonimih, ki se za razliko od avtomatsko izdelanega gesla zdaj razporedijo glede na posamezne (pod)pomene besede v iztočnici.

Ključna razlika med trenutnimi na korpusni analizi temelječimi leksikografskimi metodami in predlagano je v tem, da leksikografu ni več treba veliko gledati v korpus, razen pri preverjanju specifičnih informacij oziroma pri podrobnejši analizi frazeoloških enot, stalnih zvez ipd.

SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

globalen pridevnik **P 3000**

1. svetovni; mednarodni

- 1.1 splošno veljaven; razširjen
- 2. zemeljski; planetarni
- 3. ki zadeva celoto; celostni

DODANI POMENSKI MENI, RAZČLENITEV POMENOV, STALNE ZVEZE, RAZLAGE IN OZNAKE

1. svetovni; mednarodni

globalni procesi, zlasti gospodarski in politični, zajemajo ves svet

- 40 V New Yorku naj bi državniki in podjetniki razpravljali o **globalni** varnosti.
- 40 Tudi največje svetovne firme, ki danes obvladujejo **globalni** trg, so se razvile iz malih podjetij in obratov.
- 40 Motorola je zaradi **globalne** recesije v visokotehnoloških gospodarskih panogah lani odpustila 43.400 zaposlenih.

1.1 splošno veljaven; razširjen

če postanejo neke dejavnosti ali lastnosti globalne, jih upošteva vedno več ljudi ali držav po svetu

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

2. zemeljski; planetarni

globalne spremembe v okolju vplivajo na celoten zemeljski planet

- 40 Eden najbolj preprostih in praktično izvedljivih načinov za zmanjšanje **globalnega** segrevanja je sajenje novih dreves.
- 40 Krčenje ledenikov in spremembe v pokrajini sta zelo očitni in lahko razumljivi posledici **globalnega** ogrevanja.
- 40 Vodik bo tudi občutno skrčil emisije ogljikovega dioksida in učinke **globalne** otoplitve.

3. ki zadeva celoto; celostni

- 40 S pomočjo plakata postanejo učencem misli bolj jasne in končno odkrijejo **globalni** pomen besedila.
- 40 Ta študija zlasti opozarja na vrsto vprašanj, ki se zastavljajo glede oblikovanja **globalne** nacionalne kulturne politike.
- 40 Gre za **globalni** pristop do človeka, ki mu pomaga, da vzame življenje v svoje roke in ga znova progamira, kot si ga je sam zamislil.

stalne zveze

- globalno segrevanje/ogrevanje
- globalna vas
- globalni nomad
- globalna spremenljivka
- globalno potrdilo o lastništvu

ekonomija izziv

gospodarstvo
kapitalizem
komunikacija
konkurenčnost
korporacija
kriza
lokalen mreža
ogrevanje otoplitev
politika problem raven
razsežnost recesija
regionalen segrevanje
sklad sprememba
temperatura trend trg
vojna

ekologija

© Trojina 2013 - SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA

SLIKA 33: RUMENA FAZA – POMENSKO RAZČLENJENO GESLO

⁴³ Podrobni izračuni so pokazali, da je kompleksnejših večpomenskih gesel v slovarju večjega obsega približno 25 %.

4. MODRA FAZA – GESLO Z DODANIMI »ZUNAJBAZNIMI« PODATKI

Četrta – modra – faza stopenjske izdelave slovarskega gesla je namenjena dodajanju t. i. zunajbaznih podatkov in dopolnjevanju na spletnem portalu že prikazanih podatkov, kjer se predvideva specializirano znanje jezikoslovcev in strokovnjakov drugih področij, zlasti terminologov.

Med tipi podatkov, ki se besedi oz. njenim pomenom dodajajo na novo, so informacije o izvoru besede oz. etimološke posebnosti, za katere predvidevamo, da utegnejo biti za uporabnika zanimive. Te zagotovi strokovnjak, ki je specializiran za to področje. Poleg tega so v tej fazi dodane informacije o izgovoru, in sicer studijski posnetki ter zapis izgovora posamezne iztočnice, skupaj z naglasom – te podatke zagotovi strokovnjak za fonetiko. Poleg tega bodo dopolnjene informacije o besednih oblikah in oblikoslovnih variantah, če jih ni bilo mogoče avtomatsko zgenerirati iz obstoječe leksikonske baze, skupaj z naglasom pri posameznih oblikah. Modra faza je namenjena tudi dopolnitvi normativnih podatkov in osvetlitvi vprašanj standardizacije. Skladno s temi dopolnitvami se širi tudi baza pravopisnih in stilističnih vprašanj, na kateri temelji spletni servis Slogovni priročnik. Podatke za ta namen zagotavlja strokovnjak oz. skupina strokovnjakov za standardizacijo. Pomemben del modre faze v stopenjski izdelavi slovarskega gesla predstavlja tudi pregled pomenov, podpomenov in stalnih zvez, vezanih pretežno na določeno specializirano oz. terminološko področje. V ta namen je organizirana mreža strokovnjakov in vzpostavljena spletna platforma, ki omogoča spremljanje in usklajevanje dela.

5. ZELENA FAZA – KONČANO GESLO

V peti zaključni – zeleni – fazi stopenjske izdelave se nahajajo gesla, ki so v celoti pregledana s strani leksikografa-redaktorja in po potrebi dopolnjena z vidika vseh predvidenih slovarskih podatkov. Zadnja faza je torej namenjena predvsem pregledu konsistentnosti gesla in ugotavljanju skladnosti podatkov z izpričanim realnim jezikovnim stanjem. To sicer pomeni, da je v tej fazi geslo mogoče vrniti tudi v katero od predhodnih faz, če redaktor ugotovi večje spremembe v pomenski členitvi besede, pomanjkljive podatke v segmentu pripisovanja terminoloških podatkov ipd.

3 LEKSIKALNA BAZA ZA SLOVENŠČINO KOT VIR PODATKOV ZA *SLOVAR SODOBNEGA SLOVENSKEGA JEZIKA*

3.1 MOTIV IN OZADJE

Z namenom zapolniti vrzel na področju sodobnega slovenskega leksikalno-gramatičnega opisa in digitalnih skladiščno-semantičnih virov je bila ob upoštevanju potreb različnih končnih uporabnikov jezikovnih virov in opisov pri projektu Sporazumevanje v slovenskem jeziku (SSJ) med letoma 2008 in 2012 zasnovana Leksikalna kot baza za slovenščino. Na obsegu 2.500 gesel so bili za vse segmente določeni tipi leksikalnogramatičnih podatkov, preverjeni so bili leksikografski postopki ter postopki avtomatičnega pridobivanja in urejanja podatkov, ki bi predstavljali optimalen vir za izdelavo Sodobnega slovarja slovenskega jezika in hkrati zadovoljevali jezikovnotehnološke potrebe po procesljivih leksikalnih podatkih.

3.2 NAMEN IN UPORABNIK

Leksikalna baza je bila že v izhodišču zasnovana za dva tipa uporabnikov: kot baza slovarskih podatkov, namenjenih človeškemu uporabniku, in kot baza, v kateri so podatki strukturirani in kodirani na način, ki omogoča nadaljnje računalniško procesiranje.

Z vidika **človeškega** oz. **splošnega uporabnika** je v leksikalno bazo zajeto jedrno besedišče sodobne slovenščine (tj. v obsegu zadnjih 20 let), za bolj specializirane uporabnike, npr. šolarje in učence slovenščine kot tujega jezika, pa je poudarek na besedišču osnovno- in srednješolskih učbenikov. Za potrebe sodobnega slovarja je v nadaljevanju v leksikalno bazo smiselno zajeti in analizirati besedišče, za katerega se ugotovi hitrejšo spreminjanje, npr. na prehodu iz terminološke v splošno rabo, in besedišče, po katerem se v jezikovni skupnosti ugotavlja večje »povpraševanje«, npr. novo nastale besede, pomeni in prevzete besede, ki se v jeziku šele uveljavljajo.

Sodobne leksikografske prakse in digitalno okolje narekujeta izdelavo sodobnih slovarskih priročnikov na podlagi baz, kjer so vsi v njej strukturirani podatki na voljo tudi **jezikovnotehnološki skupnosti**. Leksikalna baza je zato izdelana tudi kot računalniško berljiv jezikovni vir, v katerem je vsak besedni pomen identifikacijsko povezan z vrsto specifičnih leksikalnih in skladiščnih podatkov v svoji besedilni okolici. Tako strukturirane podatke je mogoče povezati z drugimi bazami podatkov in jih uporabiti pri avtomatičnem luščenju informacij iz besedil, oblikovanju sistemov odgovorov na vprašanja (npr. pri izdelavi jezikovnih aplikacij za starostnike ali hendikepirane), v avtomatskih prevajalnih sistemih in aplikacijah, ki temeljijo na bazah procesljivih jezikovnih podatkov. Uporabiti jih je mogoče za avtomatsko označevanje slovenskih besedil na oblikoskladišnji, skladišnji in pomenski ravni ter za izboljšanje orodij, kot je npr. skladišnji razčlenjevalnik in označevalnik za slovenščino. Taka zasnova omogoča tudi avtomatično pridobivanje kolokacijskih podatkov iz elektronskih besedilnih korpusov in z njimi povezanih korpusnih zgledov, kar izredno pohitri postopek izdelave slovarja in omogoča izdelavo v različnih redakcijskih fazah, ki so uporabnikom na voljo na spletu.

3.3 RAZMERJE MED LEKSIKALNO BAZO IN SLOVARJEM

Leksikalna podatkovna baza je na splošno razumljena kot vir leksikalno-gramatičnih podatkov v elektronski obliki in je vedno specifična glede na namen, za katerega je izdelana, hkrati pa je vedno sekundarni jezikovni vir, kar pomeni, da so podatki v njej hierarhizirani in kategorizirani, podvrženi torej določenim jezikoslovnim opisom in interpretacijam. Običajno vsebuje leksikalna baza več informacij, kot jih dejansko vključujejo iz nje izhajajoči izdelki.

Slovar ima drugačno funkcijo. Navadno predstavlja izbor le določenega tipa jezikovnih podatkov in je – zlasti če govorimo o knjižnih izdajah – omejen tako v količini zajetega besedišča kot v časovnem obdobju, ki mu to besedišče pripada. Enojezični slovar, če ni primarno namenjen učenju jezika, je vedno tudi dokument o bogatosti besedišča določene jezikovne skupnosti in ima družbeno povezovalno vlogo. Kot tak slovar nikoli ni le inventar besed, njihovih pomenov in sintagmatskih vzorcev, pač pa se od njega pričakuje, da bo podal tudi sodbo o tem, kaj je v različnih kontekstih sprejemljivo in kaj ne, leksikalna baza pa naj bi mu omogočala, da take opredelitve podpre z ustreznimi podatki. Pomembno je namreč, kot pravi Hanks (2009: 8), da sodbe v slovarju temeljijo na empirični analizi dejanske rabe v besedilih, zato je interakcija med jezikoslovno ustrezno zasnovano leksikalno bazo in slovarjem kot družbenim artefaktom neizogibna.

Ena bistvenih razlik med elektronsko bazo kot skladiščem podatkov in končnim slovarskim izdelkom, ki je bil do nedavnega še knjiga, temelji na dihotomiji elektronskost : knjižnost in na dojemanju družbene vloge jezika, njegove prestižnosti in znanstvenosti. Leksikalni podatkovni bazi se namreč običajno pripisuje atribut računalniške robustnosti, ki ne upošteva (dovolj) jezikoslovnih argumentov, zato se ji očita zgolj jezikovnotehnološka uporabnost. V obdobju zadnjih 20 let, ko so v evropskem jezikovnem prostoru hkrati nastajale elektronske leksikalne baze in slovarji (elektronski in knjižni), sta se tako leksikalna baza kot slovar razvijala in spreminjala. Spremembe so se dogajale tako pod vplivom razvoja jezikoslovne znanosti kot komunikacijskih potreb in možnosti uporabnikov. Kot smo lahko videli iz pregleda obstoječih leksikalnih baz evropskih jezikov, so se projekti oblikovanja obsežnih, strukturiranih baz jezikovnih podatkov prek ustrezne vizualizacije teh podatkov prelevili v eno samo kompleksno enoto, v jezikovno-informacijske spletne portale, ki presegajo in nadgrajujejo še do nedavna veljavne pristojnosti slovarja. Tudi nadaljnje smernice v sodobnem leksikografskem opisu in razvoju leksikografije, zlasti razvitejših evropskih jezikov, kažejo, da se razlika med slovarjem kot končnim izdelkom in elektronsko zbirko jezikovnih podatkov zabrisuje. Elektronske zbirke leksikalnih podatkov tako združujejo najboljše obeh do sedaj ločenih polov: jezikoslovno znanje in realen opis jezika na način, ki je računalniško berljiv in večkrat uporabljen za različne namene.

3.4 ZGRADBA IN ORGANIZACIJA LEKSIKALNO-GRAMATIČNIH PODATKOV

Leksikalna baza za slovenščino je oblikovana kot mreža medsebojno povezanih leksikalno-gramatičnih podatkov, ki so organizirani v šest nivojev. Notranja hierarhična ureditev za razliko od sintaktično orientiranih slovarjev (Hanks 2013), kamor sodi tudi SSKJ, temelji na **semantičnem izhodišču**, kar

pomeni, da so vsi podatki, ki jih predvidevajo posamezni nivoji, razvrščeni pod vse registrirane pomene obravnavane besede.

I. LEMA	• iztočnica • besedna vrsta	grmeti <i>glagol</i>	
II. POMEN	• indikator • pomenska shema • oznaka	1 oddajati glasen zvok 1.1 o nevihti ko grmi, se sliši glasen bobneč zvok, navadno zato, ker se bliža nevihta v 3. osebi	2 glasno govoriti če ČLOVEK grmi, zelo glasno in odločno govori, navadno zato, ker se s čim ne strinja
III. SKLADNJA	• vzorec • struktura	----- ----- gbz Inf-GBZ rbz Inf-GBZ	▪ kdo grmi ▪ kdo grmi s česa ▪ kdo grmi nad kom/čim rbz GBZ
IV. KOLOKACIJE	• kolokacija	[začeti] grmeti [zunaj, močno] grmi	[glasno] grmeti grmeti z [odra]
V. ZGLEDI	• zgled	<i>Bili so ravno v restavraciji, ko je začelo grmeti.</i>	<i>Grmel je z govorniškega odra, dokler mikrofoni ni nenadoma umolknili.</i>
	• stalna besedna zveza		
VI. FRAZEOLOGIJA	• frazeološka enota • indikator	Če malega travna grmi, slane se kmet več ne boji. <i>pregovor</i>	

SLIKA 34: ORGANIZACIJA LEKSIKALNO-GRAMATIČNIH PODATKOV V LEKSIKALNI BAZI ZA SLOVENŠČINO

I. Hierarhično najvišja je **lema**, tj. **iztočnica v osnovni obliki**, ki zastopa vse **pomene** in **podpomene**, **stalne besedne zveze** in **frazeološke enote**, v katerih se obravnavana iztočnica pojavlja kot sestavni element. Na iztočnico so v bazi pripeti še: podatek o besedni vrsti in podatek o frekvenci v korpusu. Besednovrstna opredelitev je zaradi možnosti nadaljnjega procesiranja podatkov skladna s postopki lematizacije in oblikoslovnimi oznakami, uporabljenimi pri oblikoskladenjskem označevanju korpusa Gigafida.

Na istem nivoju kot iztočnica se v leksikalni bazi nahajajo **frazeološke enote**, kar pomeni, da jih ne podrejam posameznim (pod)pomenom, ampak so z iztočnico povezane zgolj oblikovno. Bistvena elementa frazeološkega razdelka sta frazeološka enota v osnovni obliki ali več pojavnih oblikah, pomenski indikator in/ali definicija ter korpusni zgledi.

Stalne besedne zveze so v leksikalni bazi obravnavane kot samostojne leksikalne enote, in sicer na ravni gesla v primeru, da so od pomenov besede v iztočnici neodvisne, in na ravni posameznih pomenov, če obstaja med besedno zvezo in pomenom besede v iztočnici razvidna pomenska povezava. Tako kot besede imajo tudi stalne zveze razlago pomena in/ali kratki pomenski indikator, navedeno področje rabe, zlasti v primeru, ko pripadajo določenemu terminološkemu področju, tipično besedilno okolje v obliki kolokacij, če se to izkazuje v realnih besedilih, in korpusne zglede.

II. Na **pomenskem nivoju** so posamezni (pod)pomeni opredeljeni najprej s kratkimi pomenskimi indikatorji, ki oblikujejo **pomenski meni** besede za lažjo navigacijo po geslu. Drugi del pomenske informacije predstavlja **pomenska shema**, kjer je zabeležena udeleženska zgradba posameznega pomena v najbolj tipičnem stavčnem vzorcu. Ta naj bi pokazal, kateri so obvezni udeleženci in okoliščine – oboji zapisani z velikimi črkami kot semantični tipi zaradi možnosti avtomatskega pridobivanja in nadaljnjega procesiranja – za konkretni pomen. Pomenska shema je oblikovana kot stavčna razlaga, ki vsebuje pomenski scenarij (zlasti okoliščine in pragmatiko) konkretnega pomena in predstavlja izhodišče za oblikovanje slovarske definicije. Posamezni (pod)pomen ima lahko tudi stilno, registrsko, področno ali kako drugo pojasnilo, ki ga v leksikalni bazi beležimo v elementu <oznaka>, s pripisom ustreznega atributa, kot npr. področje, kontekst, register, slovnica itd.

III. Eksplicitno je skladiščnim podatkom namenjen **skladiščni nivo**. Na tem nivoju so za vsak registrirani (pod)pomen besede v iztočnici beleženi besednozvezni in stavčni vzorci v obliki formaliziranih zapisov, ki služijo avtomatičnemu luščenju podatkov iz korpusov in organizaciji kolokacij.

IV. Sledi **kolokacijski nivo**, na katerem so za vsak konkretni (pod)pomen besede registrirani besednozvezni in stavčni vzorci potrjeni s tipičnimi leksikalnimi zapolnitvami – kolokacijami.

V. Vloga **korpusnih zgledov**, ki so navedeni pri posameznih pomenih, stalnih zvezah in frazeoloških enotah, je ponazoriti in potrditi vse predhodne informacije ter hkrati pokazati obnašanje besede oz. njenega pomena v njenem najbolj naravnem in tipičnem besedilnem okolju.

Leksikalna baza za slovenščino je ob zaključku projekta SSJ obsegala 2.500 gesel oz. nekaj več kot 10.900 leksikalnih enot, ki vsebujejo pomensko razlago in druge podatke po posameznih nivojih. Jezikovnim tehnologom in jezikoslovcem, zlasti leksikografom, je na voljo v formatu XML, skupaj z DTD shemo, ki definira formalno strukturo leksikalne baze in omogoča prilagoditev informacij za različne tipe slovarskih priročnikov. Informacije, ki so strukturirane v Leksikalni bazi za slovenščino, so poskusno – kot ena od možnih spletnih vizualizacij – na ogled tudi v testni verziji na spletu.⁴⁴

Lastnik leksikalne baze za slovenščino je Ministrstvo za izobraževanje, znanost in šport Republike Slovenije. Pogodba med Ministrstvom in izvajalci projekta določa, da se za prenos baz podatkov na tretje osebe in označevanje avtorskih del uporabi licenca »priznanje avtorstva« + »nekomercialno« + »deljenje pod istimi pogoji«, ki dovoli uporabnikom avtorsko delo in njegove predelave reproducirati, distribuirati, dajati v najem, priobčiti javnosti in predelovati pod pogojem, da se navede avtorja, da ne gre za komercialno uporabo in da se izvirna dela ali predelave naprej širijo pod istimi pogoji.

3.5 SHEMA DTD

Za potrebe leksikalne baze je bila izdelana DTD shema, ki je prilagojena delu v slovarskih urejevalnikih, kot so na primer DPS (Dictionary Writing System) francoskega podjetja IDM,⁴⁵ iLex danskega podjetja EMP,⁴⁶ TLex podjetja TshwaneDJe iz Južne Afrike,⁴⁷ Lingvo.Pro ruskega podjetja ABBYY⁴⁸ in podobnih.

⁴⁴ Dostopno na: <<http://www.slovenscina.eu/spletni-slovar>>.

⁴⁵ DPS Dictionary Writing System: http://www.idm.fr/products/dictionary_writing_system_dps/27/

⁴⁶ iLex: <http://www.emp.dk/ilexweb/index.jsp?toc=100000025,1>

⁴⁷ TLex: <http://tshwanedje.com/tshwanelex/>

Schema omogoča prilagoditev informacij v smislu dodajanja in spreminjanja kategorij podatkov in hierarhičnih odnosov za različne tipe slovarskih priročnikov.

Spodnja preglednica opisuje vsebino posameznih elementov XML sheme in zapis elementa. V vsebinskem delu je na kratko opisan tip leksikalnega, leksikalno-gramatičnega, stilnega, slovničnega in drugih podatkov, namen opisa in razmerje glede na druge podrejene ali nadrejene podatke v shemi

Vsebina LBS	DTD element
Vsebuje celotni geselski članek z elementoma glava in geslo .	<clanek></clanek>
Vsebuje elemente oblika in zaglavje .	<glava></glava>
Vsebuje elemente zapis , korpus ter iztočnica .	<oblika></oblika>
Znotraj elementa oblika: vsebuje zapis iztočnice za namene iskanja po bazi in nekatere druge (interne) podatke.	<zapis></zapis>
Znotraj elementa oblika: vsebuje podatke o frekvenci v korpusu Gigafida, ki se navezujejo na iztočnico v osnovni obliki.	<korpus></korpus>
Znotraj elementa oblika: vsebuje zapis izočnice obliki leme.	<iztocnica></iztocnica>
Vsebuje elemente: besedna vrsta in oznaka .	<zaglavje></zaglavje>
Znotraj elementa zaglavje: vsebuje zapis besedne vrste, ki ustreza besednovrstni oznaki leme v korpusu Gigafida	<besvrs></besvrs>
Znotraj elementa zaglavje: vsebuje opredelitev leme glede na področje rabe, konotacijskih, registrskih in slovničnih posebnosti.	<oznaka tip="attribute"></oznaka>
Vsebuje elemente, ki so določeni kot leksikalne enote . Sem sodijo: pomeni s podpomeni , stalne zveze in frazeološke enote .	<geslo>

⁴⁸ Lingvo.Pro: <http://lingvopro.abbyyonline.com/en/Home/Authors>

Vsebuje elemente, ki opredeljujejo posamezno leksikalno enoto . Obvezno vsebuje elementa indikator in pomenska shema .	<pomen></pomen>
Znotraj elementov pomen, podpomen, stalne zveze in frazeološke enote: vsebuje kratek pomenski indikator, katerega namen je ustvariti asociacijo o pomenskem dosegu pomena in oblikovati pomenski meni.	<indikator></indikator>
Znotraj elementa indikator: vsebuje pragmatično pojasnilo v zvezi s pomenom besede, stalne zveze ali frazeološke enote.	<pr></pr>
Znotraj elementov pomen ali podpomen: vsebuje ustaljeno obliko besede, značilno za konkretni pomen.	<ustaljena_oblika></ustaljena_oblika>
Znotraj elementov pomen ali podpomen: vsebuje opredelitev pomena glede na področje rabe, konotacijskih, registrskih in slovničnih posebnosti.	<oznaka tip="attribute"></oznaka>
Znotraj elementov pomen ali podpomen: vsebuje argumentno zgradbo konkretnega pomena, zapisano v obliki stavčne definicije s SEMANTIČNIMI TIPI kot abstraktnimi zastopniki tipičnih zapolnitev na posameznem vezljivostnem mestu.	<pomenska_shema></pomenska_shema>
Znotraj elementov pomen ali podpomen, stalne zveze in frazeološke enote: vsebuje definicijo, ki na uporabniku prijazen način ubesedi osnovne – na podlagi korpusnih rab – ugotovljene pomenske tendence konkretnega pomena.	<definicija1></definicija1> <definicija2></definicija2>
Vsebuje element skladenjske strukture s pripadajočimi kolokacijami, vzorci in korpusnimi zgledi.	<skladenjske_skupine></skladenjske_skupine>
Vsebuje obvezno vsaj eno skladenjsko strukturo in korpusne zglede . V večini primerov tudi kolokacije in vzorke .	<skladenjska_struktura></skladenjska_struktura>
Znotraj elementa skladenjska struktura: vsebuje zapis besedne zveze v obliki besedne vrste strukturnega elementa in podatka o ustreznem sklonu. Besedna vrsta, ki zastopa iztočnico v strukturi, je zapisana z	<struktura></struktura>

velikimi črkami.	
Znotraj elementa struktura: vsebuje opozorilo o slovnični omejitvi elementa v strukturi, npr. na določeno število, glagolsko obliko ipd.	<r></r>
Znotraj elementa skladenjska struktura: pri glagolskih iztočnicah vsebuje lahko več zaporednih realizacij prototipičnega vezljivostnega vzorca, izraženega v pomenski shemi.	<vzorec></vzorec>
Vsebuje vsaj eno kolokacijo in pripadajoče zgled. V večini primerov tudi razširjene kolokacije.	<kolokacije></kolokacije>
Znotraj elementa kolokacije: vsebuje zapis iztočnice in niza pomensko ali oblikovno sorodnih kolokatorjev v elementu <k></k>.	<kolokacija><k></k></kolokacija>
Znotraj elementa kolokacije: vsebuje zapis iztočnice in niza kolokatorjev, ki je lahko razširjen z nizom lastnih kolokatorjev v elementu <k></k>.	<r_kolokacija></r_kolokacija>
Znotraj elementov skladenjske skupine, skladenjske zveze, stalne zveze in frazeološke enote. Vsebuje vsaj en obvezenzgled.	<zgledi></zgledi>
Znotraj elementa zgledi: lahko vsebuje več zaporednih korpusnih zgledov, ki potrjujejo predhodne kolokacije, razširjene kolokacije in vzorce. Znotraj zgleda je v elementu <i></i> v krepkem tisku izpisana iztočnica.	<zgled><i></i></zgled>
Vsebuje vsaj en element skladenjska zveza s pripadajočimi zvezami, kolokacijami, vzorci in korpusnimi zgledi.	<skladenjske_zveze></skladenjske_zveze>
Vsebuje obvezno vsaj eno zvezo in pripadajoče zgled . Redkeje tudi kolokacije, in vzorce.	<skladenjska_zveza></skladenjska_zveza>
Znotraj elementa skladenjska zveza: vsebuje zapis pomensko prozornih, strukturno ustaljenih delčkov jezika, pogosto s semantično ali/in oblikovno predvidljivim prostim mestom v elementu <k></k>.	<zveza><k></k></zveza>
Znotraj elementa pomen: vsebuje elemente, ki	<podpomen></podpomen>

opredeljujejo posamezno leksikalno enoto (razen pomena). Obvezno vsebuje elementa indikator in pomenska shema .	
Vsebuje vsaj en element stalna zveza .	<stalne_zveze></stalne_zveze>
Vsebuje obvezno element zveza , indikator , struktura in inzgledi . V nekaterih primerih tudi kolokacije in razširjene kolokacije.	<stalna_zveza></stalna_zveza>
Znotraj elementa stalna zveza: vsebuje zapis stalne zveze, vključno z variantami posameznih elementov, ločenih s poševnico. Različne sklonske oblike stalne zveze kot celote ali različne oblike njene rabe so navedene v zaporednih elementih <zveza></zveza> .	<zveza></zveza>
Vsebuje vsaj en element frazeološka enota .	<frazeloske_zveze></frazeloske_zveze>
Vsebuje obvezno element enota in indikator . Pogosto tudi kolokacije, zglede in oznake.	<frazeloska_enota></frazeloska_enota>
Znotraj elementa frazeološka enota: vsebuje zapis frazeološke enote, vključno z variantami posameznih elementov, ločenih s poševnico. Različne sklonske oblike frazeološke enote kot celote ali različne oblike njene rabe so navedene v zaporednih elementih <enota></enota> .	<enota></enota>

TABELA 2: ELEMENTI V FORMALNI SHEMI LEKSIKALNE BAZE

3.6 ORODJA ZA PRIDOBIVANJE IN ANALIZO LEKSIKALNO-GRAMATIČNIH PODATKOV

V postopku snovanja in gradnje leksikalne baze je bilo potrebno določiti način pridobivanja podatkov, ki bo učinkovit glede na njen namen, hkrati pa ne preveč zamuden in zapleten za izdelavo. Kot temeljni vir leksikalnih informacij smo sprva uporabili korpus FidaPLUS, ob zaključku gradnje korpusa Gigafida pa smo prešli na nov 1,2 milijardni slovenski referenčni korpus. Drugo izhodišče, ki je bilo znano že pri zasnovi leksikalne baze, je izraba orodja Sketch Engine za analizo korpusnih podatkov. Med pomembnejšimi spremembami, ki jih je na področje leksikografije prinesla gradnja več milijardnih korpusov je namreč ta, da jezikovnih podatkov iz tako obsežnih elektronskih besedilnih zbirk ni mogoče več pridobivati in analizirati brez pripomočkov, ki omogočajo prepoznavanje jezikovne regularnosti in posebnosti in jih hkrati posredujejo v obliki, ki je razmeroma lahko uporabljiva za leksikografa.

3.6.1 ANALIZA KONKORDANČNEGA NIZA

Pri pridobivanju in analizi korpusnih podatkov smo želeli postopke čim bolj optimizirati: naš končni cilj je bil pridobiti leksikalno-gramatične podatke iz korpusa čim bolj avtomatsko, hkrati pa v procesu ne izgubiti ničesar, kar bi bilo glede na zastavljen namen leksikalne baze in njene uporabnike relevantnega. Ta cilj je bilo mogoče doseči postopoma, zato je bil prvi korak v opisu leksikalne enote analiza konkordančnega niza, na podlagi katere smo ugotavljali osnovno obnašanje obravnavane enote v realnih besedilih. Analiza konkordančnega niza je potekala v več fazah. Najprej smo s pomočjo konkordančnika v orodju Sketch Engine izdelali naključni ali po določenih kriterijih filtrirani vzorec najmanj 150 in navadno ne več kot 300 konkordanc, ki smo jih nato s posebnim programom prenesli v Excelovo datoteko, kjer so bile pripravljene za nadaljnjo obdelavo. Analiza je bila primarno namenjena pomenski členitvi in registriranju udeleženske zgradbe ter stavčnih vzorcev pri posameznih pomenih, očitne pa so postale tudi podrobnosti skladijskega okolja preučevane besede, kar nam je koristilo pri oblikovanju pomenske sheme, npr. pri določanju tipičnih semantičnih tipov na posameznem udeleženskem mestu, kot prikazuje Slika 35. Poleg tega smo na ta način prepoznavali slovnično obnašanje leksikalne enote, npr. prevladujočo tretjeosebno rabo ali množinsko obliko, prevladujočo upovedovalno možnost (npr. zanikanje), besedilni tip, govorni položaj ter pragmatične in druge slovnično-pomenske lastnosti.

