



Univerza v Ljubljani
**FILOZOFSKA
FAKULTETA**

POLONA GANTAR

**LEKSIKOGRAFSKI OPIS
SLOVENŠČINE
V DIGITALNEM OKOLJU**

Ljubljana 2015

Leksikografski opis slovenščine v digitalnem okolju

Zbirka Sporazumevanje

Urednik zbirke: Simon Krek

Avtorica: Polona Gantar

Recenzenta: Nataša Logar, Vojko Gorjanc

Prevod povzetka: Simon Krek

Tehnično urejanje in prelom: Aleš Cimprič

Oblikovna zasnova naslovnice: VBG d. o. o.

Založila in izdala: Znanstvena založba Filozofske fakultete Univerze v Ljubljani

Za založbo: Branka Kalenič Ramšak, dekanja Filozofske fakultete

Ljubljana, 2015

Prva izdaja, elektronska izdaja

DOI: 10.4312/9789612377922

Publikacija je brezplačna in dostopna na: <https://e-knjige.ff.uni-lj.si>

Knjiga je izšla s podporo Javne agencije za raziskovalno dejavnost Republike Slovenije.



To delo je ponujeno pod licenco Creative Commons

Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna licenca.

CIP - Kataložni zapis o publikaciji

Narodna in univerzitetna knjižnica, Ljubljana

81.163.6'374:004.9(0.034.2)

GANTAR, Polona

Leksikografski opis slovenščine v digitalnem okolju [Elektronski vir] / Polona Gantar ; [prevod povzetka Simon Krek]. - 1. izd., elektronska izd. - El. knjiga. - Ljubljana : Znanstvena založba Filozofske fakultete, 2015. - (Zbirka Sporazumevanje / Trojina, zavod za uporabno slovenistiko)

ISBN 978-961-237-792-2 (pdf)

282009600

Kazalo vsebine

Predstavitev	12
Zahvala	13
Uvod	15

A. LEKSIKALNA PODATKOVNA BAZA 17

1 Leksikalne baze in sorodni projekti evropskih jezikov. 20

1.1 Jezikovnotehnološke leksikalne podatkovne baze	20
1.1.1 FrameNet (FN)	21
1.1.2 Analiza korpusnih vzorcev (CPA).	23
1.1.3 Leksikalna zbirka za razvoj danskih jezikovnih tehnologij (STO)	25
1.1.4 Norveška leksikalna podatkovna baza (LDB)	26
1.1.5 Švedska jezikovna podatkovna banka Språkbanken	27
1.1.6 Italijanska leksikalna podatkovna baza CLIPS	29
1.1.7 Vezljivostni leksikon čeških glagolov Vallex	31
1.1.8 Distribucijski profil italijanskih glagolov, samostalnikov in pridevnikov Lexlt	32
1.1.9 Špansko-katalonski jezikovnotehnološki podatkovni bazi SenSem in ADESSE	34
1.1.10 WordNet in SloWNet	36
1.1.11 Slovenska podatkovna zbirka ASES	38
1.2 Jezikovnoprostočniške ali slovarske leksikalne podatkovne baze	39
1.2.1 Slovarska podatkovna baza projekta Collins Cobuild.	40
1.2.2 Podatkovna baza na podlagi analize angleških besedil Dante	41
1.2.3 Splošni nizozemski slovar (ANW)	43
1.2.4 Leksikalna podatkovna baza za francoščino (BFL).	45
1.2.5 Nemški jezikovni portal Elexiko	47
1.2.6 Veliki slovar poljskega jezika	48
1.2.7 Slovar za učenje španščine kot tujega jezika Daele	50
1.2.8 Skandinavski večjezični spletni slovarski projekt LEXIN	51
1.3 Obstoječe slovenske leksikalne podatkovne baze	54
1.3.1 Spletni slovar slovenskega jezika – demo različica	56

2 Leksikalna baza in slovar. 58

B. LEKSIKALNA BAZA ZA SLOVENŠČINO	60
3 Osnovna izhodišča	62
3.1 Namen	62
3.2 Glavni poudarki	63
3.3 Uporabnik	66
3.3.1 Človeški uporabnik	66
3.3.2 Jezikovnotehnološka skupnost	69
3.4 Obseg	70
4 Zgradba Leksikalne baze za slovenščino	72
4.1 Strukturiranost podatkov	72
4.2 Shema DTD	74
5 Orodja za pridobivanje, analizo in urejanje leksikalnogramatičnih podatkov	79
5.1 Pridobivanje in analiza leksikalnogramatičnih podatkov	79
5.1.1 Korpusa FidaPLUS, Gigafida in geslovník	81
5.1.2 Analiza konkordančnega niza	82
5.1.3 Leksikalnogramatični profil leksikalne enote in orodje Sketch Engine	84
5.1.3.1 Besedne skice	84
5.1.3.2 Izbor dobrih zgledov in klišikografija	87
5.1.4 Avtomatsko pridobivanje leksikalnogramatičnih podatkov iz korpusa	89
5.1.4.1 Slovníca besednih skic, različica ALLP 0.8	90
5.1.4.2 GDEX za selekcionirano luščenje korpusnih zgledov	91
5.1.4.3 Izdelava skripte API	91
5.1.4.4 Rezultati in evalvacija	92
5.2 Urejanje jezikovnih podatkov (programa za izdelavo podatkovne baze)	93
5.2.1 IDM Dictionary Production system	94
5.2.2 iLex (Integrated Lexicography Editing and XML)	96
C. LEKSIKALNOGRAMATIČNI PODATKI V LEKSIKALNI BAZI ZA SLOVENŠČINO	99

I. IZRAZNI NIVO	100
6 Iztočnica	102
6.1 Vrste iztočnic.	102
6.1.1 Enobesedne iztočnice	103
6.1.2 Večbesedne iztočnice	103
6.1.3 Lastna imena	103
7 Zapis	104
7.1 Oblika	104
7.1.1 Osnovna oblika	105
7.1.1.1 Samostalnik	106
7.1.1.2 Glagol	107
7.1.1.2.1 Zapis glagolov z glagolskim morfemom <i>se/si</i>	108
7.1.1.2.1.1 Glagoli z morfemom <i>se/si</i> v osnovni obliki	108
7.1.1.2.1.2 Glagoli z morfemom <i>se/si</i> v katerem od svojih pomenov	109
7.1.1.2.1.3 Glagoli z morfemom <i>se/si</i> kot upovedovalno možnostjo	109
7.1.1.2.2 Zapis glagolov z glagolskim morfemom <i>ga</i> ali <i>jo</i> .	110
7.1.1.3 Pridevnik	110
7.1.1.4 Prislov	111
7.1.2 Homonimne oblike	112
7.1.3 Besedotvorno povezane oblike	112
7.2 Variantnost	113
7.2.1 Domače – prevzeto	113
7.2.2 Skupaj – narazen	114
7.2.2.1 Skupaj in narazen	114
7.2.2.2 Skupaj in z vezajem.	114
7.2.2.3 Sklopi	115
8 Besedna vrsta	116
8.1 Osnovne besedne vrste	116
8.1.1 Samostalnik (in glagolnik)	116
8.1.2 Glagol	116
8.1.3 Pridevnik (in deležnik)	117
8.1.4 Prislov (in deležje)	117
8.2 Konverzija besednih vrst	117
8.2.1 Posamostaljeni pridevniki	118

8.2.2	Pridevniki v samostalniški rabi	118
8.2.3	Samostalniki v pridevniški rabi.	120
8.2.4	Samostalniki, pridevniki in prislovi v povedni rabi (obravnavava povedkovnika)	120
8.2.5	Drugi primeri	121

II. POMENSKI NIVO 124

9 Pomenska členitev besede. 127

9.1	Členjenje pomenov	128
9.1.1	Enopomenske iztočnice	128
9.1.2	Večpomenske Iztočnice	129
9.1.2.1	Kazalci večpomenskosti	130
9.1.2.1.1	Slovnični kazalci večpomenskosti	130
9.1.2.1.1.1	Samostalnik	131
9.1.2.1.1.2	Glagol	133
9.1.2.1.1.3	Pridevnik	134
9.1.2.1.1.4	Prislov	135
9.1.2.1.2	Sobesedilni in pomenski kazalci večpomenskosti	136
9.1.2.1.2.1	Kolokatorji	136
9.1.2.1.2.2	Udeleženska zgradba in udeleženci	137
9.1.2.1.2.2.1	Glagol.	138
9.1.2.1.2.2.2	Samostalnik	139
9.1.2.1.2.2.3	Pridevnik.	140
9.1.2.1.2.2.4	Prislov.	141
9.1.2.1.2.3	Pomenski elementi	142
9.1.2.1.2.3.1	Samostalnik	143
9.1.2.1.2.3.2	Glagol.	143
9.1.2.1.2.3.3	Pridevnik.	144
9.1.2.1.2.3.4	Prislov.	145
9.1.2.2	Razmerja med pomeni.	146
9.1.2.2.1	Glavni pomen in osnovni pomeni	146
9.1.2.2.1.1	Dobesedni in preneseni pomen.	147
9.1.2.2.1.2	Konkretni in abstraktni pomen	148
9.1.2.2.1.3	Splošni in specifični pomen	149
9.1.2.2.2	Krovni pomen	149

9.1.2.2.3 Podpomeni	150
9.1.3 Razvrstitev pomenov	153
9.2 Združevanje pomenov	156
9.2.1 Razpršena pomenska raba	156
9.2.2 Sistemska polisemija – problem »haremu«	157
9.2.3 Individualna ali ustvarjalna raba	159
10 Pomenski opis besede	161
10.1 Pomenski meni	162
10.2 Pomenski indikator	164
10.2.1 Vrste pomenskih indikatorjev	166
10.2.1.1 Sinonimni indikatorji	166
10.2.1.2 Opisni indikatorji	168
10.2.1.3 Nanašalni indikatorji	168
10.2.1.4 Indikatorji za vzpostavljanje pomenskih mrež – neposredne nadpomenke	169
10.2.2 Oblikovanje pomenskih indikatorjev	169
10.2.2.1 Razumljivost	169
10.2.2.2 Kratkost	171
10.2.2.3 Razločevalnost	174
10.3 Pomenska shema	177
10.3.1 Oblikovanje in zapis pomenske sheme	178
10.3.1.1 Udeleženci in okoliščine pomena	181
10.3.1.1.1 Glagol	182
10.3.1.1.1.1 (Ne)izraženost udeležencev	184
10.3.1.1.1.1.1 Pripomočki in orodja	187
10.3.1.1.1.1.2 Stavčni udeleženci in okoliščine	189
10.3.1.1.1.1.3 Pomensko neobvezni udeleženci	192
10.3.1.1.1.1.4 Inherentni udeleženci	193
10.3.1.1.1.2 Spremembe v udeleženski zgradbi pomena (uporaba skladenjskega ali)	194
10.3.1.1.1.3 Strukture s prostim glagolskim morfemom <i>se</i> oz. <i>si</i>	195
10.3.1.1.1.3.1 Povratni glagoli	195
10.3.1.1.1.3.2 Prava povratnost	196
10.3.1.1.1.3.3 Neprava povratnost	197
10.3.1.1.1.3.5 Splošni vršilec in trpna raba	198

10.3.1.1.1.4	Dodajanje oz. odvzemanje udeležencev (stisniti pest – stisniti roko v pest)	199
10.3.1.1.1.5	Možnost različnih predložnih realizacij	202
10.3.1.1.2	Samostalnik	203
10.3.1.1.2.1	Besednozvezna vezljivost samostalniških pomenov	204
10.3.1.1.2.1.1	(Iz)glagolski samostalniki	204
10.3.1.1.2.2	Neglagolski in količinski samostalniki	205
10.3.1.1.2.2.1	Razmerje med zapisom pomenske sheme in razlage	206
10.3.1.1.2.3	Stavčna vezljivost samostalniških pomenov	207
10.3.1.1.3	Pridevnik	208
10.3.1.1.3.1	Udeleženci pridevniške vezljivosti	209
10.3.1.1.3.2	Kolokabilnost pridevniških pomenov	210
10.3.1.1.4	Prislov	211
10.3.1.2	Semantični tipi.	212
10.3.1.2.1	Povezanost semantičnih tipov z elementi LBS za namene strojnega procesiranja	215
10.3.1.2.2	Vrste semantičnih tipov	218
10.3.1.2.3	Zapis semantičnih tipov	230
10.3.1.2.3.1	Večbesedni semantični tipi	231
10.3.1.2.3.2	Ujemanje	232
10.3.1.2.3.2.1	Stavčno ujemanje	234
10.3.1.2.3.2.2	Anaforično ujemanje	237
10.3.1.3	Pomenski scenarij.	239
10.4	Slovarske razlage	239
10.4.1	Klasične oz. analitične slovarske razlage	241
10.4.2	Stavčne razlage	243
10.4.2.1	Razlagalne strategije	247
11	Večbesedne leksikalne enote	252
11.1	Stalne zveze	253
11.1.1	Merila za določanje stalnih zvez	253
11.1.1.1.1	Pomenske lastnosti	253
11.1.1.1	Oblikovne lastnosti	253
11.1.1.2	Glede na skladenjsko vlogo	254
11.1.2	Merila za uvrščanje stalnih zvez v LBS	254
11.1.3	Opis stalnih zvez v LBS	257

11.1.3.1	Glede na pomen	258
11.1.3.2	Glede na zgradbo	261
11.1.3.3	Glede na skladijsko vlogo	262
11.1.4	Stalne zveze v odnosu do drugih večbesednih enot	262
11.1.4.1	Stalna zveza – samostojni pomen	263
11.1.4.2	Stalna zveza – frazeološka enota	264
11.1.4.3	Stalna zveza – kolokacija	264
11.2	Frazeološke enote	265
11.2.1	Merila za določanje frazeoloških enot	266
11.2.1.1	Glede na zgradbo	267
11.2.1.2	Glede na skladijsko vlogo	268
11.2.2	Opis frazeoloških enot	270
11.2.2.1	Osnovna oblika, variante in oblikoskladijske pretvorbe	270
11.2.2.2	Tipično besedilno okolje frazeološke enote	271
11.2.2.3	Pomenski opis	272
11.2.3	Merila za uvrščanje frazeoloških enot v LBS.	276
11.2.3.1	Glede na zgradbo	276
11.2.3.2	Glede na pomen	278
12	Oznake	281
12.1	Oznake v slovarski bazi in slovarju	283
12.2	Avtomatsko pridobivanje oznak in opozoril o rabi iz korpusa	285
12.3	Moč oznak	287
12.4	Vrste oznak v LBS	288
12.4.1	Področne oznake	288
12.4.2	Diskurzne ali kontekstualne oznake	291
12.4.3	Slovnične oznake	293
12.4.3.1	Skladijske oznake.	293
12.4.3.2	Oblikoskladijske oznake	297
12.4.4	Stilne oznake	301
12.4.4.1	Konotacijske oznake	301
12.4.4.2	Registrske oznake.	303
12.4.4.3	Pragmatične oznake	305

III. SKLADENJSKI NIVO	307
13 Skladenjske strukture	310
13.1 Pridobivanje skladenjskih struktur iz korpusa	310
13.2 Obravnava skladenjskih struktur po besednih vrstah	313
13.2.1 Posebne strukture	314
13.2.1.1 Strukture s pomožnim glagolom	314
13.2.1.2 Strukture s pomožnim glagolom in z nedoločnikom	315
13.2.1.3 Strukture z nedoločnikom	316
13.2.1.4 Strukture s količinskimi izrazi	317
13.2.1.5 Strukture z nikalnimi elementi	318
13.2.2 Regularne strukture	321
13.2.2.1 Samostalniške strukture	321
13.2.2.2 Glagolske strukture	325
13.2.2.3 Pridevniške strukture	326
13.2.2.4 Prislovne strukture	328
14 Skladenjske zveze	330
14.1 Prepoznavanje skladenjskih zvez na podlagi korpusa	331
14.2 Merila za določanje skladenjskih zvez	333
14.3 Vrste skladenjskih zvez	333
14.4 Opis skladenjskih zvez v LBS	337
14.5 Skladenjske zveze v razmerju do drugih večbesednih enot v LBS	338
15 Stavčni vzorci	340
15.1 Vrste stavčnih vzorcev v LBS	345
15.2 Prepoznavanje stavčnih vzorcev na podlagi korpusa	347
15.2.1 Analiza konkordančnega niza	347
15.2.2 Avtomatsko pridobivanje stavčnih vzorcev iz korpusa	348
17 Restrikcije	351
17.1 Vrste restrikcij v LBS	352
IV. KOLOKACIJSKI NIVO	354
18 Kaj so kolokacije	356
18.1 Kolokacija kot statistični pojav	357
18.2 Kolokacija kot pomensko-skladenjski pojav	357
19 Pridobivanje kolokacij iz korpusa	359
20 Kolokacije v LBS	360

20.1	Vrste kolokacij v LBS	361
20.1.1	Regularne kolokacije	361
20.1.2	Razširjene kolokacije	364
20.1.2.1	Razširjene kolokacije z vidika zgradbe	364
20.1.2.2	Razširjene kolokacije s pomenskega vidika.	365
20.1.2.3	Razširjene kolokacije v razmerju do stalnih zvez	366
20.1.2.4	Razširjene kolokacije v razmerju do skladenjskih zvez.	367
20.1.3	Pridobivanje razširjenih kolokacij iz korpusa	367
20.2	Zapisovanje kolokacij v LBS	369
V.	PONAZARJALNI NIVO	372
21	Zgledi rabe v slovarski bazi in slovarju	374
22	Vloga slovarskih zgledov.	375
22.1	Avtentični ali ustvarjeni slovarski zgledi.	376
23	Merila za določanje dobrih slovarskih zgledov.	379
24	Avtomatski postopki luščenja korpusnih zgledov.	384
	Sklep	387
	Povzetek.	389
	Summary	400
	Literatura	412
	Imensko kazalo	429

Predstavitev

V monografiji je opisana zasnova, zgradba in vsebina Leksikalne baze za slovenščino (LBS), tj. slovarske oz. leksikalne podatkovne zbirke (Arhar Holdt 2011: 19/20; Arhar in Holozan 2009: 25/26; Gorjanc et al. 2005: 139), ki temelji na dobrih praksah leksikalnih baz in sorodnih projektov evropskih jezikov, obenem pa upošteva specifično stanje slovenščine tako glede jezikovnih lastnosti kot glede obstoječih slovarskih priročnikov in trenutno razpoložljivih računalniško procesljivih leksikalnih virov. Kot taka je LBS izhodišče za izčrpen na korpusu temelječ leksikalni opis sodobne pisne slovenščine s poudarkom na tipični rabi besed v skladenjskem, pomenskem in besedilnem kontekstu. Besedilo je zasnovano kot vodič po posameznih nivojih v bazi strukturiranih podatkov, hkrati pa prinaša podroben opis leksikografskih postopkov, ki so bili uporabljeni pri njeni izdelavi in ki jih narakujejo aktualni leksikografski trendi. Poseben poudarek je namenjen postopkom, ki odkrivajo nove, glede na slovensko leksikalno ter slovaropisno teorijo in prakso še neuveljavljene pristope.

Monografija je primarno namenjena leksikologom in leksikografom, ki jih zanima analiza korpusnih podatkov, njihova jezikoslovna interpretacija in organizacija z vidika slovarske zasnove v digitalnem okolju. Glede na elektronsko naravo leksikalne baze bo za jezikoslovne tehnologe, ki jih zanima procesiranje jezikoslovno interpretiranih podatkov, koristen opis skladenjskih struktur, pomenskih shem in predvidene možnosti pri avtomatizaciji leksikografskih postopkov. Monografija nagovarja tudi jezikoslovce, ki jih zanima ozadje korpusnih analiz in proces nastajanja slovarske informacije na poti od korpusa prek leksikografa do končnega uporabnika. Obsežen del podatkov, ki jih baza vsebuje, pa omogoča tudi nadaljnje jezikoslovne analize.

Vsebina temelji na Specifikacijah za izdelavo LBS (Gantar et al. 2009; Gantar et al. 2009a) in na Navodilih avtorjem (v interni uporabi), ki so nastajala v času zasnove in v procesu izdelovanja posameznih gesel. Skozi delovni proces so se postopoma izrisovali relevantni podatki, njihova medsebojna povezanost in strategije opisa z vidika slovarja, jezikovnotehnoških potreb, pa tudi z vidika preverjanja leksikografskih rešitev, ki za slovenščino še niso bile uporabljene. Končne rešitve v LBS, ki je rezultat projekta, se zato ponekod ne ujemajo s tu prikazanimi, saj je bilo, da smo lahko našli pravo, marsikdaj potrebno najprej preizkusiti več različnih rešitev. V monografiji so opisane predvsem te.

Ljubljana, maj 2015
Avtorica

Zahvala

Pri pregledu evropskih jezikovnotehnoloških in slovarskih projektov so od junija do decembra 2008 sodelovali: Špela Arhar Holdt, Gaja Červ, Nuša Drinovec Sever, Tadeja Dušej, Polona Gantar, Katja Grabnar, Nanika Holz, Mateja Jemec Tomazin, Polonca Kocjančič, Simon Krek, Nataša Logar, Nina Modrijan, Sara Može, Olga Yerošina, Primož Ponikvar, Anja Strajnar, Mojca Stritar, Mojca Šorli, Kristina Šter, Simon Šuster, Blaž Trebar in Petra Zaranšek. V času od januarja do junija 2009 je ožja skupina sodelavcev: Polona Gantar, Katja Grabnar, Polonca Kocjančič, Simon Krek, Olga Yerošina, Mojca Šorli, Simon Šuster in Petra Zaranšek izdelala zasnovo slovarske baze za slovenščino in predstavila izhodišča za analizo korpusnih podatkov ter standarde za izdelavo posamezne iztočnice v bazi. V izdelavo gesel, ki je potekala od septembra 2009 do junija 2012, so največ dela, potrpežljivosti in znanja vložile Nina Drstvenšek, Katja Grabnar, Polonca Kocjančič, Mojca Šorli, Petra Zaranšek in Olga Yerošina. Večina sodelavcev, med katerimi bi rada izpostavila predvsem Kristino Bizjak, Živo Blaževič, Petro Guna, Urško Kamenšek, Andrejo Košir, Alenko Laharnar, Barbaro Modrijan, Tanjo Novak, Andraža Polončič Ruparčič, Evo Pori, Tanjo Radovič, Urško Vranjek in Petro Žalodec, pa se je z leksikografskim delom in metodologijo korpusnih analiz seznanjala šele pri izdelavi slovarske baze ob sicer rednih študijskih obveznostih. Vsem skupaj in vsakemu posebej se na tem mestu iskreno zahvaljujem za požrtvovalnost, potrpežljivost in predvsem za znanje, ki sem ga črpala iz vas. Verjamem, da lahko vse leksikografske tegobe, ki ste jih prostovoljno vzeli nase, odtehta že večkrat izkazana uporabnost vloženega truda in s tem dejstvo, da delo ni bilo opravljeno zaman.

Za tehnično podporo, izdelavo internih programov, pridobivanje podatkov iz korpusa in izpolnjevanje najrazličnejših želja so skrbeli Polonca Kocjančič, Iztok Kosem in Rok Rejc, za organizacijo in administrativno delo pa Karmen Kosem. Hvala vsem, bili ste nepogrešljivi!

Simon Krek, koordinator projekta Sporazumevanje v slovenskem jeziku, znotraj katerega je baza nastajala, in urednik zbirke Sporazumevanje, je dal pobudo za pričujoče delo in vztrajal, da je spoznanja vredno predstaviti strokovni javnosti. Sodelavcem Znanstvene založbe Filozofske fakultete UL, zlasti dr. Matevžu Rudolfu in Juretu Preglau, se zahvaljujem za sprejem monografije v uredniški program: celoten postopek sta izpeljala nadvse profesionalno, predvsem pa neproblematično, kar mi je bilo v času zadnjih popravkov besedila v veliko olajšanje.

Iskreno se zahvaljujem tudi recenzentoma doc. dr. Nataši Logar in prof. dr. Vojku Gorjancu za poglobljeno kritično branje. Njune konstruktivne pripombe so me s

številnih stranpoti vedno znova vračale v jedro problematike in v jezikovne specifičnosti slovenščine, zaradi česar je predstavljena podoba leksikografskega opisa aktualnejša in bolj zanesljiva.

Kolegom in prijateljem Damjanu Popiču, Darinki Verdonik, Darji Fišer, Heleni Dobrovoljc, Kaji Dobrovoljc, Marku Stabeju, Moniki Kalin Golob, Nataši Logar, Simonu Kreku, Špeli Arhar Holdt, Tomažu Erjavcu, Vojku Gorjancu, sodelavcem Konzorcija za jezikovne vire in tehnologije v okviru istoimenskega Centra UL in vsem, ki jih na tem mestu ne navajam, so pa z nasveti, podporo in pogledom »od zunaj« v marsičem prispevali k obravnavani temi, se iskreno zahvaljujem za izkazan entuziazem in konstruktivne pogovore. Bilo mi je v veselje in prepričana sem, da bomo to, kar smo delali, sanjali in živeli, tudi uresničili.

Uvod

Stanje v leksikografiji, kot tudi njen nadaljnji razvoj in pogled na slovarsko informacijo, v veliki meri določajo spremembe, ki so se v zadnjih 25 letih zgodile na področju digitalnih medijev. V ospredju je uporabnik, ki je več uporabe digitalnih medijev in zna s pridom izrabljati njihove prednosti, s tem pa posredno odloča tudi o vrsti, obliki in načinu predstavitve slovarskih in drugih jezikovnih podatkov. Knjižne oblike so z vidika priročnikov postale manj zanimive, na novo pa so se postavila tudi vprašanja, kaj vse lahko dejansko ponudi novodobni slovar in v kakšni obliki je mogoče najbolj učinkovito predstaviti za jezikovnega uporabnika zanimive slovarske vsebine (Rundell 2009: 9). Uporabniki za iskanje informacij uporabljajo predvsem splet, zato je v porastu ponudba spletnih slovarjev, ki skušajo optimizirati ponujeno informacijo in jo v čim večji meri narediti uporabniku razumljivo in prijazno. Med pričakovanji uporabnika so na vsebinski ravni najvišje ovrednotena zanesljivost vsebine, jasnost informacije, možnost posodabljanja, hitrost in dostopnost, prilagodljivost informacij, možnosti večpredstavnih vsebin in povezave z jezikovnimi viri, drugimi slovarji, podatkovnimi bazami in enciklopedičnimi podatki (Müller-Spitzer et al. 2011). K temu se dodajajo še alternativne možnosti dostopanja do podatkov, izraba avdio in video aplikacij ter specializirane iskalne možnosti (Prinsloo et al. 2011).

Slovenščina je, upoštevajoč obstoječe jezikovne priročnike ter razvitost in dostopnost jezikovnih virov za slovenščino, v nezavidljivem položaju (prim. Krek 2012: 33/34). Če jo vključimo v evropski jezikovni prostor ter primerjamo z jeziki, ki imajo približno enako število maternih govorcev, lahko ugotovimo, da je nezadostno opremljena predvsem s prosto dostopnimi jezikovnimi viri, da obstaja izrazita vrzel pri enojezičnih in dvojezičnih leksikalnih, skladenjskih in semantičnih virih (Krek 2009: 98/99), da ne obstajajo sistematične študije o sodobni slovenščini in da je posledično zlasti velika diskrepanca med realno slovenščino in slovenščino, ki se ohranja pri življenju znotraj jezikovnih priročnikov in šolskih učbenikov (Arhar Holdt 2012). Slovar slovenskega knjižnega jezika (SSKJ) kot trenutno najbogatejši vir leksikalnogramatičnih podatkov o slovenskem jeziku je z vidika sodobnega slovenskega jezika zastarel in pomanjkljiv. Ker informacije v njem niso izražene eksplicitno, jih je za računalniško procesiranje težko uporabljati, njegova največja pomanjkljivost pa je v tem, da kot podatkovna zbirka ni dostopen za raziskovalne namene (Erjavec 2014). V času nastajanja te monografije je izšla druga, dopolnjena in deloma prenovljena izdaja SSKJ (SSKJ2), ki vključuje večino iztočnic in pomenov Slovarja novejšega besedja slovenskega jezika (SNB), na vsebinski ravni pa se osredotoča na zamenjavo nekaterih izrazov v slovarskih razlagah in zgledih ter na

spremembe pri kvalifikaciji besedišča.¹ Gre zlasti za izraze, ki so z današnjega vidika bodisi zastareli bodisi družbenopolitično nekorektni. Ker ostaja izhodiščni opis v SSKJ2 v smislu gradivne in metodološke realizacije koncepta, kot se je izoblikoval ob nastanku SSKJ v 60. letih, nespremenjen, ne moremo govoriti o sodobnem slovarju slovenskega jezika niti o aktualnem opisu slovenske leksike. Še več, z novo izdajo SSKJ ni presegel niti očitkov o zastarelosti niti o prosti dostopnosti tako slovarja kot slovarske baze, pač pa je zabrisal legitimen odsev slovenske leksike in z njo pojmovnega sveta, kot ga je odražalo gradivo, na katerem je temeljil.

Izdela LBS je bila v okviru projekta Sporazumevanje v slovenskem jeziku (SSJ; 2008–2013) primarno namenjena prav zapolnitvi vrzeli na področju sodobnega leksikalnogramatičnega opisa. Tak opis je mogoče izrabiti za namene različnih spletnih slovarskih priročnikov in za računalniško obdelavo slovenščine, kamor sodi cela vrsta podpodročij, kjer se srečujeta računalniško procesiranje in naravni jezik. Glede zadnjega smo imeli pri snovanju leksikalne baze v mislih predvsem izboljšanje avtomatskega označevanja slovenskih besedil na oblikoskladenjski, skladenjski in pomenski ravni, uporabo leksikalne baze v procesih pomenskega razdvoumljanja ter razvoj različnih jezikovnotehnoloških aplikacij za slovenščino. Zaradi težnje po čim boljši opremljenosti z jezikovnimi viri, ki vsebujejo procesljive jezikovne podatke, se vse bolj pomembno zdi tudi združevanje in dopolnjevanje z drugimi obstoječimi jezikovnimi viri, kot je denimo slovenski WordNet (SloWNet; Fišer 2009) in posamezni deli označenih besedil po vzoru projekta FrameNet (prim. Može 2013).

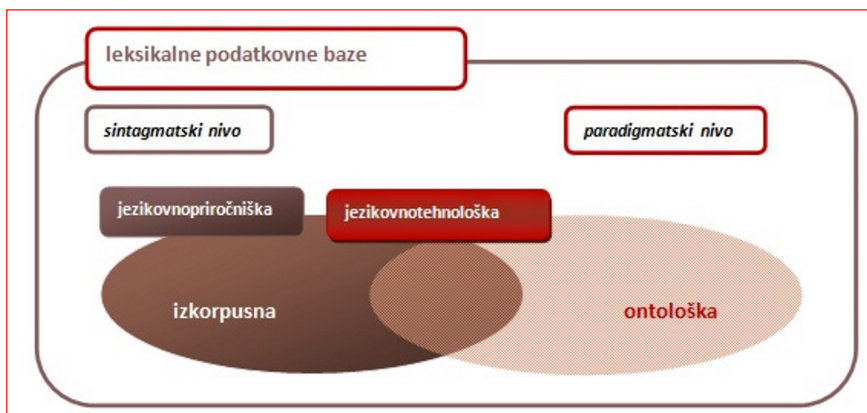
LBS, kot je opisana v nadaljevanju in kot jo določajo zgoraj omenjena izhodišča, zato ne more biti v določenem trenutku zaključen projekt, pač pa je zasnovana kot vir leksikalnih podatkov, ki se dopolnjuje in posodablja glede na jezikovni razvoj in potrebe jezikovne skupnosti. Kot taka zahteva na eni strani kontinuirano finančno podporo, na drugi pa usmeritev prizadevanj leksikografske stroke za avtomatizacijo postopkov, ki omogočajo nenehno posodabljanje ob čim manjšem finančnem in človeškem vložku.

1 Podroben pregled sprememb, ki jih prinaša SSKJ2 glede na SSKJ, je mogoče najti v Krek 2014.

A. LEKSIKALNA PODATKOVNA BAZA

V prvem delu opredelimo pojem leksikalne podatkovne baze, predstavimo različne vrste leksikalnih baz glede na namen njihove izrabe in na kratko opišemo sorodne jezikovnotehnološke in jezikovnopriročniške projekte v posameznih evropskih jezikih. Poglavje zaključimo z opredelitvijo razmerja med leksikalno bazo in slovarjem kot njenim najverjetnejšim produktom.

Leksikalna podatkovna baza² (ang. *lexical database*), kot jo razumemo v pričujočem delu, je organiziran opis leksike določenega jezika v elektronski obliki³ ali na kratko: elektronska zbirka podatkov o leksikalnih enotah. Vprašanje, katere informacije naj bi leksikalna baza vsebovala, je v prvi vrsti povezano s tem, za katere namene je oblikovana (Ooi 1998: 11). Glede na vrsto podatkov, ki jih podatkovna baza zajema, način organiziranosti, namen gradnje itd. je na splošno mogoče ločevati med t. i. **jezikovnotehnološkimi leksikalnimi bazami**, ki so namenjene obdelavi naravnega jezika in razvoju jezikovnih tehnologij, in **jezikovnopriročniškimi ali slovarskimi leksikalnimi bazami** (Arhar in Holozan 2009; Arhar Holdt 2011: 19), ki so vir jezikovnih opisov, namenjenih človeškemu uporabniku (gl. Sliko 1).



Slika 1: *Vrste leksikalnih podatkovnih baz*

2 Izraz podatkovna baza (in ne zbirka) uporabljamo iz dveh razlogov: prvi je dejstvo, da je bila izdelava repozitorija leksikalnogramatičnih podatkov kot izhodišča za različne tipe leksikografskih opisov slovenščine sprejeta v okviru projekta SSJ pod naslovom: »Izdelava leksikalne baze za slovenščino«, poimenovanje pa se je nato ustalilo tudi v strokovnih prispevkih. Drugi razlog je vsebinski. Podatki, ki jih vsebuje LBS, so z vidika obsega, izbora, ureditve in načina predstavitve dejansko osnova (baza) za izgradnjo različnih slovarskih priročnikov. Pri posameznih projektih, ki jih opisujemo v nadaljevanju, ohranjamo izvorno poimenovanje.

3 Glossary Of Linguistic Terms:
<http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsALexicalDatabase.htm>.

Skupne lastnosti obeh temeljnih tipov leksikalnih podatkovnih baz so deloma prekriven nabor podatkov o jeziku, elektronska oblika in »dejstvo, da sami po sebi nista končni produkt, ampak vmesna stopnja na poti do določene jezikovne tehnologije oz. jezikovnoopisnega vira za človeškega uporabnika« (Arhar Holdt 2011: 19). Kljub ločevanju na jezikovnotehnološke in jezikovnopriročniške leksikalne baze pa se v praksi meja med njima briše. Prav možnosti, ki jih z novimi mediji ponuja splet, zastavljajo vprašanja, kje se pravzaprav informacija slovarskega tipa konča in preide v jezikovnotehnološki vir, še posebej, ker vsaka jezikovnopriročniška baza, namenjena primarno izdelavi (spletnih) slovarjev, omogoča tudi dostop do surovih podatkov, uporabnih za nadaljnje procesiranje.

1 **Leksikalne baze in sorodni projekti evropskih jezikov**

V času snovanja LBS smo po posameznih evropskih jezikih pregledali vrsto korpusnih in jezikovnotehnoloških projektov in leksikalnih virov ter se seznanjali z metodologijo izdelave, teoretičnimi in praktičnimi pristopi, načini morfosintaktičnega označevanja elektronskih virov, uporabo jezikovnotehnoloških orodij, načini jezikovnega opisa, vizualizacijo podatkov itd. Čeprav seznam projektov nikakor ni bil izčrpen že v času snovanja LBS in sorodni leksikalni porojekti nastajajo vedno znova, se dopolnjujejo in posodabljaajo, predstavljamo posamezne projekte v nadaljevanju nekoliko podrobneje, ker vsebujejo nekatere ključne lastnosti in rešitve, po katerih smo se zgledovali tudi v okviru slovenskega projekta.⁴

1.1 **Jezikovnotehnološke leksikalne podatkovne baze**

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 zvezi s tem pa so se začele z evropskimi jezikovnotehnološkimi projekti v 90. 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, uporabljivih za zagotavljanje jezikovnih potreb več jezikovnih skupnosti znotraj EU. V ta namen je bilo potrebno izdelati digitalne leksikalne vire na podlagi enotnih standardov, orodja za njihovo gradnjo in obdelavo ter modele hranjenja in distribucije. Razvoj jezikovnotehnoloških podatkovnih baz se je nato razvijal v smeri sistematičnega združevanja izkušenj z razvojem virov in orodij za potrebe računalniške obdelave naravnih jezikov ter oblikovanja standardov in modelov za izdelavo elektronskih slovarjev evropskih jezikov. V tem kontekstu je treba omeniti zlasti projekte v okviru GENELEX I in GENELEX II (GENEriC LEXicon, 1990–1994), katerih temeljni cilj je bil ustvariti celostno infrastrukturo jezikovnih virov, ki bo upoštevala enotne standarde pri označevanju korpusov in izgradnji leksikonov za več evropskih jezikov. Od tod je bil način strukturiranja informacij po posameznih ravnanjih, oblikoskladenjski, skladdenjski in semantični, prenesen v druge evropske projekte, med katerimi naj izpostavimo zlasti PAROLE (Preparatory Action for Linguistic Resources Organization for Language Engineering, 1996–1998), katerega rezultat so usklajene (izdelane z isto metodologijo in opisane v enotni shemi DTD) morfološko in sintaktično označene elektronske leksikalne zbirke za

⁴ Opisi podatkovnih baz so povzeti na podlagi poročil in predstavitev, izdelanih v okviru aktivnosti Leksikalna baza za slovenščino pri projektu SSJ (interno gradivo).

13 jezikov takratne EU, in SIMPLE (Semantic Information for Multifunctional Plurilingual Lexicons, 1998–2002) kot nadgradnja projekta PAROLE s semantičnimi informacijami, pri čemer vse podatkovne zbirke temeljijo na skupnem jedru osnovnih konceptov, povzetih po EuroWordNetu. Tretji projekt v tem okviru je projekt DELIS (Descriptive Lexical Specifications and Tools for Corpus-based Lexicon Building, 1993–1995), v okviru katerega je bil zgrajen korpus za pridobivanje leksikalnih podatkov in razvit model za semantično označevanje. V ta namen je bil prirejen in dopolnjen model, ki temelji na teoriji generativnega leksikona J. Pustejovskega (1995).

Jezikovnotehnoške leksikalne podatkovne baze in sorodni projekti po posameznih evropskih jezikih, ki jih nekaj na kratko opišemo v nadaljevanju, so v večji ali manjši meri nasledki ali nadgradnje zgoraj omenjenih projektov.

V času snovanja slovenske baze smo podrobneje analizirali zlasti projekta FrameNet, ki ga je zasnoval Charles J. Fillmore, in Corpus Patterns Analysis (CPA) Patricka Hanksa, ki zaradi združitve skladišnih in pomenskih informacij najbolj ustrežata predvidenemu na korpusu temelječemu opisu sodobne slovenske leksike. Opis, ki sledi v nadaljevanju, vključuje tudi primerjavo glede na rešitve, ki smo jih upoštevali v LBS.

1.1.1 FrameNet (FN)

Projekt FN (2003), ki je zasnovan kot ontološka leksikalna baza, je nastajal pod vodstvom Charlesa J. Fillmora na Mednarodnem inštitutu za računalništvo na Kalifornijski univerzi v Berkeleyju. Njegov namen je za vsak pomen besede ali besedne zveze analizirati in beležiti celotni spekter pomenskih in skladišnih razmerij. Vsaka beseda, ki je vključena v FN, pripada določenemu pomenskemu okviru ali shemi (ang. *frame*). Če je beseda večpomenska, pripada več različnim pomenskim shemam. Da bi bila taka analiza mogoča, so avtorji izoblikovali niz oznak, ki poimenujejo pomenske vloge (t. i. shemske elemente, ang. *frame elements*) in omogočajo podrobno dokumentacijo korpusnega sobesedila, v katerem se nahaja obravnavana beseda oz. njen pomen. Pomensko povezane besede, npr. glagoli premikanja, govorjenja, zaznavanja itd., se tako pojavljajo znotraj istih pomenskih shem. Vsaka pomenska shema je vnaprej določena, ima svoje ime in definicijo, v kateri so izraženi jedrni shemski elementi (ang. *core elements*). Poleg jedrnih lahko shema vključuje tudi nejedrne elemente. Znotraj pomenske sheme 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 shemama: Cause_to_move_in_place (povzročiti premik v prostor(u)) in Manipulation (upravljanje). Poleg glagola *squeeze* se znotraj prve sheme nahajajo tudi glagoli kot: *rotate*, *seesaw*, *shake*, *spin*, *swing*, *swirl*, *turn* ipd., v drugi shemi pa glagoli kot: *caress*, *clasp*, *claw*, *clutch* ipd. Slika 2 prikazuje opis glagola *squeeze* v pomenski shemi Manipulation, kamor sodijo vse besede oz. njihovi pomeni, ki opisujejo upravljanje z neko entiteto, ki ga izvede aktant, navadno z roko ali glede na funkcijo s podobnim delom telesa. Iz tega izhaja, da so entiteta, aktant in del njegovega telesa jedrni elementi pomenske sheme Manipulation.

Manipulation

[Lexical Unit Index](#)

Definition:

The words in this frame describe the manipulation of an **Entity** by an **Agent**:

Jan **SQUEEZED** **Maria's hand**.

A number of lexemes listed in this frame also occur in the frame Cause-to-move (for example, push, yank). Only the non-motion uses of these words belong in this frame, e.g., combinations such as push at, push on, yank at.

FES:

Core:

Agent [Agt] Semantic Type: Sentient	The Agent manipulates an object. Agent typically refers to animate beings, but is also extended to mechanical and other inanimate objects. Agent is most frequently expressed as the external argument of verbs: A robot arm GRASPS the box on the conveyor belt and seals it.
Bodypart of agent [AgtBP] Semantic Type: Body_part	The part of the Agent's body being used to manipulate the Entity may also be expressed. While the default is for an Agent to use his/her hands, other body parts may be specified. This frame element usually occurs as a PP-complement. Kim GRASPED the needle between finger and thumb.
Entity [Enty]	The Entity being manipulated often occurs as the direct object of verbs: SQUEEZE the ball as many times as you can. Pat SQUEEZED Jo's hand.

Slika 2: Opis jedrnih elementov pomenske sheme Manipulation za glagol *squeeze* (stisniti) v *FrameNetu*

Delo pri projektu, v katerem poteka označevanje vzorčnih stavkov polavtomatsko, je v celoti računalniško podprto. Nastajajoča podatkovna baza omogoča vrsto različnih avtomatskih prikazov in iskanj podatkov, ki so namenjeni podrobnim jezikoslovnim analizam. Podatkovna baza je prosto dostopna, njen obseg pa se nenehno širi: trenutno jo uporablja na stotine raziskovalcev, učiteljev in študentov jezikoslovja po vsem svetu. Podobne analize, ki so tesno povezane z izhodiščnim projektom v Berkeleyju, nastajajo tudi za nemščino, španščino, japonščino in slovenščino (prim. Lönneker-Rodman et al. 2008; Može 2009, 2013). Kot pravita Atkins in Rundell (2008: 145),

je framenetovski pristop najbolj primeren za raziskovanje obnašanja besed na podlagi korpusnih podatkov, ker omogoča dosledno analizo besedilnega okolja, hkrati pa ne spregleda nobenega ključnega pomenskega dejstva.

Ker je eden od poudarkov FN pomagati leksikografom identificirati pomembna pomensko-skladenjska dejstva o leksikalnih enotah na podlagi korpusa, smo se pri oblikovanju informacij v LBS v obliki t. i. pomenskih shem zgledovali po FN, hkrati pa smo ohranili nekatere pomembne razlike. Tako kot pri projektu FN temelji tudi v LBS opis pomenskega scenarija v pomenski shemi konkretnega pomena na analizi velikega števila primerov iz korpusa. Za razliko od FN pa naš namen ni opisati pomena določene besede na podlagi obstoječega ali vnaprej izdelanega seznama pomenskih shem in pomenskih elementov znotraj njih. Za razliko od FN, ki za vsak shemski element predvideva ime, opis in semantično kategorijo, poimenovanje udeležencev in okoliščin v LBS ni fiksno in vnaprej določeno, pač pa so poimenovanja udeležencev v obliki semantičnih tipov podrejena opisu konkretnega pomena besede. Kljub temu težimo tudi v LBS k navajanju tistih elementov, ki omogočajo razločevanje med posameznimi pomeni oz. podpomeni istega glagola, medtem ko je posplošitev in poenotenje v smislu framenetovskih shemskih elementov mogoče izpeljati postopoma z rastjo števila izdelanih pomenskih shem. Za razliko od FN, kjer formalizirane pomenske sheme in elementi znotraj njih omogočajo medsebojno povezovanje in vzpostavljanje pomenskih razmerij med leksikalnimi enotami, primarni namen LBS ni združevati pomene na podlagi pomenskih shem, čeprav način organizacije podatkov to omogoča. Konkretno pri FN to pomeni, da v primeru, če seznam ne vsebuje pomenske sheme, ki bi ustrezala opisu določene nove besede, te besede ni mogoče vključiti v pomensko-skladenjski opis, dokler se ne pokaže zadostno število besed, ki bi ustrezale novi pomenski shemi. V nasprotju s takim sistemom je pomene v LBS mogoče obravnavati neodvisno od njihove pomenske (ne)povezanosti, kar omogoča vključitev besedišča neodvisno od pomenskega povezovanja med leksikalnimi enotami.

1.1.2 Analiza korpusnih vzorcev (CPA)

Drugi projekt, ki izhaja iz korpusne analize ter prepoznavanja stavčnih vzorcev v povezavi s pomeni obravnavane besede in smo ga vzeli kot izhodišče pri oblikovanju pomenskih shem v LBS, je projekt analize korpusnih vzorcev (ang. *Corpus Pattern Analysis*; Hanks 2004, 2008; Hanks in Pustejovsky 2005). Projekt temelji na projiciranju pomena iz sobesedila na posamezno besedo in izhaja iz teorije jezikovnih konvencij ter možnosti njihove izrabe (ang. *Theory of Norms and Exploitations*, Hanks 1994, 2013; Hanks in Pustojevsky 2004, 2005). Analiza korpusih vzorcev

se v veliki meri spogleduje s projektom FN, vendar pa se za razliko od FN, kjer so v ospredju možnosti združevanja posameznih glagolov v sheme na podlagi njihovega podobnega skladenjskega in pomenskega obnašanja v sobesedilu, CPA osredotoča na sistematično analizo tipičnih pomenskih vzorcev posameznega glagola, manj pa je za ta model zanimiva možnost njihovega medsebojnega povezovanja. V postopku analize so posamezne besede (poleg glagolov so na ta način obravnavani tudi samostalniki)⁵ 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. Identifikacija semantičnega vzorca nikakor ni preprost in avtomatiziran postopek, pač pa zahteva veliko leksikografske spretnosti in prakse: med najtežjimi leksikografskimi odločitvami je prav gotovo določitev ustrezne stopnje posplošitve, znotraj katere je mogoče razločevati posamezne pomene.

Za ponazoritev (Gantar 2008: 23) smo za glagol *vložiti* vzeli primer iz korpusa Gigafida (Logar et al. 2012):⁶ *Pivovarna Laško bo vložila pritožbo*. Pomen glagola *vložiti* bi v izkazanem kontekstu, tj. s predmetnim dopolnilom *pritožba*, lahko vseboval naslednje implikature: [Pivovarna Laško = ustanova] bo vložila [pritožbo = dokument], ki jih lahko izrazimo takole: »če ustanova vloži pritožbo, potem posreduje dokument, s katerim sproži pravni postopek«. Isti glagol v drugačnem kontekstu, npr. *v novo dejavnost smo vložili veliko denarja*, bi seveda zahteval drugačne implikature.

Izhajajoč iz opisanih skupnih točk in razlik med projektoma FN in CPA ter rešitev v LBS je za nas zanimiva Hanksova (2008) kritična primerjava. Kot pravi Hanks, skuša CPA za razliko od FN določiti sintagmatska merila za razlikovanje pomenov večpomenskih besed na t. i. površinski semantični ravni (ang. *semantically shallow way*), medtem ko FN izraža globinsko semantiko situacij (ang. *deep semantics*), izraženih v pomenskih shemah. FN leksikalne enote analizira glede na obstoječe sheme in ne glede na naključni seznam besed; situacije v pomenskih shemah so v FN analizirane z vidika shemskih elementov; FN preučuje pomenske razlike in povezave med različnimi besedami v shemi, ne ukvarja pa se eksplicitno s pomenskimi razlikami večpomenskih besed. Po Hanksovem mnenju FN ne analizira korpusnih podatkov sistematično, pač pa v korpusu išče primere, ki potrjujejo hipoteze, izražene v shemah. Poleg tega ima FN problem z vključevanjem novih besed v sheme, zaradi česar se marsikatera beseda oz. njen pomen izgubi.

5 Logika FN in CPA je usmerjena v predstavitev glagolov in samostalnikov, ki so znotraj stavčne zgradbe organizatorji pomena, tj. predvidevajo, napovedujejo in določajo skladenjske in oblikoskladenjske lastnosti drugih elementov stavka.