A	B	C	D	E
Strategija je jasna: že tedne izčrpavani in izolirani Dubrovnik bo pač	klonil	brez orožja ali pa se bo predal skupinskemu samomoru.	biti zavzet	kaj klone
Bunker je	klonil	že po prvi resnejši akciji, čeprav je bilo vmes tudi nekaj sreče.	biti zavzet	kdo/kaj klone (kdaj)
Manj uspešen je bil sinoči Marko Pečovnik (ŽBK AC Krajnc, do 91 kg), ki je	klonil	pred Ukrajincem Vitalijem Mihijjenkom .	izgubiti tekmo	kdo klone pred kom
Izraelci namreč v suproligi še niso	klonili	v domači dvorani: na 7 tekmah so tekmece v povprečju odpravili z 20 točkami razlike.	izgubiti tekmo	kdo kloone (kje)
Edinstvena pa bi bila tudi nesreča, če bi mreža	klonila	pod težo kamenja.	popustiti pod težo	kaj klone pod čim
Gradbeni velikan je te dni	klonil	in končno izplačal dolg.	popustiti pred zahtevami	kdo klone (kdaj)
Marsikdo bi že mnogo prej	klonil	in se predal.	popustiti pred zahtevami	kdo klone (kdaj)
Med zastižanjem je Mikložič	klonil	pod težo argumentov in zatem sam povedal, kam je skril Dolherjevo truplo.	popustiti pred zahtevami	kdo klone pod težo česa
Pa vendar, včasih	klonemo	in popustimo.	popustiti pred zahtevami/po dleči težavam	kdo klone
Andreja	klonila	v polfinalu	izgubiti tekmo	kdo klone (kdaj)

SLIKA 35: ANALIZA KONKORDANČNEGA NIZA ZA GLAGOL *KLONITI* Z VIDIKA DOLOČANJA SEMANTIČNIH TIPOV UDELEŽENCEV IN STAVČNIH VZORCEV

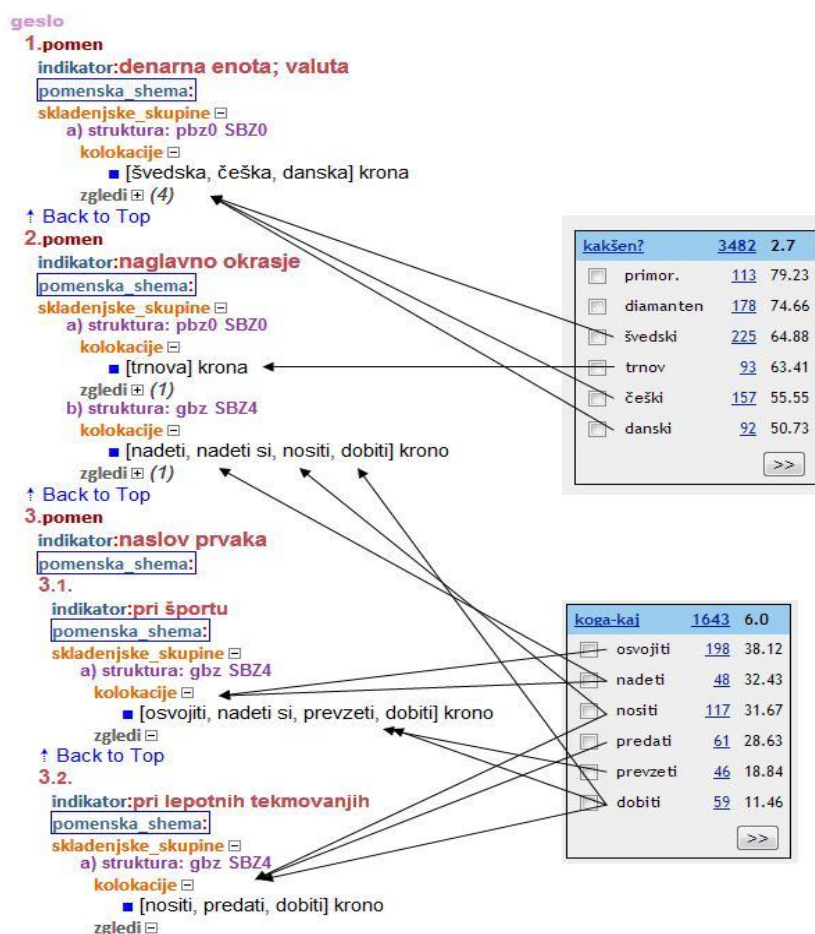
3.6.2 LEKSIKALNO-GRAMATIČNI PROFIL LEKSIKALNE ENOTE IN ORODJE SKETCH ENGINE

Orodje Sketch Engine⁴⁹ (SkE) smo uporabili kot osnovni pripomoček za pridobivanje in analizo leksikalno-gramatičnih podatkov iz korpusa in za postopek avtomatizacije, ki ga opišemo v nadaljevanju. Orodje vsebuje več aplikacij, ki omogočajo analizo posamezne leme v zelo obsežnem lematiziranem in oblikoslovno označenem korpusu. Poleg označenega korpusa potrebuje orodje za ustrezno delovanje

⁴⁹ Sketch Engine: <http://www.sketchengine.co.uk/>.

tudi pripravo slovnice besednih skic, tj. specifikacijo slovničnih relacij, na katerih temeljijo bistvene funkcije orodja SkE.

Najpomembnejša funkcija SkE je izdelava besedne skice (ang. *Word Sketch*), ki za obravnavano besedo pokaže, katere besede (t. i. kolokatorji) se v definiranem besedilnem okolju tipično povezujejo z njo. Besedilno okolje je definirano s slovnično relacijo, npr. z zvezo glagola in samostalnika v tožilniku. Če je na primer analizirana beseda *krona*, bomo pod slovnično relacijo "koga-kaj" našli tipične glagolske kolokatorje, ki se sopoljavljajo s samostalnikom *krona* v tožilniku, npr. *osvojiti*, *nadeti*, *nositi* itd., pod slovnično relacijo "kakšen" pa pridevniške kolokatorje kot: *diamanten*, *švedski*, *trnov* ipd. V procesu izdelave posameznega gesla smo na podlagi podatkov iz besedne skice izločili za posamezni pomen in skladienjsko strukturo relevantne kolokacije in jih prenesli v slovarski vmesnik, kot prikazuje Slika 36:



SLIKA 36: PRENOS KOLOKACIJ IZ BESEDNE SKICE V SLOVARSKI VMESNIK

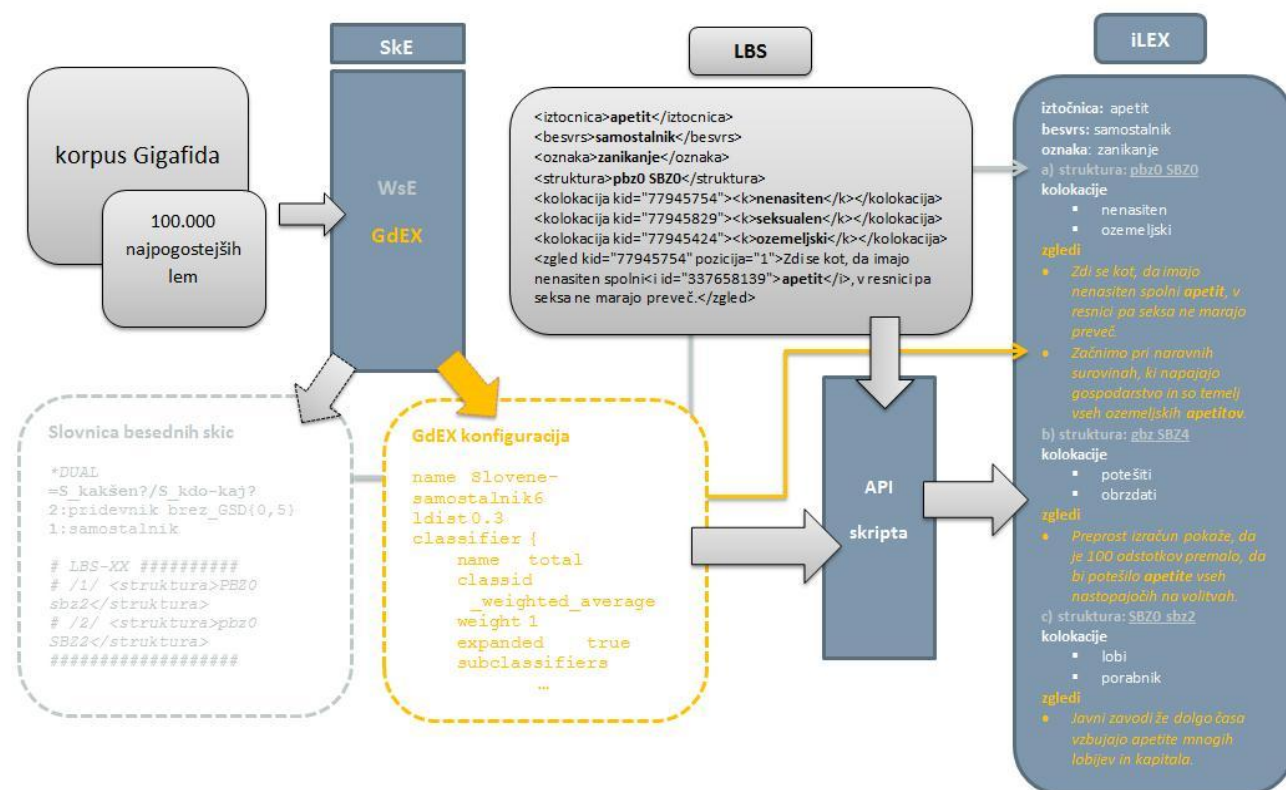
Skupaj s kolokacijo, ki smo jo umestili pod ustrezni pomen ali podpomen in skladienjsko strukturo, smo v slovarski vmesnik prenesli tudi korpusne zgleds. V orodju Sketch Engine je za ta postopek predviden sistem za avtomatično ocenjevanje slovarskih zgledov GDEX (ang. *Good Dictionary EXamples*), ki določa primernost oz. neprimernost stavkov iz korpusa za opravljanje vloge korpusnih zgledov v slovarju. Funkcija deluje tako, da posamezne konkordance – tj. potencialne zgleds – razvršča na lestvici od bolj do manj primernih, in sicer na podlagi vnaprej določenih merljivih kriterijev, med katerimi so

najpomembnejšimi dolžina stavka in dolžina ter pogostost besed v stavku. Višje so ovrednoteni stavki, v katerih se ciljna kolokacija pojavi v glavnem stavku in stavki s t. i. razširjenimi kolokacijami, ki presegajo klasično dvodelno zgradbo, npr. *pasti pod [določeno, magično, psihološko, kritično] mejo*. Ustreznejši so tudi stavki, v katerih se pred kolokacijo pojavi dovolj sobesedila, del s kolokacijo pa tipično sledi proti koncu stavka. Stavki so vrednoteni tudi glede na to, ali vsebujejo oz. ne vsebujejo besede s črne oz. bele liste, pri čemer na črno listo sodijo zlasti številke in internetni naslovi. Poleg omenjenih se pri vrednotenju stavkov uporabljajo tudi druga merila, ki so na voljo za določen jezik, npr. oblikoskladenjski podatki o besedah, stopnja njihove večpomenskosti ipd.

3.6.3 AVTOMATSKO PRIDOBIVANJE PODATKOV IZ KORPUSA

V zaključni fazi oblikovanja Leksikalne baze za slovenščino smo del gesel izdelali s postopkom avtomatskega luščenja leksikalnih podatkov iz korpusa Gigafida prek orodja Sketch Engine neposredno v program za izdelavo leksikalne baze iLex (Kosem idr. 2012).

Izhajajoč iz zgradbe leksikalne baze smo s postopkom avtomatizacije pridobili podatke na ravni skladišnih struktur, pripadajočih kolokacij in relevantnih korpusnih zgledov, poleg tega pa še podatke o tipičnem skladišnem ali besedilnem obnašanju leme v korpusu, kot je denimo sopojavljanje z lastnimi imeni ali količinskimi izrazi, možnost tretjeosebne rabe glagola ali nastopanje v se-glagolskih ali citatnih konstrukcijah. Prenos omenjenih podatkov je potekal iz korpusa Gigafida prek aplikacije Besedne skice in v leksikalni bazi registriranih skladišnih struktur, in sicer z uporabo API skripte (ang. *Application Programming Interface*) v program iLex (Erlandsen 2004), v katerem smo izdelovali leksikalno bazo, kot prikazuje Slika 37:



SLIKA 37: POSTOPEK AVTOMATSKEGA PRIDOBIVANJA PODATKOV IZ KORPUSA

3.7 LEKSIKOGRAFSKI POSTOPKI – OPIS IN TEORETIČNI PREMISLEKI

V nadaljevanju opišemo leksikografske postopke ter teoretične premisleke pri oblikovanju gesel in obravnavanju leksikalno-gramatičnih podatkov, ki jih vsebuje Leksikalna baza za slovenščino. Glede na to da je zasnova leksikalne baze, tako glede zgradbe in organizacije podatkov kot glede teoretičnega ozadja in sprejetih leksikografskih postopkov, neposredno povezana s končno podobo predlaganega Slovarja, je potrebno že v izhodišču izpostaviti, da v nadaljevanju opisani leksikografski postopki in odločitve temeljijo na:

- ▶ preverjenih dobrih leksikografskih praksah evropskih jezikov in sodobni leksikografski teoriji;⁵⁰
- ▶ podrobni večstopenjski korpusni analizi pribl. 2.500 lem sodobne slovenske leksike;
- ▶ vključevanju avtomatskih postopkov tako pri pridobivanju relevantnih leksikalno-gramatičnih podatkov iz korpusa kot pri njihovem urejanju znotraj baze;
- ▶ upoštevanju specifik spletnega medija;
- ▶ upoštevanju potreb slovenskih uporabnikov.

3.7.1 LEMA – IZTOČNICA

Način leksikografske predstavitve in izbor podatkov, vezanih na iztočnico, sta pri gradnji leksikalne baze za slovenščino zahtevala vrsto pomembnih jezikoslovnih in leksikografskih odločitev, ki se nanašajo na

⁵⁰ Glej razdelek SODOBNA LEKSIKOGRAFSKA TEORIJA v poglavju Bibliografija.

določitev različnih tipov iztočnic, zapis iztočnice v osnovni obliki ter njeno besednovrstno opredelitev. Poleg tega je bilo že na ravni iztočnice potrebno premisliti, kako reševati problem homonimije, besedotvorno povezanih oblik v odnosu do večpomenskosti (npr. v primeru manjšalnic), problem zapisovanja iztočnic glede na razmerje domače – prevzeto, glede na variantni zapis skupaj in narazen ter njihovo besednovrstno opredelitev, zlasti z vidika povezanih oblik (glagolnik, deležnik, deležje) in besednovrstne konverzije.

V Leksikalni bazi za slovenščino smo obravnavali dve vrsti iztočnic: **eno-** in **večbesedne**, pri čemer je velika večina iztočnic enobesednih, kot take pa smo obravnavali tudi glagole s prostima glagolskima morfemoma *se/si*, kjer je morfem sestavni del glagola v vseh njegovih pomenih, npr. *muditi se, odkrižati se, podvizati se, potegovati se, svitati se* ipd. Razen v redkih primerih, npr. *državni zbor, nivojski pouk*, večbesednih leksikalnih enot na ravni iztočnice nismo obravnavali, saj smo zanje predvideli samostojne razdelke gesla, kjer je kot naslovna enota izpostavljena iztočnica, ki je vsebovana v kateri od stalnih zvez ali frazeoloških enot.⁵¹

Na ravni iztočnice tudi nismo beležili **lastnih imen**. Pogosto večbesedna vrstna poimenovanja, ki so enakoglasna z lastnoimenskimi, smo obravnavali bodisi pri stalnih zvezah, npr. *društvo slepih in slabovidnih, bolnišnica za ginekologijo in porodništvo*, ali kot del kolokacijskega niza: *slovenske [železnice, železarne]*. Primere prehoda samostalnika iz občnega v (neosebno) lastno ime, kot npr. pri iztočnici *otok* -> *Otok* – 'Velika Britanija', smo obravnavali kot samostojne pomene.

Iztočnice so v leksikalni bazi zapisane v **osnovni obliki**, brez roditeljske končnice pri samostalniki in pridevniki ali oblike za 1. osebo ednine pri glagolih, ker smo v ta namen predvideli povezavo z drugimi bazami podatkov.⁵² Pri zapisu iztočnice v osnovni obliki smo upoštevali posebnosti pri posamezni besedni vrsti ter sledili oblikoskladenjskim oznakam v korpusu Gigafida. Na ravni iztočnice nismo želeli obravnavati skladenjskih posebnosti ali upoštevati jezikoslovnih meril, ki jih ne bi bilo mogoče enoznačno in z veliko mero gotovosti pripisati obliki leksikalne enote. Sem sodi odločanje o besednovrstni pripadnosti besed, ki v določenih skladenjskih položajih prevzamejo tudi drugačno skladenjsko vlogo, npr. *poslušati jazz* : *jazz* glasba ali *celofan papir* : zaviti v *celofan* ali *žvečilni gumi* : kupiti *žvečilni*. Za tovrstne skladenjske in pomenske podatke smo predvideli druge dele geselskega članka, kjer je informacijo mogoče opisati glede na obnašanje v besedilnem okolju. Prav tako se je podrejanje posameznih besednih oblik v podgesla, npr. manjšalnice, deležnike, deležja ipd., ali ločevanje na več samostojnih gesel pri homonimiji z vidika uporabniške prijaznosti in konsistentnega prikazovanja podatkov pokazalo kot neučinkovito.⁵³ Pri različnih zapisih/oblikah iztočnice na ravni leme smo se zato odločali za samostojno obravnavo (tj. dve samostojni iztočnici) le v primeru spremembe kategorije spola, npr. *prst* – ž. sp. : *prst* – m. sp., spremembe besedne vrste, npr. *naglas* – prislov : *naglas* – samostalnik, ali spremembe, tudi naglasne, v oblikoskladenjski (tj. spregatveni ali sklanjatveni) paradigmi, npr. *molíti* – *mólim* : *molíti* - *molím*. Ob tem pogoju je moral biti za obravnavanje oblikovno prekrivne besede kot

⁵¹ O obravnavi stalnih zvez in frazeologije v leksikalni bazi gl. poglavji 3.7.2.3.5 (Stalne zveze) in 3.7.2.3.6 (Frazeološke enote).

⁵² Gl. poglavje 1.1. (Slovenski leksikalni viri in orodja) ter vsebino zavirkov *Oblike* in *Izgovor* pri opisu spletnega slovarja.

⁵³ Nemogoče je denimo pričakovati, da bo slovarski uporabnik dejansko uporabil (oz. imel) etimološko vedenje, ki je edini zanesljiv ločevalec med homonimijo in večpomenskostjo, še preden bo iskal odgovor v zvezi s pomenom konkretne besede v slovarju (prim. Moon 1987: 89).

samostojne iztočnice izpolnjen tudi frekvenčni pogoj, tj. oblika je morala presegati prag uvrstitve v leksikalno bazo, torej imeti najmanj 400 oz. 600 pojavitev v korpusu FidaPLUS/Gigafida ter izkazovati zadostno razpršenost virov.

Homonime, tj. besede, ki so izrazno prekrivne znotraj istih oblikoskladenjskih kategorij (besedne vrste in spola, in sicer v celotni oblikoskladenjski paradigmi), smo v leksikalni bazi obravnavali kot večpomenske znotraj enega samega gesla pod skupno iztočnico, kot denimo v primeru samostalnika *bor*, kjer sta znotraj ene, homonimne iztočnice, obravnavana dva pomena, znotraj druge iztočnice, ki je besednovrstno opredeljena kot pridevnik, pa samostojni pomen:

bor samostalnik	bor pridevnik
1 drevo	1 majhen po količini; slab
2 kemijski element	

Kadar prihaja pri enakopisnih (ne pa tudi enakoglasnih) prekrivnih oblikah znotraj iste besednovrstne kategorije do spremembe paradigme zgolj na ravni naglasa, npr. *častíti – častím* in *částiti – částim; téma – téme* in *tèma – tème* oz. *temè*, smo besedi opisali v ločenih geslih:

častiti glagol	častiti glagol
1 izkazovati spoštovanje	1 povabiti na pijačo
1.1 v religiji	

Kadar prihaja do pisne ne pa tudi naglasne prekrivnosti, npr. *poročèn – poróčen; omrežèn – omréžen*, pri osnovni obliki, ne pa tudi v celotni sklanjatveni ali spregatveni paradigmi, smo obravnavali besede v samostojnih geslih, poleg tega pa smo pri samostalnikih in pridevniki zapisali še rodilniško obliko (*poročen – poročenega*), pri glagolih pa obliko za 1. osebo ednine (*pasti – padem* in *pasti – pasem*):

poročen-poročenega pridevnik	poročen-poročnega pridevnik
file-fileja samostalnik	file/fajl-fajla samostalnik
pasti-padem glagol	pasti-pasem glagol

Iztočnic, ki imajo prekrivno osnovno obliko in različno oblikoslovno paradigmo ter hkrati nimajo prekrivnih slovničnih kategorij, kot sta besedna vrsta ali slovnični spol, npr. *bor* – pridevnik, *bor* samostalnik, *plesen* – samostalnik, *plesen* – pridevnik; *prst* – sam ž. sp. in *prst* – sam. m. sp., nismo obravnavali znotraj večpomenskosti, ampak smo izdelali samostojni iztočnici. V primeru različnega spola znotraj iste besedne vrste (*prst, smuč*) smo poleg osnovne imenovalniške zapisali še edninsko rodilniško obliko:

smuč-smuči samostalnik	smuč-smuča samostalnik
1 pripomoček za smučanje	1 sladkovodna riba

Med besedotvorno povezanimi oblikami je treba posebej izpostaviti obravnavo **manjšalnic** (oz. besed, ki po obliki delujejo kot manjšalnice, npr. *barvica, cvetica*), ki smo jih v leksikalni bazi obravnavali kot samostojne iztočnice, prim. *hlačke, jopica, koščica, mačica, stolček*. Pomene kot 'majhne hlače', 'majhen stol', če so obstajali, smo obravnavali kot samostojne pomene. Če so bili taki pomeni osamljeni, npr.

trobentica – 'majhna trobenta', ali v realni rabi niso bili izkazani, npr. *marelica* 'majhna marela', jih v leksikalni bazi nismo registrirali.

Variantne **zapise** samostalnikov na **-lec/-vec**, ki so glede na obstoječo normo pogosto različno vrednoteni, smo, če sta obe obliki presegle frekvenčni prag za uvrstitev v leksikalno bazo, prav tako obravnavali kot samostojni iztočnici, saj je mogoče predvidevati tako pomensko razlikovanje kot različnost besedilnega okolja in rabe, ki je navadno pogojena s frekvenco posamezne oblike. Normativnih podatkov v zvezi s takimi zapisi v leksikalni bazi nismo posebej beležili, ker jih je mogoče pridobiti iz Slogovnega priročnika.⁵⁴

zajedavec samostalnik

1 parazit

2 kdor kaj neupravičeno jemlje ali dobiva

zajedalec samostalnik

1 parazit

V leksikalni bazi kot samostojnih iztočnic nismo obravnavali predpon tipa *evro-* *pro-*, *anti-* *e-* ipd., pač pa bodisi kot samostojna gesla (npr. *evro*) bodisi kot celotne besede, na katere se take predpone pripenjajo (*biomasa*).

V primeru različnih zapisov leme glede na razmerje **domače** – **prevzeto** oz. **citatno**, npr. *piercing* – *pirsing*, *file* – *file/fajl*; *džezovski* – *jazzovski* itd., smo v leksikalni bazi izdelali dve oz. več samostojnih gesel, če so imele posamezne oblike zadostno število pojavitev v korpusu. Odločitev je podprta s predvidevanji, da lahko v povezavi z različnimi zapisi pričakujemo tudi drugačne leksikalno-gramatične podatke na domala vseh ravneh leksikografske obravnave, kot je npr. različno kolokabilno okolje, izbira registra, besedilnega tipa, stilne vrednosti, tvorjenje stalnih zvez, frazeologije ipd.

Besede oz. zveze, ki v korpusu izkazujejo zapis tako **skupaj** kot **narazen**, pri čemer imata obe obliki zadostno število pojavitev za uvrstitev v leksikalno bazo, so predstavljene na dveh mestih, in sicer zapis skupaj kot samostojna iztočnica, npr. *odkod*, *naglas* (prislov), in kot skladenjska zveza *na glas* (oz. kot FE, če potrebuje razlago) pri iztočnici *glas* (samostalnik), če tako možnost izkazuje realna raba.

Pri zapisovanju **besedne vrste** iztočnice smo v leksikalni bazi sledili njeni formalni in ne vsebinski podobi, kar pomeni, da smo besedo v iztočnici besednovrstno opredelili glede na to, kako se obliko(slo)vno obnaša (sklanja, spreaga, stopnjuje), medtem ko smo njeno skladenjsko-pomensko vlogo v stavku opisali na ravni pomena. Naprimer, beseda *odvisen* je na ravni iztočnice opredeljena kot pridevnik, znotraj gesla pa so opisane tudi njene druge skladenjske rabe, npr. samostalniška: zdravljenje *odvisnih*. Na ta način je besednovrstna opredelitev tako osnovnih (samostalnik, glagol, pridevnik, prislov itd.) kot podkategorij (deležniki, deležja, glagolniki) usklajena z oblikoskladenjskimi oznakami v korpusu Gigafida in omogoča optimalno rabo Besednih skic in nadaljnje procesiranje v bazi strukturiranih podatkov.

Spremembe v besedni vrsti, ki nastanejo, ko beseda v stavku spremeni svoj skladenjski položaj in s tem prevzame tudi skladenjske in pomenske lastnosti, ki so za ta položaj značilne, t. i. **konverzija besednih vrst**, se kažejo v slovenščini kot univerzalni slovnični pojav, vendar pa so hkrati pri posameznih besedah oz. njihovih pomenih različno zastopane tako v smislu tipičnosti kot opazne pomenske vloge. Med

⁵⁴ Gl. poglavje 1.1 (Slovenski leksikalni viri in orodja) v tem dokumentu.

primeri, kjer besede v določenih skladenjskih položajih spremenijo besednovrstno kategorijo, smo v leksikalni bazi obravnavali naslednje možnosti:

Posamostaljene pridevnike, torej pridevnike, ki v določenih položajih opravljajo vlogo samostalnika, smo v leksikalni bazi obravnavali v samostojnem pomenu ali podpomenu in s slovnično oznako »v samostalniški rabi«, npr.:

rujen *pridevnik*

1 rdeč

rujno [lice]

1.1 o vinu

v samostalniški rabi

[kozarec, kozarček, kapljica ...] rujnega

1.2 dober

rujno [vince, vino]

Nekatere besede, ki se v večini primerov skladenjsko in pomensko obnašajo kot samostalniki, lahko dobijo v položaju pred samostalnikom vlogo nesklonljivega prilastka – t. i. **samostalniki v pridevniški rabi**. Taka vloga je lahko povezana s pomenskimi lastnostmi samostalnika, kot v primeru *jeans* pomen 1.2, ali pa ostaja na ravni skladnje, kar smo v leksikalni bazi označili v elementu restrikcija v skladenjski strukturi:

jeans *samostalnik*

1 vrsta tkanine

SBZ1 <r>**v pridevniški rabi**</r> sbz0: jeans [hlače, jakna]

1.1 o oblačilih

gbz SBZ4: [nositi] jeans

SBZ1 <r>**v pridevniški rabi**</r> sbz0: jeans [moda]

1.2 o barvi

v pridevniški rabi

SBZ1 sbz0: jeans [barva, modra]

Besede kot *latino*, *afro*, *mini* ipd. smo obravnavali besednovrstno kot **nesklonljive pridevnike** glede na njihovo prevladujočo skladenjsko vlogo, samostalniško rabo pa smo opisali v samostojnih pomenih in z oznako »v samostalniški rabi«:

latino *pridevnik*

1 južnoameriški

1.1 o glasbi in plesu

latino [glasba, ritmi]

1.2 o ljudeh

latino [lover, zapeljivec, ljubimec, lepotec]

latino [diva, lepotička, lepotica]

2 južnoameriška glasbena zvrst ali ples

v samostalniški rabi

[pop, hip hop] in latino

3 južnoameričan

v samostalniški rabi

*Za pultom me pričaka nizkorasli **latino** s priimkom Lopez.*

Besede kot *balzam*, *blagor*, *mraz*, *tema*, *groza*, *škoda* ipd., ki se pojavljajo tudi v povedni rabi (*biti pravi balzam/blagor za koga*), so v leksikalni bazini ravni iztočnice opredeljene kot samostalniki, na ravni skladenjske strukture pa je v restrikciji opozorjeno na povedno rabo samostalnika:

tema *samostalnik*

1 pomanjkanje svetlobe

Vez-gbz SBZ1<r>**v povedni rabi**</r>: [biti tema], biti [trda, čista, popolna] tema

Primerih, kjer prihaja do pomenskih osamosvojitve, kot npr. pri samostalniku *kos*: *biti kos komu/čemu*, smo osamosvojeno povedkovniško pozicijo v obravnavali kot frazeološko enoto, torej pod samostalniško iztočnico v frazeološkem razdelku:

kos *samostalnik*

1 del celote

2 enota

3 ptica

biti kos komu/čemu / biti kos nalogam/nalogi

biti sposoben rešiti nek problem, obvladati določeno znanje ali se znajti v konkretni situaciji

Najbolj tipične in hkrati heterogene povedkovniške besede, kot sta npr. *rad* in *všeč*, smo v leksikalni bazi obravnavali v skladu z oblikoslovnimi oznakami v korpusu kot prislove. Prav tako tudi besede kot *poleti*, *jeseni*, *pozimi*, npr. v kolokacijah: [letos, pozno] jeseni, lani [jeseni, poleti], ali: [končan] poleti, kar smo ločevali od samostalnika *jesen* v rodilniku, npr. [konec, začetek] jeseni.

V strukturi geselskega članka smo na take tipične skladenjske rabe opozorili v slovničnih oznakah neposredno za indikatorjem pomena, kot npr. pri pridevniki *aktiven* in *kriv*, kjer je povedna skladenjska vloga razvidna tudi iz pomenske sheme:

aktiven *pridevnik*

1 dejaven

v povedni rabi

ČLOVEK je **aktiven** na nekem PODROČJU ali v neki DEJAVNOSTI, če se vanjo dejavno vključuje in v njej redno sodeluje

ali na ravni kolokacij, npr. [dan] je deževen, [primeri] so ilustrativni, [vprašanje] je logično, [mnogi] so skeptični; dialog je [potreben, vzpostavljen], in: [biti] pasji, [biti] tešč itd.

3.7.2 POMEN

Pomenski nivo je v leksikalni bazi hierarhično nadrejen vsem podatkom, ki jih beležimo v zvezi s posamezno besedo v iztočnici. To, za razliko od iztočniške ravni, kjer smo izhajali iz formalne oblike besede, pomeni, da je na pomenskem nivoju izhodišče za beleženje jezikovnih podatkov vedno konkretni, na podlagi korpusne analize registrirani pomen besede.

Bistveno za razumevanje koncepta pomena v leksikalni bazi in v predlaganem spletnem slovarju je ločevanje med pomenskim potencialom, ki ga ima beseda v svojem realnem življenju, in leksikografsko razlago, ki je metaslovarski konstrukt. V prvem primeru izhajamo iz stališča, da besede same na sebi nimajo pomenov, pač pa le tendence, da nekaj pomenijo (Hanks 2009: 11), leksikografski opis pomena oz. pomensko razlago pa razumemo kot poskus približevanja pomenskih tendenc naslovniku, ali kot pravi Hanks (1987: 135; 2000: 8): »Slovarji niso namenjeni definiranju, pač pa je njihova naloga ustvariti vrsto namigov in asociacij, ki pomagajo povezati neznano z znanim«.

Poleg ločenega razumevanja pomenov znotraj komunikacije in znotraj slovarja smo pri oblikovanju leksikalne baze izhajali iz stališča, da pomenske tendence besed vplivajo na izbire v besedilnem okolju, kar se odraža v stavčnih vzorcih, skladenjskih strukturah in kolokacijah, hkrati pa tudi v prevzemanju nekaterih slovničnih kategorij, kot so denimo (ne)števnost, slovnično število, ustaljenost v določeni glagolski obliki ipd. S tem priznavamo povezanost pomena in skladnje oz. t. i. leksikalnogramatični pristop v opisu leksikalnih enot. Ta povezanost se kaže že v sami zgradbi leksikalne baze, kjer je vsak pomen besede predstavljen z vidika vseh predvidenih informacij na posameznem nivoju, eksplicitno pa tudi v zapisu pomenske sheme, ki predstavlja izhodišče za oblikovanje stavčnih definicij.

3.7.2.1 POMENSKA ČLENITEV

Pri pomenski členitvi besede, ki je zastopana v iztočnici, smo izhajali iz predpostavke, da večina pomenske informacije konkretne besede izhaja iz njenega bodisi neposrednega bodisi širšega besedilnega okolja, z drugimi besedami, pomeni besed so povezani z določenimi nizi sintagmatskih vzorcev, ki jih je mogoče odkriti s podrobno korpusno analizo Hanks (2009: 8).

Pri pomenskem členjenju besed smo v leksikalni bazi upoštevali naslednja splošna načela:

- Besede smo pomensko členili v dveh stopnjah, tj. na *osnovne pomene* (oz. na *glavni pomen*) in na *podpomene*.
- Podpomeni so glede na glavni pomen ali osnovne pomene v hierarhičnem razmerju.
- Osnovne pomene razvrščamo na podlagi ocene pogostosti kolokatorjev, ki sodijo k obravnavanemu pomenu.
- Za določanje pomenov in podpomenov je odločilno stanje, kot ga izkazuje referenčni korpus prek Besednih skic v orodju Sketch Engine oz. analiza najmanj 150 konkordanc za obravnavano besedo.
- Pomenske lastnosti obravnavane besede smo členili čim bolj podrobno, pri čemer smo težili k manjšemu številu osnovnih pomenov in večjemu številu podpomenov.
- Hkrati smo sledili tudi beleženju tipičnosti – tj. prepoznavni količini analognih pomenov – in ne k individualnim posebnostim in enkratnim pomenskim rabam.