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

V nadaljevanju predstavimo tri skandinavske in italijansko leksikalno podatkovno bazo, za katere velja, da so izrazito jezikovnotehnološko usmerjene. Za vse štiri jezikovne skupnosti je značilno, da so že zelo zgodaj prepoznale nujnost računalniške obdelave jezika in se v skladu s tem ustrezno organizirale. To se danes kaže ne samo kot izrazita prednost v tehnološki opremljenosti jezika pač pa tudi v njegovi živosti in prisotnosti v evropskem prostoru (vključenost v prevajalne sisteme ter jezikovne aplikacije, ki zahtevajo prepoznavanje in sintezo govora, avtomatsko pridobivanje informacij iz besedil, izdelava elektronskih oz. spletnih slovarjev ipd.). Za skandinavske jezikovne skupnosti je značilno tudi, da izkazujejo visoko stopnjo jezikovne variantnosti, zato je za proces jezikovne standardizacije in normativizacije izredno pomembna možnost sprotnega in kontinuiranega preverjanja jezikovnega stanja ter posledično hitrega odziva nanj.

1.1.3 Leksikalna zbirka za razvoj danskih jezikovnih tehnologij (STO)

Gradnja leksikalne baze STO (Sprogteknologisk ordbase)⁷ na Centru za jezikovne tehnologije na Filozofski fakulteti univerze v Kopenhagnu sega že v leto 1996, ko je v sklopu projekta PAROLE nastal nabor morfološko ter deloma skladenjsko označenih leksikalnih enot. Semantično označevanje je bilo nato izpeljano pri projektu SIMPLE. Leta 2001 so v okviru novega projekta zbirko nadgradili, tako da so sistem PAROLE prilagodili danščini, povečali količino podatkov z vključitvijo polspecializiranega besedišča iz spletnih besedil ter dodali morfološki ter deloma skladenjski opis leksikalnih enot.

Elektronski slovar (ang. *Computational Dictionary*), kot podatkovno bazo imenujejo avtorji, vsebuje jezikovne podatke, ki so primarno namenjeni računalniškim aplikacijam, hkrati pa je bazo mogoče uporabljati tudi kot jezikovni vir za pripravo učnih gradiv pri poučevanju danščine in pri pripravi leksikografskih projektov. Jedrno besedišče podatkovne baze sestavlja splošni jezik, približno 68.000 besed, in temelji na korpusu časopisnih besedil. Specializirano besedišče, približno 13.500 besed, pa izhaja iz besedil s področja informacijskih tehnologij, okoljskih ved, zdravja ter zdravstvenih storitev, bančništva, menedžmenta ter s področja gospodarstva in marketinga. Danska leksikalna baza vključuje tudi v elektronskem formatu izdelan opis stalnih besednih zvez in frazeologije.

⁷ STO: http://cst.ku.dk/english/sto_ordbase/.

Podatki, ki jih vključuje STO, so v osnovi treh vrst:

- Morfološki del vsebuje podatke o besedni vrsti, pregibnih oblikah, izgovornih variantah, pri samostalnikih pa tudi o tipičnih tvorbenih oblikah, zlasti zloženkah.
- Skladenjski del vsebuje podatke o možnih skladenjskih zgradbah, predložnih zvezah, o pomožnih vlogah nekaterih glagolov in vezljivostnih vzorcih, ki so potrjeni s korpusnimi zgledi. Za vsako registrirano besedo, ki napoveduje vezljivostna mesta (upoštevana je glagolska, samostalniška, pridevniška in prislovna vezljivost), je določeno njihovo število in obveznost oz. neobveznost določila. Za vsako določilo je navedeno, kateri tip skladenjske funkcije predvideva, npr. samostalniško ali predložno besedno zvezo, stavek itd. v vlogi osebka, prislovnega določila itd.
- Pomenskemu opisu besed so v različnih stopnjah podrobnosti dodane informacije o medpomenskih besednih razmerjih, podatki o udeleženski zgradbi glagolskega pomena z opisi restrikcij v leksikalnih izbirah.

Podatkovno zbirko nameravajo v prihodnjih projektih dopolniti predvsem na semantični ravni (Sandford Pedersen 2012). Lotili so se združevanja glagolov glede na sorodne pomensko-skladenjske lastnosti s pripisom oznake za določen semantični tip, vzporedno pa bazo dopolnjujejo z uvozom podatkov iz danskega WordNeta DanNet.⁸

Izkušnja STO uči, da je pri enkodiranju semantičnih informacij smiselno stremeti k veliki robustnosti kategorij ter možnosti njihovega povezovanja na različnih ravneh, kar je ključno za potrebe računalniškega procesiranja. Drug poudarek je v pomembnosti vključevanja podatka o pogostosti katerega koli opisovanega pojava (ne le na ravni pojavitve leksema v korpusu, ampak tudi na ravni posameznega tipa skladenjskega vzorca).

1.1.4 Norveška leksikalna podatkovna baza (LDB)

Med večjimi skandinavskimi jezikovnotehnološkimi leksikalnimi projekti je z vidika zasnove (projekt je namreč ostal nedokončan) smiselno omeniti izdelavo leksikalne baze za moderni norveški bokmål.⁹ Za glavna cilja so si avtorji z Inštituta za jezikoslovje in nordijske študije Fakultete za humanistiko v Oslu zadali izdelavo sistematičnega in formaliziranega semantičnega opisa za 20.000 besednih enot, ki bi služil kot vir za slovarje in jezikovnotehnološke projekte, ter opis dejanskega (tj.

⁸ DanNet: <http://wordnet.dk/dannet/menu?item=0&lang=1>.

⁹ Bokmål je najpogostejše uporabljana različica norveškega jezika, ki ga govori približno 85 % norveškega prebivalstva in je navadno jezik, ki se ga učijo tuji študenti norveščine.

na korpusu temelječega) norveškega oblikoslovja (Vatvedt Fjeld 2000: 671/672). Celotna baza naj bi imela glede na namen in način predstavitve podatkov dve različici: *človeško* za uporabnike in *tehnološko* za računalniško obdelavo. Temeljni vir, poleg v ta namen zgrajenega 40-milijonskega časopisnega in govornega korpusa, naj bi bil obstoječi norveški slovar knjižnega jezika Bokmålsordboka ter danski leksikon, zgrajen v okviru projekta SIMPLE. Predvideno vsebinsko izhodišče za vsako iztočnico naj bi torej predstavljal slovar knjižne norveščine, besedišče pa bi bilo nato dopolnjeno z iztočnicami iz drugih virov, pri čemer bi bile v celoti vključene tudi funkcijske in slovnične besedne vrste. Semantični opisi iztočnic naj bi izhajali iz leksikona SIMPLE, nadgrajeni pa bi bili s sinonimi, antonimi, hiponimi in meronimi iz norveškega WordNeta.

Do konca leta 2008, ko se je projekt prekinil, je bil izdelan uravnoteženi korpus s 40 milijoni besed, avtorji pa so se ukvarjali tudi z možnostjo avtomatskega pridobivanja kolokacij iz njega. V sodelovanju z danskimi in švedskimi kolegi so sestavili delovno različico norveškega leksikona SIMPLE z 10.000 iztočnicami in formaliziranimi semantičnimi opisi. Poleg tega so izdelali norveški WordNet z 90.000 semantičnimi povezavami med vnosi, ki temeljijo na definicijah samostalnikov v slovarju Bokmålsordboka.

1.1.5 Švedska jezikovna podatkovna banka Språkbanken

Språkbanken, švedska jezikovna podatkovna zbirka, temelji na leksikalni podatkovni bazi (The Göteborg Lexical DataBase – GLDB),¹⁰ ki je bila zasnovana že v 60. letih. Namen baze je bil – in ostaja – zbrati jezikovne vire in slovarje ter razviti jezikovnotehnološka orodja in jih narediti dostopna raziskovalcem in celotni jezikovni skupnosti. Baza je bila že v izhodišču semantično zasnovana in vključuje 61.050 lem, ki so razdeljene na posamezne pomene. Lema vključuje podatke, kot so: osnovna oblika, izgovorna varianta, besedna vrsta, pregibne oblike, izgovor in naglas, morfemska razdelitev in besedna meja v zloženkah, okrajšava ter t. i. glagolski samostalniki pri glagolih. Posamezni leksemi so kot pomenske enote znotraj leme razdeljeni v dve osnovni kategoriji: na obvezni jedrni pomen in na enega ali več neobveznih pomenov. Obe kategoriji vključujeta variante, definicijo, gramatični komentar ter morfološke in skladenjske primere. Jedrni pomeni so navadno opredeljeni tudi glede na terminološko področje, podpomenski nizi pa vsebujejo podatke o tipičnem področju rabe in tip podpomena. Poleg slovarske definicije aristotelskega tipa (z genus proximum in differentia specifica) vsebuje večina leksemov tudi dodatne pomenske podatke, npr. nabor omejitev pri glagolskih in

10 GLDB: <http://www.ilc.cnr.it/EAGLES96/rep2/node19.html>.

pridevniških udeleženskih zgradbah. Jedrni pomeni in podpomenski nizi so povezani tudi glede na pomenska razmerja, kot so sopomenskost, nadpomenskost in protipomenskost, kar je v kasnejših nadgradnjah baze omogočilo povezavo z WordNetom in izdelavo ontologij in pomenskih mrež za švedski jezik, vključno s starejšimi švedskimi besedili. GLDB je bila uporabljena pri temeljnem švedskem enojezičnem slovarju *Svensk ordbok* (1986)¹¹ in pri nacionalnem enciklopedičnem slovarju *Nationalencyklopedins ordbok* (1995–1996).

Podatkovna zbirka se od leta 1975 pod imenom Språkbanken razvija kot pomemben nacionalni projekt in danes vsebuje celotno pomembnejšo švedsko (in finsko, kjer je švedščina priznana kot eden od uradnih jezikov) besedilno produkcijo, hkrati pa je ambicija sestavljavcev vključiti tudi vire, ki bodo odražali celotno švedsko jezikovno zgodovino, od srednjega veka do danes. Švedska Språkbanken vključuje dve platformi: vmesnik za iskanje po različnih korpusih Korp¹² (Borin et al. 2012) ter leksikalno infrastrukturo Karp¹³ (Borin et al. 2012a), ki ima dve osnovni funkciji: (1) nuditi podporo pri oblikovanju in izrabi različnih leksikalnih virov ter (2) redno (tj. dnevno) posodablja jezikovne vire ter omogočiti iskanje po njih in dostop v obliki, primerni za nadaljnje procesiranje. Poleg tega Språkbanken vzdržuje dvosmerno povezavo med podatkovno bazo in korpusno infrastrukturo. Oblikovanje leksikalne infrastrukture Karp je ključni del projekta SweFN++, tj. izdelave švedskega FrameNeta, katerega namen je integrirati leksikalne vire za švedščino, ki bodo prosto dostopni ter uporabni kot osnovna infrastruktura za jezikovnotehnoške raziskave in razvoj jezikovnih aplikacij za švedščino, ter evropskega projekta META-NORD, katerega namen je nadgraditi in uskladiti jezikovne vire ter orodja v celotnem evropskem jezikovnem prostoru. Infrastruktura trenutno vsebuje 15 švedskih leksikalnih virov, vključno z zgodovinskimi (npr. digitalizirani starejši švedski slovarji). Osrednji del zbirke je obsežna morfološka leksikalna baza SALDO, ki je povezana z drugimi leksikalnimi viri in korpusi, razdeljenimi na sodobno švedščino (9 virov), kamor sodijo denimo Swesaurus, švedski WordNet, SIMPLE, PAROLE, Kelly itd., švedščino 18. stoletja (2 vira) in staro švedščino (4 viri).

Za švedsko jezikovno skupnost, zlasti za empirične in kontrastivne raziskave ter za poučevanje švedščine, je leksikalna infrastruktura, kot jo zagotavlja Språkbanken, nepogrešljiva. Temeljna naravnost, ki se kaže v odprtosti, prosti dostopnosti in možnosti doprinosov jezikovnih uporabnikov k oblikovanju in izboljšanju jezikovnih virov, je zgled, ki se ga zdi vredno posnemati.

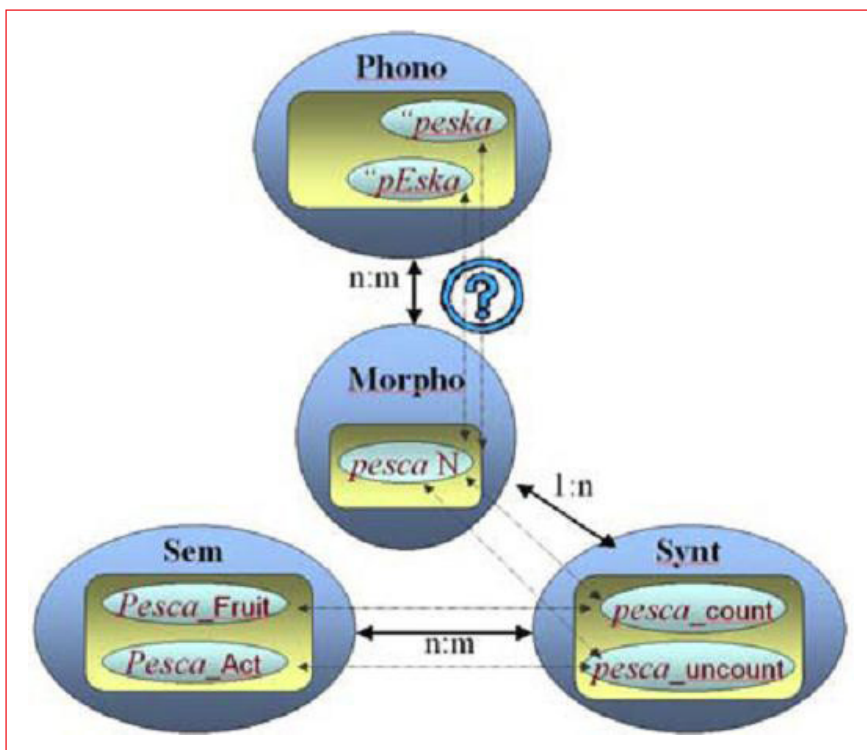
11 Elektronska različica: <http://g3.spraakdata.gu.se/saob/>.

12 Korp: korpusni vmesnik: <http://spraakbanken.gu.se/korp/#lang=eng>.

13 Karp: vmesnik za iskanje po švedskih leksikalnih virih:
[http://spraakbanken.gu.se/karp/#lang=en&textsearch=\(wf+%3D+%22%22\)+sortBy=null+wf](http://spraakbanken.gu.se/karp/#lang=en&textsearch=(wf+%3D+%22%22)+sortBy=null+wf).

1.1.6 Italijanska leksikalna podatkovna baza CLIPS

CLIPS (Corpora e Lessici dell'Italiano Parlato e Scritto)¹⁴ je trenutno najobsežnejša leksikalna baza za italijanščino (pribl. 55.000 lem oz. semantično zakodiranih pomenov besed), izdelana na Inštitutu za računalniško lingvistiko Antonio Zampolli v Pisi, namenjena pa je primarno izdelavi jezikovnih aplikacij in procesiranju naravnega jezika. Gre za leksikalno bazo, ki je neposredno nadaljevanje in nadgradnja v okviru projekta PAROLE zgrajene ter morfološko in skladenjsko označene baze, ki je bila nato v projektu SIMPLE semantično označena, v okviru nacionalnega projekta CLIPS pa so semantične oznake dopolnili s pomeni iz italijanskega WordNeta (ItalWordNet)¹⁵ in korpusa PAROLE. Danes vsebuje podatkovna baza CLIPS štiri medsebojno povezane nivoje jezikovnega opisa: glasoslovnega, oblikoslovnega, skladenjskega in pomenskega.



Slika 3: Ravni opisa v italijanski podatkovni bazi CLIPS

¹⁴ CLIPS: http://www.ilc.cnr.it/clips/CLIPS_ENGLISH.htm.

¹⁵ ItalWordNet: http://www.ilc.cnr.it/iwndb_php/.

Čeprav je povezava med posameznimi nivoji predvidena, pa prihaja, kot prikazuje Slika 3, pri enakopisnicah (enotna oblikoslovna paradigma), npr. za lemo *pesca* (*ri-bolov, breskev*), do prekinitev – zaradi ene same oblikoslovne enote, ki zastopa dve fonološki enoti – med fonološkim in pomensko-skladenjskim nivojem.

Fonološki nivo prinaša podatke o naglasnem mestu in kakovosti naglašene leme ter nekaterih oblik (npr. edninske in množinske), prek sklica na ustrezno oblikoslovno enoto pa tudi osnovne oblikoslovne podatke. Na oblikoslovnem nivoju so navedeni podatki o osnovnih kategorijah (npr. samostalni, pridevniki) in podkategorijah (npr. skupno ime), o tvorjenju oblik (sklanjate, spregatve, tvorba množinskih oblik) ter sklic na ustrezno skladijsko enoto. Osnovni podatki na skladijskem (natančneje pomensko-skladijskem) nivoju so treh vrst. Ponazarjalni del vključuje definicijo v obliki opisa ali sinonima in primere rabe. V razdelku t. i. notranjih (lastnih) lastnosti leksikalne enote so navedeni: besednovrstna kategorija, tipična upovedovalna možnost, kot je npr. pasivizacija, nikalna oblika, in drugi kategorialni podatki za posamezno besedno vrsto (npr. števnost pri samostalnikih, pomožnost pri glagolih), nadalje tipična skladijska vloga (npr. atributivna ali predikatna) oz. pomensko-skladijska funkcija pri prislovih. V strukturnem razdelku je za vsako pomensko-skladijsko enoto določena udeleženska zgradba, v kateri so posamezni elementi ali položaji določeni z zaporednim mestom, glede na obveznost/poljudnost zapolnitve in glede na značilnost strukture, ki položaj zaseda. Na pomenskem nivoju je posamezna leksikalna enota definirana kot struktura, ki jo sestavljajo različne pomenske komponente. Tradicionalni enodimenzionalni pristop v pomenskem opisu (že v projektu SIMPLE) je bil v nekaterih primerih nadgrajen z večdimenzionalnim pristopom, ki omogoča podrobno predstavitev medsebojne povezanosti leksikalnih enot na podlagi štirih osnovnih funkcij t. i. *qualia struktur* (Pustejovsky 1995), in sicer formalne, ki določa: *Kaj je X*, sestavne: *Iz česa je X*, izvirne: *Kako je nastal X*, in namenilne: *Čemu je namenjen X*. Pomenski del leksikalne baze je hkrati tudi ontološka baza, izdelana po vzoru EuroWordNeta, posebnost pa predstavlja za vsak semantični tip izdelana predloga, ki predvideva organizacijo podatkov (npr. področje, semantični razred, kot je npr. čas, znotraj tega pa še časovno obdobje, trajanje ipd., ter tip dogodka, npr. stanje, proces, prehajanje), ki so pomembni za njegovo prepoznavnost (definicijo) in povezujejo ontologijo z besediščem.

Prednosti leksikalne baze CLIPS so predvsem v njeni celovitosti in bogati opremljenosti z različnimi jezikovnimi podatki ter v možnosti povezav z drugimi, tudi tujejezičnimi podatkovnimi bazami. Zadnje omogoča postopno izboljševanja in prilagajanje sheme podatkom oz. različnim jezikom. Slabost podatkovne zbirke je

v njeni kompleksnosti in že omenjena nepovezanost med fonološkimi in pomensko-skladenjskimi podatki. Žal zbirka, kot tudi orodja, ki so bila uporabljena pri izdelavi, niso prosto dostopna, malo pa je tudi informacij o metodologiji dela.

Jezikovnotehnoške leksikalne baze, ki jih predstavimo v nadaljevanju, se osredotočajo zgolj na določen segment leksike, zlasti na glagole in v manjši meri na samostalnike ter pridevnike, ki so z vidika strojnega procesiranja zanimivi zaradi svoje osrednje pomensko-skladenjske vloge v jeziku. Gre torej za podatkovne baze, ki so v prvi vrsti namenjene skladenjskim in pomenskim analizam jezika, potrebnim za izdelavo kompleksnejših prevajalskih orodij, orodij za označevanje korpusov na skladenjski in semantični ravni ter za izdelovanje specializiranih jezikovnih priročnikov, kot so denimo vezljivostni ali kolokacijski slovarji.

1.1.7 Vezljivostni leksikon čeških glagolov Vallex

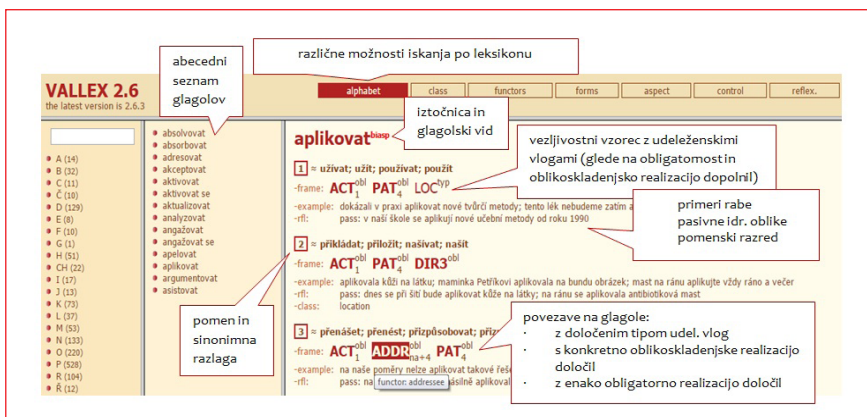
Vallex 2.6.3 (The Valency Lexicon of Czech Verbs with Complex Syntactic-Semantic Annotation) je trenutno zadnja različica¹⁶ vezljivostnega slovarja čeških glagolov, ki ga pripravljajo na Inštitutu za formalno in aplikativno jezikoslovje na Fakulteti za matematiko in fiziko Karlove univerze v Pragi. Gre za elektronsko podatkovno zbirko jezikoslovno označenih in dokumentiranih čeških glagolov, ki je bila na spletu objavljena že leta 2003, leta 2008 pa je Vallex izšel tudi v knjižni obliki. Danes vsebuje prosto dostopna zbirka približno 2.730 geselskih člankov, ki jih na ravni leme sestavljata dovršna in nedovršna oblika glagola, z okrog 6.460 pomeni. Opis glagolov temelji na Češkem nacionalnem korpusu¹⁷ in Praški odvisnostni drevesnici (ang. *PDT*),¹⁸ v izhodišču pa je zahteval veliko ročnega označevanja, kasneje pa so se izdelovalci usmerili v avtomatizacijo, zlasti v razvoj avtomatskega prepoznavanja glagolske večpomenskosti in pripisa ustrezne pomensko-skladenjske oznake.

Zadnja različica vezljivostnega leksikona, ki je uporabnikom na voljo v formatu xml, html in pdf, vsebuje t. i. valenčne okvire ali vzorce z osnovnimi pomensko-skladenjskimi značilnostmi za najpogostejše češke glagole oz. njihove posamezne pomeni (število vezljivostnih mest, njihova morfološka oblika in (ne)obveznost zapolnitve), poleg tega pa še kratko sinonimno opredelitev pomena, primere rabe in semantični razred za konkretni glagolski pomen (Slika 4). Leksikon v posameznih primerih navaja tudi idiomatične zveze ter podatke o refleksivnih in recipročnih oblikah.

16 Vallex 2.6.3: <http://ufal.mff.cuni.cz/vallex>. Različice se nenehno posodablja.

17 ČNK: <http://ucnk.ff.cuni.cz/index.php>.

18 The Prague Dependency Treebank 2.0: <http://ufal.mff.cuni.cz/pdt2.0/>.



Slika 4: Leksikalnogramatični podatki v vezljivostnem leksikonu čeških glagolov Vallex (različica 2.6)

Vallex je s svojo vsebino in zgradbo prototip jezikovnotehnoške pomensko-skladenjsko orientirane podatkovne baze, ki vsebuje širok spekter informacij slovarskega tipa, hkrati pa sta zasnova in način predstavitve namenjena procesiranju naravnega jezika ter jezikovnotehnoškim aplikacijam za češčino. Kot tak je Vallex tipičen primer jezikovnega vira, namenjenega tako jezikoslovno kompetentnemu uporabniku, jezikoslovcu in leksikografu, ki je pomensko-skladenjski podatek sposoben pretvoriti v splošno uporabniško informacijo, kot jezikovnotehnoški skupnosti, ki potrebuje kodirane ter računalniško procesljive jezikovne podatke za izboljšavo obstoječih in pripravo novih jezikovnotehnoških aplikacij.