Pri določanju pomenske zgradbe besed smo si pomagali s t. i. **formalnimi** in **vsebinskimi kazalci večpomenskosti**. Med prvimi smo upoštevali besednovrstno opredelitev besede, kjer se se kot pomenskorazločevalne kažejo za posamezno besedno vrstno značilne slovnične kategorije (Vidovič Muha 2000: 32). Pri samostalniki so to zlasti spol in podkategoriji živost in človeškost, poleg tega števnost ter število, pri glagolih spremembe na ravni prehodnosti in v udeleženski zgradbi pomena ter glagolski vid, pri pridevniki in prislovih pa zlasti možnost stopnjevanja. Pomembni in hkrati najbolj očitni formalni pokazatelji večpomenskosti so kolokatorji, pri glagolih pa zlasti udeleženska zgradba, ki se spreminja glede na njegove pomenske tendence, npr. v primeru glagola *hvaliti*:

hvaliti *glagol*

1 poudarjati dobre lastnosti

KDO **hvali** KOGA/KAJ

če ČLOVEK hvali drugega ČLOVEKA, USTANOVO, njegova DEJANJA ali DELO, poudarja njegove dobre lastnosti, sposobnosti ali dosežke

1.1 poudarjati lastne sposobnosti in dosežke

če se ČLOVEK hvali, pretirano poudarja svoje LASTNOSTI, SPOSOBNOSTI ali DOSEŽKE, predvsem z namenom, da bi ustvaril dober vtis pred drugim ČLOVEKOM

KDO **se hvali**
KDO **se hvali** s ČIM

Med vsebinskimi kazalci večpomenskosti smo obravnavali pomenske lastnosti, ki smo jih določili kot relevantne za povezovanje posameznih podpomenov v okviru nadrejenega osnovnega pomena in za ločevanje med osnovnimi pomeni ali znotraj podpomenov. V slovenski leksikalni teoriji se je za »vsebinske lastnosti s pomensko razpoznavno vrednostjo« (Vidovič Muha 2000: 21) uveljavil izraz pomenske sestavine (Toporišič 1980, 1981; Müller 1980) ali semantične komponente (Orešnik 1972) po vzoru t. i. komponentne analize, katere namen je bil razgraditi pomen leksikalne enote na atomske koncepte (Katz in Fodor 1963). Vendar pa v leksikalni bazi teh lastnosti v smislu predvidljivih izbir nismo določali vnaprej niti nismo izhajali iz njihove hierarhične ureditve, kot jo določa zgradba klasične slovarske definicije, pač pa nam je vsebinske lastnosti, kot so entiteta, namen, lastnost, oblika, dejanje ipd. sugerirala vsakokratna analiza dejanske rabe konkretne leksikalne enote v realnem kontekstu, kot v primeru glagola *znebiti se*, kjer je v Tabeli 3 prikazana določitev relevantnih pomenskih lastnosti, ki nato rezultirajo v končno pomensko zgradbo v obliki pomenskega menija:

ENTITETA	NAMEN	LASTNOST	DEJANJE	POMENSKI MENI
predmeti (konkretno)	odstranitev; ne imeti več	odvečen, nepotreben, moteč	zavreči	znebiti se <i>glagol</i> 1 odstraniti
človek, žival, ustanova	ne obstajati več	moteč, vsiljiv, ki ogroža; škodljiv	odsloviti, usmrčiti	
kilogrami - človek	zmanjšanje	odvečen, prekomeen,	shujšati	

		debel; vitek		1.1 o predmetih 1.2 o ljudeh 1.3 o telesni teži 1.4 o oblačilih 2 spremeniti 2.1 o občutkih in stanjih 2.2 o mnenju, prepričanju
oblačila - človek	odstranitev		sleči se	
negativni občutki, stanja - človek	spremeniti stanje; doseči izboljšanje	neprijeten, obremenjujoč	truditi se, izboljšati	
mnenje, prepričanje - človek	spremeniti mnenje	negativen, obremenjujoč	ukrepati	

TABELA 3: DOLOČITEV POMENSKIH LASTNOSTI

3.7.2.2 RAZVRSTITEV POMENOV

Pri razporeditvi osnovnih pomenov smo v leksikalni bazi na prva mesta uvrstili pomene, ki so glede na analizo konkordanc in besedne skice v orodju Sketch Engine predstavljali osrednje in hkrati najpogostejše rabe. Izhodiščno delitev osnovnih pomenov glede na relativno pogostost prikazuje 1. faza pomenske členitve za glagol *brusiti*. V nadaljevanju smo upoštevali možnost, da so isti kolokatorji, ki izstopajo po frekvenci in statistikah, lahko različno razporejeni med posamezne pomene in podpomene, in možnost, da je kateri od podpomenov, stalna zveza ali frazeološka enota dejansko pogostejši od osnovnega pomena, kar prikazuje 2. faza pomenske členitve in dokončno oblikovanje pomenskega menija z identifikacijo stalnih zvez in frazeoloških enot:

1. faza: razporeditev osnovnih pomenov

brusiti *glagol*

- 1 delati, da kaj postane ostro
- 2 obdelovati hrapavo površino
- 3 izražati mnenje, nestrinjanje
- 4 urediti, utrjevati se

2. faza: vključitev podpomenov, SZ in FE

brusiti *glagol*

- 1 delati, da kaj postane ostro
 - 1.1 o orodju
 - 1.2 o delih telesa
- 2 obdelovati hrapavo površino
- 3 urediti, utrjevati se

brusiti jezike/brusiti si jezike *opravljati; obrekovati*

brusiti nože *pripravljati se na besedni spopad, konflikt*

3.7.2.3 POMENSKI OPIS

Postopka oblikovanja pomenske razlage ali definicije ni mogoče ločiti od pomenske razčlenitve besede, saj zahteva pomenska členitev, če želimo pomene med seboj hierarhizirati in smiselno ločevati, tudi premislek o tem, kako posamezne pomenske tendence obravnavane besede ubesediti. Opis pomena je v LBS vezan na štiri segmente: na **pomenske indikatorje**, s katerimi oblikujemo **pomenski meni**, na **pomenske sheme**, ki predstavljajo pomen besede v obliki stavčne razlage z izpostavljenimi semantičnimi tipi na udeleženskih mestih in na **razlage**.

3.7.2.3.1 POMENSKI MENI

Pomenski meni je produkt sodobne slovarske prakse, zlasti elektronskih in spletnih slovarjev, kjer je pomemben hiter dostop do želene informacije. Z njim prikazujemo **pomensko zgradbo večpomenskih gesel** in logična razmerja med pomeni in podpomeni ter podajamo osnovno **predstavo o pogostosti pomenov**. V spletnih slovarjih je pomenski meni namenjen lažji navigaciji uporabnika po večpomenskem geslu, zato je pomembno, da je razporeditev in opredelitev pomenov znotraj njega logična in na prvi pogled razumljiva: uporabnika mora napotiti na iskani pomen, ne da bi moral preverjati posamezne pomenske razlage in iskati ustrezne zglede rabe. Pomenski meni tudi ne sme biti zavajajoč v smislu premajhne medsebojne razločevalnosti ali prekrivnosti in ne sme vzbujati napačnih predstav o pomenskih tendencah besede ali dajati neustreznih pomenskih asociacij.

V leksikalni bazi smo z namenom oblikovati pomenske menije, ki ustrezajo zgornjim zahtevam, sledili naslednjim smernicam:

- ▶ Indikatorji, s katerimi oblikujemo pomenski meni, so ubesedeni na način, da znotraj pomenske zgradbe besede vzbujajo zadostno razlikovalnost med dvema ali več pomeni.
- ▶ Indikatorji so ubesedeni tako, da ne sugerirajo prekrivnih pomenskih vsebin.
- ▶ Pri besedah, kjer je pomenska členitev izdelana tako, da ima osnovni pomen le združevalno vlogo, je to iz menija neposredno razvidno, indikator pa je ubeseden tako, da ne vzbuja pomenske prekrivnosti s katerim od svojih podpomenov ali z drugimi pomeni znotraj gesla.
- ▶ Pomenske vsebine, ki jih izražajo indikatorji, se v meniju ne ponavljajo.

Nekaj primerov ustreznih pomenskih menijev ponazarjajo spodnji primeri:

čestitati <i>glagol</i>	glasovati <i>glagol</i>	skodelica <i>samostalnik</i>	cimetov <i>pridevnik</i>
1 izraziti lepe želje	1 izraziti mnenje	1 posodica; šalica	1 o drevesu
1.1 ob pomembnem dosežku	1.1 na volitvah	2 količina	1.1 o začimbi
1.2 ob veselem dogodku	1.2 na tekmovanju	2.1 merska enota	2 o barvi
		3 del naprave ali predmeta	

3.7.2.3.2 POMENSKI INDIKATORJI

Naloga pomenskih indikatorjev, kot jih razumemo v leksikalni bazi, je **na kratko** in **na razumljiv način** opisati ali bolje določiti **pomenski obseg** ali **pomensko področje** obravnavane besede, pri večpomenskih besedah pa poleg tega vzpostaviti še **razliko glede na druge pomene** ali **podpomene**. Kot smo izpostavili že v prejšnjem poglavju, oblikujemo z indikatorji **pomenske menije**, za katere velja, da so samostojna slovarska informacija in morajo biti razumljivi sami na sebi.

Indikatorje smo glede na primernost uporabe pri posamezni besedni vrsti obravnavane besede razdelili v 4 osnovne tipe, pri izbiri posameznega tipa pa smo upoštevali:

- ▶ razmerje med osnovnim pomenom in podpomeni oz. nadrejenost osnovnega pomena;
- ▶ stopnjo pomenske razčlenjenosti besede v iztočnici;

- ▶ porazdelitev pomenske informacije med indikator in pomensko shemo oz. razlago, pa tudi razmerje med indikatorjem in oznako.

Najpogosteje smo v leksikalni bazi uporabljali enobesedne opredelitve pomenov oz. t. i. **sinonimne indikatorje**, in sicer pri vseh besednih vrstah, kot prikazuje Tabela 4:

IZTOČNICA	SINONIMNI INDIKATOR
arhiviranje	shranjevanje
emigrant	izseljenec
titula	naziv
zelenec	novinec; začetnik
crkljati	1 ljubkovati, 2 razvajati
arhivirati	1 shraniti, 2 zabeležiti
suveren	neodvisen
absoluten	brezpogojen; nesporen
blažen	srečen; zadovoljen
hote	namenoma
implicitno	posredno

TABELA 4: SINONIMNI INDIKATORJI

Pri izbiri sinonima v vlogi indikatorja smo težili k temu, da se je enobesedna sopomenka v čim večji meri, tj. glede na kolokacijske, stilne, zvrstne, pogostnostne in druge lastnosti, približala besedi v iztočnici, hkrati pa smo bili pozorni na to, da sinonimni indikator ni zajel pomenskega polja iztočnice preširoko ali preozko. Pri tem smo si pomagali z možnostjo »razdelitve« pomenske informacije med druge pomenske elemente leksikalne baze, tj. med pomensko shemo oz. razlago in oznako. V leksikalni bazi smo predvideli tudi možnost dopolnjevanja indikatorjev, sinonimnih, opisnih in drugih, z nadaljnjimi sinonimi. V takih primerih smo sinonime zaradi možnosti naknadnega pridobivanja iz leksikalne baze navajali za podpičjem, kot npr. v spodnjih primerih:

čednost *samostalnik*

lepa človeška lastnost; vrlina; krepost

kodirati *glagol*

o besedilu; šifrirati

čeden *pridevnik*

lep; privlačen

istočasno *prislov*

hkrati; obenem

Opisne indikatorje smo uporabljali pretežno za opis osnovnih pomenov pri pomensko bolj razvejanih besedah in v primerih, kjer ni bilo mogoče določiti ustreznega sinonima, npr.:

spomin *samostalnik*

1 pomnjenje česa iz preteklosti

2 sposobnost pomnjenja

3 avtobiografsko literarno delo

4 zmogljivost računalnika

okupirati *glagol*

1 zavzeti z vojaško silo

1.1 množično obiskati

2 zahtevati veliko pozornosti

V nekaterih primerih lahko opisni indikator prevzame dejansko vlogo krajše razlage:

bucika *samostalnik*

kovinska igla z glavico

feministka *samostalnik*

ženska, ki se zavzema za pravice žensk

Nanašalne indikatorje smo v leksikalni bazi uporabljali za opisovanje podpomenov ali za ločevanje krovnih pomenov od podpomenov, in sicer primarno pri pridevniških iztočnicah, npr.:

afriški *pridevnik*

1 o Afriki

2 ki izhaja iz Afrike

2.1 o človeku

2.2 o rastlinah in živalih

2.3 o kulturi

avtomobilski *pridevnik*

1 o avtomobilih

1.1 o industriji in oglaševanju

1.2 o športu

2 del avtomobila

2.1 namenjen avtomobilu

Z indikatorji v obliki **neposredne nadpomenke**, npr. *angina* – bolezen; *arnika* – rastlina; *arheologija* – znanstvena veda, *jogurt* – mlečni izdelek, *vino* – alkoholna pijača, smo želeli v leksikalni bazi vzpostaviti razmerja med pomensko in tematsko povezanimi iztočnicami oz. oblikovati t. i. semantične mreže. Kot smo predvideli pri izdelavi spletnega slovarja je taka pomenska razmerja mogoče pri prenosu na splet nadgraditi in povezati z obstoječimi ontološkimi bazami in semantičnimi leksikoni.

Na splošno smo se pri ubeseditvi indikatorjev izogibali rabi večpomenskih besed in besed v katerem od obrobnihi ali atipičnih pomenov, manj zastopanih besednih oblik in besed, ki so redko zastopane v običajni rabi, navajanju indikatorjev v nikalni obliki in indikatorjem, ki izražajo frazeološki ali metaforični pomen. Težili smo k čim bolj preprostim in nerigidnim ubeseditvam, kar smo skušali doseči z izbiro indikatorjev, ki najbolje asociirajo ključno pomensko lastnost iztočnice, ki izražajo ustrezno besedno vrsto in glagolski vid iztočnice (izjema so nanašalni indikatorji, ki te možnosti nimajo) in so hkrati čim bolj kratki in razumljivi.

3.7.2.3.3 POMENSKE SHEME

Zapis pomenskih shem predstavlja v Leksikalni bazi za slovenščino stičišče leksikalno-gramatičnih podatkov, hkrati pa je izhodišče za oblikovanje definicij stavčnega tipa, kot so prikazane v spletnem slovarju (glej poglavje 2.2.1, Zavihek: pomen). Z vidika sodobnih pristopov k opisovanju pomena besed temelji na simulaciji naravnega diskurza pri pojasnjevanja besednih pomenov v šolski situaciji ter prinaša bistveno novost v način pomenskega opisovanja slovenske leksike (Gantar in Krek 2009).

Teoretično temelji zapis pomenskih shem na smernicah projekta Frame Net, ki omogoča računalniško podprto podrobno dokumentacijo korpusnega sobesedila, v katerem se nahaja obravnavana beseda oz. njen pomen. Kot podatkovna baza, ki je prosto dostopna na spletu, omogoča Frame Net vrsto različnih avtomatičnih prikazov in iskanj podatkov, ki so namenjeni podrobnim človeškim in računalniškim analizam. Podobne analize, ki so povezane z izhodiščnim projektom v Berkeleyju, nastajajo za nemščino, španščino, japonsščino pa tudi slovenščino (Može 2009, Lönneker-Rodman 2008).

FRAMENET

Namen projekta FrameNet⁵⁵ je za vsak pomen besede ali besedne zveze analizirati in beležiti celoten spekter pomenskih in skladenjskih razmerij. Vsaka beseda, ki je vključena v FrameNet, tako pripada določenemu pomenskemu okvirju. Če je beseda večpomenska, pripada več različnim pomenskim okvirjem. Pomensko povezane besede, npr. glagoli premikanja, govorjenja, zaznavanja itd., se tako pojavljajo znotraj istih pomenskih okvirjev. Vsak pomenski okvir je vnaprej določen, ima svoje ime in definicijo, v kateri so izraženi jedrni shemski elementi. Poleg jedrnih lahko okvir vključuje tudi nejedrne elemente. Znotraj pomenskega okvirja ima vsak element svoje ime, opis in je opredeljen glede na semantični tip. Jedrni elementi so za evociranje konkretnega pomena glagola (ali samostalnika) obvezni, vendar v nekaterih natančno določenih primerih lahko tudi neizraženi oz. izpuščeni. Npr. angleški glagol *squeeze* (stisniti) pripada dvema pomenskima okvirjema: *Cause_to_move_in_place* (povzročiti premik v prostor(u)) in *Manipulation* (upravljanje). Poleg glagola *squeeze* se znotraj prvega okvirja nahajajo tudi glagoli kot: *rotate*, *seesaw*, *shake*, *spin*, *swing*, *swirl*, *turn* ipd., v drugem okvirju pa glagoli kot: *caress*, *clasp*, *claw*, *clutch*, *diddle* ipd. V pomenski okvir *Manipulation* poleg glagola *squeeze* sodijo torej vse besede (oz. njihovi pomeni), ki opisujejo upravljanje z neko entiteto, ki jo izvede aktant, navadno z delom svojega telesa, iz česar izhaja, da so entiteta, aktant in del njegovega telesa jedrni elementi pomenskega okvirja *Manipulation*.

Drugi projekt, ki izhaja z korpusne analize in prepoznavanja stavčnih vzorcev v povezavi s konkretnimi pomeni obravnavane besede, in smo ga vzeli kot izhodišče pri oblikovanju pomenskih shem v LBS, je projekt Corpus Pattern Analysis (CPA; sln. analiza korpusnih vzorcev; Hanks 2004, 2008; Hanks in Pustejovsky 2005).

CORPUS PATTERN ANALYSIS

Projekt temelji na projiciranju pomena iz sobesedila na posamezno besedo in izhaja iz teorije jezikovnih konvencij in možnostih njihove izrabe (ang. Theory of Norms and Exploitations, Hanks 1994, 2013; Hanks in Pustejovsky 2004, 2005). CPA se osredotoča na sistematično analizo tipičnih pomenskih vzorcev glagolov, kjer so posamezni pomeni na podlagi primerov iz realne rabe povezane v semantično motivirane sintagmatske vzorce. Tem vzorcem je nato pripisan pomen, ki je izražen s t. i. implikaturami, na njihovi podlagi pa je mogoče postopoma oblikovati pomensko ontologijo. Postopek je mogoče pojasniti na korpusnem primeru rabe glagola *vložiti*: *Pivovarna Laško bo vložila pritožbo*. Pomen glagola v izkazanem kontekstu, tj. s predmetnim dopolnilom *pritožba*, vsebuje naslednje implikature: [Pivovarna Laško = ustanova] bo vložila [pritožbo = dokument], ki jih lahko izrazimo takole: »če neka ustanova vloži pritožbo, posreduje dokument, s katerim sproži pravni postopek«. Isti glagol v drugačnem kontekstu, npr. *V novo dejavnost smo vložili veliko denarja*, seveda zahteva drugačne implikature.

Temeljna naloga pomenskih shem, kot smo jih zasnovali v LBS, je pojasniti pomene glagolov IN nekaterih samostalnikov v zanje prototipičnih stavčnih realizacijah, kjer so posamezni udeleženci na pomenskosladdenjsko določenih mestih zastopani s semantičnimi tipi. Pomenske sheme smo oblikovali v treh korakih. V prvem smo za konkretni glagolski pomen določili obvezna udeleženska mesta, npr. za glagol *anketirati* v pomenu 'pridobivati podatke': KDO anketira KOGA o ČEM, kar predstavlja prototipični vzorec konkretnega glagolskega pomena. V drugem koraku smo udeleženska mesta zapolnili s t. i. semantičnimi tipi, ki so abstraktni zastopniki besed, ki se tipično pojavljajo na teh mestih: ČLOVEK ali USTANOVA anketira LJUDI o določeni TEMI ali PROBLEMATIKI. V tretjem koraku smo prototipični stavčni vzorec s semantičnimi tipi ustavčili in dodali t. i. »pomenski scenarij«. Njegova vloga je vključiti razlagalni del in dodati okoliščine (tudi pragmatične), ki so potrebne za ustrezno evociranje pomena: če ČLOVEK ali

⁵⁵ Projekt nastaja pod vodstvom Charlesa J. Fillmoreja na Mednarodnem inštitutu za računalništvo na Kalifornijski univerzi v Berkeleyju. Teorija in praksa projekta FrameNet je podrobno predstavljena v posebni številki International Journal of Lexicography, Vol. 16, No. 3, september 2003, Special Issue: Framenet and Frame Semantics, z vidika slovenščine pa v prispevku S. Kreka (2008).

USTANOVA anketira **LJUDI**, jim zastavlja vnaprej pripravljena vprašanja o določeni **TEMI** ali **PROBLEMATIKI**, o kateri želi pridobiti javno mnenje ali zbrati določene podatke. Predvideli smo, da bo iz tako oblikovanih pomenskih shem mogoče izdelati stavčne definicije, ki bodo odražale najverjetnejšo pomenskoskladenjsko realizacijo pomena besed, stalnih zvez in frazeoloških enot v realni komunikaciji.

V LBS smo pomenske sheme (v spodnjih primerih v okvirčkih) zapisovali pri vseh pomenih in podpomenih glagolov ter pri »vezljivih« pomenih nekaterih samostalnikov in pridevnikov:

mreža samostalni

6 povezan sistem

mreža **USTANOV** ali **NAPRAV** je javni sistem, ki s svojim delovanjem zagotavlja storitve na določenem območju

neprilagojen pridevnik

3 neprimeren za bivanje

če je **ČLOVEK**, **RASTLINA** ali **PREDMET** neprilagojen **RAZMERAM** v okolju, nima za bivanje ali delovanje v njem ustreznih lastnosti ali pogojev

Posebno pozornost smo namenjali ugotavljanju prototipične udeleženske zgradbe, kjer smo upoštevali tudi okoliščine, če so se pokazale kot tipični del stavčne zgradbe posameznega pomena. V spodnjih primerih so udeleženci obarvani modro, okoliščine pa oranžno. Ustreznost zapisa pomenske sheme potrjujejo tudi korpusni zgledi:

izkašljati glagol

2 izročiti denar

če **ČLOVEK** ali **USTANOVA** izkašlja **DENAR** drugemu **ČLOVEKU** za konkretni **NAMEN** ali iz določenega **VZROKA**, ga izroči ali izplača, navadno proti svoji volji ali pod prisilo

- 5 milijonov odpravnine, ki jih je moral ob ločitvi izplačati prvi ženi Barbari, in 5 milijonov enkratne preživnine za otroka, ki jih je izkašljal Angeli.

bučati glagol

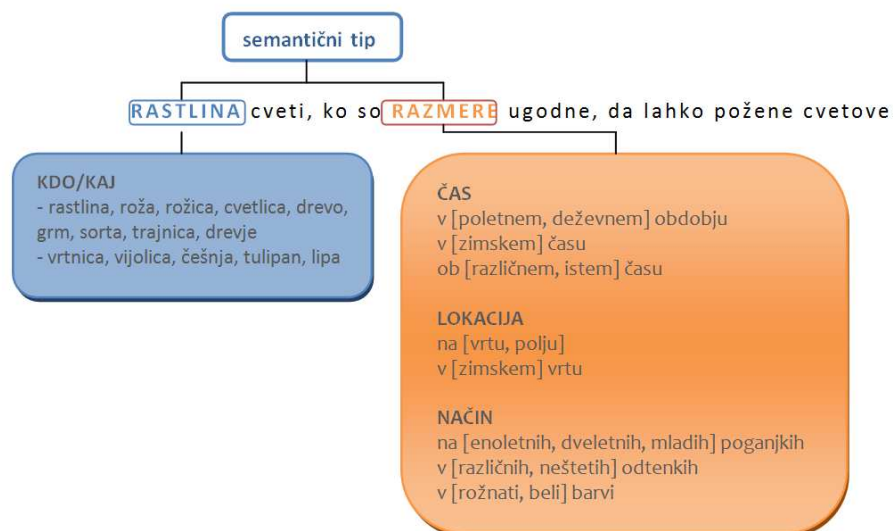
1 oddajati glasen zvok

1.3 o množci

če **MNOŽICA** ljudi ali živali buči, oddaja glasen zvok, navadno zaradi živahnega **DOGAJANJA** ali **VZNEMIRJENJA**

- Delavnica je kar bučala od silnega tekanja in prekladanja paketkov.

Kot je razvidno iz zgornjih primerov, smo udeležence in okoliščine, ki predstavljajo argumentno zgradbo posameznega (pod)pomena, v pomenski shemi zapisovali z velikimi črkami zaradi možnosti avtomatičnega pridobivanja prototipičnih vzorcev in zaradi povezovanja elementov pomenske sheme s kolokacijami in stavčnimi vzorci v procesu računalniške obdelave jezika, kot ponazarja Slika 38:



SLIKA 38: POVEZAVA MED ELEMENTI POMENSKE SCHEME IN KOLOKACIJAMI, ZABELEŽENIMI V LBS

Z določitvijo števila udeležencev so bile določene tudi njihove sklonske oblike in razporeditev v stavčni zgradbi, ki naj bi – z namenom približati se čim bolj naravni diskurzni situaciji – odražala tipično skladijsko realizacijo iztočnice v stavku:

obupan pridevnik

če je **ČLOVEK** obupan se počuti nesrečnega in nemočnega, ker je izgubil upanje in voljo, da bi karkoli spremenil na bolje

lahkomiseln pridevnik

ČLOVEK je lahkomišel, če ne razmišlja dovolj o posledicah svojih dejanj ali mu zanje ni mar

Od prototipičnih stavčnih vzorcev smo na podlagi Hanksove teorije jezikovnih konvencij in možnosti njihove izrabe (Hanks 2013) ločevali alternacijske vzorce, ki smo jih zapisovali v ločenem segmentu geselskega članka (gl. poglavje 3.7.2.3.3 – Pomenske sheme), in ustvarjalne ter »napačne« rabe, ki z vidika beleženja v leksikalni bazi niso zanimive, saj ne predstavljajo splošno sprejetih vzorcev izražanja konkretnega besednega pomena.

3.7.2.3.4 DEFINICIJSKE STRATEGIJE

Opisi pomena (razlage ali definicije), ki ne vsebujejo obveznih elementov pomenske sheme, kar pomeni da iztočnica ni nujno integralni del pomenskega opisa in da udeleženci niso navedeni oz. niso izpostavljeni v obliki semantičnih tipov, so v leksikalni bazi vezani primarno na samostalniške iztočnice, na primer:

morala samostalnik

1 družbene norme

norme in pravila, ki uravnavajo življenje ljudi v družbi tako, da določajo, katero ravnanje je sprejemljivo in katero ne

2 trdno prepričanje; zavest

pri športu ali v vojski *trdno prepričanje, pogum in pripravljenost človeka za delovanje*

3 nauk

morala ZGODBE je nauk ali sporočilo, ki iz nje izhaja in je namenjen temu, da bralca ali poslušalca o čem pouči ali s čim seznani

V LBS smo pri oblikovanju pomenskih shem in razlag upoštevali razlagalne strategije, ki temeljijo na že omenjenih projektih Frame Net in CPA ter na analizi stavčnih razlag, kot so bile zasnovane pri slovarskih projektih založniške hiše Collins Cobuild. Bistvo celostavnih razlag, ki so jih v slovarjih, primarno namenjenih tujcem za učenje angleščine kot tujega jezika, zasnovali J. Sinclair in sodelavci, je v tem, da če/ko-stavek postavi definicijo v metajezikovni okvir, v katerem je najbolj običajna raba iztočnice vključena v besedilo razlage, zaradi česar jo uporabnik lažje uzavesti (Barnbrook 2002: 178). Leksikografska enačba temelji na dvodelni zgradbi razlage, v kateri je pomen iztočnice (definiendum) opisan s pomočjo drugega dela (definiens). Levi in desni del definicijske strukture sta med seboj povezana bodisi z glagolom *biti* bodisi s kakim drugim veznim glagolom, ali pa z veznikom tipa *če/ko*. Vloga levega dela je predvsem v tem, da izpostavi tipične skladienske vzorce in besedilno okolje, v katerih se pojavlja iztočnica (kar je še posebej odločilnega pomena pri glagolu), desni, razlagalni del pa prispeva tisto, kar želimo (novega) povedati o pomenu konkretne iztočnice v prepoznanem besedilnem in skladienskem okolju.

Pri oblikovanju stavčnih razlag smo v LBS težili k ohranjanju razvidne pomenskoskladienske zgradbe iztočnice, še posebej, kadar je bila ta pomensko odločilna, vendar pa smo v primerih, ko je zapis sicer za za rabo konkretnega pomena besede ustreznejše skladienske strukture pomenil manj razumljivo ali za uporabnika težje procesljivo razlago, sledili merilom berljivosti. V spodnjem zgledu je tak primer naveden v okvirčku na desni strani:

če se ČLOVEK **krega** z drugim ČLOVEKOM |ali| če se LJUDJE **kregajo** med sabo o **VSEBINI česa**, /o tem navadno ježno in odločno izražajo drugačno mnenje

če se ČLOVEK **krega** z drugim ČLOVEKOM |ali| če se LJUDJE **kregajo** med sabo, / navadno ježno in odločno izražajo drugačno mnenje o **VSEBINI česa**

Pri oblikovanju pomenskih opisov, tako na ravni pomenskih shem kot razlag, smo na podlagi Cobuildove definicijske taksonomije, ki je primarno služila računalniški obdelavi definicij in ustrezni formalizaciji za potrebe skladienskega razčlenjevanja definicij, določili posamezne tipe razlag, ki smo jih uporabljali glede na razlikovalne lastnosti, ki se nahajajo praviloma v prvem, tj. desnem delu definicije, pa tudi glede na besednovrstno pripadnost besede v iztočnici.

Tip razlage **brez če/ko stavka**, v kateri sta levi in desni, razlagalni del povezana med seboj s preprostim veznim členom, najpogosteje z glagolom *biti* ali z drugim glagolom v tej vlogi (npr. biti sestavljen iz, vključevati, nanašati se na), smo najpogosteje uporabljali pri pomenskem opisu samostalniških pomenov, pri čemer smo iztočnico iz razlage – kadar torej nismo navajali pomenske sheme – izločili:

branjevka *samostalnik*

prodajalka

ženska, ki prodaja sadje in zelenjavo na trgu

ohladitev *samostalnik*

2 zmanjšanje intenzivnosti

2.3 o navdušenju, zanimanju

ohladitev je upad NAVDUŠENJA ali ZANIMANJA, navadno zaradi neprijetnega DOGODKA

Tip razlage **s če/ko stavkom**, ki oblikovno ustreza pogojnemu stavku, smo uporabljali predvsem za opisovanje glagolskih pomenov ter za opisovanje samostalniških in pridevniških pomenov, ki se tipično aktivirajo v povedni rabi

svitati se *glagol*

2 dojemati

2.1 nekaj vedeti

če se ČLOVEKU svita o nekem DEJSTVU, potem o tem nekaj ve ali sluti

siliti *glagol*

4 segati

4.4 o delu obraza, telesa

DEL TELESA ali OBRAZA sili v neko SMER, če se tja obrača ali pomika

Tip stavčne razlage, ki vključuje dodaten metajezikovni element in ima obliko t. i. **projekcije** (Barnbrook 2002: 152), smo uporabili pri opisu pomenov, kjer je bilo posebej pomembno izpostaviti okoliščine rabe:

cviliti *glagol*

2 oddajati cvileč zvok

2.2 o glasbilih

če rečemo, da GLASBILO cvili, želimo povedati, da iz njega prihaja neprijeten zvok, navadno zato, ker kdo ne zna igrati

Tip razlage, v kateri smo zaradi elegantnejše in/ali bolj berljive razlage **iztočnico postavili v neimenovalniško obliko** smo uporabili predvsem pri pridevniških pomenih:

deževen *pridevnik*

1 o dežju

1.1 z veliko dežja

v deževnem OBDOBJU ali KRAJU veliko dežuje

Tako imenovani semantičnocentrični princip organizacije leksikalnogramatičnih podatkov v LBS predvideva, kot smo omenili že v uvodu razdelka o leksikalni bazi, opis leksikalnih enot na vseh nivojih leksikalnobazne zgradbe. Bistveno pri tem je, da kot leksikalne enote razumemo:

- posamezne pomene in podpomene besed
- stalne besedne zveze in
- frazeološke enote

Vse enote besedišča torej, za katere glede na analizo korpusnih podatkov ugotavljamo, da potrebujejo pomensko razlago. Posledično to pomeni, da njihovega pomena navadno ni mogoče ugotoviti neposredno iz posameznih sestavin, pač pa se ta ustvarja na novo. Podobno kot pri posameznih (pod)pomenih tudi večbesedne enote predvidevajo vse ključne elemente geselskega članka, in sicer:

- ▶ večbesedno leksikalno enoto (bodisi na ravni iztočnice ali kot del geselskega članka)
- ▶ njene oblikovne in skladenjske variante
- ▶ pomenski indikator
- ▶ različne tipe oznak
- ▶ razlago
- ▶ tipično besedilno okolje v obliki kolokacij
- ▶ zglede rabe

Kot večbesedne leksikalne enote v LBS obravnavamo **stalne zveze** in **frazeološke enote**. Za oboje velja, da so strukturno in pomensko ustaljene, pri čemer taka ustaljenost ni absolutna, pač pa se kaže v različnih stopnjah v odnosu do enobesedne leksike oz. do pomenskih lastnosti sestavnih elementov, obravnavanih kot samostojne iztočnice.