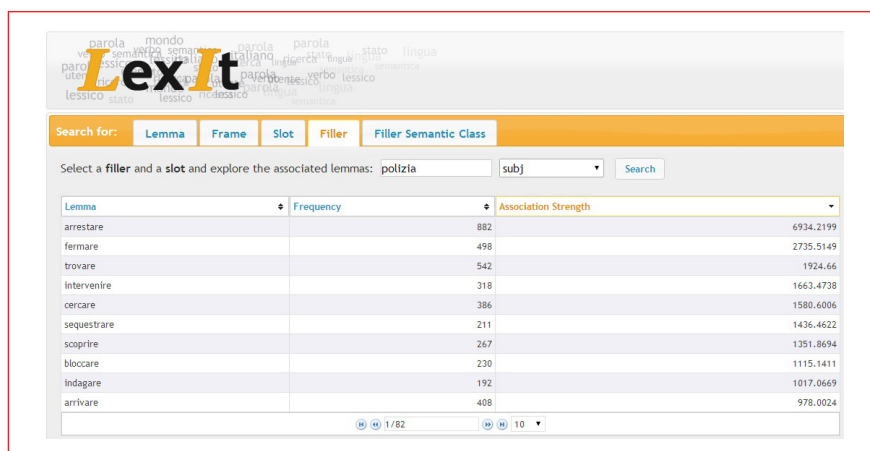
1.1.8 Distribucijski profil italijanskih glagolov, samostalnikov in pridevnikov LexIt

V nizu elektronskih podatkovnih zbirk (Vallex, Sensen Verb Lexicon ipd.), ki so prosto dostopne na spletu in združujejo pomensko-skladenjske podatke, pridobljene z analizo korpusov, njihovi rezultati pa so namenjeni predvsem jezikoslovnim analitikom in jezikovnotehnoški skupnosti, je potrebno omeniti tudi projekt LexIt,¹⁹ ki vsebuje avtomatsko pridobljene korpusne podatke o udeleženski zgradbi italijanskih glagolov, v manjši meri pa tudi samostalnikov in pridevnikov.

¹⁹ LexIt: <http://sesia.humnet.unipi.it/lexit/>.

Spletni iskalnik omogoča iskanje po osnovni obliki izbranih glagolov, samostalnikov in pridevnikov, po vnaprej določenih stavčnih vzorcih, v katerih se ti glagoli (samostalniki in pridevniki) pojavljajo, po tipičnih leksikalnih zapolnitvah posameznih mest v stavčnem vzorcu ter po skupnih semantičnih tipih besed, ki se tipično pojavljajo na posameznih mestih udeleženske zgradbe. Vsak podatek je opremljen s frekvenco pojavitve v izbranem korpusu in statistično vrednostjo povezovalne trdnosti.

Tako lahko npr. dobimo seznam vseh glagolov, ki se v izbranem korpusu realizirajo v vzorcu osebek + brezpredložno dopolnilo, ter tipične besede, ki se pojavljajo na teh mestih (Slika 5). Kot izhodišče lahko postavimo tudi konkretno besedo, določimo skladenjsko pozicijo znotraj stavka (npr. osebek) in ugotovimo, pri katerih glagolih je konkretna leksikalna zapolnitev mogoča in kako pogosto se v zvezi z njim pojavlja.



Slika 5: Del frekvenčnega seznama glagolov s predvidljivim osebkovim mestom v podatkovni bazi LexIt

Projekt LexIt je rezultat prostovoljnega dela sodelavcev in študentov Laboratorija za računalniško jezikoslovje na Oddelku za jezikoslovje Univerze v Pisi. Pomembno pri tej bazi je, da so bili t. i. leksikalni profili za posamezne glagole iz korpusov²⁰ pridobljeni povsem avtomatsko, torej zgolj s pomočjo sodobnih orodij in metod računalniškega jezikoslovja, brez kakršnegakoli ročnega preverjanja. Kot taka vključuje podatkovna baza nujno tudi napake, povezane z napačno lematizacijo, oblikoskladenjskim

²⁰ Uporabljena korpusa sta La Repubblica (331 milijonov besed) in Wikipedia (pribl. 152 milijonov besed).

označevanjem in skladenjskim razčlenjevanjem. Posledično tako ni mogoče pričakovati razločevanja med glagolskimi dopolnili in določili znotraj udeleženske zgradbe, saj avtomatizacija omenjenih postopkov zahteva formalizacijo takih položajev, ki pa lahko temelji le na jasno formuliranih parametrih v jezikoslovni skupnosti. Leksikalna baza italijanskih glagolov je dober pokazatelj tega, v kolikšni meri je mogoče informacije slovarskega tipa avtomatizirati in jih v pregledni obliki ponuditi naprednemu jezikovnemu uporabniku, kot je leksikograf, jezikovni analitik ali učitelj jezika. V nadaljevanju projekta želijo avtorji med drugim razširiti število informacij, ki jih je mogoče avtomatsko pridobiti iz označenega korpusa – dobrodošli bi bili denimo korpusni zgledi za izluščene podatke – s čimer želijo leksikalno bazo še za spoznanje bolj približati splošnemu in ne zgolj jezikoslovno podkovanemu uporabniku.

1.1.9 Špansko-katalonski jezikovnotehnoški podatkovni bazi SenSem in ADESSE

Med katalonskimi in španskimi jezikovnotehnoškimi leksikalnimi viri je treba omeniti projekt SenSem – Podatkovno bazo skladenjsko in pomensko označenih španskih stavkov (SenSem Corpus)²¹ v pristojnosti Katalonske meduniverzitetne raziskovalne skupine za aplikativno lingvistiko GRIAL in ADESSE, leksikalno bazo španskih glagolov, ki temelji na predhodnem projektu sintaktičnega označevanja španskih glagolov BDS,²² ki je potekal na Univerzi Santiago de Compostela med leti 1990–1999.

Pri projektu SenSem je bil v letih 2004–2006 izdelan korpus na pomenski in skladenjski ravni označenih približno 25.000 stavkov (100 stavkov za vsakega od 250 najpogostejših španskih glagolov). Vsak stavek je označen glede na pomen vsebovanega glagola, glede na vrsto določil in dopolnil, njihovo skladenjsko kategorijo in funkcijo ter glede na udeležensko vlogo. Vsi stavki so v okviru svojih semantičnih lastnosti označeni tudi v razmerju do glagolskega vida, modalnosti in tipa skladenjske konstrukcije, v kateri se nahajajo.

Na podlagi tako označenega korpusa je bila izdelana glagolska leksikalna baza, ki povezuje vse predhodne informacije, pri čemer je osnovna enota opisa posamezni pomen glagola. Ta vsebuje njegovo udeležensko zgradbo in povezane alternacijske vzorce, frekvenco posameznega vzorca in informacije, ki se nanašajo na stavčno semantiko.

Slika 6 prikazuje spletni vmesnik, ki omogoča na prvi ravni iskanje po glagolih z izbiro jezika (španščina ali katalonščina) ali iskanje po splošnih merilih, in sicer glede na pomen (seznam in opis glagolov, katerih udeleženci, npr. različni tipi delovalnikov,

21 SenSem Corpus in Verbal Lexicon: <http://grial.uab.es/sensem/corpus?idioma=en>.

22 Projekt BSD - Base de Datos Sintácticos del español actual: <http://www.bds.usc.es>.

okolščin itd., vključujejo določene semantične vloge) ter glede na razred glagolov, ki pripadajo določenemu glagolskemu vidu (npr. proces, dogodek, stanje). Poleg tega je izbor glagolov mogoče omejiti še na naklon, čas in druge kategorije. Seznam glagolov oz. njihovih pomenov, ki vključujejo dani pogoj ali kombinacijo pogojev, je opredeljen s pomensko razlago, klik na povezavo (druga raven) pa omogoča podroben opis glagola v danem pomenu, ki med drugim vključuje sinonime, povezavo z WordNetom, frekvenčni podatek iz korpusa, prevod v katalonščino oz. španščino ter povezavo na podroben opis argumentne zgradbe (ali na več argumentnih zgradb, če obstajajo za posamezni glagolski pomen) na tretji ravni.

Grial SenSem Corpus Verbal Lexicon

Search criteria

Search by verb

Search by common criteria

Language: Spanish

Semantic: Choose Agent (A) Process

Aspectual class: Process

Results (Spanish)

101 verbs (182 senses) found with the criteria (Semantic Agent, Aspectual class Process)

Sentido	Definición	Roles	Event.	Frec.
13	Confiar los problemas, sentimientos u opiniones a otra persona.	Agent, Goal	process	process
15	Mostrar disposición o conformidad hacia algo.	Agent, Theme	process	process

abrir

abrir_13

Realization Frames "SN V SP" [1]

abrir_13 - Confiar los problemas, sentimientos u opiniones a otra persona.

- Typifications of the logical subject:
 - 1 - PERSONA (1)
 - 2 - PERSONA (1)
- Typifications of the logical object:
 - 1 - PROBLEMA (1)
 - 2 - PROBLEMA (1)

Topicalization of the logical subject

1

Categories:

- Personal phrase (relative phrase)
- Preposition + relative clause

Syntactic function: Subject

Semantic Roles:

- Agent
- Goal

Detopicalization of the logical subject

There are no examples

Slika 6: Leksikalnogramatični podatki v leksikalni bazi španskih glagolov Grial

Med jezikovnotehnološkimi leksikalnimi podatkovnimi bazami je potrebno posebej izpostaviti ontološke leksikalne baze, ki pripisujejo t. i. pomenske koncepte posameznim leksikalnim enotam, ti pa omogočajo povezavo na podlagi prepoznanih leksikalnih razmerij (npr. že omenjeni FrameNet). Skupno vsem ontološkim leksikalnim bazam, katerih izdelava je postala v času pospešene gradnje najrazličnejših računalniško procesljivih leksikalnih virov izredno priljubljena, je, da si na različne načine prizadevajo predstaviti hierarhijo jezikovno neodvisnih pojmov in uporabljajo podobne relacije za ustvarjanje povezav med posameznimi pojmi (Fišer 2005: 18).

1.1.10 WordNet in sloWNet

Tipičen predstavnik ontološke ali semantične podatkovne zbirke je WordNet (Fellbaum 1998). Gre za semantični leksikon angleškega jezika, ki so ga pod vodstvom Georgea A. Millerja pričeli razvijati že leta 1985 na Univerzi Princeton v ZDA, njihov oddelek za kognitivne znanosti pa to zbirko angleških besed tudi danes redno posodablja. Temeljno organizacijsko načelo WordNeta je razvrščanje leksike v sopomenske nize (t. i. *sinsete*) in ugotavljanje medbesednih pomenskih razmerij, kot so sopomenskost ter nad- in podpomenskost. Zaradi proste dostopnosti je WordNet izredno priljubljena zbirka za razvoj različnih postopkov za obdelavo naravnega jezika. Podatkovno bazo, ki vsebuje že preko 150.000 besed, lahko uporabniki brezplačno namestijo na računalnik²³ ali pa uporabljajo kar spletno različico.²⁴

Na temeljih angleškega WordNeta je izdelan tudi slovenski WordNet – sloWNet (Fišer 2009, 2009a; Fišer in Erjavec 2010).²⁵ Tako kot angleški WordNet je tudi sloWNet leksikalna zbirka, ki ustreza opredelitvi ontologija, zgrajen pa je s pomočjo avtomatskih postopkov iz različnih jezikovnih virov, kot so dvojezični slovarji in tezavri ter vzporedni korpusi (Fišer 2009). Za pregledovanje in nadaljnjo gradnjo sloWNeta je na voljo tudi programsko orodje SloWTool²⁶ (Fišer in Novak 2011), ki omogoča iskanje, urejanje in vizualizacijo leksikalnih podatkov v obliki hiperboličnih grafov in slik. Razvita pa je tudi aplikacija za izboljšanje slovenskega WordNeta SloWCrowd²⁷ (Tavčar et al. 2012) s pomočjo t. i. množičenja (ang. *crowd-sourcing*).

Da bi ugotovili, na kakšen način je mogoče obe bazi, tj. LBS, ki je primarno namenjena izdelavi slovarskih priročnikov, in sloWNet, katerega temeljni namen je izboljšanje strojnega prevajanja (Fišer in Vintar 2010) in avtomatskega razdvajanja besednih pomenov, medsebojno obogatiti, smo izdelali poskus avtomatskega prekrivanja leksikalnih podatkov v obeh bazah (Fišer et al. 2012). Raziskava je pokazala, da je uspešnost pomenskega prekrivanja tako v številu pomenov (prekrivnih je 90 % pomenov v LBS in skoraj 80 % pomenov v SloWNetu) kot tudi v pomenski natančnosti zlasti za samostalniške besede precej visoka (63 %). Primerjava je tudi pokazala, da je največ prekrivnosti in s tem tudi možnosti za medsebojno plemenitenje obeh baz mogoče pričakovati na ravni sinonimov, hiperimov in holonimov ter na ravni področnih oznak. Prekrivnosti obeh leksikalnih virov, vključno z analizo podatkov, je razvidna iz Preglednice 1.

23 WordNet: A lexical database for English: <http://wordnet.princeton.edu/wordnet/download/>.

24 Spletni iskalnik za WordNet: <http://wordnetweb.princeton.edu/perl/webwn>.

25 Spletna stran: <http://lojze.lugos.si/~darja/slownet.html> prinaša opis gradnje slovenskega WordNeta s povezavami na relevantno literaturo.

26 SloWTool: <http://nl.ijs.si/slowtool/>.

27 SloWCrowd: http://nl.ijs.si/slowcrowd/select_project.php.

Preglednica 1: Rezultati pomenskega prekrivanja med LBS in sloWNetom za samostalni jezik

LBS		sloWNet		Analiza	
Iztočnica	Pomen	Sinset ID	Sinonimi/Razlaga	Prekrivne besede	Evalvacija
jezik (language, tongue)	1. organ	eng-30-05301072-n	clapper, glossa, lingua, tongue / <i>a mobile mass of muscular tissue covered with mucous membrane and located in the oral cavity</i>	jezik, organ, usten, votlina	OK
	1.1 v kulinariki	eng-30-07652995-n	tongue / <i>the tongue of certain animals used as meat</i>	meso, organ, uporabljati, žival	OK
	2. sredstvo za sporazumevanje	eng-30-05808557-n	language, linguistic process / <i>the cognitive processes involved in producing and understanding linguistic communication</i>	jezik, razumevanje	CLOSE
	2.1 miselna sposobnost			jezik, razumevanje	OK
	2.2. komunikacijski proces			jezik, proces	OK
	3. način izražanja	eng-30-06282651-n	language, linguistic communication / <i>a systematic means of communicating by the use of sounds or conventional symbols</i>	način	OK
		eng-30-07082198-n	tongue / <i>a manner of speaking</i>	način	CLOSE
	3.1 v govoru	eng-30-07082198-n	tongue / <i>a manner of speaking</i>	govor	CLOSE
		eng-30-05650820-n	language, speech / <i>the mental faculty or power of vocal communication</i>	govor	OK
	4. računalništvo	/ni prekrivanja/	/ni prekrivanja/	/	/
5. geografija		eng-30-09442595-n	spit, tongue / <i>a narrow strip of land that juts out into the sea</i>	geografija	OK
	6. del čevlja	eng-30-04450994-n	tongue / <i>the flap of material under the laces of a shoe or boot</i>	čevljev	OK

1.1.11 Slovenska podatkovna zbirka ASES

Med slovenskimi jezikovnotehnološkimi podatkovnimi zbirkami naj omenimo še zbirko ASES (Amebisov skupni elektronski slovar), ki nastaja na podjetju Amebis, d. o. o., Kamnik. V izhodišču gre za predkorpusno leksikalno podatkovno zbirko jezikovnotehnološkega tipa, saj njen nastanek leta 1988 sega v čas, ko za slovenščino korpusi še niso bili na voljo, se pa od njihovih nastankov dalje posodablja izkorpusno (Arhar in Holozan 2009: 30). Na podlagi zbirke so nastali pomembni jezikovnotehnološki izdelki za slovenščino, npr. lematizator in oblikoskladenjski označevalnik, črkovalnik, slovnični pregledovalnik, strojni prevajalnik, sintetizator govora, programirani sogovornik, ki jih je razvilo omenjeno podjetje.

Sistem ASES sestavljajo med seboj povezane iztočnice različnih tipov, in sicer *besede*, ki predstavljajo oblikovni del jezikovnega znaka, zato so nanje z oblikoskladenjskimi oznakami pripeti tudi podatki o pregibnih oblikah, statistični podatki ter podatki o izgovoru (Romih in Holozan 2002: 166). Ta tip iztočnic prinaša še podatke o *pomenih* in *uporabah*, kamor sodijo besedne zveze, v katerih se beseda pojavlja. Prek pomenov, ki so razumljeni kot formalizacija semantičnega dela posameznega (enobesednega) jezikovnega znaka (Arhar in Holozan 2009: 32), se slovenske besede povezujejo bodisi s tujejezičnimi ustrezniki, t. i. *izvori*, bodisi z enim ali več delnimi tezavri, besednimi zvezami, sopomenkami in protipomenkami. Pomeni so v zbirki definirani glede na potrebe prevajanja, in sicer ločevalno, kadar ima beseda več možnih prevodov, npr. *moder* – ang. *blue* za pomen 'barva' in *wise* za pomen 'pameten', ter združevalno, kadar se več besed prevaja na enak način, npr. *moder* in *sinji* – oba za pomen 'barva' – se v angleščini prevajata z *blue* (Arhar in Holozan 2009: 32). Posamezni pomeni, ki predstavljajo povezave med besedami, vstopajo v t. i. *odnose*, ki so definirani z medpomenskimi razmerji, kot so sopomenskost, nadpomenskost in protipomenskost, oz. so kvalificirani glede na rabo in/ali normo, npr. napačno, pogovorno, strokovno, redko ipd. Sistem ASES sestavljajo še iztočnice tipa *zveza* s podatkom o sestavljenosti, tipa *skupina*, ki združuje besede, ki so za namene strojnega procesiranja združene, npr. dovršna in nedovršna oblika glagola, ter *glagolska predloga*, ki vsebuje vezljivostni vzorec glagola, natančnejše podatke o oblikoskladenjskih in v nekaterih primerih tudi pomenskih realizacijah udeležencev na posameznih mestih stavčnega vzorca ter podatek o njihovi obligatornosti.

ASES je torej razmeroma kompleksna leksikalna zbirka, ki v medsebojne odnose postavlja različne tipe iztočnic in z njimi povezane oblikoskladenjske in leksikalne podatke, kot so pomenski opisi v obliki indikatorjev, uporabni predvsem na ravni onotoloških pomenskorazmernih in jezikovnokontrastivnih

povezav. Ambicija zbirke v nadaljevanju je vključitev nabora stalnih besednih zvez, nadgradnja informacij o glagolski vezljivosti, vključitev kolokacijskih informacij ter kompleksnejših povezav med leksikalnimi odnosi (Arhar in Holzan 2009: 49).

1.2 Jezikovnopriločniške ali slovarske leksikalne podatkovne baze

Jezikovnopriločniške ali slovarske leksikalne baze so primarno namenjene izdelavi konkretnih slovarjev, pri čemer končni izdelek ni nujno – oz. v večini primerov ni – knjižni slovar. Danes smo priča vrsti hibridnih jezikovnotehnoških in slovarskih podatkovnih baz, ki so prosto dostopne na spletu v obliki jezikovnoinformacijskih portalov in ponujajo informacije za različne profile končnih uporabnikov: od možnosti pridobivanja podatkov v surovi obliki za nadaljnje procesiranje, kar je zanimivo predvsem za računalniško skupnost, do splošnega, šolskega ali specializiranega uporabnika, ki ga zanima bodisi zgolj določen tip informacij ali zgolj določen način predstavitve informacij, npr. stopnja zahtevnosti, podrobnosti, način vizualizacije ipd. Govoriti je mogoče o neke vrste jezikovnih spletnih servisih, kjer je uporabnik lahko kdorkoli, informacije pa so strukturirane na način, da jih je mogoče prikazovati samostojno ali v različnih kombinacijah oz. po potrebi. Spletni medij ponuja tu tako rekoč neskončno število možnosti.

Pri izdelavi LBS smo se zgledovali zlasti po tistih projektih, ki združujejo jezikovnotehnoške cilje, tj. uporabnost za namene nadaljnjega jezikovnega procesiranja, in uporabniškost, tj. vključevanje slovarskega tipa informacij, namenjenih človeškemu uporabniku.

Med slovarskimi leksikalnimi bazami v nadaljevanju predstavimo zasnovo interne Cobuildove podatkovne baze s konca 80. let in sodobno podatkovno bazo za angleški jezik DANTE ter nekatere slovarske baze v obliki spletnih portalov, kot so nizozemski spletni slovar, francoska leksikalna podatkovna baza, nemški leksikološko-leksikografski informacijski sistem, veliki spletni poljski slovar, španski spletni slovar za učenje španščine kot tujega jezika in večjezični norveško-švedski spletni slovarski projekt. Na koncu na kratko opišemo predhodnico slovenske leksikalne baze in obrnjeno angleško-slovensko slovarsko podatkovno bazo, ki je nastala na podlagi Oxford DZS-jevega Velikega angleško-slovenskega slovarja. Omenjeni slovarski bazi sta nastali v okviru korpusnih slovarskih projektov založbe DZS med letoma 1995 in 2006.

1.2.1 Slovanska podatkovna baza projekta Collins Cobuild

Slovarski projekt Collins Cobuild English Language Dictionary je v 80. letih prejšnjega stoletja z uvedbo korpusov in računalniške tehnologije v leksikografski proces popolnoma spremenil pogled na metodologijo izdelave slovarjev, posledično pa tudi na slovarje in slovarsko informacijo samo. Čeprav so 80. leta veljala še za dobo tiskanih slovarjev, je predstavljal Cobuildov projekt prelomnico, brez katere ne bi bila mogoča revolucija elektronskih in spletnih slovarjev, ki smo ji priča 30 let pozneje. Čeprav oblikovanje slovarske podatkovne baze ob začetku projekta še ni bilo zastavljeno v elektronski obliki (Clear 1987: 42–62), je organizacija jezikovnih podatkov v njej pomembna zaradi nekaterih splošnih načel, ki v veliki meri veljajo za slovarske leksikalne baze še danes:

- Leksikalna podatkovna baza naj bi bila skladišče leksikalnih podatkov in ne elektronska različica natisnjene slovarja: njen primarni namen je podrobna analiza splošnega besedišča, iz katere je mogoče oblikovati različne referenčne jezikovne priročnike bolj ali manj avtomatsko. Iz tega sledi, da mora leksikalna podatkovna baza vsebovati več informacij kot iz nje izhajajoči končni izdelek, hkrati pa končni format priročnika ni nujno identičen formatu leksikalne podatkovne baze.
- Glede na to, da se kot primarni končni produkt leksikalne podatkovne baze predvideva slovar, je smiselno prilagoditi organizacijo leksikalne baze slovarskemu formatu, ki preverljivo najbolje funkcionira pri uporabnikih.
- Leksikalna baza naj bi bila čim bolj fleksibilna. V praksi to pomeni, da je mogoče organizacijo in zgradbo leksikalne baze sproti prilagajati glede na vsebino in naravo jezikovnih podatkov.²⁸

Vsebinsko sestavljata Cobuildovo podatkovno bazo dva osnovna tipa informacij, in sicer: (a) za vsak pomen besede: definicija ter osnovne skladišjske, leksikalne in pomenske informacije; (b) za vsak korpusni zgled h konkretnemu pomenu besede: podrobna skladišjska in pomenska informacija oz. opis obnašanja besednega pomena v stavčnem vzorcu.

Zapis posamezne informacije v leksikalni bazi, zlasti pregibnih oblik, skladišjskih konstrukcij, kolokacij, besednih zvez, sinonimov itd. je bil pri snovanju podatkovne baze natančno premišljen v skladu z namenom, da bo mogoče vse podatke v bazi avtomatsko obdelovati, pridobivati in kombinirati. V ta namen je bil izdelan nabor

²⁸ To načelo se je izkazalo kot izredno pomembno tudi pri oblikovanju LBS. Natančna in strogo določena zgradba podatkovne baze lahko vodi leksikografe v prilagajanje jezikovnega podatka strukturi podatkovne baze in ne obratno. Hkrati je zelo pomembno tudi obrnjeno načelo: bolj ko ostaja struktura leksikalne baze nezapletena, večjo konsistentnost podatkov zagotavlja.

kod, ki so omogočale prenos podatkov v računalniško berljivo obliko, hkrati pa je bil sistem oblikovan dovolj preprosto, da so ga leksikografi lahko hitro in učinkovito uporabljali.

Zgradbo podatkovne baze so pri prenosu v elektronsko obliko določale zgoraj omenjene vsebine, združene v tri medsebojno povezane enote: (a) kategorije (iztočnica in definicija pomena), (b) zgledi (iztočnica, številka pomena, korpusni zgledi) in (c) znotrajbazne reference, ki so omogočale povezavo vseh gesel, ki denimo vsebujejo določeno besedo, ki se pojavlja kot sinonim, kolokator ali del razlage v drugih geslih oz. delih gesel znotraj baze.