3.7.2.3.5 STALNE BESEDNE ZVEZE

Kot ugotavljajo leksikološke analize (Cowie 1998: 9), predstavljajo stalne zveze obsežen del besedišča posameznega jezika. Zlasti produktivno naj bi bilo ustvarjanje večbesednih leksikalnih enot pri poimenovanju nove predmetnosti, hkrati pa je prepoznavanje stalnih zvez in njihov pomenski opis zelo pomemben pri učenju tujega jezika. Raziskava uporabniškega obnašanja pri rabi slovarjev (prim. Rozman 2010: 30), je denimo pokazala, da slovarske prakse, ki prikazujejo posamezne sestavine stalnih zvez (npr. *čarterski za zvezo čarterski polet*) bodisi izolirano (gl. spodaj način predstavitve v SSKJ) bodisi kot del podgesla, niso usklajene s potrebami uporabnikov, ki jih zanima pomen in raba zveze kot celote:

čárter in chárter -ja [čár-] m (á) *pogodba o najemu ladje ali letala za določen prevoz ali listina o tej pogodbi*: registracija čarterjev // *prevoz na osnovi takih pogodb*: letalski čarter se močno razvija; neskl. pril.: uvedli so čarter polete med Evropo in Afriko; čarter prevozi ♪

čárterski tudi chárterski -a -o [čár-] prid. (á) *nanašajoč se na čarter*: čarterski polet, prevoz / letalsko čartersko podjetje **čártersko** tudi chártersko prisl.: podjetje posluje tudi čartersko ♪

Rešitev, ki jo ponuja SSKJ, namreč od uporabnika zahteva, glede na to da zveza kot celota v slovarju ni navedena, da o pomenu besede (*čarterski*) sklepa na podlagi definicije pomena druge besede (*čarter*), pri čemer je ta sama na sebi večpomenska. V sodobnih leksikografskih pristopih k problematiki večbesednih leksikalnih enot je zato navajanje stalnih besednih zvez, ki v splošni rabi izstopajo po pogostosti ali sugerirajo samostojni pomen, kot samostojnih enot s celostnim pomenskim opisom, običajna praksa.

Kot stalne besedne zveze smo v LBS prepoznavali večbesedne enote, za katere je bilo na podlagi analize korpusnih podatkov mogoče ugotoviti:

- ▶ da so strukturno in pomensko ustaljene oz. da njihove variantne sestavine pripadajo istemu/sorodnemu pomenskemu polju (*debeli/krepki tisk; morilska čebela/čebela ubijalka; odzivni/reakcijski čas*);
- ▶ da so vsaj delno pomensko neprozorne – tj. imajo samostojni pomen (*spalna srajca, glasbeni stolp, osnovna šola; veliki petek*), zaradi česar jih je smiselno obravnavati na enak način kot enobesedno leksiko;
- ▶ da njihov celostni pomen (ne pa nujno tudi motivirajoči pomen oz. način pomenskega združevanja sestavnih elementov) ni metaforičen, po čemer se ločujejo od frazeoloških enot (*tiskarski škrat, rumeni tisk*);
- ▶ da so lahko, kot tudi sicer leksikalne enote, večpomenske (*topla greda* – 1 ‘del vrta’; 2 ‘proces v zemeljskem ozračju’);
- ▶ da je njihova raba pogosto omejena na določeno terminološko oz. specializirano področje (*kisli dež* – ekologija; *pomožni glagol* – jezikoslovje; *notranji glas* – psihologija);
- ▶ da imajo pogosto ob dobesečnem (besednozveznem ali kolokacijskem) še osamosvojeni besednozvezni pomen (*orjaški dežnik* – ‘velik dežnik’ : ‘goba’) oz. konkretnega referenta (*rdeči križ, kisl juha*);
- ▶ da imajo kot samostojne leksikalne enote pogosto tudi lastno kolokabilno okolje (učinek *tope grede*, povzročati *toplo gredo*; dokončanje, izgradnja, gradnja *avtocestnega križa*);
- ▶ da njihovi tujejezični ustrezniki pogosto niso dobesečni prevodi ali večbesedne enote (*jedilna/prava/temna čokolada* – ang. plain/cooking chocolate; nem. echte/reiner Schokolade; *osnovna šola* – ang. elementary/primary school, nem. Grundschule).

Znotraj geselske strukture smo v LBS beležili stalne zveze na dva načina. V primeru izkazane pomenske povezave med pomenom zveze kot celote in katerim od pomenov besede v iztočnici, smo zvezo navedli v okviru danega pomena. V primeru, kjer take pomenske povezave na podlagi analize jezikovne rabe ni bilo mogoče ugotoviti, pa smo stalne zveze navajali kot samostojne, od posameznih registriranih pomenov iztočnice, neodvisne enote.

V spletni postavitvi slovarja, kjer prikazujemo stalne zveze v samostojnem meniju pri iztočnicah, kjer se realno pojavljajo, se tak način obravnave kaže v tem, da se ob odprtih pomenih, ki vsebujejo stalne zveze, ustrezno filtrira tudi seznam vseh stalnih zvez: tako so na ravni celotnega gesla prikazane vse stalne besedne zveze, na ravni posameznih “odprtih” pomenov pa le tiste, ki se nanj nanašajo (gl.

poglavje 2.2.1, Slika 15). Pri stalnih zvezah znotraj registriranih pomenov na ta način v leksikalni bazi in v spletnem slovarju ohranjamo tudi informacijo o pomenu iztočnice, ki je z besedno zvezo v pomenski soodvisnosti. Spodnji primer prikazuje obe opisani rešitvi v geselskem članku pridevnika *bazičen*:

bazičen *pridevnik*

1 o kemični sestavi; alkalen

2 osnovni; temeljni

bazična industrija *industrija, ki se ukvarja s predelavo surovin v materiale*

bazična znanost *znanost, ki preučuje temeljne naravne značilnosti in nima neposredne uporabne vrednosti*

bazična raziskava *raziskava, ki predstavlja osnovo za razvoj znanja na določenem področju in v trenutku, ko poteka, nima neposredne uporabne vrednosti*

bazično cepljenje *zdravstvo osnovno, začetno cepljenje, ki navadno zahteva tri vnose cepiva v določenem časovnem razmaku*

2.2 v športu

bazični šport *šport*

bazična ocena *gimnastika*

bazična točka / bazična vrednost indeksa

bančništvo

vrednost, ki znaša 1/100 odstotne točke ali 0,01 % in izraža razliko v donosu med obveznicami

3.7.2.3.6 FRAZEOLŠKE ENOTE

V LBS za razliko od stalnih besednih zvez, frazeološke enote (FE) vedno navajamo od posameznih pomenov besede v iztočnici neodvisno, pri čemer predvidevamo slovarski opis v celotni vertikali, torej z vsemi pripadajočimi elementi geselske zgradbe, kot je prikazano v spodnjem primeru:

frazeološka iztočnica in variantne oblike

naliti/natočiti komu čistega vina | naliti si čistega vina | naliti/natočiti komu čistega vina o čem/glede česa

navadno v političnem kontekstu *če rečemo, da smo komu nalili čistega vina, želimo poudariti, da smo mu povedali resnico ali dej tva o kaki neprijetni stvari ali dogodkih*

- *Nalijmo si čisteg a vina, v Muri je nekaj vredna le blagovna znamka.*
- *Francija bo morala domači javnosti naliti čistega vina glede Evrope.*
- *Guverner pa je na koncu nalil vsem čistega vina z besedami, da razpoložljivega kapitala za privatizacijo bank v*

razlaga

kontekstualna
oznaka

korpusni zgledi

Frazeološke enote v LBS pojmujeemo v skladu s frazeološko teorijo, ki mednje šteje tiste enote besedišča, ki glede na svoje sestavne elemente izkazujejo samostojni, od sestavnih elementov neodvisni – t. i. *frazeološki pomen*. Ker pa ta pogoj velja tudi za prepoznavanje stalnih besednih zvez, smo za FE upoštevali še metaforičnost (v nekaterih teorijah pojmovano kot ekspresivnost, slikovitost ipd.) pomena enote kot celote. Med FE, ki ustrezajo zgornjim merilom, smo pri izdelavi LBS izpostavili dva osnovna tipa:

- FE, ki opravljajo v stavku vlogo stavčnega člena, tj. pretežno, a ne nujno, z zgradbo nestavčne besedne zveze, npr. *rešilna bilka, trojanski konj, princ na belem konju, občutiti kaj na lastni koži; zob časa načne koga/kaj, mrak pade na oči komu* ipd.;
- FE, ki so same na sebi »zaključena« besedila in imajo navadno strukturo stavka. Sem sodijo zlasti pregovori in reki (*kadar/ko mačke ni doma, miši plešejo; tresla se je gora, rodila se je miš; ljubezen je slepa* itd.) ter izrazi s s pretežno pragmatično vlogo, ki se kaže v nanašanju na vsebino sporočanega, npr. *če me spomin ne vara*, ali v vrednotenju udeležencev, okoliščin, splošnih nazorov in prepričanj, vezanih na govorca oz. udeležence v govorni situaciji, npr. *kaj iščeš/iščete tukaj?!, boli koga kurac (za koga/kaj), jok, brate/bratce* itd. V zadnjo skupino prištevamo tudi ustaljene konverzacijske formule in izraze, kot so pozdravi, opozorila, kletvice ipd.: *čestitamo, hvala bogu, korajža velja, mater!, fant moj ...* ipd.

Tipična značilnost frazeoloških enot je, kljub »pregovorni« pomenski in strukturni trdnosti, variantnost, in sicer tako pri posameznih sestavinah kot pri besednih oblikah in »zunanjih« udeležencih: *luč na koncu tunela/predora, stopicati/capljati na mestu; stisniti pest/pesti; biti trn v peti komu/koga, pasti v kremplje komu/čemu/koga*. Te vrste variante smo v LBS navajali znotraj ene same frazeološke enote za poševnico.

Nekoliko drugače smo v LBS obravnavali zapis FE v primeru **variantnih oblik rabe** in **pretvorbenih možnosti**. V prvem primeru gre za variantnost besednih oblik, ki so posledica oblikoslovnega prilagajanja znotraj konkretne skladenjske kombinacije. Tipični primeri so, ko katera od sestavin FE, npr. variantni predlog, glagol ipd., zahteva drugačno sklonsko obliko določila, npr. *stisniti skozi zobe* in *stisniti med zobmi*, *stisniti koga v kot* in *stisniti koga ob zid*, *zaiti v slepo ulico* in *znajti se v slepi ulici*, ali ko variantnost oblike vpliva na razporeditev udeleženskih vlog, ki jih predvideva FE: *srce se stisne komu* in *kaj stisne srce komu*. O pretvorbenih možnostih pa govorimo v primerih, ko gre za:

- (a) različne skladenjske oz. stavčnočlenske funkcije FE, npr. *zdrava pamet – biti zdrave pameti – pri zdravi pameti – (narediti kaj) po zdravi pameti*;
- (b) različne realizacijske možnosti,⁵⁶ ki jih omogoča zgradba FE, zlasti posamostaljenje, npr. *prati možgane – pranje možganov, stopicati na mestu – stopicanje na mestu*, in potrpnjenje, npr. *dvigati prah – prah se dviga*;
- (c) različno število in razporeditve udeležencev, npr. *prepustiti se na milost in nemilost komu/čemu – prepustiti koga na milost in nemilost komu/čemu – biti prepuščen na milost in nemilost komu/čemu*;
- (d) tipične realizacije v kateri od upovedovalnih oz. naklonskih možnosti, kot so zanikanje, velelnost, vprašalnost ipd., npr. *kaj pade na pamet/misel/glavo/um komu – kaj ne pade niti na kraj pameti komu – da ti na kraj pameti ne pade*;
- (e) fakultativne sestavine FE: *pasti kot (zrela) hruška* in *pasti komu (v naročje) kot (zrela) hruška* ipd.

V LBS smo take primere beležili znotraj enega nadelementa frazeološka enota zaporedoma v elementih enota, če ni prihajalo do bistvenih pomenskih sprememb (gl. tudi poglavje 2.2.1, Slika 18). Pretvorbene

⁵⁶ V primeru, ko se katera od pretvorb izkaže kot edina možna, npr. *kopja se lomijo* (*lomiti kopja), navajamo samo to obliko.

možnosti in vseh vrst variante smo seveda beležili le, če so se kazale kot tipične in za posamezno FE relevantne.

Pomembna novost, ki jo pri obravnavi frazeoloških enot vnašamo v slovensko leksikografsko prakso, je vključevanje stavčnih razlag in metajezikovnih komentarjev pri opisu frazeoloških enot, kjer so pragmatične komponente sestavni del pomenske informacije:

absoluten *pridevnik*

- ♦ **pasti na absolutno ničlo**

če ugled ali človekovo počutje pade na absolutno ničlo, je v tako slabem stanju, da bi se težko še poslabšalo

alkohol *samostalnik*

- ♦ **utopiti/utapljati kaj v alkoholu | utopiti se/utapljati se v alkoholu**

če se človek utaplja v alkoholu ali utaplja v alkoholu svoje skrbi in težave, se omamlja z alkoholnimi pijačami, da bi na težave pozabil

fasati *glagol*

- ♦ **fasati minuse**

neformalno če rečemo, da kdo faše minuse, želimo povedati, da se mu to, kar je rekel ali naredil, šteje v slabo

3.7.2.4 OZNAKE IN RESTRIKCIJE

Pojasnila o tem, kako je besedo, njeno varianto, pomen ali podpomen smiselno rabiti v kontekstu v smislu, katere slovnične ali skladenjske omejitve oz. preference predvideva, na katerem strokovnem področju se rabi, kakšen odnos do vsebine ali udeležencev izraža, na katero časovno obdobje, besedilni tip, regionalne ali kulturološke posebnosti je vezana ipd., se v jezikovnih priročnikih slovarskega tipa navadno navajajo v obliki krajših pojasnil, kot so kvalifikatorji ali oznake (ang. *labels*).

V LBS, ki temelji na korpusni analizi velike količine realnih jezikovnih podatkov, čim večji avtomatizaciji leksikografskih postopkov in je primarno zasnovana za izdelavo slovarja na spletu, se pridobivanje, beleženje in prikazovanje tovrstnih podatkov v marsičem razlikuje od tradicionalnih leksikografskih pristopov, vezanih na analizo listkovnega gradiva, zanašanje na intuicijo in knjižni format. Velik del informacij tega tipa, zlasti slovničnih, ki so oblikovno prepoznavne, je namreč iz korpusa mogoče pridobiti avtomatsko. Na eni strani gre za informacije, ki leksikografa opozarjajo na določeno kontekstualno obnašanje besed, kot je denimo tipično sopojavljanje s števili (*čakra, finalist*), količinskimi izrazi (*denar*) ali lastnimi imeni (*film, mesto*), tipično pojavljanje v povratnih se/si-stavčnih konstrukcijah (*izvoliti*), v nedoločniku (*blažiti, brneti*) in v zvezi z glagolom *biti* (*všeč*). Na drugi strani pa za informacije, ki jih je mogoče neposredno prepisati v slovarske oznake, npr. tipično v 3. os. (*boleti, brneti*), navadno v množini, pogosto zanikano (*nakloniti*) ipd.

Tovrstne Informacije je iz korpusa mogoče pridobiti na dva načina: prek gramatičnih relacij v slovnici besednih skic v orodju Sketch Engine, kjer posamezno relacijo v osnovi sestavljajo trije elementi: (procesna) direktiva, ki določa tip odnosa v gramatični relaciji, ime in poizvedba. Znotraj poizvedbe je mogoče skladenjske odnose definirati glede na en sam element (t. i. direktiva *UNARY), ki izpostavlja en pojav v odnosu do vseh ostalih elementov v korpusu. Za namene pridobivanja oznak slovničnega tipa je

glede na poizvedbo – tj. konkretno besedo v iztočnici – mogoče pridobiti podatek o tem, da je nek pojav, npr. množinska oblika, sopojavljanje s količinskimi izrazi statistično izstopajoč. Drugi način avtomatičnega pridobivanja se nanaša na področne oznake, kjer je mogoče v orodju Sketch Engine s pomočjo t. i. semenskih besed (*seed words*) in spletnega iskalnika sestaviti specializiran korpus za določeno področje,⁵⁷ posledično pa je glede na področje mogoče ustrezno označiti tudi specializirano leksiko (Rundell in Kilgariff 2011).

V leksikalni bazi predvidevamo več tipov oznak: področne, diskurzne oz. kontekstualne, slovnične oz. skladijske, stilne, pragmatične in časovne. Vsak tip oznak ima znotraj slovarske baze svoj atribut, kar nam je prišlo prav pri poenotenju oznak znotraj baze, kjer nabor ni bil vnaprej določen, pri avtomatski ekstrakciji določenega tipa besedišča (npr. terminologije določenega področja), predvsem pa je razločevanje med posameznimi tipi oznak omogočilo uporabo različnih slogov pri prikazovanju oznak v spletnem slovarju.

Največ besedišča je v LBS označenega glede na tipično področje rabe, in sicer smo **področno oznako** uporabili v zvezi z besedami, besednimi zvezami ali njihovimi pomeni, pri katerih je bila raba omejena na določeno strokovno področje:

blokirati *glagol*

1 omejiti gibanje

1.2 zaustaviti udarec

šport

blokirati [strel, napad], blokirati z [roko]

aktiven *pridevnik*

1 dejaven

1.5 ki vključuje različne dejavnosti

aktivna varnost

avtomobilizem

več lastnosti in naprav, ki so namenjene zmanjšanju tveganja avtomobilskih nesreč

Diskurzne ali **kontekstualne** oznake za razliko od vezanosti na določeno strokovno področje označujejo omejitev rabe besede ali pogostejše katerega od njenih pomenov na tipične komunikacijske situacije, besedilne tipe ali diskurze:

bobnati *glagol*

3 veliko govoriti

v političnem kontekstu

če ČLOVEK bobna, veliko govori, navadno o PROBLEMATIČNI TEMI

zbor *samostalnik*

1 skupina igralcev

zlasti v grški drami

S **slovničnimi oznakami** smo v LBS označevali omejitve besede, zveze ali pomena glede na (a) skladijsko vlogo, npr. »v samostalniški rabi« (*latino*), »v povedni rabi« (*balzam, tema*); (b) določeno

⁵⁷ Več o tem (Baroni & Bernardini, 2004) in (Baroni idr. 2006).

slovnično obliko, npr. »navadno v množini« (*brzica, ikra*), »nešteвно« (*grozdje, tangice*) in (c) določeno vrsto zapisa, npr. »navadno v velelniku« (*klikniti, paziti*), »ustaljena oblika« (*tešče, bore*):

klikniti *glagol*

1 aktivirati s pomočjo računalniške miške

računalništvo

pogosto v velelniku

če ČLOVEK klikne izbrani PREDMET ali VSEBINO na računalniškem zaslonu, ga odpre ali aktivira s pomočjo tipkovnice ali računalniške miške

- *Kar kliknite! Na zaslonu se vam bo pokazal popisan oblaček.*
- *Odprite ustrezno mapo in dvakrat kliknite ime izbrane datoteke.*

Ko smo želeli na določeno slovnično omejitev opozoriti na ravni skladenjskih struktur, smo podatek zabeležili v samostojnem elementu neposredno za tistim delom strukture, na katerega se slovnična omejitev nanaša:

izkušen *pridevnik*

2 izurjen in spreten

2.1 v športu

Kol-rbz PBZO <r>**primernik**</r>

[precej, veliko] **izkušenejši**

PBZO <r>**presežnik**</r>v sbz5

najizkušenejši v [moštvu], **najizkušenejši** v [naši] vrsti



Stilne oznake smo razdelili na (a) **registrske** za označevanje rabe besede oz. pomena, ki izkazuje v odnosu do nevtralnega besedišča posebnosti glede na vrsto komunikacijske situacije, npr. v neformalnih govornih situacijah (*kasirati, šparati*), ali besedišča, ki ga uporabljajo samo določene skupine govorcev, npr. mladostniki (*keš*), ji se uporablja v govoru z otroki (*ritka*), in na (b) **konotacijske** oznake, s katerimi smo označevali rabo besede oz. pomena glede na konotativne pomenske lastnosti, npr. odnos govorca do predmetnosti, sogovorca, vsebine besedila ipd.:

kuža *samostalnik*

1 pes

ljubkovalno

[prijazen, majhen, domač] kuža

klišé *samostalnik*

1 vzorec; ponavljanje

z negativnim vrednotenjem

[obrahljen, ustaljen, dolgočasen] klišé

Pri stalnih zvezah in frazeoloških enotah smo v kombinaciji z indikatorji uporabljali t. i. pragmatične oznake, s katerimi smo v zvezi z rabo določene besede želeli opozoriti na njeno posebno vlogo, vezano na komunikacijsko situacijo, kot je npr. raba besede kot pozdrav, opozorilo, izraz odobravanja, grožnje ipd.

fasati *glagol*

♦ **fasati jih**

kot grožnja

zelo grobo

- »A, bi jih rad fasal?« je zdaj zarjovel Semi in to tako, da je Domna Frasa in Moža, ki je pel in žel, odneslo po rumeni opečnati cesti.
- Pa kaj ta kmet bi jih rad fasal na gobec!?

Časovne oznake smo uporabili v primerih, ko beseda v razmerju do drugih, sinonimnih besed v sodobnem jeziku ohranja časovno distanco, npr.

soldat *samostalnik*

1 vojak

starinsko

- Tommy bo v filmu odigral ameriškega soldata, ki Japonce uči boja proti samurajem.
- Ena izmed legend iz tistega časa govori o nemškem soldtu, ki je zavil iz polka, da bi se olajšal, in se je ob tem zazrl naravnost v oči partizana na straži.

Moč posamezne oznake določa v LBS njena umestitev znotraj geselskega članka. Če je oznaka navedena pred vsemi pomeni in podpomeni, tj. za iztočnico, zajema tudi vse njene pomene in podpomene; če je navedena za indikatorjem osnovnega pomena, velja za vse podpomene, oz. če je navedena za indikatorjem posameznega podpomena, se nanaša zgolj nanj. Pri označevanju stalnih zvez velja oznaka vedno le za določeno stalno zvezo, zato smo oznake pri več zaporednih stalnih zvezah, ki so značilne npr. za isto strokovno področje, vedno ponovili. V spletnem slovarju to omogoča ločen prikaz vsake stalne zveze kot samostojne leksikalne enote z vsemi pripadajočimi informacijami.

3.7.3 SKLADENJSKI NIVO

Skladenjski nivo je v LBS namenjen tipu informacij, ki so pomembne za:

- ▶ avtomatsko pridobivanje leksikalnih in skladenjskih podatkov iz korpusa;
- ▶ nadaljnje računalniško procesiranje v bazi strukturiranih podatkov in uporabo v jezikovnotehnološke namene;
- ▶ raziskovanje skladenjskih lastnosti slovenske leksike v povezavi njenim semantičnim potencialom.

Uresničevanje prvih dveh ciljev zahteva vnašanje podatkov na način, ki bo omogočal računalniško obdelavo slovenščine. Konkretno to pomeni formalizacijo besednih zvez – **skladenjskih struktur**, v katerih se tipično za določen pomen pojavlja beseda v iztočnici, npr. za besedno zvezo *rdeča jagoda*: pridevnik v kateremkoli sklonu + samostalnik v kateremkoli sklonu, s kodami oz. dogovorjenimi simboli (pbz0 sbz) in na način, ki je usklajen z oblikoskladenjskimi oznakami v korpusu. Poleg tega smo za namene avtomatskega pridobivanja leksikalno-gramatičnih podatkov o slovenščini, pa tudi za namene slovničnih analiz v LBS, zapisovali **stavčne vzorce**, ki prikazujejo različne realizacijske možnosti določenega glagolskega pomena, kot se kaže na ravni stavka, npr.

IZTOČNICA	POMEN	STAVČNI VZOREC
cveteti	1 poganjati cvetove	KAJ cveti (kako/kdaj)
	2 prinašati dobiček	KAJ cveti KAJ cveti KOMU
	3 izgledati zadovoljen, zdrav	KDO cveti od ČESA

TABELA 5: STAVČNI VZORCI

Predvidevamo, da bo s tako formaliziranimi opisi in z uporabo skladijskega razčlenjevalnika za slovenščino mogoče iz LBS avtomatično pridobiti kompleksne skladijske in pomenske informacije o sodobnem slovenskem jeziku.

Vključitev skladijskih podatkov, ki zadevajo pomenski potencial slovenske leksike, je v LBS najprej vezana na ugotovitve, da so pomeni neposredno povezani z vzorci in strukturami, v katerih se pojavljajo (Hanks), da pomenske tendence besed vplivajo na izbire v svojem besedilnem okolju in da, kot kažejo korpusne analize, je jezikovna raba vzorčena in hkrati izredno variabilna, kar z drugimi besedami pomeni, da se pomen besed in zvez v stavku lahko uresničuje na različne skladijske načine: neskončno število pomenov je mogoče izraziti s končnim številom skladijskih kombinacij. Da bi bilo te skladijske možnosti izražanja pomena mogoče računalniško procesirati, smo, poleg že omenjenih pomenskih shem, predvideli dva načina formaliziranja skladijskih informacij: zapis skladijskih struktur in stavčnih vzorcev.

3.7.3.1 SKLADIJSKE STRUKTURE

Znotraj vsakega registriranega pomena besede zapisujemo skladijske strukture s pomočjo simbolov, npr.: pbz (pridevnik oz. pridevniška besedna zveza), sbz – samostalnik oz. samostalniška besedna zveza, gbz – glagol oz. glagolska besedna zveza, in rbz – prislov. oz. prislovna besedna zveza. S številko od 1 – 6 ob posameznem simbolu za besedo določene besedne vrste označujemo sklon v strukturi, npr. SBZ0 sbz2 označuje zvezo dveh samostalnikov, od katerih se prvi lahko pojavlja v kateremkoli sklonu, drugi pa obvezno v rodilniku, npr. *spekter znanja, urednik revije*. Zapis dela strukture z velikimi črkami pomeni besedo, ki je obravnavana v iztočnici:

spekter *samostalnik*

1 razpon

SBZ0 sbz2 spekter [znanja]

Oznake za besedne vrste v strukturah, lahko vsebujejo še druge elemente, ki ustrezajo oblikoskladijskim oznakam v korpusu, npr. Inf-gbz – nedoločnik, Vez-gbz – pomožni glagol, Nam-gbz – namenilnik, Neg-gbz – zanikani glagol, Kol-sbz – samostalnik ali samostalniška besedna zveza, ki izraža količino itd.

V izhodišču smo ločili dva tipa skladijskih struktur: (a) regularne strukture pri samostalnikih, pridevnikih in prislovih, ki so tipično (a1) dvodelne ali (a2) trodelne v primeru predložnih zvez, in (b)

posebne strukture, kamor sodijo (b1) glagolske strukture, ki vsebujejo pomožne glagole, nedoločnike, namenilnike ipd., ter (b2) strukture s količinskimi izrazi in (b3)zanikane strukture.

(a1) samostalniške	(a1) glagolske	(a1) pridevniške	(a1) prislovne
SBZ0 num2	rbz GBZ	Vez-gbz PBZ4	rbz RBZ
petek [trinajstega]	[nesrečno, hudo] pasti	[delati se] pomembnega	[nadvse] preprosto
Vez-gbz SBZ1	GBZ Inf-gbz	PBZ0 sbz0	sbz0 RBZ
[izpasti] bedak	dati se [pregovoriti]	svetla [prihodnost]	[odhod, pot] domov
pbz0 SBZ0	GBZ Nam-gbz	PBZ0 sbz2	RBZ sbz1
[boleč] spomin	priti [spodbujati]	poln [denarja]	pasje [mrz]
pbz0 SBZ3	GBZ Odv-da ... vedeti, da	PBZ0 sbz4	num1 RBZ
uporabniku [prijazen]		dolžen [vsoto, denar]	[dvakrat] tedensko
(a2) samostalniške	(a2) glagolske	(a2) pridevniške	(a2) prislovne
SBZ0 blizu sbz2	GBZ ali gbz	PBZ0 glede sbz2	RBZ do sbz2
vir blizu [ministrstva]	kupiti ali [najeti]	negotov glede [izida]	najpozneje do [februarja]
SBZ0 k/h sbz3	GBZ in gbz	PBZ0 do sbz2	RBZ po sbz5
pripombe k [osnutku]	spotakniti se in [pasti]	kritičen do [udeležencev]	takoj po [sestanku]
SBZ0 med sbz658	GBZ Inf-gbz sbz1	PBZ0 k/h sbz3	RBZ s/z sbz6
razlika med [spoloma]	želeti postati [manekenka]	nagnjen k [samomorom]	skladno z [inflacijo]

TABELA 6: DVODELNE IN TRODELNE REGULARNE STRUKTURE

glagolske	s količinskimi izrazi	zanikane
Vez-GBZ pbz1	Kol-SBZ0 sbz2	Neg-gbz Inf-GBZ
postajati [utrujen, naveličan]	ura [hoje, dela], kos [potice]	[ne gre, ne kaže] dvomiti
gbz Inf-GBZ	Kol-SBZ1 pbz0	sbz2 Neg-Vez-gbz Inf-za-GBZ
[nehati, začeti] boleti	meter [dolga], uro [oddaljen]	[česa/koga] ni za prepoznati
sbz0 Inf-GBZ	Kol-RBZ gbz	Neg-GBZ sbz2
[možnost, pravica] pritožiti se	veliko [se smejeti, govoriti]	ne pričakovati [čudežev]

TABELA 7: POSEBNE STRUKTURE

Ob zaključku izdelave LBS smo identificirali nekaj več kot 300 skladenjskih struktur za vse obravnavane besedne vrste. Ti vzorci so nam služili za beleženje kolokacijskih podatkov, kot bo prikazano v

⁵⁸ Samostalnik v tožilniku je v tej strukturi navadno needninski, na kar opozorimo v restrikciji.

nadaljevanju, in, kot že omenjeno, za nadgradnjo slovnice besednih skic za avtomatsko pridobivanje leksikalnih podatkov iz korpusa.

3.7.3.2 STAVČNI VZORCI

Vloga vzorcev pri slovničnem opisu glagolov je pokazati tipične realizacije konkretnega glagolskega pomena v okvirih stavčne zgradbe, ali kot pravi Hanks (2009) »Vsaka beseda, ki jo uporabi govorec, da bi z njo nekaj povedal, je uporabljena v skladu z neko normo (dogovorom), ki jo določa njen stavčni vzorec«. Na podlagi korpusne analize je v posameznih konkordancah te vzorce mogoče prepoznati, kot smo ponazorili že zgoraj s primerom glagola *cveteti*, hkrati pa ugotoviti, da je večina konkordanc dejansko realizacija prepoznavnega, tj. normalnega vzorca, da so številne druge konkordance znotraj jezikovnega sistema alternacije tega vzorca, manj pa je t. i. eksploatacij ali ustvarjalnih besedilnih rab. Za slovarski opis – in s tem tudi za vključitev v LBS – so nas zanimali le regularni vzorci in njihove alternacije, ki jih za izražanje pomenskega potenciala uporablja večina slovenskih govorcev, individualnih, ustvarjalnih in napačnih rab pa nismo upoštevali.

Možnosti alterniranja prototipičnega vzorca, kot je zabeležen v pomenski shemi, prikazujejo primeri zapisov stavčnih vzorcev za posamezne pomene glagolov *briti*, *stisniti* in *razkačiti*:

briti *glagol*

1 odstranjevati dlake

če se MOŠKI brije |ali| če si MOŠKI brije BRADO ali BRKE, si s PRIPOMOČKOM odstranjuje dlake

- **briti se**
- **kdo se brije**
- **briti si kaj**
- **briti se s čim**

stisniti *glagol*

1 uporabiti fizični pritisk

1.19 poškodovati, zmečkati

če NAPRAVA stisne ČLOVEKA |ali| stisne ČLOVEKU DEL TELESA, ga močno poškoduje

- **kaj stisne koga**
- **kaj stisne komu kaj**

razkačiti *glagol*

1 močno razjeziti

če ČLOVEK, njegova DEJANJA ali BESEDE razkačijo drugega ČLOVEKA, ga zelo razjezijo

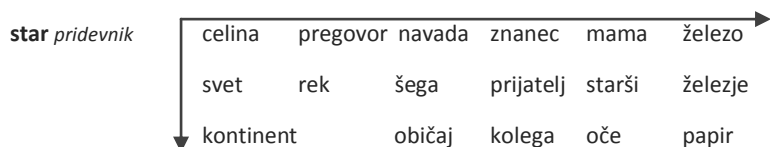
- **kdo/kaj razkači koga**
- **kdo razkači koga s čim**
- **koga razkači zaradi česa**

Stavčni vzorci na ta način potrjujejo argumentno zgradbo, zapisano v pomenski shemi pomena, hkrati pa razkrivajo možnosti opuščanja elementov kot tudi tipične izbire znotraj več skladenjskih možnosti, kar je zlasti očitno pri predložnih zvezah, ki jih beležimo pod elementi: »lokacija/smer«, »čas«, »način« ali

»vzrok«. Zelo pomembno je v zvezi s tem opozoriti, da smo v LBS razlikovali med jezikovnosistemskimi in dejansko realiziranimi možnostmi, tj. ob posameznem pomenu smo vedno navajali zgolj tiste vzorce, ki so se za konkretni pomen izkazovali kot tipični in ne tistih, za katere bi bilo mogoče na podlagi jezikovnega sistema predvidevati, da so možni.⁵⁹

3.7.4 KOLOKACIJE

Kolokacije predstavljajo najočitnejši dosežek korpusne leksikografije in hkrati največjo pridobitev za slovarske uporabnike, zato si sodobnih slovarskih priročnikov brez njih ni mogoče predstavljati. Teoretično so kolokacije statistično prepoznavne sopojava dvih (ali več) besed, pri čemer so možnosti ugotavljanja besedne sopovezovalnosti od trenutka, ko je bila prepoznana njihova neposredna povezanost z naravno rabo jezika in izražanjem njegovega pomenskega potenciala (prim. večkrat citiran stavek J. R. Firtha (1957): »*You shall know a word by the company it keeps*«,⁶⁰ dosegle zelo sofisticirane statistične izračune.⁶¹ Statistično izstopajoča medsebojna povezanost besed je navadno tudi pomensko in posledično jezikovnospecifično pogojena. T. i. izbire besed na sintagmatski – linerarni osi, vplivajo tudi na pomenske izbire na paradigmatiski, horizontalni osi:



To med drugim kaže, da so nekatere povezave v jeziku za govorce sprejemljive, druge pa ne: če je mogoče pridevnika *hud* in *težek* v kombinaciji z besedo *rana* enakovredno zamenjati in torej tvorita sinonimni par: *huda/težka rana*, takih sinonimnih parov v kombinaciji z nekaterimi drugimi besedami, vsaj ne znotraj konkretnega jezika in za izražanje enakih pomenov, ni mogoče tvoriti: **hudo vprašanje*, **težek pes*:



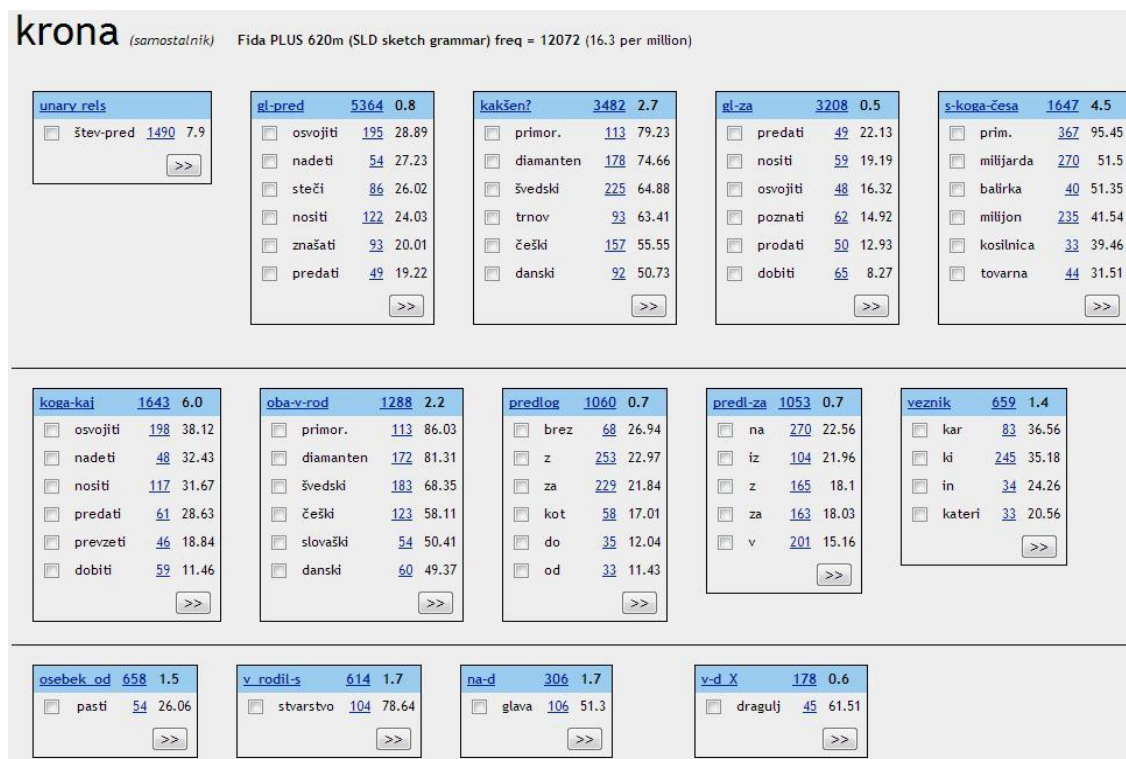
To nas pripelje do ugotovitve, da so kolokacijske povezave besed specifične znotraj jezika samega, njihovo obvladovanje oz. poznavanje pa je nujno za njegovo naravno in smiselno rabo.

V LBS smo kolokacije beležili v povezavi s skladenjskimi strukturami, in sicer na podlagi izdelanih besednih skic v orodju SkE. Spodnja slika prikazuje kolokacijski profil za besedo *krona*, na podlagi katerega so bile prenesene kolokacije pod ustrezne skladenjske strukture v slovarski vmesnik.

⁵⁹ Princip navajanja stavčnih vzorcev v LBS, ki sledi dejanskim realizacijam, se tako v temelju ločuje od jezikovnosistemskega pristopa navajanja stavčnih vzorcev v Vezljivostnem slovarju slovenskih glagolov (Žele 2008, 2011), saj menimo, da je za uporabnika, še posebej šolarja ali učenca slovenščine kot tujega jezika, jezikovnosistemska informacija, ki nima potrditve v realni rabi, napačna in zavajajoča.

⁶⁰ »Besedo poznaš po tem s kom se družiš.« Prevod P. G.

⁶¹ Gl. rubriko VII. (Jezikovne tehnologije) v poglavju Bibliografija



SLIKA 39: KOLOKACIJSKI PROFIL BESEDE KRONA V OBLIKI BESEDNE SKICE V ORODJU SKETCH ENGINE

Če predstavljajo skladijske strukture v obliki končnega števila ponavljajočih se kombinacij skelet jezika, so kolokacije mišice in tkivo, ki se nanj pripenja. Glede na skelet, ki ostaja do neke mere dokončen (končno število skladijskih kombinacij, ki jih omogoča slovnica nekega jezika), so kolokacije bolj življenjske in spremenljive, omejene le z besediščem tega istega jezika. Ko govorimo o kolokacijah, se moramo zavedati, da nam odražajo tipičnost in ne individualne ali kreativne posameznikove rabe.

Glede na stanje, ki smo ga detektirali v korpusu, smo predvideli dva tipa kolokacij: **dvočlenske** oz. **tročlenske kolokacije** v primeru predloženih kombinacij in t. i. **razširjene kolokacije**. Kot samostojni tip besednih povezav, ki presegajo strukturo tročlenske kolokacije, smo v LBS beležili tudi t. i. **skladijske zveze**.

3.7.4.1 RAZŠIRJENE KOLOKACIJE

Poleg običajnih dvočlenskih in tročlenskih kolokacij, ki zapolnjujejo »skelet« skladijskih struktur pri samostalniških, pridevniških in prislovnih iztočnicah in stavčnih vzorcev pri glagolskih, kot v spodnjih primerih,

blondinka samostalnik

1 ženska s svetlimi lasmi

pbz0 SBZ0 [prsata, dolgonoga, postavna] blondinka

sbz0 o SBZ5 [šale, vici] o blondinkah

SBZ0 z/s sbz6 blondinka z [bujnim, velikim] oprsjem

popestriti *glagol*

1 narediti zanimivejše

- kdo/kaj popestri kaj **[glasbenik, gost] popestri**
- kdo popestri kaj s čim **popestriti [ponudbo, jedilnik]**
- kdo popestri komu kaj **popestriti [otrokom, gostom]**
- kdo popestri komu kaj s čim **popestriti z/s [diapozitivi, prikazom]**
- kaj popestri komu kaj **[dodatek, glasba] popestri**

smo beležili tudi t. i. **razširjene kolokacije**, kjer se posamezni deli statistično signifikantnih dvobesednih kombinacij notranje širijo z ustaljenimi leksikalnimi elementi. Za pridobivanje tovrstnih informacij smo uporabili funkcijo »multi word links« v orodju SkE, ki za posamezne kolokacije registrira njihovo nadaljnje besedilno okolje, ki izkazuje statistično pomembne sopovezave:

mreža *samostalnik*

1 predmet

pbz0 SBZ0 [zaščitna] mreža -> **zaščitna [kovinska] mreža**

2 o ribolovu

pbz0 SBZ0 [ribiška] mreža -> **[ročna, vlečna] ribiška mreža**

gbz SBZ4 [krpati, plesti] mreže -> **plesti [ribiške] mreže**

3 športni pripomoček

gbz SBZ4 [zatresti, zadeti] mrežo -> **zatresti [nasprotnikovo] mrežo**

4 pajčevina

pbz0 SBZ0 [pajkova] mreža -> **[ujeti se, zaplesti se] v pajkovo mrežo**

5 vzorec

6 povezan sistem

pbz0 SBZ0 [diplomatska, mednarodna] mreža -> **mednarodna [teroristična, tihotapska] mreža**

SBZ0 sbz2 mreža [poslovalnic, prodajaln] -> **[široka] mreža prodajaln**

mreža [bankomatov] -> **[razvejana] mreža bankomatov**

7 niz dejanj

gbz v SBZ0 [ujeti se, zaplesti se] v mrežo -> **ujeti se v [svojo, lastno] mrežo**

3.7.4.2 SKLADENJSKE ZVEZE

Skladenjske zveze so na splošno rečeno večbesedni nizi/deli jezika, ki nimajo statusa leksikalne enote – ne predvidevajo torej pomenske razlage, za katere veljajo naslednja načela:

- ▶ so (relativno) pomensko prozorne besedne kombinacije, kar pomeni, da je njihov pomen (načeloma) kombinacija pomenov posameznih sestavin;
- ▶ hkrati so dovolj strukturno trdne (ustaljene), kar pomeni, da so posamezni elementi leksikalizirani in jih ni mogoče nadomestiti z drugimi oz. da je niz nadomestljivih elementov omejen, npr: *breme (česa) pade na [pleča, ramena] koga/česa; temperatura pade pod [ničlo, ledišče]* ipd.
- ▶ da imajo pogosto tudi semantično in/ali oblikovno predvidljivo prosto mesto znotraj stavčne zgradbe, v kateri se tipično pojavljajo, npr. *angažirati se v smeri česa*;
- ▶ da imajo lahko za razliko od kolokacij in razširjenih kolokacij znotraj svoje zgradbe več kolokabilnih nizov, označenih z oglatimi oklepaji, npr. *otroci v starosti od [x] do [x] let; cveteti od [maja, junija,*

julija] do *[avgusta, septembra, oktobra]*; *odigrati [tekmo, dvoboj] [x.] kroga končnice*. Pri avtomatskem pridobivanju podatkov iz korpusa in usklajevanju znotraj LBS nam je to omogočalo njihovo enotno obravnavo in priklic vseh tovrstni enot po potrebi.

V LBS smo registrirali različne tipe skladenjskih zvez, in sicer v stopnjah od povsem neidiomatičnih tj. s »prozorno« pomensko zgradbo, kamor smo uvrstili **(a)** zveze s številiškimi elementi (zaradi možnosti avtomatske ekstrakcije smo številske elemente označili z »x«) in **(b)** zveze, ki tipično vsebujejo lastnoimenske elemente. Ker smo želeli ohraniti možnost enotnega obravnavanja in pridobivanja iz leksikalne baze, smo te elemente označili s tremi pikami: [...]:

pomlad *samostalnik*

1 leto

◇ **šteti [x] pomladi**

datiran *pridevnik*

1 zabeležen z datumom

◇ **datiran v čas od [x.] stoletja do [x.] stoletja**

članek *samostalnik*

1 krajše besedilo v časopisu ali reviji

◇ **članek z naslovom [...]**

◇ **članek izpod peresa [...]**

Najbolj tipičen tip SKZ so **predložne zveze** v prislovni funkciji, znotraj tega pa smo ločili:

(c1) s pomensko/skladenjsko predvidljivim vezljivostnim mestom:

angažirati se *glagol*

1 dejavno se vključevati; prizadevati si

◇ **[angažirati se] v prid koga/česa**

◇ **[angažirati se] na področju česa**

(c2) pomensko neodvisne od iztočnice

spomin *samostalnik*

1 pomnjenje česa iz preteklosti

1.1 o dogodkih in ljudeh

◇ **[slikati se, ohraniti, imeti] za spomin**

(c3) pomensko odvisne od iztočnice

slep *pridevnik*

1 ki ne vidi

1.1 zaradi okvare vida

◇ **slep na eno oko**

◇ **slep na [levo, desno] oko**

V naslednji sklop sodijo **prireadne** in **primerjalne** skladenjske zveze

alkohol *samostalnik*

1 alkoholna pijača

◇ **alkohol in [tobak, cigarete, droge, nikotin]**

cviliti *glagol*

1 oglašati se s cvilečim glasom

1.1 o ljudeh in živalih

◇ **cviliti kot [pašček, prašič]**

Primerjalne skladijske zveze se pogosto približujejo FE, saj njihov pomen ni vedno na prvi pogled razviden iz posameznih sestavin. Zlasti posebne so primere, ki se uvrščajo pod katerega od metaforičnih pomenov sestavin ali pa potrebujejo kot celota pragmatična ali žanrska pojasnila, kot npr.

debel *pridevnik*

1 ki ima preveč telesne maščobe

◇ **debel kot [svinja, prašič, pujs]**

v zelo neformalnih situacijah

Odločitev, da jih v LBS uvrščamo med enote, ki jih pomensko ne razlagamo, smo sprejeli predvsem zaradi tega, ker je mogoče njihovo ustrezno rabo in pomen razbrati iz zadostnega števila zgledov, ki posamezno skladijsko zvezo potrjujejo v realnem okolju.

3.7.5 KORPUSNI ZGLEDI

Na zadnjem nivoju zgradbe LBS navajamo korpusne zglede, s katerimi želimo pojasniti vse predhodne v bazi zabeležene podatke. S primeri rabe, ki v celoti temeljijo na korpusu Gigafida in smo jih v programski vmesnik prenašali s pomočjo aplikacije GDEX v orodju Sketch Engine, tako potrjujemo podatke, ki smo jih na višjih nivojih vnesli v LBS, in sicer: s pomenskim indikatorjem opredeljen pomen, podpomen in pomensko shemo z obveznimi in opcijskimi udeleženci, skladijske strukture in skladijske vzorce. S primeri rabe vedno potrdimo tudi realno obnašanje stalnih besednih zvez in frazeoloških enot v besedilu. Pri izbiri primerov smo stremeli k relativni kratkosti. V primeru, da je bilo izbrani primer mogoče krajšati (npr. izpustiti obsežen vrinjen stavek ipd.), smo krajšanje v zgledu ustrezno označili. Če je bilo mogoče, smo se pri izbiri in urejanju korpusnih zgledov izogibali takim, ki so vsebovali lastna imena, predvsem pa smo bili pozorni na to, da smo izbirali primere, ki niso vnašali dvoma v podatke, ki smo jih želeli s primeri potrditi. Sem so sodili denimo primeri rabe, v katerih je bila obravnavana beseda uporabljena bodisi v katerem od individualnih pomenskih odtenkov, bodisi primerom, v katerih se je obravnavana beseda pojavljala v več pomenih. S primeri rabe smo želeli zajeti tudi tipično besedilno rabo (npr. diskurz, premi govor ipd.), tip besedila (npr. osmrtnice), področje rabe (šport) in skladijsko okolje. Pri navajanju števila primerov v leksikalni bazi nismo bili omejeni.

V hierarhično urejenih nivojih LBS tako predstavljajo primeri rabe najkonkretnjšo informacijo in skupaj s pomenskimi, skladijskimi in kolokacijskimi podatki zaključujejo celoto, znotraj katere so informacije med seboj povezane in se medsebojno dopolnjujejo.

4 ORGANIZACIJA DELA

Finančni in organizacijski načrt za izdelavo SSSJ obsega dva tipa stroškov:

1. človeški viri

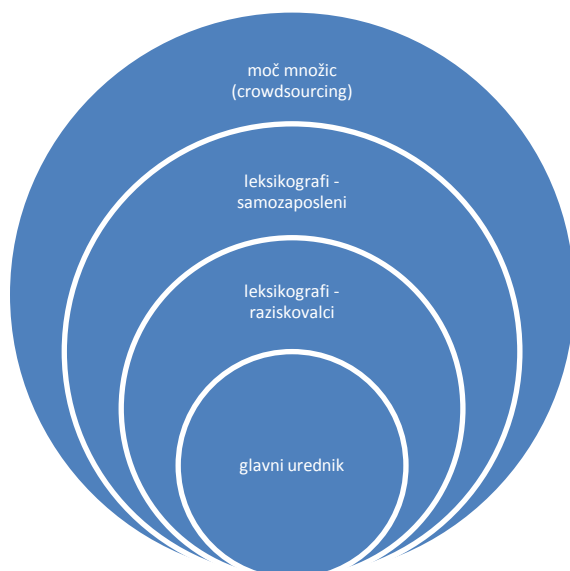
- a. menedžment (vsebinsko in finančno vodenje)
- b. leksikografi – raziskovalci
- c. leksikografi – samozaposleni v kulturi
- d. moč množic (crowdsourcing)
- e. strokovnjaki za pregled terminologije
- f. strokovnjaki za dodajanje normativnih podatkov
- g. strokovnjaki za dodajanje podatkov o izgovoru
- h. strokovnjaki za dodajanje podatkov o izvoru
- i. računalniški strokovnjaki za pridobivanje jezikoslovnih podatkov
- j. vzdrževanje korpusa
- k. računalniška podpora in vzdrževanje
- l. administracija

2. materialni stroški

- a. strojna in programska oprema
- b. posredni stroški

4.1 ORGANIZACIJA LEKSIKOGRAFSKEGA DELA

Vodenje projekta: Projekt vodi glavni urednik, ki je odgovoren tako za organizacijske in finančne kot za vsebinske vidike projekta. Po predvidevanju polovico delovnega časa namenja organizacijskim in finančnim nalogam, polovico vodenju vsebinskega dela. V projektu je za glavnega urednika predvidena ena zaposlitev oz. FTE (*Full-Time Equivalent*), letni bruto strošek delodajalca je 50.000 € na FTE.



SLIKA 40 ORGANIZACIJA LEKSIKOGRAFSKEGA DELA

Leksikografi – raziskovalci: Pričakuje se, da imajo zaposleni leksikografi status raziskovalca (ARRS), opravljajo pa najzahtevnejše leksikografske naloge: pomenska delitev in pisanje definicij pri zahtevnih geslih, končna redakcija slovarskih gesel in podobno. Petino svojega časa (en dan na teden) posvečajo drugim nalogam, ki niso del projekta (raziskave, znanstveno publiciranje ipd.). Sestavljajo tudi ožjo redaktorsko ekipo, ki sodeluje pri odločanju

o vsebinskih rešitvah. V projektu so predvidene štiri polne zaposlitve oz. FTE-ji (*Full-Time Equivalent*), ki se lahko razdelijo na več leksikografov–raziskovalcev. Eden od leksikografov–raziskovalcev se specializira za problematiko sinonimije in skrbi za ta del slovarskih podatkov. Eden iz skupine je imenovan za namestnika glavnega urednika in po potrebi namenja del svojega časa tudi vodstvenim nalogam. Letni bruto strošek delodajalca je 43.000 € na FTE.

Leksikografi – samozaposleni v kulturi: Drugi del leksikografske ekipe sestavljajo leksikografi, za katere ni predvidena pogodba o zaposlitvi, temveč imajo status samozaposlenih v kulturi (poklic: leksikograf). Plačani so honorarno po učinku, bodisi po urni postavki, po številu obdelanih gesel ali po številu znakov. Zanje je predvideno plačilo, ki ustreza štirim FTE in se razdeli med poljubno število sodelavcev. Pri organizaciji tega dela ekipe je potreben dogovor z Ministrstvom za kulturo glede statusa samozaposlenih v kulturi. Letna višina honorarjev na en FTE je 43.000 EUR.

Moč množic (*crowdsourcing*): Pomemben segment dela na slovarju sestavlja množično delo oz. izkoriščanje moči množic, s pomočjo katerega v začetnem delu obdelave gesel čistimo avtomatsko izločene podatke iz korpusa. Za ta namen v projektu uporabljamo spletne servise za izkoriščanje moči množic. Za ta tip dela je predvideno plačilo, sistem je podoben kot pri drugih sorodnih spletnih storitvah, npr. Amazon Mechanical Turk. Predvideno je plačilo na posamezno odločitev, za eno končno odločitev so potrebne tri oz. pet odločitev različnih uporabnikov. Sistem je organiziran tako, da za sodelovanje pri tem delu ni potrebno posebno učenje, zato lahko katerikoli sodelavec takoj vstopi v proces in kadarkoli iz njega izstopi. Predvidevamo, da bo v teku projekta pri tem delu sodelovalo večje število sodelavcev, tipično ne skozi celotno obdobje, temveč le v krajših nekajmesečnih intervalih. Predvideni so tudi popularizacijski ukrepi z namenom, da se v to delo pritegne čim večje število sodelujočih. Cena ene odločitve je 2 centa oz. 0,02 € za manj zahtevne odločitve, kar po predvidevanjih ustreza finančni kompenzaciji približno 7 € na uro (= 350 odločitev). Za zahtevne odločitve je cena 5 centov oz. 0,05 EUR, kompenzacija na uro pa je podobna kot pri lažjih odločitvah, ker je proces odločanja daljši. Pri delu na LBS se je izkazalo, da je potrebnih povprečno 60 končnih odločitev na geslo za zglede ter 20 za kolokatorje in za pripisovanje kolokacij/zgledov posameznim pomenom:

	število končnih odločitev na geslo	število odločitev za eno končno odločitev	število gesel	število vseh odločitev	cena za odločitev	skupni strošek
zgledi	60	3	100.000	18.000.000	0,02 €	360.000€
kolokacije	20	3	100.000	6.000.000	0,02 €	120.000€
pomeni	20	5	25.000	2.500.000	0,05 €	125.000€
SKUPAJ						605.000€

TABELA 8 IZRAČUN STROŠKOV ZA IZKORIŠČANJE MOČI MNOŽIC

4.2 ORGANIZACIJA DELA – DRUGI ČLOVEŠKI VIRI

Strokovnjaki za pregled terminologije: Za pregled terminološko opredeljenih iztočnic ali (pod)pomenov



SLIKA 41 ORGANIZACIJA DELA - DRUGI ČLOVEŠKI VIRI

glavni urednik organizira mrežo terminoloških strokovnjakov za vsa predvidena področja. Mreža je ohlapno povezana s pomočjo sodobnih spletnih platform za potrebe usklajevanja. Finančna sredstva za ta namen ustrezajo enemu FTE (43.000 € na leto), razdelijo pa se na večje število honorarnih sodelavcev – terminoloških strokovnjakov.

Strokovnjaki za dodajanje normativnih podatkov: Za pregled iztočnic ali (pod)pomenov, ki tako ali drugače izkazujejo težave glede pravopisnih oz. standardizacijskih vprašanj, glavni urednik organizira in koordinira manjšo skupino svetovalcev za standardizacijska vprašanja.

Delo na tem segmentu poteka v soodvisnosti s procesi izdelave portala Slogovni priročnik⁶²

in drugih standardizacijskih priročnikov, katerih izdelava poteka v času projekta. Finančna sredstva za ta namen ustrezajo tretjini FTE (43.000 € na leto = 14.300 € na leto), razdelijo pa se na manjše število honorarnih sodelavcev – strokovnjakov za standardizacijo.

Strokovnjaki za dodajanje podatkov o izgovoru: Med stroški, ki jih predvideva dodajanje podatkov o izgovoru, je predvsem snemanje izgovora iztočnic ter strokovno svetovanje glede zapisa izgovora pri iztočnicah. Projekt ne financira nadgradnje spletnega slovarja znakovnega jezika in strojnega sintetizatorja govora, ki je uporabljen na portalu. Finančna sredstva za namen dodajanja podatkov o izgovoru ustrezajo tretjini FTE (43.000 € na leto = 14.300 € na leto), razdelijo pa se na enega ali manjše število honorarnih sodelavcev, poleg tega pokrivajo stroške, povezane s snemanjem izgovora iztočnic.

Strokovnjaki za dodajanje podatkov o izvoru: Podatki o izvoru izhajajo v osnovi iz dveh virov: iz rezultatov projekta Jezikovni viri starejše slovenščine IMP,⁶³ ki vključuje tudi podatke iz drugih virov starejše slovenščine (Wikivir itd.), ter iz obstoječih etimoloških priročnikov. Stroški za ta del vsebine slovarja so namenjeni prenosu informacij iz obstoječih virov in svetovanju glede jezikoslovnih podatkov o starejši slovenščini. Del sredstev je lahko namenjen za digitalizacijo virov starejše slovenščine. Finančna sredstva za namen dodajanja podatkov o izgovoru ustrezajo tretjini FTE (43.000 € na leto = 14.300 € na leto), razdelijo pa se na enega ali manjše število honorarnih sodelavcev, poleg tega pokrivajo stroške, povezane z digitalizacijo virov.

⁶² Slogovni priročnik: <<http://slogovni.slovenscina.eu/>>.

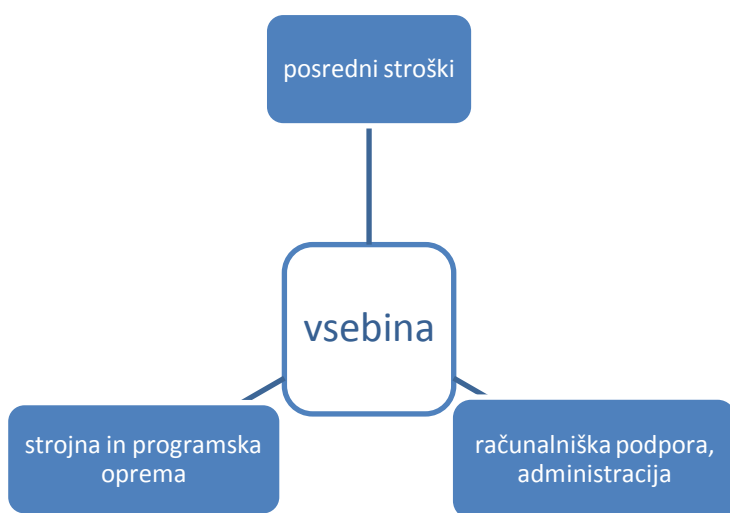
⁶³ Jezikovni viri starejše slovenščine IMP: <<http://nl.ijs.si/imp/>>.

Računalniški strokovnjaki za pridobivanje jezikoslovnih podatkov: Zelo pomemben del projekta je stalna računalniška podpora leksikografskemu delu pri izločanju podatkov iz raznih virov (korpusov itd.) ter njihova obdelava za namen vključitve v slovar. Za to delo je predvidena bodisi zaposlitev ali honorarno plačilo, finančna sredstva za ta namen pa ustrezajo enemu FTE (43.000 € na leto), ki so namenjena bodisi enemu stalnemu zaposlenemu sodelavcu ali se razdelijo na večje število honorarnih sodelavcev – strokovnjakov za jezikovne tehnologije.

Vzdrževalec korpusov: Pomembno je, da so leksikografom in drugim sodelavcem ves čas na voljo sveži in relevantni podatki o sodobnem slovenskem jeziku. Ker korpusi kot ključni vir tovrstnih podatkov hitro zastarijo, je treba v okviru projekta zagotoviti njihovo sprotno dopolnjevanje in vzdrževanje. Osnovni vir podatkov o pisnem jeziku je korpus Gigafida, o govorjenem jeziku korpus Gos, ki se v okviru projekta povečujeta po sistemu, kakor sta nastala v okviru projekta Sporazumevanje v slovenskem jeziku, torej s pogodbami o predaji besedil, usmerjenim spletnim pajkanjem ter snemanjem in transkripcijo (v primeru korpusa Gos). Za izvajanje in koordinacijo tega dela je predvidena bodisi zaposlitev ali honorarno plačilo, finančna sredstva za ta namen pa ustrezajo enemu FTE (43.000 € na leto), ki so namenjena bodisi enemu stalnemu zaposlenemu sodelavcu ali se razdelijo na večje število honorarnih sodelavcev – strokovnjakov za sestavljanje pisnih in govorjenih korpusov.

4.3 ORGANIZACIJA DELA – DRUGO

Računalniška podpora, administracija: Za vzdrževanje računalniških sistemov ter tehnično računalniško podporo leksikografski ekipi je predvidena bodisi redna zaposlitev ali ustrezno honorarno plačilo, finančna sredstva za ta namen pa ustrezajo enemu FTE (36.000 € na leto). Za administrativna dela in podporo leksikografski ekipi pri nezahtevnih delih je predvidena zaposlitev enega administrativnega delavca oz. en FTE (30.000 € na leto).



SLIKA 42: ORGANIZACIJA DELA - DRUGO

Strojna in programska oprema: Strojna oprema, potrebna za delo s specializiranimi leksikografskimi programi, zajema predvsem en močnejši strežnik ter količino osebnih računalnikov, ki ustreza številu zaposlenih na projektu. Predpostavlja se, da honorarni sodelavci uporabljajo svoje računalnike. Programska oprema zajema ustrezno število licenc za program iLex⁶⁴ ter licenčnino za uporabo programa Sketch Engine⁶⁵ za ustrezno število let. Poleg tega so vključeni še stroški za operacijske sisteme in manjše nakupe druge programske opreme. Za

⁶⁴ <<http://www.emp.dk/ilexweb/index.jsp>>.

⁶⁵ <<http://www.sketchengine.co.uk/>>.

strojno opremo je predvidenih 25.000 EUR, za programsko opremo skupaj 40.000 EUR.

Posredni stroški: Ti zajemajo stroške elektrike, ogrevanja in komunale, telekomunikacij, amortizacijo osnovnih sredstev, stroške potrošnega materiala in drobnega inventarja, najemanje prostorov in podobno. Za te vrste stroškov je predvidenih 79.000 EUR.