Zasnova leksikalne podatkovne baze v okviru projekta Cobuild, ki si je za cilj zastavila čim boljše izrabo računalniških tehnik in zmogljivosti tako pri podrobnih in objektivni analizi korpusnih podatkov kot pri njihovem zajetju in organizaciji znotraj baze, je potisnila leksikografsko znanost v novo razvojno fazo: zmožnost računalnika, da hrani, išče, povezuje, sortira ali kako drugače obdeluje veliko količino jezikovnih podatkov, je, kot pravi Jeremy Clear, vodja računalniške podpore pri projektu (Clear 1987: 61), odprla nova obzorja v leksikografiji na enak način, kot je izum mikroskopa odprl nov pogled na predmetnost v fiziki in naravoslovnih znanostih.

1.2.2 Podatkovna baza na podlagi analize angleških besedil Dante

Projekt Dante (Database of Analysed Text of English) je primarno namenjen izdelavi Novega angleško-irskega slovarja (NEID: New English-Irish Dictionary).²⁹ V obliki, ki je prosto dostopna na spletu (Slika 7), lahko trenutno iščemo po bazi, ki predstavlja končano 1. fazo³⁰ in vključuje analizo angleških besedil na podlagi v ta namen izdelanega korpusa. Analiza, ki je bila za konkretni projekt zastavljena popolnoma na novo – tj. ne da bi se naslanjala na katerokoli predhodno izdelano (korpusno) analizo angleščine ali irščine –, temelji na sistematičnem, koherentnem in podrobnem opisu jezikovnih in leksikalnih dejstev. Podatkovna baza prinaša sistematičen opis pomenov, slovničnega in kolokacijskega obnašanja približno 50.000 najpogostejših angleških besed ter njihove besedilne značilnosti. Po vseh jezikovnih podatkih, ki so v leksikalni bazi strukturirani v obliki 40 različnih podatkovnih tipov (besedna vrsta, inherentne slovnične lastnosti, kot so npr. števnost/neštevnost pri samostalniki, glagolske oblike pri glagolih, vezljivostne lastnosti pri glagolih,

²⁹ NEID: <http://www.focloir.ie/>.

³⁰ Za 1. fazo, tj. za analizo izhodiščnega jezika, je bilo zadolženo podjetje Lexicography MasterClass s svojo ekipo leksikografov.

samostalnikov in pridevnikov, skladenjski kontekst pri posameznih besednih vrstah in številni tipi slovničnih, stilnih, področnih in drugih oznak), je mogoče iskati s pomočjo iskalnika, dostopnega na spletni strani projekta.³¹

The screenshot shows the Dante lexical database interface. At the top, there's a search bar with the text "Search the DANTE database" and a "Search" button. Below the search bar, there's a navigation menu with links: HOME, GETTING STARTED, ABOUT DANTE, APPLICATIONS, SAMPLE XML TEXT, PUBLICATIONS, and CONTACT US. The main content area displays the search results for the word "crusty". On the left, there's a list of 68 documents found, with a "Search again" button and a "Back" button. On the right, there's a detailed view of the word "crusty" with its definition, examples, and collocations. The interface is annotated with several callouts: "iskalni pogoj" (search condition) pointing to the search bar, "iztočnica" (lemma) pointing to the word "crusty", "stilne, področne idr. oznake" (stylistic, domain, etc. markers) pointing to the word "crusty", "definicija" (definition) pointing to the definition of "crusty", "korpusni zgledi" (corpus examples) pointing to the examples of "crusty", "kolokatorji" (collocations) pointing to the collocations of "crusty", "seznam iztočnic, ki ustrezajo iskalnemu pogoju" (list of lemmas that match the search condition) pointing to the list of 68 documents found, and "št. pomena in bes. vrsta" (number of meanings and word class) pointing to the definition of "crusty".

Slika 7: Leksikalnogramatični podatki v leksikalni bazi za angleščino DANTE

Z leksikografskega vidika vključuje Dante metodologijo, po kateri je vsaka lema v 1,7-milijardnem korpusu analizirana v dveh fazah. Prva faza zajema identifikacijo posameznih rab ali pomenov (imenovanih leksikalne enote) za konkretno lemo, v drugi fazi pa so za vsak pomen ugotovljena relevantna jezikovna dejstva (Atkins in Rundell 2008; Atkins et al. 2010). Posamezni pomen oz. leksikalna enota 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 skladenjske lastnosti pomena, njegovo koligacijsko³² in kolokacijsko obnašanje in

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

32 Koligacija se navadno opredeljuje kot tipična zveza besed, ki jo za razliko od kolokacije, ki jo določa leksikalna sopojavitve, določa tipična skladenjska zgradba, npr. v slovenščini zveza zanikanega glagola in samostalnika v roditeljski: *ne imeti časa*.

besedilnotipske preference, kot so področje rabe, register, stil ipd. Drugi temeljni izziv, ki ga uresničuje Dante, je zagotoviti popoln opis realnega obnašanja posameznega pomena in zabeležiti vsa ugotovljena dejstva v obliki formaliziranih kod, ki jih je mogoče računalniško procesirati.

Dante je, kot pravijo avtorji, veliko več kot samo slovar: podatki, ki jih vsebuje, so namenjeni založnikom, ki izdajajo različne tipe elektronskih slovarjev, jezikovnim tehnologom za avtomatsko razdvoumljanje pomenov, izdelavo črkovalnikov, ekstrakcijo informacij ter za razvoj prevajalnih sistemov. Nenazadnje pa tudi jezikoslovcem, raziskovalcem in učiteljem, ki se ukvarjajo z besediščem angleškega jezika. Dante je na področju slovarskih leksikalnih baz prav gotovo zgledovanja vreden projekt, ki opisuje jedrno angleško besedišče v smislu detajlnosti, objektivnosti ter večnamenske uporabnosti na način, ki je v angleški leksikografski teoriji in praksi nov in edinstven.

1.2.3 Splošni nizozemski slovar (ANW)

Elektronski slovar sodobnega nizozemskega jezika (ANW: Algemeen Nederlands Woordenboek),³³ ki pokriva področje Nizozemske in Flandrije, je projekt Inštituta za nizozemsko leksikologijo in bo zajemal obdobje od leta 1970 do 2019. Gre za korpusnoleksikografski projekt, katerega cilj je prikazati jezikovno-pomenski razvoj posamezne besede (z vključitvijo pomenskih in skladenjskih opisov konkretne besede v posameznih obdobjih) 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 različica, slovar na spletu posodablja vsaka dva meseca.

Poleg pomenskega opisa besede ali zveze predvideva ANW tudi prikaz pomenskih razmerij (npr. sopomenke in nadpomenke, semagrami), izgovornih posebnosti ter pregibnih oblik (Moerdijk et al. 2008), primere rabe, tipične skladenjske zveze, stalne besedne zveze, pregovore in reke ter prikaz besedne družine (zlasti zloženke in izpeljanke). Slovar ne prinaša podatkov o etimologiji, razen pri neologizmih, ki predvidevajo izčrpne etimološke podatke. Celotna slovarska baza je interaktivna; podatki o pomenskih lastnostih besed in medpomenskih razmerjih so dopolnjeni s fotografijami, video in zvočnimi posnetki ter notranjimi in zunanjimi povezavami (npr. na Wikipedijo). Kot rečeno, se slovar redno posodablja, pri čemer so upoštevani tako odzivi uporabnikov kot sodobne pridobitve leksikografske teorije in prakse.

33 ANW: <http://anw.inl.nl/show?page=search1>.

Uporabniški vmesnik (Slika 8) 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, aktivnost, lastnost; oblika, količina itd.), ki so uporabniku na voljo v spustnem meniju 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 različnih tipih prenosnikov.

ANW | Algemeen Nederlands Woordenboek

INL | INL SCHATKAMER VAN DE NEDERLANDSE TAAL

Word — Meaning | Meaning — Word | Properties — Words | Find examples | Neologisms | Help · About the ANW

bier

1.0: alcoholische drank

- Part of speech
- Spelling and inflected forms
- Relations to other words
- Woordvorming
- Semagram
- Combinatiemogelijkheden
- Fixed expressions
- Spreekwoorden
- Word family
- 1.1: biersort
- 1.2: glas bier

bier

bier 1.0:
goudgele tot donkerbruine alcoholische drank met een schuimkraagje, die gebrouwen wordt van mout, hop, gist en water

Semagram
Bier...

- is een drank; is een vloeistof; is een stof
- is goudgeel, blond, amberkleurig, bruin, donker, donkerbruin, robijnrood of witgeel
- is meestal licht sprankelend
- smaakt bitter, bitterzoet, iets kruidig, zoet of zuur
- is helder of iets troebel
- heeft meestal een schuimkraag
- is een brouwsel van mout, hop, gist en water, en bevat in de regel alcohol en koolzuur
- dient om de dorst te lessen of wordt gedronken voor de gezelligheid
- wordt gebrouwen door gemalen mout, meestal gekiemde en geeeeste gerst, met water in een beslagkuip of roerkuip te laten trekken, dit beslag te koken, de hieruit gefilterde wort met hop te vermengen en opnieuw te koken, en het vocht dan te laten gisten
- zit in een flesje of een blikje, of wordt getapt van het vat; wordt gedronken uit een bierglas, bv. een pint, pul, fluitje of voetglas, of uit het flesje of blikje
- is licht, met een betrekkelijk laag alcoholpercentage, of zwaar, met een hoger alcoholpercentage
- is dorstlessend, maakt dronken, vrolijk of juist somber bij overmatig gebruik en kan verslavend zijn
- wordt gedronken voor of bij het eten, 's avonds thuis voor de gezelligheid, in de kroeg met vrienden of collega's, op feestjes en evenementen

Part of speech
Type: substantief
Naamtype: soortnaam
Geslacht: onzijdig
Lidwoord: het
Getal: geen meervoud
Betekenisklasse: stofnaam

Spelling and inflected forms
Enkelvoud: bier (bier)
Diminutive: (bier)tje

Relations to other words
Hyperoniem: drank

Woordvorming
Type: ongeleed
[Zoek woord in Etymologiebank](#)

Also:
Whole article
Semagrams
Word family
Relations to other words
Morphology

Proverbs
Combinations
Fixed expressions
Examples

Also search for 'bier' in:
INL-woordenboeken | Wikipedia

Combinations als subject bij een werkwoord
gisten; schuimen; vergisten; verschalen; verschralen
Het bier schuimt rijkelijk over de glazen.
- *Hete zeeprinkels*, Andreas Roels, 1984

Slika 8: Del interaktivnega gesla v Splošnem nizozemskem slovarju

Elektronski slovar sodobnega nizozemskega jezika je dober primer sproti posodabljanega slovarskega priročnika, ki je uporabnikom na voljo na spletu, hkrati pa je vir

računalniško procesljivih podatkov, ki jih je mogoče uporabiti za pripravo ali izboljšavo jezikovnotehnoloških aplikacij za nizozemščino.

1.2.4 Leksikalna podatkovna baza za francoščino (BFL)

Cilj oblikovalcev Leksikalne podatkovne zbirke za francoščino (BFL: Base lexicale du français),³⁴ ki je nastala na Jezikovnem inštitutu Univerze v Leuvenu, je združiti informacije v obstoječih elektronskih francoskih slovarjih, zlasti enojezičnega slovarja za tujce (DAFLES: Dictionnaire d'apprentissage du français langue étrangère ou seconde), 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 učinkovito dostopanje do relevantnih jezikovnih podatkov. Cilj projekta je torej ustvariti celostno, prosto dostopno spletno orodje za učenje in rabo sodobne francoščine, pri čemer je poudarek na interakciji slovarja oz. slovarskih podatkov, učnega okolja ter računalniških orodij za analizo korpusov.

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 9, 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.

Znotraj prvega modula: »poišči podatke o besedi ...« lahko uporabnik dobi informacije o tem, s katerim členom v francoščini se beseda veže, ali je pravilno zapisana, za katero glagolsko obliko in čas gre, kaj pomeni, katere so besede, ki pomenijo ravno nasprotno, kakšen je prevod te besede v nizozemščino, angleščino, nemščino, španščino ali švedščino, kako se beseda rabi v stavku, kako pogosta je v jeziku, katere so druge besede znotraj njene besedne družine, katere besede se pogosto pojavljajo ob njej (z možnostjo filtriranja glede na besedno vrsto), kako se izgovarja itd. Uporabnik lahko prek leksikalne baze dostopa tudi do elektronskega Slovarja za učenje francoščine kot tujega ali drugega jezika (DAFLES), kjer lahko za vsak pomen iskane besede (pomenske razlage so zapisane v obliki stavčnih definicij) dobi podatke o tipični okolici (kolokacije, stalne

34 BFL: <https://ilt.kuleuven.be/inlato/>.

zveze in frazeologija), tipičnih skladenjskih zgradbah in stavčnih vzorcih, v katerih beseda nastopa. Slovar prikazuje tudi medbesedna pomenska razmerja wordnetovskega tipa (sopomenke, protipomenke, nad- in podpomenke) ter s kombinacijo iskalnih pogojev uporabniku omogoča dostop do kompleksnejših podatkov, kot so npr. pravila za določanje spola v francoščini, ekstrakcijo glagolov, ki ustrezajo določenemu stavčnemu vzorcu, iskanje po posameznih besedah znotraj stavčnih definicij, kar omogoča združevanje besed znotraj določenega semantičnega razreda ipd.

KU LEUVEN Contact Wie-is-wie Organigram Biblioteken Toledo Intranet KU Leuven English

ONDERWIJS ONDERZOEK MAATSCHAPPELIJKE ROL OVER KU LEUVEN KU LEUVEN 2013

Leuven Language Institute

Lexical Database for French (Base lexicale du français - BLF) - new site

(Almost) everything you always wanted to know about... French words

Find information on

- a word
Meaning, gender, use of prepositions, translation, ...
- a word combination/expression
lors de, une ambiance règne, à tout prix, ...

Find the translation of

- a word to
- a word combination/expression to

Check

- the use of a (sequence of) word(s)
What would you use? *apparaître (à or sur) l'écran?* *espérer (de or -) faire ...* - check on the web
- whether a translation is correct
salarié > salaried person?

Help me

- understand a text
"Beta" version: work in progress
- translate a text
Coming soon...
- write a text
Coming soon...

Learn

- how to express an idea
Something starts, something is intense, a large amount of something, words and word combinations about economics, ...
- how to combine words to form sentences
salary > to earn a ~ | employer, company, colleagues, ...
- how to avoid common errors
Use of prepositions, position of the adjective, gender, ...

Do exercises

- on all aspects of French vocabulary
On verb forms, gender, use of prepositions, word combinations, ...

INTERFACE

- In English
- Coming soon
- en français
- In het Nederlands

ACCENTED LETTERS AND SPECIAL CHARACTERS

- à é i ô u
- à è i ô u
- à é i ô u
- à é i ô u
- ç
- œ
- ñ
- ü
- ä

Slika 9: Spletni vmesnik za dostop do različnih sklopov podatkov v leksikalni bazi za francoščino BLF

Francoska leksikalna podatkovna baza BFL je tipičen primer interaktivnega jezikovnega portala, ki združuje različne tipe jezikovnih informacij (slovarske, pravopisne, slovnične) na enem mestu in jih povezuje z didaktičnimi gradivi. Čeprav je baza primarno namenjena študentom francoščine kot drugega/tujega jezika, je dober primer celostnega jezikovnega portala, ki se skuša čim bolj približati uporabnikovim potrebam, pri čemer s pridom izrablja možnosti spletnega okolja in rezultate računalniško procesiranih jezikovnih podatkov.

1.2.5 Nemški jezikovni portal Elexiko

Elexiko je slovarski (leksikološko-leksikografski) informacijski sistem, ki so ga izdelali na Inštitutu za nemški jezik v Mannheimu (IDS: Institut für deutsche Sprache) in ponuja podatke o pravopisu, pomenu, slovnicu, sinonimih in posebnostih rabe besed v sodobnem nemškem jeziku. Na spletnem portalu³⁵ lahko uporabnik dobi informacije o besedah, ki so bodisi leksikografsko obdelane (zlasti visokofrekventne nemške besede), ali o besedah, ki imajo avtomatsko generirane podatke na podlagi korpusa. Vse besede je v seznamu mogoče filtrirati glede na to, ali so leksikografsko obdelane ali ne, med prvimi pa še glede na stalne zveze, neologizme, besede, značilne za šolski diskurz, ter na pregovore. Leksikografsko obdelane besede vsebujejo podatke o pravopisu, etimologiji, besedotvorju (seznam vseh tipov tvorjenk, ki vsebujejo ključno besedo, in število pojavitev posamezne tvorjenke v korpusu) in so pomensko opisane znotraj različnih semantičnih konceptov ali t. i. možnih »načinov branja« (nem. *Lesarten*). Slika 10 prikazuje opis samostalnika *Abend* (večer) znotraj treh semantičnih konceptov, in sicer kot 'zaključek dneva,' kot 'dogodek' in kot 'zahod', v okviru posameznega koncepta pa je beseda na kolokacijski ravni opredeljena glede na tipično semantično okolje, ki ga določajo vprašanja kot: *Kakšen je večer?*, *Kaj počnemo zvečer?*, *Kaj prinaša večer?*, in odgovori: *večer je hladen, lep topel; zvečer povabimo, jemo, se srečamo; večer prinaša pivo, večerjo, tekmo, kino* itd. Nadalje glede na tipično rabo, ki vključuje frekventne zveze, kot npr.: *ves večer, vsak večer, ob večerih, v času večera, do poznega večera* itd., ter glede na pomensko sorodne besede, kot so: *poldan, jutro, popoldan, dopoldan* itd.

35 Elexiko: <http://www.owid.de/wb/elexiko/start.html>.

elexiko

Abend

Lesart: 'Tagesende'

zur Übersichtsseite

Lesarten im Überblick

Bedeutungs-
erläuterung

Kollo-
kationen

Konstruk-
tionen

Sinnverwandte
Wörter

Gebrauchs-
besonderheiten

Grammatik

Kollokationen: Wörter im Kontext

Wie ist ein Abend?

kühl

lau

schön

warm

verbergen

Was macht man am Abend?

demonstrieren

einladen

essen

(sich) melden

(sich) treffen

sitzen

verbergen

Was geschieht am Abend?

beginnen

(sich) bilden

(sich) ereignen

stattfinden

verbergen

Kommentar anzeigen

Slika 10: Kolokacijski podatki v leksikološko-leksikografskem sistemu Elexiko

1.2.6 Veliki slovar poljskega jezika

Ob koncu leta 2004 se je Odbor jezikovnih znanosti Poljske akademije znanosti in umetnosti obrnil na skupino jezikoslovcev s pozivom za oblikovanje predloga za izdelavo splošnega velikega slovarja poljskega jezika (Wielki słownik języka polskiego), ki bi kot glavni vir vedenja o poljskem besedišču zamenjal že

močno zastarel Slovar Poljske akademije znanosti in umetnosti.³⁶ Rezultat teh prizadevanj je zasnova Velikega slovarja poljskega jezika, ki je celoti financiran iz državnega proračuna. V projekt je vključenih več deset strokovnjakov z 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 naj bi bil ob zaključku uporabnikom na voljo v elektronski obliki oz. bodo do njega lahko dostopali brezplačno na spletu, knjižna oblika pa ni predvidena. Trenutno je na spletu na voljo testna različica slovarja³⁷ (Slika 11), iz katere je razviden osnovni koncept, ki pa se skozi leksikografski opis še posodablja in dopolnjuje.

Slika 11: Slovarski vmesnik Velikega slovarja poljskega jezika

Preizkusna različica slovarja ponuja seznam trenutno razpoložljivih gesel. Ob kliku na izbrano geslo se odpre pomenska zgradba gesla. Posamezni deli celotnega gesla so za vsak posamezni pomen na voljo z izbiro zavirkov: definicija, tematska klasifikacija (pomenske povezave v obliki ontologije), medpomenska razmerja (npr.

36 Słownik języka polskiego w 11 zvezkih je izhajal med leti 1958–1969 pod uredniškim vodstvom Witolda Doroszevskega. Po njem je mogoče iskati tudi na spletu: <http://doroszewski.pwn.pl/>.

37 Veliki slovar poljskega jezika, demo različica: <http://www.wsjp.pl/>.

so- in protipomenke), kolokacije oz. tipično besedilno okolje, primeri rabe ter leksikon besednih oblik. Nekatera gesla imajo poleg tega še 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.

Spletni slovar poljskega jezika se kaže kot posnemanja vreden primer premišljene organizacije sodobnega slovarskega dela s konsenzom na nacionalni ravni ter osrednjim koordinatorjem. Izredno pomembno pri tem je dejstvo, da je iz državnega proračuna financiran projekt namenjen jezikovni skupnosti brezplačno.

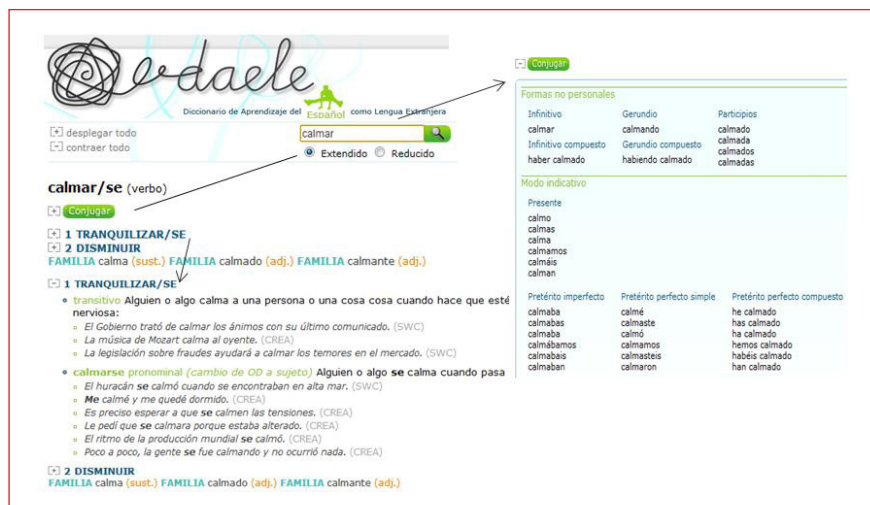
1.2.7 Slovar za učenje španščine kot tujega jezika Daele

Med spletnimi slovarji, ki vključujejo različne leksikalnogramatične podatke, se zdi smiselno omeniti tudi Daele – slovar za učenje španščine kot tujega jezika, ki je prosto dostopen na spletu³⁸ in je bil izdelan na Univerzitetnem inštitutu za aplikativno lingvistiko na Univerzi Pompeu Fabra v Barceloni. Slovar zagotavlja učencem španščine kot tujega jezika pomenske, skladenjske, leksikalne, morfološke, pravopisne, pragmatične in druge podatke za (trenutno) 363 španskih glagolov. Posamezne dele gesla, leksikonski oz. morfološki, pomensko-skladenjski in besednodružinski, lahko uporabnik po potrebi odpira/zapira, s čimer je omogočena preglednost nad celotnim geslom in hkrati osredotočenost na podatke, ki so za uporabnika v danem trenutku zanimivi.

Najobsežnejši je leksikonski del, ki vsebuje podatke o spregatvenih lastnostih konkretnega glagola, in sicer v vseh obstoječih glagolskih oblikah (osebnih in neosebnih) in glagolskih načinih (npr. povedni, železni itd.) ter ločeno znotraj posameznih glagolskih časov. Za vsak glagol (gl. Sliko 12) je izdelan pomenski opis v dveh stopnjah: v obliki kratke pomenske opredelitve glagola, ki tvori neke vrste pomenski meni v »zaprtem« načinu gesla, v »odprtem« načinu pa ponuja za izbrani pomen še pomenski opis v obliki kratke definicije, pri čemer so, zlasti za namene tekstne produkcije v španščini, ločeno pomensko opisane prehodne, refleksivne, trpne in druge skladenjske zgradbe. Vsak pomen oz. vsaka za konkretni pomen tipična skladenjska zgradba ima vsaj eno, navadno pa več, potrditve v korpusnem

38 DAELE: <http://www.iula.upf.edu/rec/daele/>.

zgledu. Avtorji so pri pridobivanju podatkov iz korpusa (Spanish Web Corpus) uporabljali orodje Sketch Engine (Kilgarriff et al. 2004) ter program za izdelavo slovarjev Tshwanelex.



Slika 12: Leksikonski in pomensko-skladenjski del spletnega slovarja Daele za učenje španščine

1.2.8 Skandinavski večjezični spletni slovarski projekt LEXIN

Med zanimivimi večjezičnimi slovarskimi projekti – zlasti z uporabniškega vidika – je švedsko-norveški projekt LEXIN,³⁹ ki zajema izdelavo dvo- oz. večjezičnih spletnih in tiskanih slovarjev manjšinskih jezikov, namenjenih primarno priseljen- cem na Švedskem in Norveškem.

V nadaljevanju na kratko predstavimo norveško spletno različico,⁴⁰ ki vsebuje eno- jezična norveška slovarja: Bokmål Dictionary in Nynorsk Dictionary ter dvojezični slovar obeh jezikovnih različic: Bokmål – Nynorsk. Poleg tega vsebuje spletni slovar še 14 jezikovnih kombinacij obeh jezikovnih različic s priseljenjskimi jeziki, kot so npr. arabski, kurdski, perzijski, poljski, ruski, tajski, vietnamski itd. Norveški projekt temelji na švedskem Lexinu,⁴¹ iz katerega so prenesli 80 % iztočnic.

39 Na podlagi odločitve Švedskega jezikovnega konzila, da se švedsko-angleški in angleško-švedski del jezikovne baze izloči iz projekta LEXIN, je na pobudo Vigga Kanna in Joachim Hollmana nastal t. i. People's dictionary: <http://folkets-lexikon.csc.kth.se/folkets/folkets.en.html>, ki v celoti nastaja s pomočjo množičenja.

40 Norveški LEXIN: <http://decentius.hit.uib.no/lexin.html?ui-lang=eng&com-prosjektet>.

41 Švedski LEXIN: <http://lexin2.nada.kth.se/lexin/#>.

L

E

X

I

N

Bokmål-Thai

Searchword: sak

Search

Norwegian characters: æ ø å Æ Ø Å
Number of hits: 395. Showing 1 to 5.
Previous 5 hits | [Next 5 hits](#)

Bokmål	dictionary entry	sak /sɑ:k/ substantiv thing, object, article noun
English		
Thai		อุปกรณ์, สิ่งของ
Bokmål	inflection	[saken (el saka) saker sakene]
Bokmål	grammar	oftest i flertall
English		most commonly used in the plural form
Thai		ปกติใช้ในรูปแบบพหูพจน์
Bokmål	explanation	en gjenstand, en ting

Bokmål		hvor har du lagt sakene mine?
English		where have you put my stuff?
Thai		คุณเอาสิ่งของส่วนตัวไปวางไว้ที่ไหน?

Bokmål		søtsaker
English		sweets
Thai		ขนมหวาน, ลูกอม

Bokmål		skriveaker
English		writing materials
Thai		อุปกรณ์การเขียน

Slika 13: Leksikalnogramatični podatki v večjezičnem spletnem slovarju LEXIN

Slovarji zajemajo 36.000 besed (iztočnic, sestavljenk in izpeljank) in so prijazni do uporabnikov, ki so pogosto manj izobraženi ljudje: besede so izbrane z mislijo na ciljne skupine (splošno besedišče, socialni izrazi). Za ljudi z manj razvito bralno zmožnostjo je del iztočnic ilustriran (2700 besed), pregibne oblike pa so v celoti izpisane. Slovar vključuje tudi avtomatsko sintezo govora, ki omogoča poslušanje izgovorjave posameznih besed, besednih oblik, branje razlag, zgledov in drugih besedilnih delov slovarja, uporabnik pa lahko izbira med moškim ali ženskim govorcem. Navodila v angleščini – kot tudi vsi metajezikovni podatki v slovarju, npr. iztočnica, razlaga,

stalna zveza, zloženka, pregibna oblika, primer itd., – so za uporabo spletnega slovarja zelo preprosta in strnjena na vsega 12 točk. Na spletni strani slovarja je mogoče najti tudi dvojezične slikovne priloge, ki prikazujejo izrazje izbranega tematskega področja, kot npr. sorodstvena razmerja, oblačila, glasbeni inštrumenti ipd.

Slika 13 prikazuje enega od pomenov samostalnika *sak* (stvar, predmet) v norveško(bokmål)-tajskem slovarju. Vsak pomen – ki je vedno prikazan kot slovarska iztočnica – ima navedeno izgovarjavo, besedno vrsto, angleško in tajsko prevodno ustreznico, sklanjatvene oblike, slovnična opozorila ter primere rabe v vseh treh jezikih.

Pregled spletnih slovarskih portalov za evropske jezike, ki imajo po ocenah Evropske mreže odličnosti META-NET (Krek 2012) dobro jezikovnotehnološko podporo na področju analize besedila in jezikovnih virov in posledično zagotovljeno preživetje v digitalni dobi, 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 (jezikoslovec, pisec, novinar, prevajalec, učitelj, lektor, oglaševalec, predstavnik odnosov z javnostmi, raziskovalec itd.) ali splošni uporabnik z različnimi (pravopisnimi, slogovnimi, pomenskimi, slovničnimi) jezikovnimi problemi. Na drugi strani je jezikovni informacijsko-servisni prosto dostopni spletni portal, namenjen celotni jezikovni skupnosti določenega jezika. Skupni imenovalec dobrih jezikovnih praks na tem področju je mogoče izpostaviti v nekaj točkah:

- Za elektronske slovarje in spletne jezikovne portale je pomembno, da so za ta namen zasnovani že v izhodišču in da ne gre za prenašanje knjižnih različic in metodologije v elektronsko obliko.
- Jezikovni podatki, ki jih izkazujejo spletni portali v uporabniško prijazni obliki, temeljijo na medsebojno povezanih elektronskih podatkovnih bazah, zato so podatki računalniško procesljivi in jih je mogoče izrabljati za razvoj najrazličnejših jezikovnotehnoloških aplikacij.
- Jezikovni podatki združujejo različne ravni jezikovnega udejanjanja in tako brišejo mejo med slovarskimi, slovničnimi, pravopisnimi oz. stilističnimi in drugimi jezikovnimi pojavi, posledično uporabniki ne razmišljajo več o kategoriji jezikovnega problema, pač pa o jezikovnem problemu kot takem.
- V bazah in na jezikovnih portalih strukturirani jezikovni podatki so hitro dostopni: odgovor na jezikovno vprašanje je mogoče dobiti razmeroma hitro in brez vnapijnjega specializiranega jezikoslovnega znanja ali poznavanja strukturiranosti podatkov v podatkovni bazi.

- Pridobljeni podatki so učinkoviti: ker sta izbor in način predstavitve jezikovnih podatkov izdelana na podlagi predhodnih analiz potreb jezikovne skupnosti oz. analiz potreb potencialnih skupin jezikovnih uporabnikov, uporabnik dejansko najde rešitev za svoj jezikovni problem.
- Jezikovni podatki na spletnih portalih so ažurirani: zagotovljeno je redno posodabljanje in dopolnjevanje glede na dejanske spremembe v jeziku.
- Dostop in uporaba jezikovnih podatkov sta za celotno jezikovno skupnost brezplačna: jezikovni spletni portali so navadno izdelani na državnih ustanovah s pomočjo proračunskih ali projektih sredstev in v skladu s potrjenim jezikovnopolitičnim načrtom za določeno obdobje.

1.3 Obstoječe slovenske leksikalne podatkovne baze

Slovenščina je, upoštevajoč obstoječe jezikovne priročnike, razvitost jezikovnih tehnologij in dostopnost jezikovnih virov med jeziki, ki imajo t. i. delno podporo (Krek 2012: 33), kar predstavlja na 5-stopenjski lestvici predzadnje mesto. Tak status imajo tudi baskovščina in katalonščina ter danščina, finščina, grščina, bolgarščina, romunščina, slovaščina in portugalsščina. Poročilo evropske mreže odličnosti META-NET za slovenščino (prav tam) med drugim ugotavlja, da:

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

Dejstvo je, da je Slovar slovenskega knjižnega jezika edinstven dosežek slovenske leksikografske teorije in prakse, zato tudi ni naključje, da je vlogo enojezičnega slovarja slovenskega jezika opravljal tako dolgo in tako uspešno. Njegova pomembnost za slovensko leksikografsko tradicijo je zlasti v naslonitvi na najpomembnejša evropska jezikoslovna spoznanja tistega časa, čeprav so bila takrat za slovenski prostor v marsičem revolucionarna (Vidovič Muha 2014: 128). Prevladujoči tok slovenskega jezikoslovja je namreč s sprejetjem slovarskega koncepta konkretiziral ideje, ki so v slovenski prostor začele prihajati že v zgodnjih 30. letih,⁴² z vidika leksikografske teorije in prakse pa je pomembno predvsem to, da je bil koncept

⁴² Gre predvsem za dela Voduška, Isačenka, Breznika in Mikuša, ki so v slovenski prostor prinašali prvine ženevskega in praškega strukturalizma (Vidovič Muha 2014: 128/129).

SSKJ splošno sprejet in realiziran v vseh petih knjigah (Vidovič Muha 2014: 130). Prav zato pomenijo dopolnitve razlag in spreminjanje zgledov, ki jih je v želji približati se sodobnemu slovenskemu jeziku mogoče opazovati v drugi izdaji SSKJ (prim. Krek 2014), ne da bi se skladno s sodobnimi leksikografskimi trendi posodobil tudi koncept v gradivnem, metodološkem in teoretičnem smislu, glede na SSKJ1 korak nazaj, saj se zabrisuje pomembna razlika med včerajšnjo in današnjo slovenščino ter njeno pomensko, skladenjsko in oblikoslovno podobo. Z vidika jezikovnih sprememb, ki se kažejo tako v pomenskem potencialu kot v vnašanju novih izraznih možnosti v jezik, je namreč v procesu jezikovnega opisa še toliko bolj pomemben jasno definiran odnos do gradiva in jezikovne stvarnosti, tako pretekla kot sodobna, saj je le na tej podlagi mogoče razmišljati npr. o slovničnih in pravopisnih spremembah ter utemeljevati njihovo (ne)upravičenost.

Ob povedanem ne moremo mimo dejstva da SSKJ (in njegova druga izdaja) ne more dohajati spremenjenih družbenih okoliščin tako v gradivnem, konceptualnem kot tudi v metodološkem smislu, s tem pa slovar tudi ne ponuja rešitev glede kontinuiranega spremljanja razvoja slovenske leksike.

Zametki izključno na korpusnih podatkih zasnovane leksikalne podatkovne baze za slovenščino, ki bi služila sodobnim in bodočim slovarskim potrebam, segajo v leto 2005, ko je bila zasnovana leksikalna baza za izdelavo večjezičnih slovenskih-x slovarjev (Gorjanc et al. 2005), ki jo je takrat še v okviru slovarskega založništva podprla DZS.⁴³ Ideja je temeljila na izkušnjah poskusne faze izdelave geslovnika za male dvojezične slovarje s slovenščino kot izhodiščnim jezikom in na korpusni analizi. Predlagana podatkovna baza je bila oblikovana kot notranje hierarhiziran leksikalni opis sodobnega slovenskega jezika s poudarkom na deskriptivizmu, uveljavljanju aktualnih leksikografskih metod in preprostejšega metaleksikografskega jezika. Tako oblikovana podatkovna zbirka naj bi služila kot izhodišče za izdelavo sodobnih slovarskih priročnikov, ki stremijo k za človeka uporabnemu jezikoslovnemu opisu besedišča.

Izdelava take leksikane baze se je začela dejansko uresničevati tri leta kasneje pri projektu SSJ v okviru aktivnosti Izdelava leksikalne baze za slovenščino, ki jo podrobno opišemo v drugem in tretjem delu monografije. Na tem mestu pa se – v kontekstu dobrih praks evropskih jezikovnotehnoloških in slovarskih projektov – zdi smiselno predstaviti testni spletni prikaz gesel LBS, ki smo ga poimenovali Spletni slovar slovenskega jezika.

43 V tem kontekstu je treba omeniti tudi obrnjeno angleško-slovensko slovarsko podatkovno bazo (Krek et al. 2008), ki je nastala na podlagi Oxford-DZS-jevega Velikega angleško-slovenskega slovarja (Grabnar in Šorli 2003; Šorli et al. 2006).

1.3.1 Spletni slovar slovenskega jezika – demo različica

Spletni slovar slovenskega jezika je poskusni prikaz relevantnih slovarskih podatkov o sodobni slovenščini na spletu in v celoti temelji na LBS. V demo različici⁴⁴ vsebuje slovar približno 2.500 gesel oz. okoli 11.000 leksikalnih enot in je v izhodišču zasnovan dvodelno (Slika 14): na pomenski meni (leva stran) in celotno geslo (desna stran), ki ga je mogoče odpirati s klikom na posamezni pomen, podpomen, stalno zvezo ali frazeološko enoto v pomenskem meniju. Pomenski meni prinaša informacijo o pomenski strukturi gesla, tj. o številu pomenov in njihovi medsebojni povezanosti, ter o obstoju stalnih zvez in frazeoloških enot, ki vsebujejo iztočniško besedo. Posamezni pomeni ali podpomeni, stalne zveze in frazeološke enote, ki se odpirajo na desni strani, vsebujejo razlago pomena, in sicer v obliki pomenskega indikatorja, lahko pa tudi v obliki pomenske sheme, ki prikazuje skladenjsko zgradbo pomena z izpostavljenimi semantičnimi tipi (velike črke) na predvidenih vezljivostnih mestih, in/ali stavčno razlago. Za vsak pomen ali podpomen so navedene kolokacije, združene v pomensko sorodne nize in uvrščene pod ustrezne skladenjske strukture. Kolokatorji znotraj kolokacijskega niza vsebujejo povezavo na korpus Gigafida, kar uporabniku, poleg velikega števila izbranih korpusnih zgledov, omogoča vpogled v rabo besede, zveze ali frazeološke enote v avtentičnem besedilu. Pod posameznimi pomeni so lahko navedene tudi ustaljene skladenjske zveze, v katerih se obravnavana beseda pojavlja, ter opozorila na tipično področje rabe, konotativne in pragmatične prvine, stilistične in slovnične posebnosti, zadnje zlasti na ravni skladenjskih struktur ali pomena kot celote. Pri glagolskih pomenih ima uporabnik možnost videti tudi stavčne vzorce, v katerih se pomen tipično realizira, ter pripadajoče kolokacije, ki so združene pod ustrezno pomenskoskladenjsko vlogo, kot npr. LOKACIJA, NAČIN, VZROK ipd. Nekateri podatki (stavčni vzorci, slovnične restrikcije), ki so bolj podrobni ali namenjeni zahtevnejšim uporabnikom, so vidni prek notranjih povezav oz. s prehodom miške prek besedila.

44 Spletni slovar slovenskega jezika (demo): <http://www.slovenscina.eu/spletni-slovar>.

2 Leksikalna baza in slovar

Leksikalna podatkovna baza je razumljena kot vir leksikalnogramatičnih podatkov v elektronski obliki, namenjenih leksikografom za izdelavo slovarskih priročnikov, jezikoslovcem za nadaljnjo analizo in jezikovnim tehnologom za računalniško obdelavo. Leksikalna baza je 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, zahtevajo torej določeno jezikoslovno interpretacijo, kot je npr. besednovrstna opredelitev iztočnice, pomenska ali skladišna hierarhizacija, pomenska razčlenitev, opredelitev do večbesednih leksikalnih enot, opredelitev leksikalnih enot glede na rabo ipd. Običajno vsebuje leksikalna baza več informacij, kot jih dejansko vključujejo iz nje izhajajoči slovarski in drugi priročniki.

Splošni enojezični razlagalni slovar, zlasti v knjižni obliki, ima drugačno družbeno vlogo. Navadno predstavlja izbor le določenega tipa jezikovnih podatkov in je omejen tako v količini zajete leksike kot v časovnem obdobju, ki mu leksika pripada. Hkrati se slovar dojema tudi kot zakladnica besedišča določene jezikovne skupnosti in dokument o njegovem obstoju, bogatosti in razvejanosti, zaradi česar ima pomembno družbenopovezovalno vlogo (Hanks 2009: 8; Béjoint 2000). Kot tak slovar nikoli ni le inventar besed, njihovih pomenov in sintagmatskih vzorcev, pač pa se od njega pričakuje, da bo podal sodbo, kaj je v različnih kontekstih sprejemljivo in kaj ne, slovarska baza pa mu omogoča, da za tako opredelitev najde ustrezne podatke. Pomembno je namreč, kot pravi Hanks (2009: 8), da sodbe v slovarju temeljijo na empirični analizi rabe v avtentičnih besedilih, zato je interakcija med jezikoslovno ustrezno zasnovano leksikalno oz. slovarsko bazo in slovarjem kot družbenim artefaktom neizogibna.

Ena bistvenih razlik med elektronsko bazo, ki je skladišče podatkov, in končnim slovarskim izdelkom, ki je bil do nedavnega še knjiga, temelji v slovenskem prostoru na dihotomiji baza – elektronskost : slovar – knjižnost, globlje pa so nanjo vezana še druga nasprotja, povezana z dojemanjem jezika, njegove družbene vloge, prestižnosti in znanstvenosti. Slovarski podatkovni bazi se pogosto pripisuje prevladujoča jezikovnotehnološka uporabnost, medtem ko se s slovarjem in slovarskostjo misli predvsem na premišljen, jezikoslovnoteoretično utemeljen izbor in opis določenih tipov jezikovnih podatkov. Pri tem se knjižna slovarska oblika upravičuje s statusom stabilnosti in nespremenljivosti, hkrati pa tudi znanstvenosti in prestižnosti v smislu ohranjanja narodovega jezikovnega zaklada in s tem povezane simbolne vloge jezika. Seveda pa knjižnost, zlasti v dobi pametnih telefonov, tabličnih računalnikov in številnih uporabniško naravnanih aplikacij, zbuja predstave o zastarelosti, neaktualnosti in rigidnosti slovarja kot jezikovnega priročnika.

V obdobju zadnjih 20 let, ko so v evropskem jezikovnem prostoru hkrati nastajale elektronske leksikalne baze in slovarji (knjižni, elektronski in spletni), sta se tako leksikalna baza kot slovar razvijala in spreminjala. Spremembe so se dogajale tako pod vplivom razvoja jezikoslovja in leksikografije kot tudi sporočanjskih potreb in zmožnosti uporabnikov v digitalni dobi. 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 jezikovnoinformacijske spletne portale, ki presegajo in nadgrajujejo prevladujoče pristojnosti slovarja, kamor sodi zlasti pomenska členitev in pomenski opis besed. Tudi nadaljnje smernice v sodobnem leksikografskem opisu in razvoju leksikografije kažejo, da se razlika med slovarjem kot končnim izdelkom in elektronsko zbirko jezikovnih podatkov zabrisuje. Glede na računalniški zapis leksikalne baze je namreč podatke v njej mogoče predstaviti izbirno in za različne končne namene, jih združiti in povezati z drugimi bazami podatkov, s slikovnim in zvočnim gradivom oz. uporabniku omogočiti, da sam določi nivo zahtevnosti. Ni si težko zamisliti možnosti, da uporabniku računalniški mehanizem na vprašanje, kaj neka beseda pomeni, ponudi zgolj niz primerov, v katerih se ta beseda pojavlja, sam pa potem razbere njene pomene glede na tipične kontekste.⁴⁹ Poleg omenjenega lastniki zares obsežnih jezikovnih baz, kot je denimo Google, že sedaj omogočajo vse bolj sofisticirano avtomatsko generirano jezikovno informacijo, zlasti na področju dvojezične leksikografije oziroma prevajanja, s postopnim dodajanjem časovnih, žanrskih ali kako drugače določenih filtrov in zahtevnejših nivojev jezikoslovne interpretacije pa bo uporabnik še lažje usmerjal svoj izbor oz. konkretiziral informacijo, ki ga zanima.

Če torej pri pojmu slovar nismo več obremenjeni s knjižnim formatom, pri pojmu leksikalna baza pa ne z njeno računalniško robustnostjo, nas razlika med njima zanima predvsem na vsebinski ravni. Elektronski vir v primerjavi z analognim načinom predstavitve podatkov slovarskega tipa namreč v marsičem vpliva na zasnov, strukturiranost in količino podatkov. Proces, v katerem sovpadajo razvoj avtomatskega luščenja podatkov iz korpusa, združevanje podatkovnih baz, avtomatsko povezovanje podatkov, izkoriščanje moči množic itd., je tako rekoč nezadržen in v marsičem spreminja razmerje med pojmom leksikalna baza in slovar, kot tudi vlogo leksikografa v odnosu do jezikovnih tehnologov in vse pomembnejše računalniške podpore pri njegovi izdelavi.

49 Na tem mestu omenimo izdelavo eksperimentalnih avtomatskih kolokacijskih slovarjev (<http://www.sketchengine.co.uk/>) za nekatere jezike s pomočjo orodja Sketch Engine, ki deluje na slovnici besednih skic in mehanizmu za izbor dobrih korpusnih zgledov.

B. LEKSIKALNA BAZA ZA SLOVENŠČINO

V drugem delu opišemo zasnovo LBS v okviru projekta SSJ,⁵⁰ predstavimo njen namen, obseg in uporabnike, ki so jim v bazi strukturirani leksikalni podatki namenjeni. Nato se osredotočimo na njeno zgradbo oz. na organiziranost in medsebojno povezanost jezikovnih podatkov, opišemo shemo DTD ter možnost njenega prilagajanja za različne tipe jezikovnopriročniških izdelkov. Posebej opišemo proces pridobivanja jezikovnih podatkov iz korpusa s poudarkom na avtomatizaciji postopkov in izdelavi mehanizmov, specifičnih za slovenščino. V samostojnem poglavju opišemo orodja za pridobivanje jezikovnih podatkov iz korpusa, nato pa še postopke pri njihovi analizi in interpretaciji.

50 Projekt, katerega rezultati so opisani na projektni spletni strani: <http://www.slovenscina.eu/>, je delno financirala Evropska unija iz Evropskega socialnega sklada ter Ministrstvo za šolstvo in šport, izvajal pa se je 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.

3 Osnovna izhodišča

Izdelava LBS je potekala od junija 2008 do junija 2012 in je v obliki datoteke xml, skupaj s shemo DTD (ang. *Document Type Definition*)⁵¹ in XSD (ang. *W3C schema*), uporabnikom na voljo pod pogoji Creative Commons: Priznanje avtorstva-Nekomercialno-Deljenje pod enakimi pogoji.⁵² Na podlagi LBS je bil izdelan poskusni Spletni slovar slovenskega jezika, kot ena izmed možnih slovarskih vizualizacij podatkov, strukturiranih v slovarski bazi. Rešitve, ki so prikazane v Spletnem slovarju, niso vedno identične z rešitvami, prikazanimi v pričujočem besedilu, saj pomeni izdelava LBS v prvi vrsti določitev koncepta, strukture in vrste podatkov, ne pa že celotnega opisa slovenskega leksikalnega fonda. Marsikatera rešitev, razvidna iz LBS ali poskusnega Spletnega slovarja, je zato predstavljala šele izhodišče za debato in leksikografski problem, ki ga je bilo potrebno šele rešiti oz. določiti ustrezno strategijo in način obravnave. V tem smislu tudi nekatere leksikografske rešitve in postopki, prikazani v pričujoči monografiji, v obstoječi slovarski bazi niso izkazani ali pa se od navedenih razlikujejo.

3.1 Namen

Namen izdelave LBS je v Specifikacijah (Gantar et al. 2009: 15) opredeljen takole:

Oblikovanje leksikalne podatkovne baze za slovenščino v okviru projekta SSJ sledi dvema temeljnima ciljema: zadostiti želi leksikografskim potrebam, konkretnije izdelavi za slovenščino najpomembnejših eno- in dvojezičnih slovarjev in specializiranih jezikovnih priročnikov ter potrebam računalniške obdelave naravnega jezika oz. uporabi v jezikovnotehnoloških aplikacijah za slovenščino.

S podatki, strukturiranimi v LBS, naj bi bilo tako mogoče izdelati konkretne, za jezikovno skupnost najbolj strateško pomembne jezikovne priročnike, zlasti slovar sodobnega slovenskega jezika,⁵³ šolski slovar, dvojezične slovarje, posamezna poglavja slovnice ter specializirane slovarje, kot so frazeološki, kolokacijski, vezljivostni itd. Poleg tega naj bi bili podatki v bazi strukturirani na način, ki bi slovenski jezikovni skupnosti zagotavljal možnost uporabe v jezikovnotehnološke namene.

51 Shema DTD je elektronsko berljiv zapis zgradbe slovarskega gesla.

52 Obe datoteki sta na voljo na povezavi: <http://www.slovenscina.eu/spletni-slovar/prenos>.

53 Na podlagi podatkov, strukturiranih v LBS, je bil ob zaključku aktivnosti strokovni in splošni javnosti predstavljen Predlog za izdelavo Slovarja sodobnega slovenskega jezika (Krek et al. 2013), ki vključuje podroben opis postopkov izdelave spletno zasnovanega slovarja in podaja način izvedbe s tehničnega vsebinskega, organizacijskega ter finančnega vidika.

Pri tem je seveda razumljivo, da količina podatkov, ki jih je bilo mogoče vključiti v slovarsko bazo v obdobju njenega nastajanja, ne more že v celoti zadosti-
ti slovarskim in jezikovnotehnološkim potrebam jezikovne skupnosti, saj se od
sodobnega splošnega slovarja pričakuje sistematičen opis vsaj 100.000 lem oz.
vsaj 150.000 leksikalnih enot. LBS, kot je bila zasnovana v okviru projekta, želi
torej ugotoviti za svoje namene najustreznejše oz. najrelevantnejše tipe leksikal-
nogramatičnih podatkov za besede z leksikalnim pomenom ter najustreznejše
načine njihovega pridobivanja in strukturiranja. Nenazadnje se proces polnjenja
slovarske baze, če govorimo o izdelavi sodobnih jezikovnih priročnikov, ne more
v določenem obdobju zaključiti, pač pa potrebuje organizirano kontinuiteto, v
kateri se podatki nenehno posodablajo v povezavi z nadgradnjo korpusa in v
kateri poteka stalno spremljanje sprememb na ravni besedišča in pomenske di-
namike znotraj njega.