5 FINANČNI NAČRT

	FTE	cena (na leto)	obdobje (št. let)	skupaj
glavni urednik	1	50.000	5	250.000
leksikografi – raziskovalci	4	43.000	5	860.000
leksikografi – samozaposleni	4	43.000	5	860.000
moč množic	-	-	-	605.000
jezikovne tehnologije	1	43.000	5	215.000
vzdrževanje korpusa	1	43.000	5	215.000
terminologija	1	43.000	4,5	193.500
normativni podatki	1	43.000	4,5	193.500
(iz)govor	1	43.000	4,5	193.500
izvor	1/3	43.000	4,5	63.855
računalniška podpora	1	36.000	5	180.000
administracija	1	30.000	5	150.000
strojna oprema	-	-	-	30.000
programska oprema	-	-	-	40.000
posredni stroški	-	-	-	79.000
SKUPAJ				4.128.355

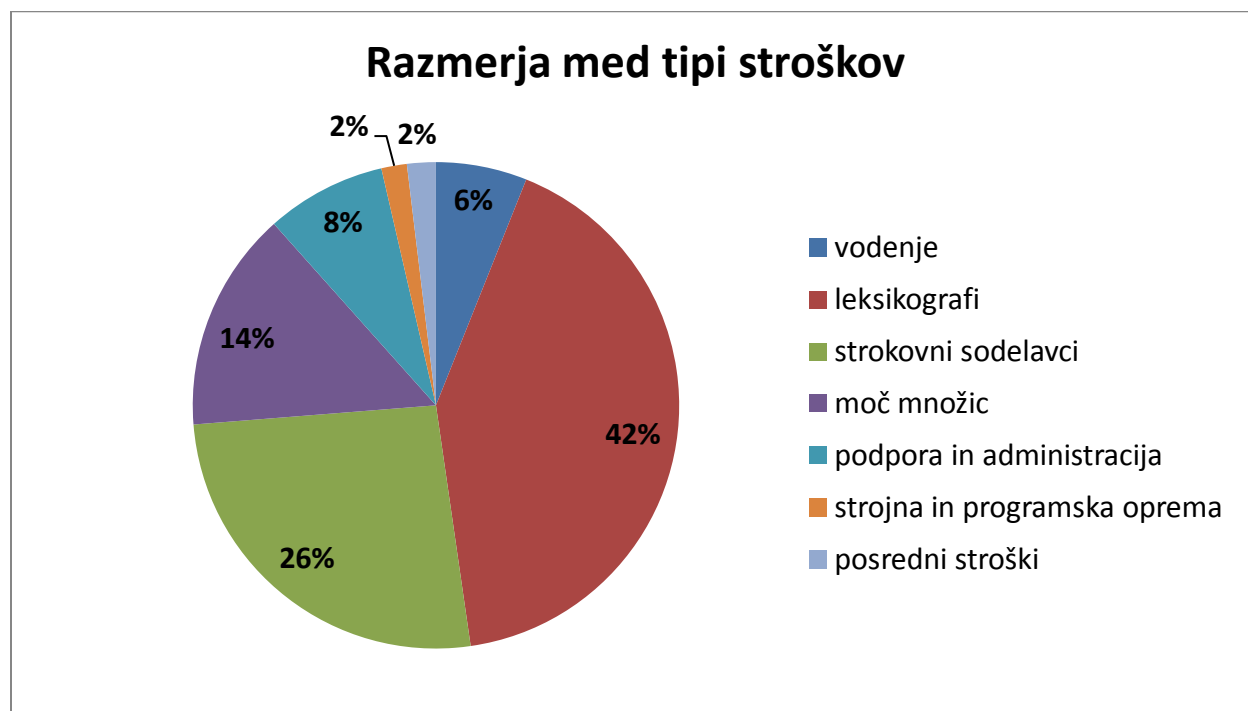
TABELA 9: FINANČNI NAČRT

5.1 RAZMERJA MED TIPI STROŠKOV

V spodnjem grafu prikazujemo razmerja med tipi stroškov, ki so predvideni v finančnem načrtu. V kategoriji *leksikografi* so združeni stroški za leksikografe-raziskovalce in samozaposlene leksikografe. Kategorija *strokovni sodelavci* zajema stroške strokovnjakov za jezikovne tehnologije, za terminologijo, normativne podatke, izgovor in izvor. Pri *podpora in administracija* sta združeni postavki za tehnično računalniško podporo in administracijo, pri *strojna in programska oprema* pa obe postavki, povezani z računalniškimi stroški.

kategorija	vsota
vodenje	250.000 €
leksikografi	1.720.000 €
strokovni sodelavci	1.074.355 €
moč množic	605.000 €
podpora in administracija	330.000 €
strojna in programska oprema	70.000 €
posredni stroški	79.000 €

TABELA 10: TIPI STROŠKOV



SLIKA 43: RAZMERJA MED TIPI STROŠKOV

Ganttov diagram – finančni načrt

Večina sodelavcev pri projektu sodeluje v celotnem obdobju trajanja projekta, med njimi glavni urednik, leksikografi-raziskovalci, jezikovni tehnologi ter tehnična računalniška in administrativna podpora. Sistem izkoriščanja moči množic je odvisen od vnaprej pripravljenih podatkov, avtomatsko izločenih iz korpusa. Priprava sistema in izločanje podatkov bosta trajala približno pol leta, zato bomo s tem začeli s polletnim zamikom, enako – in s tem povezano – tudi leksikografi-samozaposleni. S čiščenjem podatkov po sistemu izkoriščanja moči množic bomo predvidoma zaključili najkasneje pol leta pred zaključkom projekta. Strokovni sodelavci za terminologijo, normativne podatke, izgovor in izvor besed bodo začeli z delom s polletnim zamikom, ko bodo na voljo prva gesla v fazi, ko se tovrstna redakcija gesel lahko začne.

	1. leto	2. leto	3. leto	4. leto	5. leto
glavni urednik					
leksikografi – raziskovalci					
leksikografi – samozaposleni					
jezikovne tehnologije					
vzdrževanje korpusa					
moč množic					
terminologija					
normativni podatki					
izgovor					
izvor					
tehnična računalniška podpora					
administracija					

TABELA 11: GANTTOV DIAGRAM

6 KAZALNIKI

Kazalniki so formalni pokazatelji doseganja projektnih ciljev, od česar je odvisno tudi nadaljnje financiranje. Izpolnjevanje pogojev se dokazuje v dveh oblikah: (a) z oddajo ustreznega dela slovarske **baze v formatu XML** in (b) s prikazom gesel na **portalu SSSJ**, označenih z ustrezno barvo.

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K1	Računalniška infrastruktura	6	1	6
K2	Avtomatsko luščenje podatkov iz korpusa	6	1	6
K3	Geslovnik in spremljanje sprememb	6	1	6
K4	Priročnik za sestavljanje slovarja	6	1	6
K5	Faza 3 – prva četrtina (25.000 gesel)	12	6	18
K6	Faza 4 – prva četrtina (25.000 gesel)	12	9	21
K7	Faza 5 – prva četrtina (25.000 gesel)	12	12	24
K8	Faza 3 – druga četrtina (25.000 gesel)	12	18	30
K9	Faza 4 – druga četrtina (25.000 gesel)	12	21	33
K10	Faza 5 – druga četrtina (25.000 gesel)	12	24	36
K11	Faza 3 – tretja četrtina (25.000 gesel)	12	30	42
K12	Faza 4 – tretja četrtina (25.000 gesel)	12	33	45
K13	Faza 5 – tretja četrtina (25.000 gesel)	12	36	48
K14	Faza 3 – zadnja četrtina (25.000 gesel)	12	42	54

K15	Faza 4 – zadnja četrtina (25.000 gesel)	12	45	57
K16	Faza 5 – zadnja četrtina (25.000 gesel)	12	48	60

TABELA 12: KAZALNIKI

Izračun za leksikografsko delo temelji na naslednjih predpostavkah: analiza SSKJ kaže, da je približno četrtina vseh gesel večpomenskih, torej od faze 3-5 zahtevajo kompleksnejšo leksikografsko obravnavo, ne zgolj pripis definicije in/ali semantičnega indikatorja. Natančni izračun pri SSKJ kaže naslednjo sliko (upoštevamo samostalnice, pridevnike, glagole in prislove, ki sestavljajo 97 % celotnega geslovnika):

	vsa gesla	večpomenska gesla	odstotek
samostalniki	51.488	11.447	22,2 %
pridevniki (s prislovi)	20.588	3.080	15,0 %
glagoli	16.519	7.037	42,6 %
(samo) prislovi	1.720	718	41,7 %
SSKJ	93.154	22.298	23,9 %

TABELA 13: VEČPOMENSKA GESLA

Na podlagi izkušenj z delom na LBS in preteklih leksikografskih izkušenj je bil narejen spodnji izračun glede potrebnega leksikografskega dela za izdelavo celotnega slovarja. Izračun predvideva, da je mogoče dnevno učinkovito izdelovati gesla 6 ur, v povprečju 20 dni na mesec, 11 mesecev na leto. Predvidena je obdelava 70.000-75.000 korpusnih lem, ki v leksikografskem procesu (po izračunih LBS) generirajo približno 100.000 enobesednih ali večbesednih iztočnic oz. posledično slovarskih gesel. Za izdelavo enega gesla leksikograf potrebuje skupaj v vseh fazah 45 minut, kar pomeni, da izdela 9 gesel na dan oz. 1,5 gesla na uro. Pri tem upoštevamo, da v 75 % primerov mora napisati zgolj definicijo in/ali semantični indikator, ker je ostalo delo opravljeno v okviru drugih procesov (izkoriščanje moči množic, preverjanje terminologije itd.), in na koncu preveriti konsistentnost podatkov.

število lem	gesel/dan	leksikografi	delovnih dni	delovnih mesecev	delovnih let
70.000	9	8	972	78,6	4,4
71.000	9	8	986	49,3	4,5
72.000	9	8	1.000	50,0	4,5
73.000	9	8	1.014	50,7	4,6
74.000	9	8	1.028	51,4	4,7
75.000	9	8	1.042	52,1	4,7

TABELA 14: IZRAČUN DELOVNIH MESECEV/LET

Konkretno delo na geslih se začne pol leta po začetku, torej obdelava 72.000 lem omogoča izdelavo gesel v danem časovnem okviru (4,5 + 0,5 let). V primeru, da bo treba obdelati več kot omenjeno število lem, obstaja možnost stiskanja časovnih parametrov z razpršitvijo dela na večje število leksikografov – to pomeni, da gesla v danem trenutku obdeluje več kot osem leksikografov.

KAZALNIK K1

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K1	Postavitev računalniške infrastrukture	6	1	6

V prvih šestih mesecih je treba postaviti celotno računalniško infrastrukturo za uspešno delovanje leksikografske ekipe in vseh ostalih strokovnih sodelavcev. Med osnovno infrastrukturo spada predvsem:

- ureditev pogojev za delo s programom za sestavljanje slovarjev **iLex**
 - postavitev strežnika in ureditev licenc
 - izdelava sheme XML in vizualizacije
 - vključitev iLexa v predvideno kroženje podatkov (uvoz, izvoz, objava)
- ureditev vseh pogojev za delo s programom za poizvedovanje po korpusih **Sketch Engine**
 - postavitev lastne inštalacije in ureditev licenc za leksikografe
 - izbira slovnice besednih skic za avtomatski izvoz podatkov
 - vzpostavitev programskega vmesnika za avtomatski izvoz podatkov
- vzpostavitev **portala SSSJ** z vizualizacijo avtomatsko izvoženih podatkov iz korpusa in vsemi predvidenimi deli (zavijki)
- vzpostavitev **spletnega servisa** za izkoriščanje moči množic z avtomatiziranim uvozom podatkov iz orodja Sketch Engine v spletni servis in izvozom iz servisa v iLex
- vzpostavitev sprotne objavljanja podatkov, urejenih v programu iLex, na portalu SSSJ
- vzpostavitev diskusijske platforme za strokovne sodelavce, predvsem
 - terminološka platforma
 - normativna platforma
- intranet

Kazalnik je delujoča infrastruktura z vsemi naštetimi elementi – dokazuje se s (krajšo) specifikacijo.

KAZALNIK K2

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K2	Avtomatsko luščenje podatkov iz korpusa	6	1	6

Iz korpusa Gigafida se s pomočjo programskega vmesnika za izvoz kolokacij in dobrih zgledov v programu Sketch Engine na podlagi struktur iz slovnice besednih skic (*sketch grammar*) izvozi podatke za 100.000 glagolskih, pridevniških in prislovnih lem ter lem občnih samostalnikov, v razmerju 50 (občni samostalniki), 25 (pridevniki), 15 (glagoli) in 10 (prislovi). Ostale besedne vrste niso vključene v prikazovanje avtomatsko generiranih podatkov.

Kazalnik so izvoženi podatki v formatu XML in prikazani na portalu v fazi 1.

KAZALNIK K3

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K3	Geslovnik in spremljanje sprememb	6	1	6

V okviru te aktivnosti se iz frekvenčne liste korpusa Gigafida, v kombinaciji s podatki iz korpusa Kres, Gos in drugih baz podatkov, ki so na voljo (Pravopis, Sloleks, Slogovni priročnik itd.), izdela izhodiščni geslovnik, ki obsega približno četrtno več gesel, kot je predvideno končno število iztočnic. Spisek obsega okrog 100.000 korpusnih lem, kar po izračunih generira približno 130.000 slovarskih iztočnic.

Geslovnik se lahko v teku izvajanja projekta tudi spreminja, ker je treba sprotno spremljati nastajanje novih leksemov oz. dodajanje pomenov starim. Zato je treba v okviru vzdrževanja korpusa vzpostaviti mehanizem za avtomatično opozarjanje na potencialne nove iztočnice ali pomene (s pomočjo kontekstualnih podatkov).

Kazalnik sta geslovnik v formatu XML in dostopen na portalu, ter specifikacije sistema za sprotno spremljanje leksikalnih sprememb.

KAZALNIK K4

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K4	Priročnik za sestavljanje slovarja	6	1	6

Priročnik za sestavljanje slovarja vsebuje opis leksikografskega in drugih del pri sestavljanju slovarja. Zajema predvsem:

- navodila za proces izkoriščanja moči množic
- navodila za leksikografsko delo: pomenska delitev, definicije, indikatorji, oznake itd.
- navodila za terminološko delo
- navodila za delo z normativnimi podatki
- navodila za obravnavo podatkov o izgovoru
- navodila za obravnavo podatkov o izvoru

Kazalnik je priročnik za sestavljanje slovarja – specifikacija.

KAZALNIK K5

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K5	Faza 3 – prva četrtnina (25.000 gesel)	12	6	18

Faza 3 predvideva izdelano pomensko delitev gesla, pomenski meni, definicije in ustrezno delitev kolokacijskih podatkov z zgledi, v kombinaciji s skladijskimi podatki. Gesla še niso terminološko pregledana in ne vsebujejo podatkov iz drugih procesov (terminologija, norma, izgovor, izvor itd.).

Kazalnik je baza s 25.000 gesli v fazi 3 v formatu XML ter ustrezno število gesel na portalu – prvi paket.

KAZALNIK K6

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K6	Faza 4 – prva četrtina (25.000 gesel)	12	9	21

Faza 4 predvideva dodajanje informacij o normi, izgovoru, izvoru ter terminološko preverjanje iztočnic in pomenov, ki spadajo na določeno strokovno področje. Gesla še niso redigirana in dokončno potrjena.

Kazalnik je baza s 25.000 gesli v fazi 4 v formatu XML ter ustrezno število gesel na portalu – prvi paket.

KAZALNIK K7

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K7	Faza 5 – prva četrtina (25.000 gesel)	12	12	24

Faza 5 predvideva končno redakcijo (pregled) gesel, ki so s stališča vseh predvidenih postopkov potrjena kot dokončana.

Kazalnik je baza s 25.000 gesli v fazi 5 v formatu XML ter ustrezno število gesel na portalu – prvi paket.

KAZALNIK K8

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K8	Faza 3 – druga četrtina (25.000 gesel)	12	18	30

Faza 3 predvideva izdelano pomensko delitev gesla, pomenski meni, definicije in ustrezno delitev kolokacijskih podatkov z zgledi, v kombinaciji s skladijskimi podatki. Gesla še niso terminološko pregledana in ne vsebujejo podatkov iz drugih procesov (terminologija, norma, izgovor, izvor itd.).

Kazalnik je baza s 25.000 gesli v fazi 3 v formatu XML ter ustrezno število gesel na portalu – drugi paket.

KAZALNIK K9

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K9	Faza 4 – druga četrtina (25.000 gesel)	12	21	33

Faza 4 predvideva dodajanje informacij o normi, izgovoru, izvoru ter terminološko preverjanje iztočnic in pomenov, ki spadajo na določeno strokovno področje. Gesla še niso redigirana in dokončno potrjena.

Kazalnik je baza s 25.000 gesli v fazi 4 v formatu XML ter ustrezno število gesel na portalu – drugi paket.

KAZALNIK K10

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K10	Faza 5 – druga četrtina (25.000 gesel)	12	24	36

Faza 5 predvideva končno redakcijo (pregled) gesel, ki so s stališča vseh predvidenih postopkov potrjena kot dokončana.

Kazalnik je baza s 25.000 gesli v fazi 5 v formatu XML ter ustrezno število gesel na portalu – drugi paket.

KAZALNIK K11

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K11	Faza 3 – tretja četrtina (25.000 gesel)	12	30	42

Faza 3 predvideva izdelano pomensko delitev gesla, pomenski meni, definicije in ustrezno delitev kolokacijskih podatkov z zgledi, v kombinaciji s skladijskimi podatki. Gesla še niso terminološko pregledana in ne vsebujejo podatkov iz drugih procesov (terminologija, norma, izgovor, izvor itd.).

Kazalnik je baza s 25.000 gesli v fazi 3 v formatu XML ter ustrezno število gesel na portalu – tretji paket.

KAZALNIK K12

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K12	Faza 4 – tretja četrtina (25.000 gesel)	12	33	45

Faza 4 predvideva dodajanje informacij o normi, izgovoru, izvoru ter terminološko preverjanje iztočnic in pomenov, ki spadajo na določeno strokovno področje. Gesla še niso redigirana in dokončno potrjena.

Kazalnik je baza s 25.000 gesli v fazi 4 v formatu XML ter ustrezno število gesel na portalu – tretji paket.

KAZALNIK K13

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K13	Faza 5 – tretja četrtina (25.000 gesel)	12	36	48

Faza 5 predvideva končno redakcijo (pregled) gesel, ki so s stališča vseh predvidenih postopkov potrjena kot dokončana.

Kazalnik je baza s 25.000 gesli v fazi 5 v formatu XML ter ustrezno število gesel na portalu – tretji paket.

KAZALNIK K14

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K14	Faza 3 – zadnja četrtina (25.000 gesel)	12	42	54

Faza 3 predvideva izdelano pomensko delitev gesla, pomenski meni, definicije in ustrezno delitev kolokacijskih podatkov z zgledi, v kombinaciji s skladejskimi podatki. Gesla še niso terminološko pregledana in ne vsebujejo podatkov iz drugih procesov (terminologija, norma, izgovor, izvor itd.).

Kazalnik je baza s 25.000 gesli v fazi 3 v formatu XML ter ustrezno število gesel na portalu – četrti paket.

KAZALNIK K15

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K15	Faza 4 – zadnja četrtina (25.000 gesel)	12	45	57

Faza 4 predvideva dodajanje informacij o normi, izgovoru, izvoru ter terminološko preverjanje iztočnic in pomenov, ki spadajo na določeno strokovno področje. Gesla še niso redigirana in dokončno potrjena.

Kazalnik je baza s 25.000 gesli v fazi 4 v formatu XML ter ustrezno število gesel na portalu – četrti paket.

KAZALNIK K16

št.	Ime kazalnika	trajanje	Začetek (mes.)	Konec (mes.)
K16	Faza 5 – zadnja četrtina (25.000 gesel)	12	48	60

Faza 5 predvideva končno redakcijo (pregled) gesel, ki so s stališča vseh predvidenih postopkov potrjena kot dokončana.

Kazalnik je baza s 25.000 gesli v fazi 5 v formatu XML ter ustrezno število gesel na portalu – tretji paket.

Ganttov diagram – kazalniki

Diagram prikazuje časovni potek dela na slovarju. Prvi štirje kazalniki so končani in na voljo po šestih mesecih, ko se začne intenzivno leksikografsko delo. Faze 3, 4 in 5 potekajo v trimesečnih zamikih, s kazalniki po četrtinah slovarja. Avtomatsko generirana gesla s skladejskimi vzorci, kolokacijami in zgledi, ekstrahiranimi iz korpusa Gigafida so na voljo na portalu že po šestih mesecih do 54-tega meseca, ko vsa predvidena gesla preidejo skozi prvo fazo obdelave. Četrtina do konca izdelanih gesel je na voljo

po dveh letih od začetka projekta, torej na koncu 24-tega meseca, kar shematično prikazujemo na spodnjem diagramu.

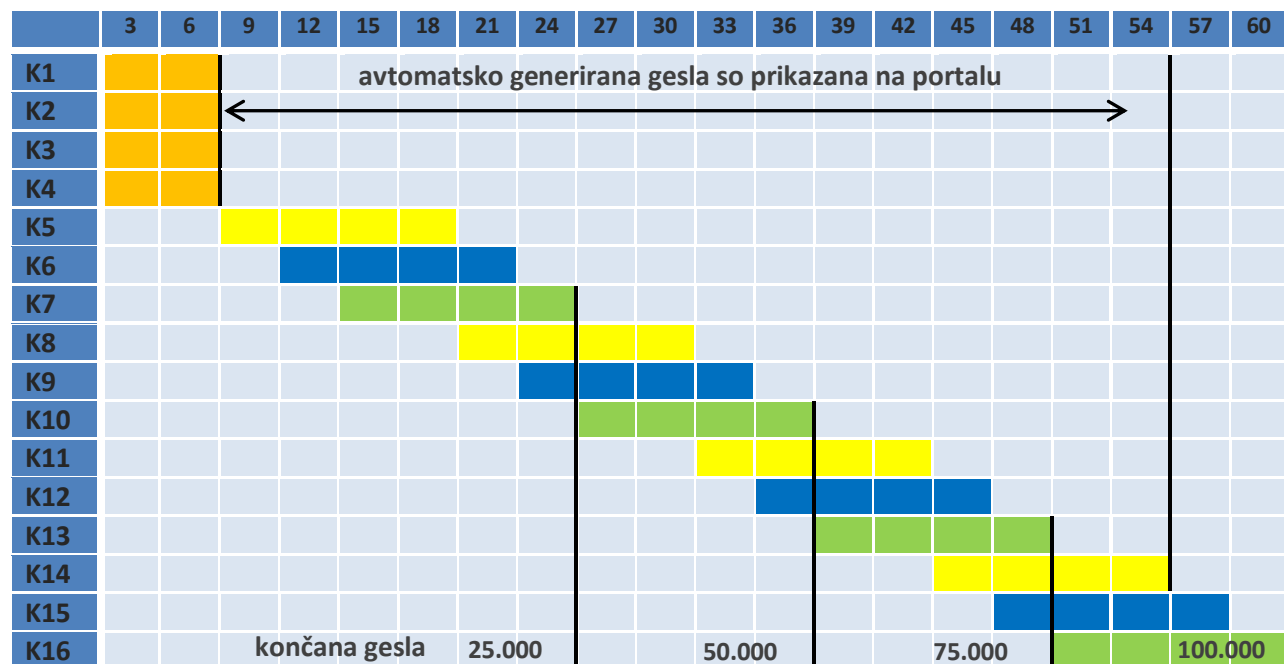


TABELA 15: TERMINSKI NAČRT

7 REFERENCE/BIBLIOGRAFIJA

REFERENCE

- Barnbrook, Geoff. 2002. »Defining language: A local grammar of definition sentences.« *Studies in Corpus Linguistics*. Amsterdam: John Benjamins Publishing Company.
- Baroni, Marco in Bernardini, Silvia. 2004. »BootCaT: Bootstrapping corpora in terms from the web.« *Proceedings of LREC 2004, Lisbon*. Lisbon: ELDA. 1313–1316.
- Baroni, Marco, Kilgarrieff, Adam, Pomikálek, Jan in Rychlý, Pavel. 2006. »WebBootCaT: a web tool for instant corpora.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marello. Torino: European Association for Lexicography, Università di Torino. 123–131.
- Bizjak Končar, Aleksandra idr. 2013. *Slovar novejšega besedja slovenskega jezika*. Ur. Marko Snoj in Aleksandra Bizjak Končar. Ljubljana: Založba ZRC, ZRC SAZU.
- Cowie, Antony. 1998. »Introduction.« *Phraseology: Theory, Analysis, in Applications*. Ur. A. P. Cowie. Oxford University Press. 1–20.
- Coxhead, Averil. 2000. »A New Academic Word List.« *TESOL Quarterly* 34(2). 213–238.
- Erlandsen, Jens. 2004. »iLex – new DWS.« *Third International Workshop on Dictionary Writing systems (DWS 2004)*. Brno, 6. – 7. september 2004. <<http://nlp.fi.muni.cz/dws2004/pres/#15>>.
- Firth, John Rupert. 1957. *Papers in linguistics, 1934-1951*. Oxford: Oxford University Press.
- Firth, John Rupert. 1968. *Selected Papers of J. R. Firth 1952-59*. London: Longman.
- Fišer, Darja, Gantar, Polona in Krek, Simon. 2012. »Using explicitly in implicitly encoded semantic relations to map Slovene wordnet in Slovene lexical database.« *LREC 2012 / 8th International Conference on Language Resources in Evaluation, 21-27 May 2012, Istanbul, Turkey*. 77–84. <<http://www.lrec-conf.org/proceedings/lrec2012/index.html>>.
- Gantar, Polona. 2007. *Stalne besedne zveze v slovenščini: korpusni pristop*. Ljubljana: Založba ZRC.
- Gantar, Polona in Krek, Simon. 2009. »Drugačen pogled na slovarske definicije: opisati, pojasniti, razložiti?« *Infrastruktura slovenščine in slovenistike*. Ur. Marko Stabej. Ljubljana: Znanstvena založba Filozofske fakultete. 151–159. <http://www.centerslo.net/files/file/simpozij/simp28/Gantar_Krek.pdf>.
- Hanks, Patrick. 1987. »Definitions in explanations.« *Looking up. An account of the COBUILD project*. Ur. J. M. Sinclair. London, Glasgow: Collins ELT. 116–136.

- Hanks, Patrick. 1994. »Linguistic norms in pragmatic exploitations, or why lexicographers need prototype theory in vice versa.« *Papers in Computational Lexicography. Complex '94*. Ur. G. Kiss, F. Keifer in J. Pajzs. Budapest: Hungarian Academy of Sciences.
- Hanks, Patrick. 2000. »Contributions of Lexicography in Corpus Linguistics to a Theory of Language Performance.« *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000 Stuttgart, Germany, August 8th – 12th, 2000*. Ur. U. Heid, C. Evert, S. Lehmann in E. Roher. Stuttgart: University of Stuttgart. 3–14.
- Hanks, Patrick. 2004. »Corpus Pattern Analysis.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud.
- Hanks, Patrick. 2008. »Mapping meaning onto use: a Pattern Dictionary of English Verbs.« *AACL 2008*. Utah.
- Hanks, Patrick. 2009. *Sestavljanje enojezičnega slovarja za domače govorce*. *Jezik in slovstvo* 54(3-4). 7-14.
- Hanks, Patrick. 2010. »Elliptical arguments: a problem in relating meaning to use.« *E-Lexicography in the 21st century: New challenges, new applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*. Ur. Magali Paquot, Sylviane Granger. UCL Press Universitaires de Louvain. 109–125.
- Hanks, Patrick. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press Ltd.
- Hanks, Patrick, Pustejovsky, James. 2004. »Common Sense About Word Meaning: Sense in Context.« *TSD 004*. Brno.
- Hanks, Patrick, Pustejovsky, James. 2005. »A Pattern Dictionary for Natural Language Processing.« *Revue Française de Linguistique Appliquée* 10(2). 63–82.
- Katz, Jerrold in Fodor, Jerry. 1963. »The structure of a semantic theory.« *Language* 39. 170–210.
- Kosem, Iztok, Gantar, Polona in Krek, Simon. 2012. »Avtomatsko luščenje leksikalnih podatkov iz korpusa.« *KONFERENCA Jezikovne tehnologije (8 ; 2012 ; Ljubljana) Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012, [Ljubljana, Slovenia]: zbornik 15. mednarodne multikonference Informacijska družba - IS 2012, zvezek C*. Ur. Jerneja Žganec Gros in Tomaž Erjavec. Ljubljana: Institut Jožef Stefan. 117–122.
- Krek, Simon. 2005. »FrameNet in slovenščina.« *Jezik in slovstvo* 53(5). 37–54.
- Krek, Simon. 2009. »Od SSKJ do spletnega portala.« *Jezik in slovstvo* 54(3-4). 95–113.
- Krek, Simon. 2012a. *Slovenski jezik v digitalni dobi = The Slovene language in the digital age*. Ur. Hans Rehm in Georg Uszkoreit. Heidelberg [etc.]: Springer.

- Krek, Simon. 2012b. »Spletni portal Slogovni priročnik.« *Slavistika v regijah - Koper / [Slovenski slavistični kongres, Koper, 27.-29. september 2012]*. Ur. Boža Krakar-Vogel. Ljubljana: Zveza društev Slavistično društvo Slovenije; Znanstvena založba Filozofske Fakultete. 225–231.
- Lorentzen, Henrik in Theilgaard, Liisa. 2012. »Online dictionaries – how do users find them in what do they do once they have?« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*, Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusén. Department of Linguistics in Scandinavian Studies, University of Oslo. 654–660.
- Lönneker-Rodman, Birte, Baker, Collin in Hong, Jisup. 2008. »The new FrameNet Desktop: A Usage Scenario for Slovenian.« *Proceedings of ICGL 2008, the First International Conference on Global Interoperability for Language Resources, 9-11 January 2008, Hong Kong*. City University of Hong Kong. 147–154.
- Moon, Rosamund. 1987. »The Analysis of Meaning.« *Looking up: An account of the COBUILD Project in Lexical Computing*. Ur. J. M. Sinclair. London, Glasgow: Collins ELT. 86–103.
- Može, Sara. 2009. »Semantično označevanje korpusa slovenščine po modelu FrameNet.« *Infrastruktura slovenščine in slovenistike*. Ur. Marko Stabej. Ljubljana: Znanstvena založba Filozofske fakultete. 265–269 <<http://www.centerslo.net/files/file/simpozij/simp28/Moze.pdf>>.
- Müller, Jakob. 1980. »Pomenske skupine in pomenska zgradba besed (samostalnikov).« *XVI. seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta. 35–47.
- Nation, Paul in Waring, Robert. 1997. »Vocabulary size, text coverage and word lists.« *Vocabulary: Description, acquisition in pedagogy*. Ur. Norbert Schmitt in Michael McCarthy. Cambridge: Cambridge University Press. 6–19.
- Orešnik, Janez. 1972. »Formalizacija semantičnih definicij najmanjših jezikovnih enot s pomenom.« *Problemi semantike, sintakse in obravnave tekstov*. Ljubljana: Institut »Jozef Štefan.« 27–33.
- »Slovar slovenskega knjižnega jezika na CD-romu z Odzadnjim slovarjem slovenskega jezika in Besediščem slovenskega jezika z oblikoslovnimi podatki.« Ljubljana: SAZU, ZRC SAZU, Inštitut za slovenski jezik, DZS. /SSKJ/.
- »Strokovni posvet o novem slovarju slovenskega jezika, 23. in 24. oktober 2008.« 2009. Ur. Andrej Perdih. Ljubljana: Založba ZRC, ZRC SAZU.
- Rozman, Tadeja. 2010. »Vloga enojezičnega razlagalnega slovarja slovenščine pri razvoju jezikovne zmožnosti: doktorska disertacija. Ljubljana: Filozofska fakulteta UL.
- Rundell, Michael in Kilgariff, Adam. 2011. »Automating the creation of dictionaries: where will it all end?« *A Taste for Corpora. A tribute to Professor Sylviane Granger*. Ur. G. Gilquin, M. Paquot, F. Meunier in S. De Cock. Benjamins. 257–281.

- Snoj, Marko. 2012. »Podgesla v novem slovarju slovenskega jezika.« *Škrabčevi dnevi 7 – zbornik prispevkov s simpozija 2011*. Ur. Franc Marušič in Rok Žaucer. Nova Gorica: Založba Univerze v Novi Gorici. 96–103.
- Suhadolnik, Stane. 1997. »Šolar-Ramovšev načrt za slovar knjižnega jezika.« *Slavistična revija* 45(3-4). 558–566.
- Toporišič, Jože. 1980. »O strukturalnem določanju besednih pomenov (ob glagolu biti).« *Linguistica*. 151–167.
- Toporišič, Jože. 1981. »Strukturalno pomenoslovje besed.« *XVII. seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta. 95–112.
- Vidovič-Muha, Ada. 2000. *Slovensko leksikalno pomenoslovje: govorica slovarja*. Ljubljana: Znanstveni inštitut Filozofske fakultete.
- Žele, Andreja. 2008. *Vezljivostni slovar slovenskih glagolov*. Ljubljana: Založba ZRC, ZRC SAZU.