3.2 Glavni poudarki

Da bi bilo mogoče namen dosledno uresničiti, so bile pri oblikovanju LBS upošte-
vane sodobne leksikografske prakse, tako v izrabi leksikografskih orodij za analizo
korpusov kot pri leksikografskih opisih, ki sledijo potrebam sodobnega uporabni-
ka. Ker pa nismo želeli praks, ki se dobro obnesejo za tuje jezike, zgolj prenesti v
slovenski jezik, je bila posebna pozornost namenjena analizi za slovenščino speci-
fičnih leksikalnogramatičnih lastnosti in prilagoditvi jezikovnotehnoloških orodij
posebnostim slovenščine.

Znotraj uporabniškega dela LBS je bilo največ pozornosti namenjene pomenske-
mu opisu besed, pri čemer smo sledili dvema ciljema. Želeli smo opisati razme-
roma pogost del besedišča na podlagi realnega stanja v jeziku, in sicer povsem od
začetka, tj. neodvisno od rešitev v SSKJ.⁵⁴ Slovarska informacija v njem je bodisi
pomanjkljiva zaradi odsotnosti sodobnega konteksta, stilistične in žanrske vloge,
spremenjenih pomenskih tendenc, v nekaterih primerih pa celo napačna in zavaja-
joča (prim. Müller 2009: 19; 1996).

Naslednji izziv, ki smo ga želeli uresničiti, je upoštevanje dejstva, da pomenskega
opisa ni mogoče ločiti od skladenjskega obnašanja besede in da je potrebno – z
vidika možnosti, ki jih ponuja nov medij oz. narekujejo potrebe sodobnega uporab-
nika – ločiti izrazno podobo besede od njene (pomenske) vsebine.

⁵⁴ Pomenska razčlenitev in opisi leksike na novo, nimajo namena kritizirati pomenskih opisov v SSKJ, saj se v nekaterih primerih pomenska členitev in opisi precej približujejo, kar je pričakovano, pač pa je namen neobremenjeno izhajati iz sodobnega stanja slovenščine (prim. analizo posameznih primerov v Gantar 2014).

Tu smo sprejeli za slovenščino največ novih leksikografskih rešitev. Prva je pojmovanje leksikalne enote kot vsebinskega dela leksema in ne leksema kot celote. Tako v LBS pojmujeemo kot leksikalno enoto vsako prepoznavno pomensko vsebino, ki v leksikografski presoji zahteva pomenski opis.

Novosti, ki jih LBS prinaša v slovenski leksikografski prostor, nadalje izhajajo iz dejstva, da sta izbor podatkov in način organiziranja znotraj baze v celoti podrejena digitalni oz. spletni predstavitvi. Izhodiščno digitalna zasnova pa odločilno vpliva na vrsto ne samo praktičnih oblikovalskih, pač pa tudi jezikoslovnih odločitev, ki se nanašajo npr. na pojmovanje homonimije v odnosu do večpomenskosti, na izkazovanje hierarhičnih razmerij med iztočnico in njenimi pomeni, besedno-vrstnimi konverzijami, stalnimi zvezami in frazeološkimi enotami. Če je knjižna oblika slovarja zahtevala skrbne premisleke o urejenosti podatkov na posamezni strani v knjigi, ki se kaže npr. v koncentraciji oblikoskladenjskih in slovničnih podatkov ter podatkov o izgovoru v zaglavju iztočnice in v sistemu podgesel, se je bilo treba pri prenosu v spletno zasnovo slovarja osredotočiti predvsem na tipe leksikalnih enot, ki niso več le besede (leme) in večbesedni izrazi, ampak tudi posamezni pomeni. Digitalno okolje zahteva prehod od razumevanja organizacije slovarskih podatkov v ploskem enodimenzijskem prostoru k njihovi predstavitvi v večdimenzijskem prostoru, v katerem je potrebno definirati le vozlišča, na katera so pripeti določeni jezikovni podatki, saj jih je v spletnem okolju mogoče v vsakem trenutku določiti kot izhodišče uporabnikovih poizvedovanj oz. kot izhodiščno točko, od katere tečejo povezave do vseh drugih slovarskih vozlišč. Princip organiziranja, kjer je vsaka leksikalna enota definirana kot nosilka določenih leksikalnih, gramatičnih, normativnih ipd. podatkov, zahteva bolj kot opredelitev hierarhičnega razmerja do iztočnice, ki je le izrazna podoba ali nosilka enega ali več pomenov, del besedne zveze ali FE, premislek o tem, katere podatke je smiselno nanjo pripeti oz. katere med seboj povezati, da bo slovarska informacija relevantna, skladna s konceptualno zasnovo in empirično utemeljena. Tak princip, ki je v temelju sinhroni, je v LBS za razliko od knjižnega diahrono-sinhronega v celoti izpeljan z mislijo na spletni medij.

Za beleženje pomenskoskladenjskih lastnosti pomena, ki jih bo mogoče uporabiti pri nadaljnjih strojnih analizah, smo po vzoru že omenjenih projektov FN in CPA uvedli zapisovanje t. i. pomenskih shem, ki predstavljajo izhodišče za oblikovanje razlag stavčnega tipa. Za stavčne razlage je namreč značilno, da posnemajo naravni diskurz pojasnjevanja pomenov besed in da predstavljajo pomen besede v njenem naravnem življenjskem okolju, vključno s tipičnimi udeleženci in okoliščinami, ki jih ta zahteva. Ker so opisi postopkov oblikovanja stavčnih

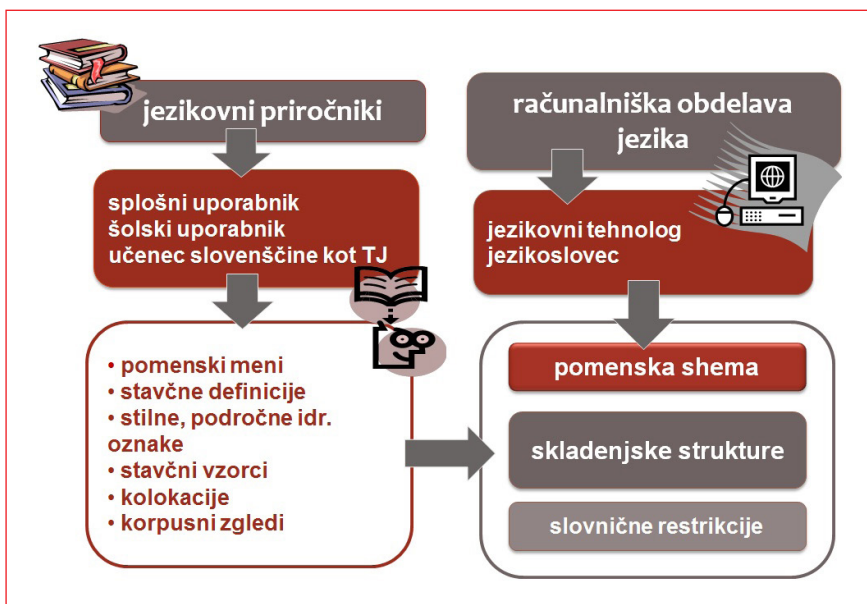
razlag v leksikografski literaturi na voljo predvsem za učenje angleščine kot tujega jezika, je bilo treba ob prenosu na slovenščino odgovoriti na vrsto jezikoslovno teoretičnih vprašanj, kot je denimo problem ujemalnih in anaforičnih konstrukcij znotraj stavčnih razlag, vključevanje pragmatičnih elementov, ubeseditiv odnosa udeležencev do vsebine izraženega pomena ipd., posledično pa razmisliti tudi o ustreznem tipu pomenske razlage in nenazadnje o terminološkem aparatu, saj obstoječe izrazje za nove leksikografske postopke in jezikovnotehnoška okolja še ne obstaja ali pa ne zadošča več.

Naslednji temeljni namen LBS – tj. izdelati opise leksikalnih enot na način, ki bo računalniško berljiv, ter s tem zagotoviti uporabnost LBS tudi za računalniško obdelavo naravnega jezika – je zahteval izbiro ustreznih orodij za pridobivanje in analizo relevantnih leksikalnogramatičnih podatkov, izbiro najprimernejšega zapisa podatkovne strukture in ustreznega programa za strukturiranje podatkov ter določitev načina njihove organiziranosti. Kot smo videli, so leksikalne baze večjih evropskih jezikov že v 90. letih omogočile izdelavo digitalnih leksikalnih virov, ki so bili nato s pridom uporabljeni pri izdelavi elektronskih slovarjev in leksikonov, njihov rezultat pa so jezikovni portali z različnimi tipi medsebojno povezanih jezikovnih informacij. Slovenščina se z aktivnostmi projekta SSJ, kamor sodi tudi izdelava LBS, vključuje na evropski vlak tako rekoč v zadnjem trenutku, kljub vsemu pa ji 20-letni zamik omogoča zgledovanje po dobrih praksah in hkrati izogibanje slepim ulicam, ki so jih bili deležni pionirski poskusi.

Nikoli prevečkrat poudarjena lastnost LBS je dejstvo, da vsi podatki, ki jih vsebuje, temeljijo na dejanski besedilni rabi, kot je zajeta v referenčnih slovenskih korpusih FidaPLUS (Arhar in Gorjanc 2007) in Gigafida (Logar et al. 2012). S tem je uporabniku zagotovljen realen jezikovni opis sodobne slovenščine, hkrati pa je na ta način mogoče detektirati probleme, ki izhajajo iz sodobne jezikovne rabe. Korpusi namreč odražajo tako regularne vzorce kot variabilnost in ustvarjalnost. Če hoče leksikalna baza zagotavljati kontinuiteto aktualnega jezikovnega opisa, potrebuje vedno znova posodobljeno podobo jezikovnega stanja, zato je posodabljanje korpusov ključnega pomena za oblikovanje slovarske informacije, ki bo zadostila potrebam uporabnika na dolgi rok. V okoliščinah, kjer je posodabljanje jezikovnih virov sistematično, je omogočen tudi hiter prenos aktualnih jezikovnih podatkov do uporabnika, kar postaja pomemben vidik učinkovitosti jezikovne informacije in celostne jezikovne infrastrukture jezika. Izraba sodobne tehnologije je pri tem v veliko pomoč, zato so trendi posredovanja jezikovnih informacij usmerjeni v celovite rešitve, ki ustrezajo tako globalni jezikovni infrastrukturi, npr. evropskih jezikov, kot potrebam konkretne jezikovne skupnosti in njej specifičnim jezikovnim problemom.

3.3 Uporabnik

Glede na izhodiščno dvonamenskost sta tudi potencialna uporabnika LBS in na njej temelječih priročnikov dva, in sicer *človek* in *računalnik*, kot prikazuje Slika 15.



Slika 15: Podatki v LBS glede na profil uporabnika

3.3.1 Človeški uporabnik

Pri zasnovi slovarske baze (še toliko pomembneje pri uporabi podatkov iz slovarske baze za slovar) je smiselno končnega uporabnika opredeliti predvsem glede na to, ali gre za odrasle ljudi ali za otroke in ali so podatki in njihovi opis namenjeni rojenemu govorcu ali učencu tujega jezika. Znotraj tega podatki, ki so v LBS namenjeni v prvi vrsti človeku – v širšem smislu bi jim lahko rekli tudi priročniški jezikovni podatki –, nagovarjajo štiri profile jezikovnih uporabnikov: splošnega, šolskega, učenca slovenščine kot tujega jezika in jezikoslovca, natančneje raziskovalca in učitelja.

Predvidevamo, da se profil splošnega jezikovnega uporabnika spreminja skladno z demografskimi in družbenopolitičnimi spremembami v določeni jezikovni skupnosti. Če upoštevamo podatek, da se več kot 60 % slovenske populacije po

končani srednji šoli ali gimnaziji vpiše na fakulteto,⁵⁵ je povprečni potencialni splošni uporabnik sodobnega enojezičnega slovarja slovenskega jezika v obdobju zadnjih 10 let materni govorec slovenščine z najmanj srednješolsko izobrazbo, za katerega lahko predvidevamo, da se pri svojem poklicnem ali vsakdanjem delu ukvarja z besedilno produkcijo in/ali interpretacijo. Predvidevamo tudi, da je informacija, ki jo pričakuje od slovarja, predvsem pomenski opis besed in zvez, ki so bodisi v slovenskem besedišču nove ali še neuveljavljene,⁵⁶ ter besed, ki so sicer specifične za določena specializirana področja, ki pa pod vplivom družbenopolitičnih, gospodarskih in kulturnih okoliščin in zlasti prek informativnih in stvarnih besedil ter sodobnih množičnih medijev, zlasti spleta, opazneje prehajajo v splošni jezik. Od splošnega uporabnika je namreč – glede na starostni in izobrazbeni profil – mogoče pričakovati, da je z osnovnim besediščem že seznanjen, saj je ob koncu srednješolskega izobraževanja usvajanje temeljnega besedišča že bolj ali manj zaključeno (prim. Rozman 2010: 21).⁵⁷ Zato je mogoče predvideti, da splošnega uporabnika glede osnovnega besedišča ne zanima toliko ustaljena pomenska vrednost, ampak predvsem ustrezna raba besede oz. njenega pomena v besedilnem kontekstu, zlasti glede na področje rabe, žanr ali stil, ter nekatere skladenjske in slovnično-pomenske lastnosti, kot so vezljivost, števnost, tipična raba določenih besednih oblik ipd.

LBS znotraj splošnega profila nagovarja predvsem uporabnika, ki ga zanimajo pomenske lastnosti osnovnega besedišča z vidika sodobne jezikovne rabe, vključno s tipičnim pomenskim in skladenjskim kontekstom (kolokacije, stalne zveze, stavčni vzorci) ter s tem povezane oblikoskladenjske in pomenske lastnosti. Za splošnega uporabnika je po predvidevanjih in glede na izkazane jezikovne probleme manj zanimivo besedišče z ustaljeno pomensko vrednostjo.

Šolski uporabnik in **učenec slovenščine kot tujega jezika** zahtevata vsak svoje ne le specifično besedišče, pač pa tudi vrsto, strukturiranost in prikaz informacij, zato ju kljub skupnemu izobraževalnemu imenovalcu ni mogoče dajati v isti koš. Leksikalna baza, ki navadno predstavlja vir leksikalnih podatkov za več različnih slovarskih priročnikov, mora tako vključevati tudi širok spekter leksikalnogramatičnih

55 Za študijsko leto 2008/09 je bil delež dijakov, ki so se vpisali v višje ali visokošolske študijske programe, celo okoli 96%-en. Po podatkih Andragoškega centra Slovenije (<http://www.acs.si/InfO-mozaiik>) za leto 2010 je imelo med aktivnim prebivalstvom Slovenije srednješolsko izobrazbo 60,3 % prebivalcev, 24,8 % je imelo dokončano nižjo ali srednjo poklicno izobrazbo, višješolsko ali visokošolsko izobrazbo pa 18,8 % odraslih.

56 To potrjuje tudi pregled realnih jezikovnih problemov, ki je bil izdelan v okviru projekta SSJ za namene izdelave Slogovnega priročnika (Dobrovoljc in Krek 2011: 91).

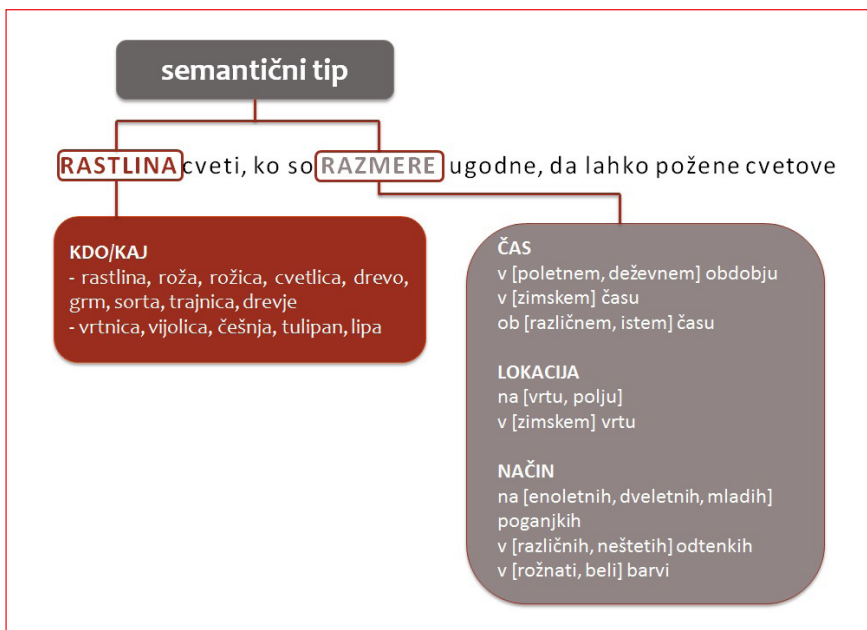
57 Usvajanje jezika sicer poteka vse življenje, saj govornici usvajamo tudi besede, zveze in pomene ter njihovo gramatično okolje, ki v jezik prihajajo na novo.

podatkov in hkrati omogočati selekcijo in predvidevati različne končne prikaze, saj so si potrebe različnih uporabniških profilov v nekaterih segmentih celo nasprotujoče, npr. podrobnost in zahtevnost pomenskih opisov, večji oz. manjši delež slovničnih, pragmatičnih in drugih informacij. Kljub temu pa je z mislimi na različne profile uporabnikov in njihove potrebe mogoče v posameznih segmentih slovarsko bazo zasnovati tako, da lahko zadosti različnim uporabniškim potrebam vsaj z nekaterimi podatki, ki so vanjo vključeni.

Z mislimi na šolskega uporabnika smo upoštevali besedišče aktualnih srednješolskih učbenikov ter prilagodili pomenski opis v obliki kratkih pomenskih indikatorjev, ki tvorijo pomenske menije za lažjo navigacijo po večpomenskem geslu, ter v obliki stavčnih razlag, ki zlasti pri glagolih prikazujejo celotni stavčni vzorec, z vsemi udeleženci in okoliščinami, ki jih potrebuje konkretni glagolski pomen. Take rešitve upoštevajo tudi slovarji za učenje angleščine kot tujega jezika.

Vključitev kolokacij pri vsaki registrirani leksikalni enoti, torej pri pomenih, stalnih zvezah in frazeoloških enotah, je v LBS namenjena vsem uporabniškim profilom. Z njimi je mogoče prikazati naravno umestitev obravnavane besede oz. njenega pomena v kontekst, kar uporabniku olajša odločanje pri izbiri za njegov sporočanski namen najustreznejše kombinacije pri tekstni produkciji, pa tudi pri prepoznavanju konteksta in rabe (do sedaj) še neznane besede ali zveze, ko gre za tekstno recepcijo in interpretacijo. V bazo vključene kolokacije in realni korpusni zgledi najboljše prikazujejo besede v njihovem realnem besedilnem okolju, kar je zlasti pomembna informacija za učence slovenščine kot tujega jezika.

Poleg podatkov, ki so v LBS namenjeni splošnemu in šolskim uporabnikom, so v njej strukturirani podatki lahko tudi izhodišče za nadaljnje analize. Pri oblikovanju pomenskih shem smo za pomensko-skladenjske analize uvedli zapisovanje udeležencev na predvidenih vezljivostnih mestih konkretnih glagolskih pa tudi pridevniških in nekaterih samostalniških pomenov s t. i. semantičnimi tipi. Ti predstavljajo skupna imena za konkretne leksikalne zapolnitve, realizirane v obliki kolokacij na posameznih vezljivostnih mestih. Na ta način smo, z možnostjo nadaljnjega procesiranja v bazi strukturiranih podatkov, jezikoslovcem in jezikovnim analitikom omogočili povezavo med prototipičnim stavčnim vzorcem besednega pomena in izpostavljenimi udeleženci ter okoliščinami, izraženimi v obliki kolokatorjev. Semantični tipi lahko tvorijo semantične mreže, povezljive z obstoječimi semantičnimi mrežami, kot je denimo WordNet, in predstavljajo izhodišče za pomensko označevanje korpusov. Primer možnosti avtomatskega pridobivanja in kombiniranja v LBS strukturiranih podatkov, ki je namenjeno pomensko-skladenjskim analizam, prikazuje Slika 16.



Slika 16: Povezava med pomenskimi in skladenjskimi podatki v LBS

Organizacija pomenskih, skladenjskih, slovničnih in kolokacijskih podatkov v LBS na način, ki ga bomo podrobneje opisali v poglavju 4.1 Strukturiranost podatkov in v tretjem delu monografije, omogoča torej tudi specializiranemu uporabniku avtomatsko pridobivanje in kombiniranje vseh vrst informacij, ki so nujno potrebne za podrobne semantične in skladenjske analize. Ker je celotna slovarska baza zgrajena na podlagi korpusnih podatkov, odražajo pridobljene informacije realno stanje sodobne pisne slovenščine in so osnova za pripravo učnih gradiv kot tudi za posodabljanje in prevrednotenje obstoječih jezikovnih priročnikov.

3.3.2 Jezikovnotehnoška skupnost

Kot je bilo že večkrat poudarjeno, je leksikalna baza računalniško berljiv jezikovni vir, v katerem je vsak besedni pomen identifikacijsko povezan z vrsto specifičnih leksikalnih in skladenjskih podatkov v svoji besedilni okolici. Različni tipi oznak in slovnične restrikcije omogočajo pomensko razdvoumljanje pri razumevanju besedila in ustrezno izbiro jezikovnih sredstev pri njegovem

tvorjenju. Podatke v LBS je mogoče povezati z drugimi bazami podatkov in jih uporabiti pri avtomatskem luščenju informacij iz besedil, oblikovanju sistemov odgovorov na vprašanja, v avtomatskih prevajalnih sistemih in aplikacijah, ki temeljijo na bazah procesljivih jezikovnih podatkov. Uporabiti jih je mogoče za avtomatsko označevanje slovenskih besedil na oblikoskladenjski, skladenjski in pomenski ravni ter za izboljšanje orodij, kot je npr. skladenjski razčlenjevalnik in označevalnik za slovenščino. Skladske strukture, ki so formalizirane dvo- ali tribesedne besedne zveze, predstavljajo ključno informacijo za avtomatsko pridobivanje kolokacijskih podatkov iz korpusov in z njimi povezanih korpusnih zgledov, kar pospeši postopek izdelave slovarske baze oz. slovarja in omogoča prenos podatkov do uporabnika v različnih redakcijskih fazah, ki se na spletu sproti posodablja.

3.4 Obseg

LBS vsebuje podatke o pomenskih, stilnih, skladskih, slovničnih, kolokacijskih, frazeoloških in drugih lastnostih dela besedišča sodobne slovenščine. Ob zaključku aktivnosti je obsegala 2.500 iztočnic oziroma 10.946 leksikalnih enot, tj. pomenov, podpomenov, stalnih zvez in frazeoloških enot. Obseg je bil določen na podlagi presoje, da slovarska obdelava 2.500 lem lahko zagotovi zadostno količino različnih leksikalnih in leksikografskih problemov, na podlagi katerih bi bilo mogoče določiti vrste leksikalnogramatičnih podatkov in način njihove notranje organizacije, ki bi učinkovito delovala tudi na obsežnejšem izboru. Številčno zastopnost posameznih tipov leksikalnih enot in drugih podatkov v LBS prikazuje Preglednica 2.

Preglednica 2: Številčna zastopnost leksikalnih enot in drugih elementov v LBS

Iztočnice	2.500	Leksikalne enote	10.946	Druge slovarske enote	
				kolokacije	44.626
samostalniki	1.288	pomeni	4.371	razširjene kolokacije	4.602
glagoli	528	podpomeni	3.076	skladske zveze	8.298
pridevniki	546	stalne zveze	2.053	stavčni vzorci	7.151
prislovi	138	frazeološke enote	1.446	zgledi	152.996
				oznake	1.197
				slovnice	716
				omejitve	

V LBS so besednovrstno zastopane samo »leksikalne« besede ali besede, ki so nosilke leksikalnih lastnosti, zlasti pomena,⁵⁸ torej samostalniki, glagoli, pridevniki in prislovi. LBS ne vsebuje slovnčnih ter funkcijskih besednih vrst, lastnih imen in kratic kot samostojnih iztočnic. Predlogi in vezniki so v bazo vključeni posredno, tj. kot deli skladenjskih struktur, stavčnih vzorcev in kolokacij, zato jih je vedno mogoče iz baze pridobiti naknadno, in sicer v povezavi s konkretno leksikalno enoto. Na enak način – torej z možnostjo avtomatskega pridobivanja – so v bazi zastopani členki, med pridevniškimi besedami tudi števniki, medtem ko medmetov in zaimkov v kolokacijah nismo beležili, so pa lahko zastopani kot deli frazeoloških enot, vendar v tem primeru niso označeni s samostojno XML kodo, zato jih iz baze ni mogoče pridobiti avtomatsko.