BIBLIOGRAFIJA

I. KRITIKE SSKJ

- Bajt, Drago. 1991. »Dohitevanje in prehitevanje.« *Delo* 33 (12. 12.). 291, 13.
- Bajt, Drago. 1992. »Zgodba o ž-ju: Odgovor Stanetu Suhadolniku.« *Naši razgledi* 41 (20. 3.)(6). 167–168.
- Kosem, Iztok. 2006. »Definicijski jezik v Slovarju slovenskega knjižnega jezika s stališča sodobnih leksikografskih načel.« *Jezik in slovstvo* 51(5). 25–45.
- Kržišnik, Erika. 1987. »Frazeološko gradivo v Slovarju slovenskega knjižnega jezika.« *Slava* 2(2). 143–162.
- Müller, Jakob. 1996. »Slovar slovenskega knjižnega jezika in kritika z bibliografijo (1960-1992).« *Razprave SAZU, Razred za filološke in literarne vede* 15. 187–234.
- Müller, Jakob. 2009. »Kritične misli in zamisli o SSKJ.« *Strokovni posvet o novem slovarju slovenskega jezika, 23. in 24. oktober 2008*. Ur. Andrej Perdih. Ljubljana: Založba ZRC SAZU, ZRC SAZU. 17–21.
- Rigler, Jakob. 1968. »Problematika naglaševanja v slovenskem knjižnem jeziku.« *Jezik in slovstvo* 13(6). 192–199.
- Rigler, Jakob. 1971. »H kritikam pravopisa, pravorečja in oblikoslovja v SSKJ.« *Slavistična revija* 19(4). 433–462.
- Rigler, Jakob. 1972. »H kritikam pravopisa, pravorečja in oblikoslovja v SSKJ.« *Slavistična revija* 20(2). 244–251.
- Suhadolnik, Stane. 1992. »SSKJ in SP: Odgovor J. Toporišiču in D. Bajtu.« *Naši razgledi* 43 (3. 4.)(7). 198.

Toporišič, Jože. 1971. »Pravopis, pravorečje in oblikoslovje v Slovarju slovenskega knjižnega jezika I.« *Slavistična revija* 19(2). 222–229.

Toporišič, Jože. 1971. »Pravopis, pravorečje in oblikoslovje v Slovarju slovenskega knjižnega jezika I.« *Slavistična revija* 19(1). 55–75.

II. LEKSIKALNA BAZA ZA SLOVENŠČINO

Arhar Holdt, Špela. 2011. *Luščenje besednih zvez iz besedilnega korpusa z uporabo dvodelnih in tridelnih oblikoskladenjskih vzorcev*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Arhar Holdt, Špela in Holozan, Peter. 2009. »Leksikalna podatkovna zbirka ASES (Amebisov skupni elektronski slovar).« *Jezikovni korpusi v medkulturni komunikaciji*. Ur. Vesna Mikolič. Koper: Univerza na Primorskem, Znanstveno-raziskovalno središče, Založba Annales; Zgodovinsko društvo za južno Primorsko. 30–51.

Fišer, Darja, Gantar, Polona in Krek, Simon. 2012. »Using explicitly in implicitly encoded semantic relations to map Slovene wordnet in Slovene lexical database.« *LREC 2012 / 8th International Conference on Language Resources in Evaluation, 21-27 May 2012, Istanbul, Turkey*. 77–84
<<http://www.lrec-conf.org/proceedings/lrec2012/index.html>>.

Gantar, Polona. 2009. »Leksikalna baza: vse, kar ste vedno želeli vedeti o jeziku.« *Jezik in slovstvo* 54(3/4). 69–94.

Gantar, Polona idr. 2009. »Opis analize referenčnega korpusa (kazalnik 5): projekt Sporazumevanje v slovenskem jeziku: specifikacije za izdelavo leksikalne baze za slovenščino.«
<http://www.slovenscina.eu/Media/Kazalniki/Kazalnik5/SSJ_Kazalnik_5_Specifikacije-opis-analize-korpusa_v1.pdf>.

Gantar, Polona idr. 2009. »Standard za izdelavo posamezne leksikalne enote v leksikalni bazi (kazalnik 6): projekt Sporazumevanje v slovenskem jeziku: specifikacije za izdelavo leksikalne baze za slovenščino.«
<http://www.slovenscina.eu/Media/Kazalniki/Kazalnik6/SSJ_Kazalnik_6_Specifikacije-leksikalna-baza_v1.pdf>.

Gantar, Polona. 2010. »K uporabniku usmerjeni slovnično-leksikalni opisi slovenskega jezika.« *Izzivi sodobnega jezikoslovja*. Ur. Vojko Gorjanc in Andreja Žele. Ljubljana: Znanstvena založba Filozofske fakultete. 35–51.

Gantar, Polona. 2011. »Leksikalna baza za slovenščino: komu, zakaj in kako (naprej)?« *Jezikoslovni zapiski* 17(2). 77–92.

Gantar, Polona. 2012. »Slovnični in pomenski opisi v leksikalni bazi za slovenščino.« *Škrabčevi dnevi 7 – zbornik prispevkov s simpozija 2011*. Ur. Franc Marušič in Rok Žaucer. Nova Gorica: Založba Univerze v Novi Gorici. 17–27.

Gantar, Polona. 2012. »Večbesedne leksikalne enote v leksikalni bazi za slovenščino« Ur. Vida Jesenšek. *Frazeologija in kultura*. 147.

Gantar, Polona idr. 2013. »Projekt Sporazumevanje v slovenskem jeziku: Leksikalna baza za slovenščino.« <<http://www.slovenscina.eu/spletni-slovar/prenos>>.

Gantar, Polona in Krek, Simon. 2011. »Slovene lexical database« Ur. Daniela Majchráková in Radovan Garabík. *Natural language processing*. 72–80.

Gorjanc, Vojko, Krek, Simon in Gantar, Polona. 2005. »Slovenska leksikalna podatkovna zbirka.« *Jezik in slovstvo* 50(2). 3–19.

Grabnar, Katja. 2010. »Slikar slika, slikarka ilustrira? Vprašanje ženskih poimenovanj oseb v opisu sodobne slovenščine.« *Slovenske korpusne raziskave (Zbirka Prevodoslovje in uporabno jezikoslovje)*. Ur. Špela Vintar. Ljubljana: Znanstvena založba Filozofske fakultete.

Kocjančič, Polonca in Zaranšek, Petra. 2009. »The Slovene Lexical Database: The Organizing Principles of the Argument Structure.« *A survey on corpus-based research = Panorama de investigaciones basadas en corpus*. 293–206.

Šorli, Mojca. 2011. »Pragmatic Components in the Slovene Lexical Database Meaning Descriptions.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 251–259.

III. SLOVARSKÉ IN SEMANTIČNE PODATKOVNE BAZE

Anastassiadis-Symeonidis, Anna in Nikolaou, Georgia. 2006. »Electronic Database of New Words in Greek.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marello. Torino: European Association for Lexicography, Università di Torino. 377–381.

Atkins, B. T. Sue, Kilgarrriff, Adam in Rundell, Michael. 2010. »Database of ANalysed Texts of English (DANTE): the NEID database project.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 549–556.

Borin, Lars, Forsberg, Markus, Kokkinakis, Dimitrios in Toporowska Gronostaj, Maria. 2010. »The Past Meets the Present in Swedish FrameNet++.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 269–281.

Braasch, Anna in Pedersen S., Bolette. 2002. »Recent Work in the Danish Computational Lexicon Project 'STO'«. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 301–314.

- Convery, Cathal, Atkins B. T. Sue, Kilgarrieff Adam in Rundell, Michael. 2012. »The DANTE Database (Database of ANalysed Texts of English).« *Electronic Lexicography*. Ur. Magali Paquot in Sylviane Granger. Oxford University Press. 293–295.
- Erjavec, Tomaž in Fišer, Darja. 2006. »Building Slovene WordNet.« *Proceedings / 5th International Conference on Language Resources in Evaluation [LREC 2006] Genoa, Italy, 22-28 May 2006*. 1678–1683.
- Erjavec, Tomaž in Fišer, Darja. 2006. »Building the Slovene Wordnet: first steps, first problems.« *Proceedings / Third International WordNet Conference, GWC 2006, Seogwipo, Jeju Islin, Korea, January 22-26, 2006*.
- Fellbaum, Christiane 1998. »*WordNet: An Electronic Lexical Database*.« Cambridge, MA: MIT Press.
- Fišer, Darja. 1995. »Pristopi k izdelavi leksikalnih podatkovnih zbirk.« *Jezik in slovstvo* 50(6). 17–32.
- Fišer, Darja. 2007. »A multilingual approach to building Slovene wordnet.« Ur. Elena Paskaleva. *A common natural language processing paradigm for Balkan languages / held in conjunction with the International Conference RANLP - 2007, September 27-29, 2007, Borovets, Bulgaria*. 1–6.
- Fišer, Darja. 2007. »Using multilingual resources for building SloWNet faster.« Ur. Attila Tanács. *GWC 2008 / The Fourth Global WordNet Conference, Szeged, Hungary, January 22-25, 2008*. 185–193.
- Fišer, Darja. 2009. »Leveraging parallel corpora in existing wordnets for automatic construction of the Slovene wordnet.« *Human language technology / Zygmunt Vetulani*. Ur. Zygmunt Vetulani in Hans Uszkoreit. Berlin; Heidelberg: Springer. 162–166.
- Fišer, Darja. 2009. »Pristopi za avtomatizirano gradnjo semantičnih zbirk.« *Terminologija in sodobna terminografija*. Ur. Nina Ledinek, Mojca Žagar Karer in Marjeta Humar. Ljubljana: Založba ZRC, ZRC SAZU. 357–370.
- Fišer, Darja. 2009. »sloWNET - slovenski semantični leksikon« Ur. Marko Stabej. *Infrastruktura slovenščine in slovenistike*. 145–149.
<<http://www.centerslo.net/files/file/simpozij/simp28/Fiser.pdf>>.
- Fišer, Darja, Broda, Bartosz in Piasecki, Maciej. 2012. »Weaving sloWNet using window-based co-occurrence features.« Ur. Tomaž Erjavec in Jerneja Žganec Gros. *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012, [Ljubljana, Slovenia]*. 63–68.
- Fišer, Darja in Erjavec, Tomaž. 2008. »Predstavitev in analiza slovenskega wordneta« Ur. Tomaž Erjavec in Jerneja Žganec Gros. *Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008*. 37–42.
- Fišer, Darja in Erjavec, Tomaž. 2010. »sloWnet: construction in corpus annotation.« Ur. Pushpak Bhattacharyya, Christiane Fellbaum in Piek Vossen. *Principles, construction in application of multilingual wordnets*. 177–183.

- Fišer, Darja in Erjavec, Tomaž. 2011. »sloWNet: slovenski semantični leksikon.«
<<http://nl.ijs.si/slownet/>>.
- Fišer, Darja in Ljubešić, Nikola. 2011. »Avtomatsko luščenje hrvaško-slovenskega leksikona iz primerljivih korpusov.« *Meddisciplinarnost v slovenistiki*. Ur. Simona Kranjc. Ljubljana: Znanstvena založba Filozofske fakultete. 137–144.
<http://www.centerslo.net/files/file/simpozij/simp30/Zbornik/Fiser_Ljubesic.pdf>.
- Fišer, Darja in Novak, Jernej. 2011. »Visualizing sloWNet.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 76–82.
- Fišer, Darja in Sagot, Benoît. 2008. »Combining multiple resources to build reliable wordnets.« Ur. Petr Sojka. *Text, speech in dialogue / 11th International Conference, TSD 2008, Brno, Czech Republic, September 8-12, 2008*. 61–68.
- Fontenelle, Thierry. 2012. »WordNet, FrameNet and Other Semantic Networks in the International Journal of Lexicography – The Net Result?« *International Journal of Lexicography* 25(4). 437–449.
- Hult, Ann-Kristin. 2010. »Lexin - a report from a recycling lexicographic project in the North.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 800–809.
- Janssen, Maarten. 2006. »Orthographic Variation in Lexical Databases.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelllo. Torino: European Association for Lexicography, Università di Torino. 167–172.
- Kalliokuusi, Virpi in Varantola, Krista. 2000. »User-Sensitive Lexical Databases: A case of lexical knowledge management.« *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000 Stuttgart, Germany, August 8th – 12th, 2000*. Ur. Stefan Evert, Ulrich Heid, Egbert Lehmann in Christian Rohrer. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. 393–402.
- Kernerman, Ilan. 2011. »From Dictionary to Database: Creating a Global Multi-Language Series.« in *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 113–121.
- Klosa, Annette, Schnörch, Ulrich in Storjohann, Petra. 2006. »A Lexical in Lexicological, Corpus - based Hypertext Information System at the Institut für Deutsche Sprache, Mannheim elexiko.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelllo. Torino: European Association for Lexicography, Università di Torino. 425–429.
- Lorentzen, Henrik in Trap-Jensen, Lars. 2011. »There And Back Again – from Dictionary to Wordnet to Thesaurus in Vice Versa: How to Use in Reuse Dictionary Data in a Conceptual Dictionary.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex*

- 2011, 10-12 November 2011, Bled, Slovenia. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 175–179.
- Meyer, Peter. 2011. »Vernetziko: A Cross-Reference Management Tool for the Lexicographer's Workbench.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 191–198.
- Müller-Spitzer, Carolin. 2008. »The Lexicographic Portal of the IDS: Connecting Heterogeneous Lexicographic Resources by a Consistent Concept of Data Modelling.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 457–461.
- Müller-Spitzer, Carolin. 2010. »OWID – A Dictionary Net for Corpus-Based Lexicography of Contemporary German.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 445–452.
- Nimb, Sanni. 2004. »A Syntactic Lexicon for Danish Adverbs.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Université de Bretagne - sud. 905–910.
- Novak, Jernej, Fišer, Darja in Erjavec, Tomaž. 2011. »slowTool: orodje za iskanje, urejanje in vizualizacijo semantičnih leksikonov.« (<http://nl.ijs.si/slowtool>)
- Renau, Irene in Battaner, Paz. 2011. »The Spanish Learner's Dictionary DAELE on the Panorama of the Spanish E-lexicography. A Short Retrospective of Spanish in European E-lexicography.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 221–226.
- Rundell, Michael in Atkins B. T., Sue. 2011. »The DANTE database: a User Guide.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 233–246.
- Ruppenhofer, Josef, Baker, Collin F. in Fillmore, Charles J. 2002. »The FrameNet Database in Software Tools.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 371–375.
- Ruppenhofer, Josef, Baker, Collin F. in Fillmore, Charles J. 2002. »Collocational Information in the FrameNet Database.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 359–369.

- Schoonheim, Tanneke in Tempelaars, Rob. 2010. »Dutch Lexicography in Progress: the Algemeen Nederlins Woordenboek (ANW).« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 718–725.
- Selva, Thierry, Verlinde, Serge in Binon, Jean. 2002. »Le DAFLES, un nouveau dictionnaire électronique pour apprenants du français.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 199–208.
- Tavčar, Aleš, Fišer, Darja in Erjavec, Tomaž. 2012. »sloWCrowd: orodje za popravljanje wordneta z izkoriščanjem moči množic.« Ur. Tomaž Erjavec in Jerneja Žganec Gros. *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012, [Ljubljana, Slovenia]*. 197–202.
- Tiberius, Carole in Niestadt, Jan. 2010. »The ANW: an online Dutch Dictionary.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 747–753.
- Van Der Vliet, Hennie, Maks, Isa, Vossen, Piek in Segers, Roxane. 2010. »The Cornetto database: Semantic issues and linking lexical units in synsets.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 477–483.
- Vatvedt Fjeld, Ruth. 2000. »An outline of Norwegian Lexical Database (LDB) in its classification of adjectives.« *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000 Stuttgart, Germany, August 8th – 12th, 2000*. Ur. Stefan Evert, Ulrich Heid, Egbert Lehmann in Christian Rohrer. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. 671–678.
- Verlinde, Serge in Binon, Jean. 2006. »The Base lexicale du français (BLF): A Multifunctional Online Database for Learners of French.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelllo. Torino: European Association for Lexicography, Università di Torino. 471–481.
- Viberg, Åke. 2002. »The Swedish WordNet Project.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 407–412.
- Vintar, Špela in Fišer, Darja. 2009. »Adding multi-word expressions to sloWNet« Ur. Tomaž Erjavec. *Research infrastructure for digital lexicography / Mondilex Fifth Open Workshop, Ljubljana, Slovenia, October 14-15, 2009*. 56–63.
- Vintar, Špela in Fišer, Darja. 2011. »Enriching Slovene WordNet with domain-specific terms.« *Translation: computation, corpora, cognition* 1(1). 29–44. <<http://www.t-c3.org/index.php/t-c3/article/view/4>>.

IV. SODOBNA SLOVARSKA TEORIJA IN PRAKSA

- Apresjan, Jurij, D. 2002. »Principles of Systematic Lexicography.« *Lexicography in Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins*. UK: Euralex. Ur. Corr  ard. 91–104.
- Atkins, B. T. Sue. 2002. »Then in Now: Competence in Performance in 35 Years of Lexicography.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi.
- Atkins, B. T. Sue. 2008. »Theoretical Lexicography and its relation to Dictionary-making.« *Practical Lexicography. A Reader*. Ur. T. Fontenelle. Oxford: Oxford University Press.
- Atkins, B. T. Sue, Rundell, Michael. 2008. *The Oxford Guide to Practical lexicography*. Oxford: Oxford University Press.
- Azarian, Rogelio in Nazar, Jenny. 2010. »The Living Lexicon: Methodology to set up Synchronic Dictionaries.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Af  k, Ljouwert: Fryske Akademy. 527–538.
- Beeken, Jeannine. 2010. »Quo Vadis Lexicography at the Institute for Dutch Lexicology?« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Af  k, Ljouwert: Fryske Akademy. 557–566.
- Bogaards, Paul. 2008. »Frequency in learners' dictionaries.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Ling  stica Aplicada Universitat Pompeu Fabra. 1231–1236.
<[http://www.euralex.org/elx_proceedings/Euralex2008/015_Euralex_2008_Paul Bogaards_Frequency in Learnes Dictionaries.pdf](http://www.euralex.org/elx_proceedings/Euralex2008/015_Euralex_2008_Paul_Bogaards_Frequency%20in%20Learners%20Dictionaries.pdf)>.
- Cavallin, Karin. 2012. »Exploring semantic change with lexical sets.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjus  n. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 1018–1022.
- Coffey, Stephen. 2010. »'Offensive' items, in less offensive alternatives, in English monolingual learners' dictionaries.« in *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Af  k, Ljouwert: Fryske Akademy. 1270–1281.
<[http://www.euralex.org/elx_proceedings/Euralex2010/119_Euralex_2010_9_COFFEY_Offensive items, in less offensive alternatives, in English monolingual learners dictionaries.pdf](http://www.euralex.org/elx_proceedings/Euralex2010/119_Euralex_2010_9_COFFEY_Offensive%20items,%20in%20less%20offensive%20alternatives,%20in%20English%20monolingual%20learners%20dictionaries.pdf)>.
- Demers, Marie-Claude in Kernerman, Ilan. 2012. »Lexicographic interchange between a specialized in a general language.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjus  n. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 750–757.
- De Schryver, Gilles-Maurice. 2012. »Lexicography in the crystal ball: facts, trends in outlook.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt

- Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo.
- Dobrovol'skij, Dmitrij. 2010. »Deiktische Konstruktionen des Deutschen aus lexikographischer Perspektive.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1282–1290.
- Drstvenšek, Nina. 2003. »Vloga besedilnega korpusa pri postavitvi geselskega članka v enojezičnem slovarju.« *Jezik in slovstvo* 48(5). 65–81.
- Dziemianko, Anna. 2008. »Noun in Verb Codes in Pedagogical Dictionaries of English: User-friendliness Revisited.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 1243–1250.
- Dziemianko, Anna. 2010. »Dictionary, lexicon, glossary, wordbook or thesaurus? The usefulness of OALDCE7 in OLT for choosing the right word.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1082–1091.
- Dziemianko, Anna in Lew, Robert. 2006. »When You Are Explaining the Meaning of a Word: The Effect of Abstract Noun Definition Format on Syntactic Class Identification.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marello. Torino: European Association for Lexicography, Università di Torino. 857–863.
- Ellis C., Nick, Frey, Eric in Jalkanen, Isaac. 2009. »The psycholinguistics reality of collocations in semantic prosody.« *Exploring the Lexis-grammar interface*. Ur. Rainer Romer in Ute Shulze.
- Fedorova V., Irina 2004. »Style in Usage Labels in Learner's Dictionaries: Ways of Optimization.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 265–272.
- Fontenelle, Thierry. 1990. »Automatic extraction of lexical-semantic relations from dictionary definitions.« *EURALEX '90 Proceedings Actas del IV congreso internacional IV International Congress*. Malaga: Biblograf.
- Gantar, Polona. 2008. »(Slovenska) leksika med leksikonom in slovnico.« *Jezik in slovstvo* 53(5). 19–35.
- Gefen, Raphael in Kernerman, Ari. 2010. »Defining Dictionary Definitions for EFL Dictionaries.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1327–1332.
- Gehweiler, Elke in Unger, Christiane. 2012. »Digital in traditional resources for the second edition of the Deutsches Wörterbuch.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 786–793.

- Granger, Sylviane in Paquot, Magali. 2008. »From dictionary to phrasebook.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 1345–1355.
- Hanks, Patrick. 2008. »Lexical patterns: From hornby to hunston in beyond.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 89–129. <http://www.patrickhanks.com/uploads/5/1/4/9/5149363/hornby_to_hunston.pdf>.
- Hanks, Patrick. 2010. »Terminology, phraseology, and lexicography.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1299–1308. <http://www.patrickhanks.com/uploads/5/1/4/9/5149363/terminology_phraseology.pdf>.
- Heid, Ulrich. 2004. »On the presentation of collocations in monolingual dictionaries.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 729–738.
- Hoekstra, Eric. 2010. »Grammatical information in dictionaries.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1007–1012.
- Jarošová, Alexandra in Benko, Vladimír. 2012. »The Dictionary of the Contemporary Slovak Language: A Product of Tradition in Innovation.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusén. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo. 257–261.
- Kilgariff, Adam. 1997. »'I Don't Believe in Word Senses'«. *Computers in the Humanities* 31(2). 91–113.
- Kola, Kjersti Wictorsén. 2012. »A study of pupils' understanding of the morphological information in the Norwegian electronic dictionary Bokmålsordboka in Nynorskordboka.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusén. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo. 672–675. <http://www.euralex.org/elx_proceedings/Euralex2012/pp672-675_Kola.pdf>.
- Kosem, Iztok. 2008. »Dictionaries for University Students: A Real Deal or Merely a Marketing Ploy?«. *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 1575–1584.
- Kovarikova, Dominika, Chlumska, Lucie in Cvrcek, Vaclav. 2012. »What belongs in a dictionary? The Example of Negation in Czech.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusén. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo. 822–827.

<http://www.euralex.org/elx_proceedings/Euralex2012/pp822-827_Kovarikova_ChlumskainCvrcek.pdf>.

Krek, Simon. 2004. »Slovarji serije COBUILD in formalizacija definicijskega jezika.« *Jezik in slovstvo* XLIX/2. 3–16.

Landau, Sidney. 2001. *Dictionaries: the art and craft of lexicography*. 2nd edition. Cambridge: Cambridge University Press.

Lane, Church in Grundy, Valerie. 2006. »Lexicographic Profiling: an Aid to Consistency in Dictionary Entry Design.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelllo. Torino: European Association for Lexicography, Università di Torino. 1097–1107.

Lew, Robert in Dziemianko, Anna. 2012. »Single-clause when-definitions: Take three.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 997–1002.

Lew, Robert in Galas, Katarzyna. 2008. »Can dictionary skills be taught? The effectiveness of lexicographic training for primary-school-level Polish learners of English.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 1273–1285.

Lorentzen, Henrik. 2004. »The Danish Dictionary at large: Presentation, Problems in Perspectives.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 285–294.

Lorentzen, Henrik. 2010. »Orthographical Dictionaries: How Much Can You Expect? The Danish Spelling Dictionary Revis(it)ed.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Fryske Akademy. 664–670.

Moerdijk, Fons. 2008. »Frames in Semagrams. Meaning Description in the General Dutch Dictionary.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 561–569.

Novotná, Renata. 2012. »Treatment of Complex Prepositions in Czech in English Dictionaries.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 743–749. <http://ucnk.ff.cuni.cz/english/doc/Novotna_Treatment-of-Complex-Prepositions-in-Czech-in-English-Dictionaries.pdf>.

Oelke, Daniela, Eklund, Ann-Marie, Marinov, Svetoslav in Kokkinakis, Dimitrios. 2012. »Visual analytics in the language of web query logs-a terminology perspective.« *Proceedings of the 15th EURALEX*

- international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 541–548.
- Parqui, Jaap, Den Boon, Ton in Hendrickx, Ruud. 2010. »Crossing borders in lexicography: How to treat lexical variance between countries that use the same language.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 704–708.
- Proost, Kristel. 2008. »Verb Class-specific Criteria for the Differentiation of Senses in Dictionary Entries.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 553–555.
- Romero Aguilera, Laura. 2008. »Colocaciones léxicas en diccionarios generales monolingües del español.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 1401–1408.
- Rundell, Michael. 1999. »Dictionary use in production.« *International Journal of Lexicography* 12(1).
- Rundell, Michael. 2006. »More than One Way to Skin a Cat: Why Full-Sentence Definitions Have not Been Universally Adopted.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th–9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelllo. Torino: European Association for Lexicography, Università di Torino.
- Rundell, Michael. 2009. »The Road to Automated Lexicography: First Banish the Drudgery then the Drudges?« in *eLexicography in the 21 st century: New challenges, new applications, Proceedings of eLex 2009, Louvain-la-Neuve, 22–24 October 2009*. Ur. M. Paquot in S. Granger. 9–10.
- Rundell, Michael. 2012. »‘It works in practice but will it work in theory?’ The uneasy relationship between lexicography in matters theoretical.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 47–93.
<[http://www.euralex.org/elx_proceedings/Euralex2012/pp47-92 Rundell.pdf](http://www.euralex.org/elx_proceedings/Euralex2012/pp47-92_Rundell.pdf)>.
- Simonsen, Henrik Køhler. 2008. »Lexicographic Document Templates: Text Genre Conventions in Corporate Lexicography.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 1065–1072.
- Šorli, Mojca 2003. »The Retrieval of Data for Slovene-X Dictionaries.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 849–854.

- Trap-Jensen, Lars. 2002. »Descriptive in Normative Aspects of Lexicographic Decision-Making: The Borderline Cases.« 503–508.
- Trap-Jensen, Lars. 2004. »Spoken Language in Dictionaries: Does It Really Matter?« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 311–318.
- Tutin, Agnès. 2008. »For an extended definition of lexical collocations.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 1453–1460.
- Van Der Meer, Geart. 2002. »Dictionary Entry and Access Trying to see Trees in Woods.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi.
- Van Der Meer, Geart. 2004. »On Defining: Polysemy, Core Meanings in Great Simllicity.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 199–200.
- Van Der Meer, Geart. 2008. »On Connotation, Denotation in All That, or: Why a Nigger Is Not a 'Black Person'. A case study: the word nigger in similar terms.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 1547–1555.
- Villada Moirón, M. Begoña. 2004. »Distinguishing Prepositional Complements from Fixed Arguments.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 935–941.
- Vrbinc, Alenka. 2010. »Part-Of-Speech Labelling in the Retrieval of Phraseological Units.« in *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*, Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy.
- Wojciechowska, Sylwia. 2010. »Metonymy representation in English monolingual learners' dictionaries: Problems in solutions.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1436–1449.

V. E-LEKSIKOGRAFIJA

- Abel, Andrea, Klosa, Annette. 2012. »The lexicographic working environment in theory in practice.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 6–11.
- Corréard, Marie-Hélène. 2002. »Are space-saving strategies relevant in electronic dictionaries?« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark*

- August 13-17, 2002. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 463–470.
- Docherty, Vincent J. 2000. »Dictionaries on the Internet : an Overview.« *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000 Stuttgart, Germany, August 8 th – 12th , 2000*. Ur. Stefan Evert Ulrich Heid in Christian Rohrer Egbert Lehmann. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. 67–74.
- Elliott, Laura in Williams Sarah. 2006. »Pasadena: A New Editing System for the Oxford English Dictionary.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelllo. Torino: European Association for Lexicography, Università di Torino. 257–264.
- Flinz, Carolina. 2011. »The microstructure of Online Linguistics Dictionaries: obligatory in facultative elements.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 83–88
- Giacomini, Laura. 2010. »A proposal for an electronic dictionary of Italian collocations highlighting lexical prototypicality in the syntactic-semantic relations between collocation partners.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1183–1192.
- Granger, Sylviane, Paquot, Magali, ur. 2012. *Electronic Lexicography*. Oxford Unyversity Press.
- Hanks, Patrick. 2012. »Corpus evidence in electronic lexicography.« in *Electronic Lexicography*, Ur. Sylviane Granger in Magali Paquot. Oxford University Press.
- Heid, Ulrich in Zimmermann, Jan Timo. 2012. »Usability testing as a tool for e-dictionary design : collocations as a case in point.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 661–671.
- Hult, Ann-Kristin. 2012. »Old in New User Study Methods Combined—Linking Web Questionnaires with Log Files from the Swedish Lexin Dictionary.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 922–928.
<http://www.euralex.org/elx_proceedings/Euralex2012/pp922-928 Hult.pdf>.
- Joffe, David in De Schryver, Gilles-Maurice. 2004. »TshwaneLex A State-of-the-Art Dictionary Compilation Program.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 99–104.
- Koplenig, Alexander. 2011. »Understining How Users Evaluate Innovative Features of Online Dictionaries – An Experimental Approach.« *Electronic lexicography in the 21st century: new applications for new*

- users. *Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 147–150.
- Kosem, Iztok in Kosem, Karmen ur. 2011. »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.
<http://www.trojina.si/elex2011/elex2011_proceedings.pdf>.
- Kwary, Deny Arnos. 2008. »From Language-Oriented to User-Oriented Electronic LSP Dictionaries: A Case Study of an English Dictionary of Finance for Indonesian Students.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 1112–1120. <http://www.euralex.org/elx_proceedings/Euralex2010/103_Euralex_2010_7_KWARY_From_Language-Oriented_to_User-Oriented_Electronic_LSP_Dictionaries_A_Case_Study_of_an_English.pdf>.
- Kwary, Deny Arnos. 2011. »Adaptive Hypermedia in User-Oriented Data for Online Dictionaries: A Case Study on an English Dictionary of Finance for Indonesian Students.« *International Journal of Lexicography* 25(1). 30–49.
- Lannoy, Vincent. 2010. »The IDM Free Online Platform for Dictionary Publishers.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 389–401.
- Lew, Robert. 2002. »Questionnaires in Dictionary Use Research : A Reexamination.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 267–271.
<<http://hdl.hinle.net/10593/350>>.
- Lew, Robert. 2008. »Can Dictionary Skills Be Taught ? The Effectiveness of Lexicographic Training for Primary-School-Level Polish Learners of English The study.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 1273–1285.
- Lew, Robert. 2010. »Users Take Shortcuts: Navigating Dictionary Entries.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1121–1132.
- Lew, Robert, in Roger Mitton. 2011. »Not the Word I Wanted? How Online English Learners' Dictionaries Deal with Misspelled Words.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 165–174
- Lorentzen, Henrik in Theilgaard, Liisa. 2012. »Online dictionaries—how do users find them in what do they do once they have?« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and

- Scinavian Studies, University of Oslo. 654–660.
<http://curis.ku.dk/ws/files/44913954/pp654_660_Lorentzen_in_Theilgaard.pdf>.
- McElvenny, James. 2010. »Mobile phone dictionaries for small languages: the Whitesins electronic dictionary.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1505–1507.
- Měchura, Michal Boleslav. 2008. »Giving them what they want: search strategies for electronic dictionaries.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 1295–1299.
- Müller-Spitzer, Carolin, Koplenig, Alexander, Töpel, Antje. 2011. »What Makes a Good Online Dictionary? – Empirical Insights from an Interdisciplinary Research Project.« *Proceedings of eLex 2011, Bled, 10-12 November 2011: Electronic Lexicography in the 21st Century - New Applications for New Users*. Ur. Karmen Kosem in Iztok Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 203–208.
- Nesi, Hilary. 2000. »Electronic Dictionaries in Second Language Vocabulary Comprehension in Acquisition : the State of the Art.« *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000 Stuttgart, Germany, August 8 th – 12th , 2000*. Ur. Stefan Evert, Ulrich Heid in Christian Rohrer Egbert Lehmann. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. 839–848.
- Pala, Karel in Horák, Aleš. 2006. »From Web Pages to Dictionary: a Language-Independent Dictionary Writing System.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelllo. Torino: European Association for Lexicography, Università di Torino. 199–204.
- Pedro A. Fuertes-Olivera, Henning Bergenholtz, ur. 2010. *E-lexicography. The Internet, Digital Initiatives in Lexicography*. London, New York: Continuum.
- Prinsloo, D. J., Heid, Ulrich, Bothma, Theo in Faaß, Gertrud. 2011. »Interactive, dynamic electronic dictionaries for text production.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 215–220.
- Rylova, Anna. 2010. »Electronic Dictionary in Dictionary Writing System: how this duo works for dictionary user's needs (ABBYY Lingvo in ABBYY Lingvo Content case).« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 463–467.
<[http://www.euralex.org/elx_proceedings/Euralex2010/030_Euralex_2010_1_RYLOVA_Electronic Dictionary in Dictionary Writing System_how this duo works for dictionary users ne.pdf](http://www.euralex.org/elx_proceedings/Euralex2010/030_Euralex_2010_1_RYLOVA_Electronic_Dictionary_in_Dictionary_Writing_System_how_this_duo_works_for_dictionary_users_ne.pdf)>.
- Sato, Hiroaki. 2000. »Multi-Functional Software for Electronic Dictionaries.« *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000 Stuttgart, Germany, August 8 th – 12th , 2000*. Ur.

Stefan Evert, Ulrich Heid in Christian Rohrer Egbert Lehmann. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. 863–870.

Schryver, Gilles-Maurice De, in David Joffe. 2004. »On How Electronic Dictionaries are Really Used.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 187–196.

Spohr, Dennis. 2008. »Requirements for the design of electronic dictionaries in a proposal for their formalisation.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 617–629.

Schoonheim, Tanneke, Tiberius, Carole, Niestadt, Jan in Tempelaars, Rob. 2012. »Dictionary Use in Language Games: Getting to Know the Dictionary as Part of the Game.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 974–979.

Trap-Jensen, Lars. 2010. »One, two, many: Customization in user profiles in internet dictionaries.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1133–1143.
<http://www.euralex.org/elx_proceedings/Euralex2010/105_Euralex_2010_7_TRAP-JENSEN_One, Two, Many_Customization in User Profiles in Internet Dictionaries.pdf>.

Verlinde, Serge in Binon, Jean. 2010. »Monitoring dictionary use in the electronic age.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 1144–1151.
<http://www.euralex.org/elx_proceedings/Euralex2010/106_Euralex_2010_7_VERLINDE BINON_Monitoring Dictionary Use in the Electronic Age.pdf>.

VI. KORPUSI IN KORPUSNA LEKSIKOGRAFIJA

Arhar Holdt, Špela. 2008. »FidaPLUS: the upgrade of the Slovene reference corpus« Ur. Vida Jesenšek. *Wörterbuch und Übersetzung / 4. Internationales Kolloquium zur Lexikographie und Wörterbuchforschung, Universität Maribor, 20. bis 22. Oktober 2006*. 286–300.

Arhar Holdt, Špela. 2009. »Uporabniška evalvacija korpusa FidaPLUS : zasnova vprašalnika, prvi rezultati.« P. Str. in *Infrastruktura slovenščine in slovenistike*. Ur. Marko Stabej. Ljubljana: Znanstvena založba Filozofske fakultete. 19–26.
<<http://www.centerslo.net/files/file/simpozij/simp28/Arhar.pdf>>.

Arhar Holdt, Špela in Gorjanc, Vojko. 2007. »Korpus FidaPLUS: nova generacija slovenskega referenčnega korpusa.« *Jezik in slovstvo* 52(2). str. 95–110.

Arhar Holdt, Špela, Gorjanc, Vojko in Krek, Simon. 2007. »FidaPLUS corpus of Slovenian: the new generation of the Slovenian reference corpus: its design in tools.« *Proceedings of the Corpus*

Linguistics Conference, CL2007, University of Birmingham, UK, 27-30 July 2007, Ur. Matthew Davies idr. Birmingham.

Arhar Holdt, Špela, Kosem, Iztok in Logar Berginc, Nataša. 2012. »Izdelava korpusa Gigafida in njegovega spletnega vmesnika.« *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012, [Ljubljana, Slovenia]*. Ur. Jerneja Žganec Gros in Tomaž Erjavec. Ljubljana: Institut Jožef Stefan. 16–21.

Avinesh, P. V. S., McCarthy, Diana, Glennon, Dominic in Pomikálek, Jan. 2010. »Domain Specific Corpora from the Web.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 336–342.

Barnbrook, Geoff. 1996. *Language in computers: A Practical introduction to the computer analysis of language*. Edinburgh . Edinburgh: Edinburgh University Press.

Barnbrook, Geoff, Pernilla Danielsson, Michaela Mahlberg, ur. 2005. *Meaningful Texts: The Extraction of Semantic Information from Monolingual in Multilingual Corpora*. Continuum.

Baroni, Marco, Kilgariff, Adam, Pomikálek, Jan in Rychlý, Pavel. 2006. »WebBootCaT: a web tool for instant corpora.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marello. Torino: European Association for Lexicography, Università di Torino. 123–131.

Braasch, Anna. 2006. »Exploitation of syntactic patterns for sense group identification.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marello. Torino: European Association for Lexicography, Università di Torino. 133–139.

Cavallin, Karin. 2012. »Exploring semantic change with lexical sets.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 1018–1022.

Clear, Jeremy. 1987. »Overview of the Role of Computing in Cobuild.« in *Looking up: An Account of the COBUILD Project in Lexical Computing*. Ur. J. M. Sinclair. London: Collins.

Čermak, František. 2005. »Jezikovni korpus: sredstvo in vir spoznanj.« *Študije o korpusnem jezikoslovju*, Ur. Simon Krek in Vojko Gorjanc. Ljubljana: Krtina. 137–172

Čermák, František. 2006. »Collocations , Collocability in Dictionary.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marello. Torino: European Association for Lexicography, Università di Torino. 929–937.

Čermák, František. 2009. »Leksikografovi zapiski o korpusnem slovarju.« *Jezik in slovstvo* 54(3/4). 25–42.

- Erjavec, Tomaž, Fišer, Darja, Krek, Simon in Ledinek, Nina. 2010. »The JOS linguistically tagged corpus of Slovene.« *Proceedings / LREC 2010, 7th International Conference on Language Resources in Evaluations, 19-21 May 2010, Valletta, Malta*. 1806–1809.
- Evert, Stefan idr. 2004. »Supporting corpus-based dictionary updating.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud. 255–264.
- Feliu, Judit, Pedemonte, Berta in Guirado, Cristina. 2010. »Corpus Exploitation Strategies for the Lexicographic Definition Task.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 307–319.
- Fišer, Darja, Pollak, Senja in Vintar, Špela. 2011. »Samodejno luščenje definicij iz specializiranih besedil.« *Meddisciplinarnost v slovenistiki*. Ur. Simona Kranjc. Ljubljana: Znanstvena založba Filozofske fakultete. 145–150
<http://www.centerslo.net/files/file/simpozij/simp30/Zbornik/Fiser_Pollak_Vintar.pdf>.
- Gantar, Polona. 2007. *Stalne besedne zveze v slovenščini: korpusni pristop*. Ur. Helena Dobrovoljc. Ljubljana: Založba ZRC, ZRC SAZU.
- Geirnaert, D. J. G. in Kruyt J. G. 2002. »Text Typology in Selection Criteria for a Balanced Corpus: the Integrated Language Database of 8th-21st-Century Dutch.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 401–406.
- Gianitsová, Lucia. 2005. »Morphological Analysis of the Slovak National Corpus.« Pp. 166–178 in *Insight into Slovak in Czech Corpus Linguistics*. Bratislava: Veda.
- Gorjanc, Vojko. 1999. »Korpusi v jezikoslovju in korpus slovenskega jezika FIDA.« *35. seminar slovenskega jezika, literature in kulture*. Ur. Erika Kržišnik. Ljubljana: Oddelek za slovanske jezike in književnosti. 47–59.
- Gorjanc, Vojko. 2003. »Odkrivanje leksikalnih sprememb s pomočjo korpusa.« *Współczesna polska i słoweńska sytuacja językowa – Sodobni jezikovni položaj na Poljskem in v Sloveniji*. Ur. S. Gajda in A. Vidovič Muha. Opole, Ljubljana: Uniwersytet Opolski, Instytut Filologii Polskiej – Univerza v Ljubljani, Filozofska fakulteta. 99–111.
- Gorjanc, Vojko. 2005. *Uvod v korpusno jezikoslovje*. Domžale: Izolit.
- Gorjanc, Vojko in Fišer, Darja. 2010. *Korpusna analiza*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Gorjanc, Vojko, Krek, Simon. 2001. »A Corpus-Based Dictionary Database as the Source for Compiling Slovene-X Dictionary.« *COMPLEX 2001. 6th Conference on Computational Lexicography in Corpus*

- Research: »Computational Lexicography in New EU Languages«*. Birmingham: The University of Birmingham. 41–47.
- Gorjanc, Vojko, Vintar, Špela. 2000. »Nekatere možnosti jezikoslovne izrabe enojezikovnih korpusov.« *36. seminar slovenskega jezika, literature in kulture*. Ur. Erika Kržišnik. Ljubljana: Oddelek za slovanske jezike in književnosti. 335–348.
- Halliday, Michael, Kirkwood, Alexander. 2005. »Jezik kot sistem in jezik kot primer: korpus kot teoretični konstrukt.« *Študije o korpusnem jezikoslovju*. Ur. Simon Krek in Vojko Gorjanc. Ljubljana: Krtina. 59–80.
- Hanks, Patrick. 2012. »The corpus revolution in lexicography.« *International Journal of Lexicography* 25(4). 398–436.
- Heid, Ulrich in Weller, Marion. 2010. »Corpus-derived data on German multiword expressions for lexicography.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 331–340.
- Horák, Alexander, Gianitsová, Lucia, Šimková, Mária, Šmotlák, Martin in Garabík, Radovan. 2004. »Slovak National Corpus.« *Text, Speech in Dialogue. 7th International Conference TSD*. Berlin-Heidelberg: Springer. 89–93.
- Jansson, H. 2010. »Working with the web as a source for dictionaries of informal vocabulary.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 357–363.
- Kennedy, Graeme. 1998. *An introduction to corpus linguistics*. London: Longman.
- Kilgarrieff, Adam, Kovář, Vojtech in Krek, Simon. 2010. »A quantitative evaluation of word sketches.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 372–379.
- Kilgarrieff, Adam, Pomikalek, Jan, Jakubíček, Miloš in Whitelock, Pete. 2012. »Setting up for Corpus Lexicography.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusén. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo. 778–785.
- Kilgarrieff, Adam in Rychlý, Pavel. 2008. »Finding the Words Which Are Most X.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 433–436.
- Křen, Michal in Hlavačova, Jaroslava. 2008. »Corpus as a Means for Study of Lexical Usage Changes.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 357–366.

- Leech, Geoffrey. 2005. »Stanje stvari v korpusnem jezikoslovju.« *Študije o korpusnem jezikoslovju*, Ur. Simon Krek in Vojko Gorjanc. Ljubljana: Krtina. 29–58.
- Logar Berginc, Nataša idr. 2012. *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES : gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.
- Mcenery, Tony, in Andrew Wilson. 2001. *Corpus linguistics: an introduction*. Edinburgh: Edinburgh University Press.
- Mcenery, Tony, Richard, Xiao in Tono, Yukio. 2006. *Corpus-Based Language Studies*. London, New York: Routledge.
- Müller-Spitzer, Carolin. 2010. »OWID – A Dictionary Net for Corpus-Based Lexicography of Contemporary German.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 445–452.
- Ooi, Vincent B. Y. 1998. *Computer corpus lexicography*. Edinburgh University Press.
- Rozman, Tadeja idr. 2012. »Šolar / korpus Šolar kot podatkovna zbirka.«
<<http://www.slovenscina.eu/korpusi/solar>>.
- Selegey, Vladimir. 2012. »On automated semantic in syntactic annotation of texts for lexicographic purposes.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 456–461.
<http://www.euralex.org/elx_proceedings/Euralex2012/pp456-461_Selegey.pdf>.
- Simonsen Køhler, Henrik. 2008. »Lexicographic Document Templates: Text Genre Conventions in Corporate Lexicography.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 1065–1072.
- Stabej, Marko in Gorjanc, Vojko. 2000. »FidaPLUS: korpus slovenskega jezika« Ur. Simon Krek idr.
<http://www.fidaplus.net/Info/Info_index.html>.
- Stabej, Marko, Gorjanc, Vojko in Erjavec, Tomaž. 1998. »Korpus FIDA.« *Jezikovne tehnologije za slovenski jezik / Mednarodna multi-konferenca Informacijska družba - IS'98, Ljubljana, 6. do 7. oktobra 1998*. Ljubljana: Institut Jožef Stefan. 124–127.
- Steiner, Petra. 2004. *Wortarten und Korpus: Automatische Wortartenklassifikation durch distributionelle und quantitative Verfahren*. Shaker Verlag. (<http://www.amazon.de/Wortarten-Korpus-Automatische-Wortartenklassifikation-distributionelle/dp/3832223800>).
- Stritar, Mojca. 2012. *Korpusi usvajanja tujega jezika*. Ljubljana: Zveza društev Slavistično društvo Slovenije.

- Summers, Della. 2005. »Corpus lexicography - The importance of representativeness in relation to frequency.«
- Teubert, Wolfgang. 2001. »Corpus Linguistics in Lexicography.« *International Journal of Corpus Linguistics* 6(1). 153.
- Thomas, Jenny in Short, Mick. 1996. *Using corpora for language research: studies in honour of Geoffrey Leech*. London: Longman Publishing Group.
- Vossen, Piek idr. 2011. »DutchSemCor: Building a semantically annotated corpus for Dutch.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 286–296
- Walter, Elizabeth, in Inrew Harley. 2002. »The role of corpus in collocation tools in practical lexicography.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 851–857.

VII. JEZIKOVNE TEHNOLOGIJE

- Arhar Holdt, Špela, in Nina Ledinek. 2008. »Oblikoskladenjske oznake JOS: revizija in nadgradnja nabora oznak za avtomatsko oblikoskladenjsko označevanje slovenščine / Špela Arhar, Nina Ledinek« Ur. Tomaž Erjavec in Jerneja Žganec Gros. *Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008*. 54–59.
- Atkins, B. T. S. in Zampolli, Antonio. 1994. *Computational approaches to the lexicon*. Oxford University Press. (<http://books.google.com/books?id=ITnirwRMIw8C&pgis=1>).
- Baroni, Marco in Bernardini, Silvia. 2004. »BootCaT: Bootstrapping Corpora in Terms from the Web.« *Proceedings of LREC 2004, Lisbon*. Lisbon: ELRA. 1313–1316.
- Bel, Nuria, Calzolari, Nicoletta in Monachini, Monica. 1995. *Common specifications in notation for lexicon encoding in preliminary proposal for the tagsets. — MULTTEXT Deliverable D1.6.1B*. Pisa.
- Boguraev, Bran in Briscoe, Ted. 1989. *Computational Lexicography for Natural Language Processing*. London, New York: Longman Sc & Tech.
- Cartoni, Bruno in Lefer, Marie-Aude. 2010. »Improving the Representation of Word-Formation in Multilingual Lexicographic Tools: the MuLeXFor Database.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 581–591.
- Cinková, Silvie, Holub, Martin in Kríž, Vincent. 2012. »Optimizing semantic granularity for NLP - report on a lexicographic.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scinavian Studies, University of Oslo. 523–531.

- Cipollone, Nick. 2008. »Lexicon Creator: A Tool for Building Lexicons for Proofing Tools in Search Technologies Building full-form lexicons Authoring morphological rules.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 359–369.
- Derouin, Marie-Jeanne. 2008. »ISO-Stinards for Lexicography in Dictionary Publishing.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 663–668.
- Derzhanski, Ivan A. in Kotsyba, Natalia. 2009. »Towards a consistent morphological tagset for Slavic languages : Extending MULTTEXT-East for Polish, Ukrainian in Belarusian.« *MONDILEX Third Open Workshop Bratislava, Slovakia, 15–16 April, 2009*. Bratislava: Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences. 9–26.
- Derzhanski, Ivan in Kotsyba, Natalia. 2008. »The Category of Predicatives in the Light of the Consistent Morphosyntactic Tagging of Slavic Languages.« *MONDILEX First Open Workshop Moscow, Russia, 3–4 October, 2008*. 68–79.
- Dobrovoljc, Kaja, Krek, Simon in Rupnik, Jan. 2012. »Skladenjski razčlenjevalnik za slovenščino = Dependency parser for Slovene« Ur. Tomaž Erjavec in Jerneja Žganec Gros. *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012, [Ljubljana, Slovenia]*. 42–47.
- Elliott, Laura in Williams, Sarah. 2006. »Pasadena: A New Editing System for the Oxford English Dictionary.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelllo. Torino: European Association for Lexicography, Università di Torino. 257–264.
- Erjavec, Tomaž. 1998. »The MULTTEXT-East Slovene Lexicon.« *Proceedings of the 7th Electrotechnical Conference ERK '98, Portorož*. Ljubljana: Institut Jožef Stefan. 189–192.
- Erjavec, Tomaž. 2003. »Označevanje korpusov.« *Jezik in slovstvo* 48(3/4). 61–76.
- Erjavec, Tomaž idr. 2003. »The MULTTEXT-east morphosyntactic specifications for Slavic languages.« *ACL Workshops*. 25–32.
- Erjavec, Tomaž in Džeroski, Sašo. 2000. »Learning to lemmatise Slovene words.« *Learning language in logic*. Springer. 69–88.
- Erjavec, Tomaž in Džeroski Sašo. 2000. »Strojno učenje lematizacije neznanih slovenskih besed.« *Jezikovne tehnologije*, Ur. Tomaž Erjavec in Jerneja Žganec Gros. Ljubljana: Institut Jožef Stefan. 14–30.
- Erjavec, Tomaž in Džeroski, Sašo. 2004. »Machine Learning of Morphosyntactic Structure: Lemmatizing Unknown Slovene Words.« *Applied Artificial Intelligence* 18(1). 17–41.

- Erjavec, Tomaž, Fišer, Darja, Krek, Simon in Ledinek, Nina. 2010. »The JOS linguistically tagged corpus of Slovene.« *Proceedings of the Seventh conference on International Language Resources in Evaluation (LREC'10)*. Malta: European Language Resources Association (ELRA). 1806–1809.
- Erjavec, Tomaž in Krek, Simon. 2008. »The JOS morphosyntactically tagged corpus of Slovene.« *Proceedings of the Sixth International Language Resources in Evaluation (LREC'08)*. Marrakesh: European Language Resources Association (ELRA). 322–326.
- Erjavec, Tomaž, Krek, Simon, Fišer, Darja in Ledinek, Nina. 2011. »Projekt JOS: jezikoslovno označevanje slovenskega jezika = Project JOS.« <<http://nl.ijs.si/jos/>>.
- Erjavec, Tomaž in Ledinek, Nina. 2009. »Odvisnostno površinskostkladenjsko označevanje slovenščine: specifikacije in označeni korpusi.« *Infrastruktura slovenščine in slovenistike*. Ur. Marko Stabej. Ljubljana: Znanstvena založba Filozofske fakultete. 219–224.
- Fišer, Darja. 2010. »Semantično označevanje korpusov.« *Slovenske korpusne raziskave*. Ur. Špela Vintar. Ljubljana: Znanstvena založba Filozofske fakultete. 110–130.
- Fišer, Darja. 2012. »Language resources and tools for semantically enhanced processing of Slovene / Darja Fišer.« *Multilingual processing in eastern in southern EU languages*. Ur. Christina Vertan in Walther Hahn. [S.l.]: Cambridge Scholars. 92–118.
- Fitschen, Arne in Gupta, Piklu. 2008. »Lemmatising in morphological tagging.« *Corpus Linguistics. An International Handbook* 1. 552–564.
- Fontenelle, Thierry. 2006. »Developing a Lexicon for a New French Spell-checker.« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marelli. Torino: European Association for Lexicography, Università di Torino. 151–158.
- Forróová, Martina, Garabík, Radovan, Gianitsová Lucia, Horák, Alexander in Šimková, Mária. 2005. »Návrh morfológického tagsetu SNK.« *Slovanské jazyky v počítačovom spracovaní. Slovko 2003. Zborník z medzinárodnej konferencie*. Bratislava: Veda. 65–83
- Forróová, Martina in Horák, Alexander. 2004. »Morfológická anotácia korpusu.« *Slovenčina na začiatku 21. storočia*. Prešov: Prešovská univerzita, Fakulta humanitných a prírodných vied. 174–186.
- Gala, Nuria in Lafourcade, Mathieu. 2011. »NLP lexicons: innovative constructions in usages for machines in humans.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 89–95.
- Garside, Roger, Leech, Geoffrey in Mcenery, Tony. 1997. *Corpus Annotation: Linguistic Information From Computer Text Corpora*. Harlow: Addison Wesley Longman Limited.
- Geyken, Alexander, Didakowski, Jorg in Siebert, Alexander. 2008. »Generation of Word Profiles on the Basis of a Large in Balanced German Corpus.« *Proceedings of the XIII EURALEX International*

- Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 371–383.
- Gianitsová, Lucia. 2005. »Morphological Analysis of the Slovak National Corpus.« *Insight into Slovak in Czech Corpus Linguistics*. Bratislava: Veda. 166–178
- Grčar, Miha, Juršič, Matjaž, Rupnik, Jan, Krek, Simon in Dobrovoljc, Kaja. 2012. »Obelisk: statistični oblikoskladenjski označevalnik za slovenski jezik.« in *KONFERENCA Jezikovne tehnologije (8 ; 2012 ; Ljubljana) Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012, [Ljubljana, Slovenia]: zbornik 15. mednarodne multikonference Informacijska družba - IS 2012, zvezek C = Proceedings of the Eig*. Ur. Jerneja Žganec Gros in Tomaž Erjavec. Ljubljana: Institut »Jožef Stefan«, Ljubljana. (<http://oznacevalnik.slovenscina.eu/>).
- Grčar, Miha, Juršič, Matjaž, Rupnik, Jan, Krek, Simon in Dobrovoljc, Kaja. 2013. »Statistični oblikoskladenjski označevalnik za slovenski jezik.« (<http://razclenjevalnik.slovenscina.eu/>).
- Greene, Barbara in Rubin, Gerald. 1971. *Automatic grammatical tagging of English*. Providence, Rhode Island.
- Hajič, Jan. 2004. *Disambiguation of Rich Inflection: Computational Morphology of Czech*. Karolinum Press, Charles University.
- Hrovat, Goran. 2010. »Izdelava oblikoslovnega označevalnika.« Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru.
- Ignatova, Kateryna in Abel, Andrea. 2008. »The Use of Context Vectors for Word Sense Disambiguation within the ELDIT Dictionary.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 403–408.
- Jakopin, Primož in Bizjak, Aleksandra. 1997. »O strojno podprtem oblikoslovnem označevanju slovenskega besedila.« *Slavistična revija* 45(3-4). 513–532.
- Jiménez, Marisa. 2002. »Computational Lexicography in Lexicology NLP Use of Multiple Sources of Information Available In A Spanish Machine-Readable Dictionary.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 129–135.
- Jurafsky, Daniel in Martin, James H. 2009. *Speech in language processing: an introduction to natural language processing*. Prentice Hall <<http://books.google.com/books?id=fZmj5UNK8AQC&pgis=1>>.
- Juršič, Matjaž. 2007. »Implementacija učinkovitega sistema za gradnjo, uporabo in evaluacijo lematizatorjev tipa RDR.« Fakulteta za računalništvo in informatiko, Univerza v Ljubljani.

- Kilgarriff, Adam. 2006. »Collocationality (in how to measure it).« *Proceedings of the XII EURALEX International Congress, Torino, Italia, September 6th-9th, 2006*. Ur. Cristina Onesti, Elisa Conno in Carla Marello. Torino: European Association for Lexicography, Università di Torino. 997–1004. <[http://www.euralex.org/elx_proceedings/Euralex2006/122_2006_V2_Adam KILGARRIFF_Collocationality \(in how to measure it\).pdf](http://www.euralex.org/elx_proceedings/Euralex2006/122_2006_V2_Adam KILGARRIFF_Collocationality (in how to measure it).pdf)>.
- Kilgarriff, Adam, Avinesh, P. V. S. in Pomikálek, Jan. 2011. »Comparable Corpora BootCaT.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 122–128.
- Kilgarriff, Adam, Kovář, Vojtech in Krek, Simon. 2010. »A quantitative evaluation of word sketches.« *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Ur. Anne Dykstra in Tanneke Schoonheim. Afûk, Ljouwert: Fryske Akademy. 372–379 <http://www.euralex.org/elx_proceedings/Euralex2010/019_Euralex_2010_1_KILGARRIFF KOVAR KREK SRDANOVIC TIBERIUS_A Quantitative Evaluation of Word Sketches.pdf>.
- Kilgarriff, Adam, Mcadam, Katy, Rundell, Michael in Rychlý, Pavel. 2008. »GDEX: Automatically Finding Good Dictionary Examples in a Corpus.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 425–432.
- Kilgarriff, Adam in Rychlý, Pavel. 2008. »Finding the words which are most X.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 433–436 <http://www.euralex.org/elx_proceedings/Euralex2008/098_Euralex_2008_Adam Kilgarriff_Pavel Rychly_Finding the Words Which are Most X.pdf>.
- Kilgarriff, Adam, Rychly, Pavel, Smrz, Pavel in Tugwell, David. 2004. »The Sketch Engine.« *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6-10, 2004*. Ur. Geoffrey Williams in Sandra Vessier. Lorient: Universite de Bretagne - sud.
- Kilgarriff, Adam, Rundell, Michael. 2002. »Lexical Profiling Software in its lexicographic applications: a case study.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 807–818.
- Kilgarriff, Adam, Rychly, Pavel, Kovar, Vojtech in Baisa, Vit. 2012. »Finding Multiwords of More Than Two Words.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusen. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo.
- Koeva, Svetla, Krstev, Cvetana, Obradović, Ivan in Vitas, Duško. 2005. »Resources for Processing Bulgarian in Serbian – a brief overview of Completeness, Compatibility, in Similarities electronic

- dictionaries of Bulgarian in Similarities.« *Recent Advances in Natural Language Processing*), 21-23 September 2005, Borovets, Bulgaria. 31–38.
- Kosem, Iztok, Milos Husak in McCarthy, Diana. 2011. »GDEX for Slovene.« *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ur. Iztok Kosem in Karmen Kosem. Ljubljana: Trojina, Institute for Applied Slovene Studies. 151–159.
- Kotsyba, Natalia, Shypnivska Olga in Turska, Magdalena. 2009. »Principles of organising a common morphological tagset for PolUKR (Polish-Ukrainian Parallel Corpus).« *Intelligent Information Systems 2008*. 471–480.
- Krek, Simon. 2012. »New Slovene sketch grammar for automatic extraction of lexical data.« *SKEW3, tretja mednarodna delavnica orodja Sketch Engine, Brno, Češka, 21-22. marec 2012*.
- Krek, Simon in Erjavec, Tomaž. 2008. »Oblikoskladenjske specifikacije in označeni korpusi JOS.« *Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008*. Ur. Tomaž Erjavec in Jerneja Žganec Gros. Ljubljana: Institut Jožef Stefan.
- Krek, Simon in Kilgariff, Adam. 2006. »Slovene Word Sketches.« *Jezikovne tehnologije 5. IJS*. 62–65
- Krenn, Brigitte. 2000. »Empirical Implications on Lexical Association Measures.« in *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000 Stuttgart, Germany, August 8th – 12th, 2000*, Ur. Stefan Evert Ulrich Heid in Christian Rohrer Egbert Lehmann. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Leech, Geoffrey. 1997. »Introducing Corpus Annotation.« *Corpus Annotation* Ur. Roger Garside, Geoffrey Leech in Tony Mcenery. Longman.
- Leech, Geoffrey in Eyes, Elizabeth. 1997. »Syntactic Annotation: Treebanks.« *Corpus Annotation*. Harlow: Longman. 34–52
- Leech, Geoffrey, Mcenery, Tony in Wynne, Martin. 1997. »Further Levels of Annotation.« *Corpus Annotation*. Harrow: Longman. 85–101.
- Lönneker Rodman, Birte in Jakopin, Primož. 2004. »Checking POSBeseda, a Part-of-Speech tagged Slovenian corpus.« *Jezikovne tehnologije: zbornik B 7. mednarodne multi-konference Informacijska družba IS 2004, 9. do 15. oktober 2004*. Ljubljana: Institut Jožef Stefan. 48–55.
- Luder, Marc in Distl, Bernhard. 2008. »Ein elektronisches Lexikon im OLIF-Format für die Erzählanalyse Einleitung Die Erzählanalyse JAKOB Implementierung des Lexikons im OLIF-Format.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Lingüística Aplicada Universitat Pompeu Fabra. 729–735.
- Manning, Christopher in Schütze, Hinrich. 1999. *Foundations of statistical natural language processing*. Cambridge: MIT Press.

- Marcus, Mitchell P, Marcinkiewicz, Mary Ann in Santorini, Beatrice. 1993. »Building a Large Annotated Corpus of English: The Penn Treebank.« *Computational Linguistics* Volume 19(2). 313–330.
- Mikheev, Andrei. 2005. »Text segmentation.« *The Oxford handbook of computational linguistics*. Ur. Ruslan Mitkov. Oxford: Oxford University Press. 201–218.
- Mladeníć, Dunja. 2002. »Automatic word lemmatization.« in *Jezikovne tehnologije / Information Society Multi-Conference, IS 2002, 14.-15. oktober 2002*. Institut Jožef Stefan.
- Oakes, Michael P. 1998. *Statistics for Corpus Linguistics*. Edinburgh University Press.
<<http://www.amazon.com/Statistics-Linguistics-Edinburgh-Textbooks-Empirical/dp/0748608176>>.
- Pedersen, Sinford, Bolette. 2012. »Lexicography in Language Technology (LT).« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusén. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 31–46.
<[http://www.euralex.org/elx_proceedings/Euralex2012/pp31-46 Pedersen.pdf](http://www.euralex.org/elx_proceedings/Euralex2012/pp31-46%20Pedersen.pdf)>.
- Plisson, Joël, Lavrač, Nada in Mladeníć, Dunja. 2004. »A Rule based Approach to Word Lemmatization.« *Zbornik C 7. mednarodne multi-konference Informacijska družba IS 2004, 9. do 15. oktober 2004*. Institut Jožef Stefan. 83–86.
- Przepiórkowski, Adam. 2009. »A Comparison of Two Morphosyntactic Tagsets of Polish.« *Representing Semantics in Digital Lexicography: Innovative Solutions for Lexical Entry Content in Slavic Lexicography, MONDILEX Fourth Open Workshop Warsaw, Polin, 29 June – 1 July, 2009*. Warsaw: Institute of Slavic Studies, Polish Academy of Sciences. 138–144.
- Rupnik, Jan, Grčar, Miha in Erjavec, Tomaž. 2008. »Improving Morphosyntactic Tagging of Slovene Language through Meta-tagging.« *Informatica* 32. 437–444.
- Schmid, Helmut. 2008. »Tokenizing.« *Corpus Linguistics. An International Handbook*. Berlin: Mouton de Gruyter. 1–17.
- Sedláček, Radek. 1999. »Morfologický analyzátor češtiny.« Masarykova universita fakulta informatiky.
- Selegey, Vladimir. 2012. »On automated semantic in syntactic annotation of texts for lexicographic purposes.« *Proceedings of the 15th EURALEX international Congress 7–11 August, 2012 Oslo*. Ur. Ruth Vatvedt Fjeld in Julie Matilde Torjusén. Oslo: Department of Linguistics and Scininavian Studies, University of Oslo. 456–461.
<[http://www.euralex.org/elx_proceedings/Euralex2012/pp456-461 Selegey.pdf](http://www.euralex.org/elx_proceedings/Euralex2012/pp456-461%20Selegey.pdf)>.
- Sharoff, Serge, Kopotev, Mikhail, Erjavec, Tomaž, Feldman, Anna in Divjak, Dagmar. 2008. »Designing in Evaluating a Russian Tagset.« *Proceedings of the Sixth International Language Resources in Evaluation (LREC'08)* 279–285.
- Tichy, Ondrej in Čermak, Jan. 2008. »An Anglo-Saxon Dictionary in a Morphological Analyzer of Old English Digitization.« *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July*

2008). Ur. Janet DeCesaris in Elisenda Bernal. Barcelona: Institut Universitari De Linguística Aplicada Universitat Pompeu Fabra. 483–488.

Voutilainen, Atro. 2003. »Part-of-speech tagging.« *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press. 219–232.

Wanner, Leo, Klatt, Stefan, Kapanadze, Oleg in Kapanadze Nunu. 2002. »Towards a Semantically Motivated Organization of a Valency Lexicon for NLP: The GREG-Proposal.« *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002 Copenhagen, Denmark August 13-17, 2002*. Ur. Anna Braasch in Claus Povlsen. Copenhagen: Center for sprogteknologi. 231–236.

Wilson, Andrew in Thomas, Jenny. 1997. »Semantic Annotation.« *Corpus Annotation*. Harrow: Longman. 53–65.