Kot omenjeno, LBS na ravni iztočnic ne vsebuje lastnih imen, kar opišemo v samostojnem poglavju, lahko pa so v nekaterih primerih lastna imena v bazo vključena kot stalne zveze ali njihovi deli, kot kolokatorji in kot sestavni elementi frazeoloških enot.

58 Namenoma se izogibamo izrazoma polno- ali predmetnopomenske besede, saj ne odražata semantičnega koncepta, kot ga upoštevamo v monografiji. V zvezi s tem prim. poglavji o pomenski členitvi in pomenskem opisu.

4 Zgradba Leksikalne baze za slovenščino

Izbor in še prej strukturo relevantnih leksikalnogramatičnih podatkov narekujejo v LBS njen namen, uporabnik in zasnovanost za spletni medij. Pri izboru in notranji organiziranosti smo zato težili k ureditvi, po kateri si podatki sledijo od abstraktnega k vse bolj konkretnemu, kar pomeni, da so najbolj abstraktne informacije na pomenskem in skladijskem nivoju in da se konkretizirajo od kolokacijskega k ponazarjalnemu nivoju. Prva dva nivoja, z izjemo indikatorjev in pomenskih razlag, sta zato primarno namenjena jezikoslovnemu uporabniku in nadaljnjemu procesiranju, kjer je mogoče podatke na novo organizirati ter dopolnjevati z drugimi bazami podatkov. Na preostalih nivojih, tj. na kolokacijskem in ponazarjalnem, pa so informacije oblikovane tako, da jih je mogoče razumeti tudi brez poznavanja posebnih leksikografskih ali slovničnih konvencij.

Tako ureditev podpira tudi semantično izhodišče, na podlagi katerega so podatki na posameznem nivoju vedno podrejeni vsakemu posameznemu registriranemu pomenu in ne npr. izraznim ali skladijskim lastnostim besede.

4.1 Strukturiranost podatkov

Pri organizaciji podatkov v bazi, ki je namenjena elektronski oz. spletni postavitvi, uporabnikovo poznavanje slovarske makro- in mikrostrukture, čeprav seveda tudi spletni slovar zahteva notranjo urejenost, ki omogoča konsistentnost in medsebojno povezanost podatkov, ni več potrebno. Pri spletni zasnovi slovarja je fokus s prikazovanja hierarhičnih razmerij med iztočnico in leksikalnimi enotami, ki iztočnico vsebujejo, premaknjen na leksikalno enoto samo in na odločitev o tem, kaj vse je oz. ni leksikalna enota, katere podatke nanjo pripeti in kako jih prikazati z upoštevanjem možnosti elektronskega medija, kot so navigacijske strategije, povezave znotraj in zunaj slovarske baze, pojavna okna, večpredstavnost ipd.

LBS je oblikovana kot mreža med seboj povezanih pomenskih in skladijskih podatkov o posamezni besedi⁵⁹ na šestih nivojih, med katerimi so najpomembnejši pomenski, skladijski in kolokacijski. Kot prikazuje Slika 17, je hierarhično najvišja lema, tj. iztočnica v osnovni obliki, ki zastopa vse pripadajoče leksikalne enote, tj. pomene in podpomene, stalne zveze in frazeološke enote.

⁵⁹ V celotnem besedilu uporabljamo izraz beseda, kadar imamo v mislih splošni pomen, tj. izrazno enoto ali iztočnico. Za razliko od besede uporabljamo izraz leksikalna enota predvsem takrat, ko želimo poudariti pomenske enote, ki jih obravnavamo v LBS, tj. posamezne pomene, stalne zveze in frazeološke enote.

Na pomenskem nivoju smo beležili pomensko razčlenjenost obravnavane besede v iztočnici (osnovne pomene in podpomene). Vsak pomen smo opisali s kratkim pomenskim indikatorjem, ki oblikuje pomenski meni gesla, in/ali razlago. Drugi del pomenske informacije predstavlja pomenska shema v obliki stavčne razlage. S pomenskimi shemami smo pri glagolih, samostalniki in pridevniki beležili udeležensko zgradbo posameznega pomena. Udeležence in okoliščine smo v pomenski shemi povezali s pomočjo t. i. pomenskega scenarija v smiselno celoto, ki odraža sporočanje situacijo, npr. za glagol grmeti v pomenu 'razburjati se; kritizirati': če rečemo, da ČLOVEK grmi, želimo povedati, da zelo glasno izraža svojo jezo ali negativno mnenje o nekem ČLOVEKU, DOGAJANJU ali TEMI. Konkretna realizacija pomenske sheme predstavljajo posamezni deli stavčnega vzorca: kdo grmi | s česa | o čem | nad kom/čim, ki se potrjujejo s kolokacijami. Pomenska shema s prikazanimi udeleženskimi mesti in tipičnimi leksikalnimi zapolnitvami tako predstavlja povezovalno med pomenom in skladenjskimi pogoji za njegovo realizacijo.

Čeprav so skladenjski podatki vključeni že v pomensko shemo, jim je posebej namenjen še skladenjski nivo. Na tem nivoju so za vsak registrirani pomen ali podpomen zabeležene skladenjske strukture, kot skladenjske podlage kolokacijam, in skladenjske zveze, pri glagolskih iztočnicah pa tudi stavčni vzorci.

Stalne zveze in frazeološke enote so obravnavane kot leksikalne enote, zato prve vključujemo pod posamezni pomen ali podpomen besede v iztočnici, če je pomensko razmerje izkazano, druge pa na konec geselskega članka v samostojni razdelek. Tako stalne zveze kot frazeološke enote predvidevajo vse ključne elemente pomenskega opisa: indikator, razlago, kolokacije in zglede.

Vloga korpusnih zgledov je ponazoriti in potrditi vse predhodne podatke ter hkrati pokazati obnašanje leksikalne enote v njenem najbolj naravnem besedilnem okolju. Narava jezikovnih podatkov, kot je izkazana v realni rabi, vedno omogoča, ne glede na izbran sistem urejanja, več različnih interpretacij in hierarhičnih ureditev ter možnosti hkratnega uvrščanja v več različnih kategorij. Pri strukturiranju iz korpusa pridobljenih jezikovnih podatkov v leksikalno bazo smo zato predvideli možnost obravnave posameznih jezikovnih podatkov na različnih nivojih leksikalne zgradbe hkrati, npr. kot kolokacije in kot stalne zveze, kot samostojne pomene in – skupaj s tipičnim besedilnim okoljem – kot frazeološke enote. Na ta način smo uresničevali vodilo, da je potrebno uporabljati sistem leksikalne baze za čim bolj smiselno uvrstitev jezikovnega podatka in ne prilagajati podatkov sistemu leksikalne baze za vsako ceno. Hkrati smo se jasno zavedali tudi nasprotnega vidika: strukturiranje jezikovnih podatkov vedno zahteva določeno brušenje robov in posploševanje, kar je nujno potrebno za omejevanje živega jezika s slovarskimi okvirji.

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 je ali 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	Bili so ravno v restavraciji, ko je začelo grmeti.	Grmel je z govorniškega odra, dokler mikrofona ni nenadoma umolknil.
• stalna besedna zveza			
VI. FRAZEOLOGIJA	• frazeološka enota • indikator	Če malega travna grmi, slane se kmet več ne boji. <i>pregovor</i>	

Slika 17: Nivojska zgradba LBS

4.2 Shema DTD

Za potrebe LBS je bila izdelana shema DTD, ki je prilagojena delu v slovarskih urejevalnikih, kot so na primer DPS – Dictionary Writing System francoskega podjetja IDM,⁶⁰ iLex danskega podjetja EMP,⁶¹ Tlex, podjetja TshwaneDJe iz Južne Afrike,⁶² Lingvo.Pro,⁶³ ruskega podjetja ABBYY in podobnih. Shema omogoča prilagoditev informacij v smislu dodajanja in spreminjanja kategorij podatkov in hierarhičnih odnosov za različne tipe slovarskih priročnikov.

Da bomo v nadaljevanju lažje sledili opisu in vrsti leksikalnogramatičnih podatkov v LBS, si je potrebno pogledati nekatere izraze, ki so uporabljeni v shemi. Celotni geselski članek je sestavljen iz dveh delov: glave in gesla. V glavi geselskega članka lahko pričakujemo osnovne oz. tehnične podatke o obravnavani besedi, kot je njen zapis v bazi, zapis osnovne oblike, besedna vrsta, oblikoslovni podatki in podatek

60 DPS: <http://dps.cw.idm.fr/>.

61 iLex: http://issuu.com/jens.erlandsen/docs/ilex_brochure_120dpi/2.

62 Tlex: <http://tshwanedje.com/tshwanelex/>.

63 Abby: <http://lingvopro.abbyyonline.com/en/Home/Authors>.

o frekvenci. Vsebinski del gesla – tj. leksikalnogramatični podatki, ki so predvideni pri vsakem pod(pomenu), sledijo v drugem delu gesla, ki ga tako tudi imenujemo: geslo. Pomemben del geselskega članka, ki ni del sheme DTD, ampak je iz nje avtomatsko generiran, je t. i. pomenski meni, ki je zgrajen iz kratkih pomenskih opredelitev pomenov in podpomenov – indikatorjev, in omogoča hitro navigacijo po celotnem geslu z možnostjo izbire posameznega pomena ali podpomena.

Preglednica 3 opisuje vsebino posameznih elementov sheme v formatu XML in zapis elementa v bazi. V vsebinskem delu je na kratko opisan tip leksikalnogramatičnega podatka, namen opisa in razmerje glede na druge podrejene ali nadrejene podatke v shemi. Obarvana polja označujejo hierarhično višje elemente, ki vsebujejo različne podelamente in/ali attribute.

Preglednica 3: Opis sheme DTD v LBS

Vsebina LBS	Element DTD
Vsebuje celotni geselski članek z elementoma glava in geslo .	<clanek>
Vsebuje elemente oblika in zaglavje .	<glava>
Vsebuje elemente zapis , korpus ter iztočnica .	<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>
Znotraj elementa oblika: vsebuje zapis iztočnice v obliki leme.	<iztocnica></iztocnica> </oblika>
Vsebuje elemente: besedna vrsta in oznaka .	<zaglavje>
Vsebuje pomenski meni, ki ga generirajo indikatorji pri posameznih (pod)pomenih.	<senseToc> <deeplink>diocerf="besedilo" senseref="> <deeplink>diocerf="besedilo" senseref="> </senseToc>
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, konotacijske, registrske in slovnične posebnosti.	<oznaka tip="attribute"></oznaka> </zaglavje> </glava>

Vsebina LBS	Element DTD
Vsebuje elemente, ki so določeni kot leksikalne enote . Sem sodijo: pomeni s podpomeni, stalne zveze in frazeološke enote .	<geslo>
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 zvezo, v kateri se pojavlja beseda v konkretnem pomenu.	<ustaljena_oblika></ustaljena_oblika>
Znotraj elementov pomen ali podpomen: vsebuje opredelitev pomena glede na področje rabe, konotacijske, registrske in slovnične posebnosti.	<oznaka tip="attribute"></oznaka>
Znotraj elementov pomen ali podpomen: vsebuje argumentno zgradbo konkretnega pomena, zapisano v obliki stavčne razlage 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 razlago, ki na uporabniku prijazen način ubesedi osnovne – na podlagi korpusnih rab – ugotovljene pomenske tendence konkretnega pomena. Za morebitne različne poskuse izdelave razlag na podlagi pomenskih shem, pa tudi z vidika možne povezave z drugimi bazami podatkov, sta predvidena dva samostojna elementa za razlage.	<definicija1></definicija1> <definicija2></definicija2>
Vsebuje element skladenjske strukture s pripadajočimi kolokacijami, vzorci in korpusnimi zgledi.	<skladenjske_skupine>

Vsebina LBS	Element DTD
Vsebuje obvezno vsaj eno skladenjsko strukturo in korpusne zglede . V večini primerov tudi kolokacije in vzorke .	<skladenjska_struktura>
Znotraj elementa skladijska 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 velikimi črkami.	<struktura></struktura> </skladenjska_struktura>
Znotraj elementa struktura: vsebuje opozorilo o slovnični omejitvi elementa v strukturi, npr. na določeno število, glagolsko obliko ipd.	<r></r>
Znotraj elementa skladijska struktura: pri glagolskih iztočnicah vsebuje lahko več zaporednih realizacij prototipičnega vezljivostnega vzorca, izraženega v pomenski shemi.	<vzorec></vzorec>
Vsebuje vsaj eno kolokacijo in pripadajoče zglede. V večini primerov tudi razširjene 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><k></k></r_kolokacija> </kolokacije> </skladenjska_struktura>
Znotraj elementov skladijske skupine, skladijske zveze, stalne zveze in frazeološke enote. Vsebuje vsaj en obvezen zglede .	<zgledi>
Znotraj elementa zgledi: lahko vsebuje več zaporednih korpusnih zgledov, ki potrjujejo predhodne kolokacije, razširjene kolokacije in vzorce. Znotraj zgleda je v elementu <i> v krepkem tisku izpisana iztočnica.	<zgled><i></i></zgled> </zgledi> </skladenjske_skupine>
Vsebuje vsaj en element skladenjska zveza s pripadajočimi zvezami, kolokacijami, vzorci in korpusnimi zgledi.	<skladenjske_zveze>
Vsebuje obvezno vsaj eno zvezo in pripadajoče zglede . Redkeje tudi kolokacije, in vzorce.	<skladenjska_zveza>

Vsebina LBS	Element DTD
Znotraj elementa skladijska	
zveza: vsebuje zapis pomensko prozornih, strukturno ustaljenih besednih nizov, pogosto s semantično ali/in oblikovno predvidljivim prostim mestom v elementu <k></k> .	<zveza><k></k></zveza> </skladijska_zveza> </skladijske_zveze>
Znotraj elementa pomen: vsebuje elemente, ki opredeljujejo posamezni podpomen (razen elementa <pomen>). Obvezno vsebuje elementa indikator in pomenska shema .	<podpomen></podpomen> </pomen>
Vsebuje vsaj en element stalna zveza .	<stalne_zveze>
Vsebuje obvezno element zveza , indikator , struktura in zgledi . V nekaterih primerih tudi kolokacije in razširjene kolokacije.	<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> </stalna_zveza> </stalne_zveze>
Vsebuje vsaj en element frazološka enota .	<frazoloske_zveze>
Vsebuje obvezno element enota in indikator . Pogosto tudi kolokacije, zglede in oznake.	<frazoloska_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> </frazoloska_enota> </frazoloske_zveze> </clanek>

Stisnjena datoteka z Document Type Definition (DTD) in W3C schema (XSD), ki definirata formalno strukturo LBS v formatu XML, je na voljo na povezavi, ki se nahaja na spletni strani: <http://www.slovenscina.eu/spletni-slovar/prenos>.

5 Orodja za pridobivanje, analizo in urejanje leksikalnogramatičnih podatkov

V postopku snovanja in gradnje LBS je bilo potrebno določiti način pridobivanja podatkov, ki bi bil učinkovit glede na namen leksikalne baze, hkrati pa ne preveč zamuden in zapleten za izdelavo ter seveda izvedljiv v okviru projektnih finančnih postavk. Že v izhodišču je bilo jasno, da bomo kot temeljni vir leksikalnih informacij uporabili referenčni in oblikoskladenjsko označeni korpus FidaPLUS ter da bomo ob zaključku gradnje korpusa Gigafida, ki je bila prav tako predvidena v okviru projekta SSJ (Logar in Šuster 2009), prešli na nov 1,2-milijardni slovenski referenčni korpus. Drugo izhodišče, ki je bilo znano že pri zasnovi, je bilo izraba orodja Sketch Engine (Kilgarriff et al. 2004) za analizo korpusnih podatkov. Uporaba besednih skic in nekaterih drugih možnosti, ki jih ponuja orodje, je bila namreč že preizkušena na korpusu in uporabljena pri izdelavi leksikalne podatkovne zbirke za izdelavo dvojezičnih slovenskih-x slovarjev pri založbi DZS (Gorjanc et al. 2005). Nadalje je bilo treba sprejeti zelo pomembno odločitev o najprimernejšem slovarskem orodju za urejanje pridobljenih podatkov, saj smo želeli omogočiti hkratno izdelavo gesel več leksikografom ter imeti ustrezen uredniški nadzor nad leksikografskim delom, vključno z možnostjo naknadnega pridobivanja in analiziranja v bazi strukturiranih podatkov, prenosa na spleto in povezovanja z drugimi bazami jezikovnih podatkov.

V nadaljevanju opišemo vire ter celotno tehnično ozadje izdelave LBS. Najprej oba referenčna korpusa ter način pridobivanja podatkov iz njih s poudarkom na avtomatizaciji postopkov, nadalje postopke analize ter slovarska orodja, ki smo jih uporabljali za analizo podatkov in pri njihovem organiziranju znotraj slovarske baze.

5.1 Pridobivanje in analiza leksikalnogramatičnih podatkov

V digitalni dobi si izdelave kateregakoli jezikovnega priročnika ni mogoče zamisliti brez obsežnega referenčnega in jezikoslovno označenega besedilnega korpusa. To potrjuje dolgoletni razvoj gradnje in izrabe korpusov, ki so od začetkov v 60. letih, kamor štejemo pionirske korpuse, kot je denimo korpus Brown,⁶⁴ doživeli vrsto korenitih in za razvoj jezikoslovja pomembnih sprememb. V izredno hitro napredujočem tehnološkem razvoju je z njim rastoča jezikovnotehnološka skupnost spoznala, da je obvladovanje resnično obsežnih jezikovnih virov, če ne želi

64 Vir: Korpus Brown: http://www.dh.informatik.uni-erlangen.de/IMMD8/Services/sammlung_korpora/Brown.html.

računalniške obdelave jezikovnih podatkov povsem prepustiti velikim korporacijam, kot je na primer Google, neizogibno. Kot poročata urednika zbornika prispevkov z delavnice Web as Corpus Baroni in Kilgarrieff (2006), je bilo že leta 2005 eno najbolj odmevnih predavanj na konferenci Združenja za računalniško jezikoslovje predavanje Franza-Josefa Ocha, v katerem je predstavil obetavne statistične rezultate strojnega prevajanja, ki so temeljili na 200-milijardnem korpusu v lasti Googla. Preprosto dejstvo, večja kot je količina podatkov, boljši so rezultati, je tako postalo smernica za vse, ki se kakorkoli ukvarjajo s procesiranjem jezikovnih podatkov in njihovo uporabnostjo v digitalni dobi.

Ogromna količina jezikovnih podatkov, ki jo obsežni besedilni korpusi prinašajo, pa je na področje jezikoslovja in zlasti leksikografije, ki nas v tem kontekstu zanima, vnesla še druge spremembe. Med pomembnejšimi je ta, da jezikovnih podatkov iz tako obsežnih elektronskih besedilnih zbirk ni več mogoče pridobivati in analizirati brez pripomočkov, ki omogočajo prepoznavanje jezikovne regularnosti in posebnosti (prim. Čermák 2009: 26) in ki jih znajo posredovati v obliki, ki je razmeroma lahko uporabna za jezikoslovca oz. leksikografa. Še do nedavnega povsem zanesljiva analiza konkordančnega niza nas danes pri analizi besede, ki ima naprimer več kot 100.000 pojavitev v korpusu, vrne tako rekoč na stopnjo pregledovanja listkovne kartoteke, kjer je bilo hkrati mogoče pregledati toliko listkov, kot je omogočala prostornost delovne mize. To ima posledice za celotni leksikografski proces, v katerem se leksikografova vloga tesno povezuje z jezikovnotehnološkim znanjem in poznavanjem tehnoloških zmožnosti sodobnih komunikacijskih naprav.

Hkrati je pospešen razvoj obsežnih elektronskih besedilnih zbirk in orodij za njihovo analizo občutno skrajšal čas oblikovanja leksikografske informacije in pot do končnega uporabnika ter s tem pocenil celotni postopek izdelave slovarja. Jezikovne informacije so se v okoliščinah naglo razvijajoče digitalne družbe preselile na splet in pod vplivom prepričanja, da predstavlja jezik skupno dobrino in da je kot tak last celotne jezikovne skupnosti (prim. Grønvik in Ore 2013), postale prosto dostopne. Velik doprinos k temu so naredile evropske iniciative in projekti, kot npr. Mondilex,⁶⁵ Meta-Net,⁶⁶ Clarin,⁶⁷ Dariah⁶⁸ ipd. ter napredne jezikovne politike v državah, kot so Nizozemska, Danska in Estonija, ki so razumele, kako pomembna za razvoj in ohranjanje jezikovne samobitnosti je opremljenost jezikovne skupnosti s prosto dostopnimi jezikovnimi viri in priročniki.

65 Mondilex: <http://www.mondilex.org/index.php?page=home>.

66 Meta-Net: <http://www.meta-net.eu/meta/about>.

67 Clarin: <http://www.clarin.nl/page/about/2>.

68 Dariah-eu: <https://www.dariah.eu/>.

5.1.1 Korpusa FidaPLUS, Gigafida in geslovník

Opis leksikalnih enot v LBS temelji na interpretaciji podatkov, pridobljenih iz referenčnih besedilnih korpusov FidaPLUS⁶⁹ in Gigafida.⁷⁰ Oba korpusa imata svoja lastna konkordančnika, ki omogočata iskanje po korpusu. Oba korpusa sta tudi vgrajena v orodje Sketch Engine (SkE), ki smo ga uporabljali kot osnovno orodje pri pridobivanju in interpretaciji jezikovnih podatkov iz korpusa.

Ob začetku izdelave LBS sta bili na podlagi korpusa oblikovani dve listi besed, t. i. širši in ožji geslovník. Širši geslovník je bil izdelan na podlagi 620-milijonskega referenčnega korpusa FidaPLUS v obsegu 5000 najpogostejših leksikalnopomenskih lem, ožji geslovník, 2500 lem, pa je bil uravnotežen glede na zastopanost posamezne besedne vrste v dejanski jezikovni rabi, kot prikazuje Preglednica 4, in je služil kot merilo za uravnoteženje leksikalne baze v posameznih fazah njenega nastajanja. Predviden odstotek lem znotraj besedne vrste omogoča primerjave z dejansko realizacijo ob zaključku aktivnosti, kot je prikazana v zadnjem stolpcu.

Preglednica 4: Zastopanost besednih vrst v LBS

	Odstotek	Število	Realizacija (število)
Samostalnik	40	1.000	1.288
Glagol	30	750	528
Pridevnik	25	625	546
Prislov	5	125	138
Skupaj	100	2.500	2.500

Poleg tega je bil izdelan učbeniški podkorpus, ki je obsegal osnovno- in srednješolska učbeniška besedila, vključena v korpus Gigafida. Iz omenjenega korpusa je bil nato izločen seznam s približno 1000 lemmi, kjer smo poleg frekvence upoštevali še večpomenskost, npr. strokovne izraze, ki imajo tudi splošnejši pomen (*aplikacija, izolirati, bazičen, toleranca*), strokovne izraze, ki prehajajo v splošni jezik (*mandat, recesija, biomasa*), nadalje splošno uporabljane pogoste prevzete besede (*subjektiven, sterilen, suveren, trivialen, ambicija*), pa tudi besede, ki so za šolskega uporabnika zanimive z vidika učnih vsebin (*ikona, tranzicija, sakralen, humanističen, emancipacija*), besede, ki poimenujejo novo predmetnost (*reciklirati, habitat, domena, naravovarstvenik*), in besede, ki so bile po našem mnenju zanimive za šolajočo se mladino z vidika generacijske pripadnosti (npr. *adolescenca, občevati, legalizirati, intimen, frustracija*).

⁶⁹ FidaPLUS: <http://www.fidaplus.net/>.

⁷⁰ Gigafida: <http://www.gigafida.net/>.

Posamezne leme so bile z vseh treh seznamov izbrane v slovarsko obdelavo na podlagi presoje avtorja ali urednika, pri čemer je bila odločitev pogojena z oceno težavnostne stopnje in časa, ki bi ga bilo potrebno nameniti za izdelavo gesla, ter s stopnjo usposobljenosti avtorja in količino izdelanih gesel v okviru projekta. Dodatno smo leme izbirali tudi glede na predvidene deleže po besednih vrstah pa tudi glede na predvidene slovarskospecifične probleme, kot je npr. homonimija, besednovrstna konverznost, obravnava povedkovnika, različnih besednih oblik ipd. Če beseda ni bila izbrana iz katerega od obstoječih geslovníkov, ustrezala pa je denimo problematiki, ki smo jo želeli obravnavati z leksikografskega vidika, je morala v korpusu FidaPLUS imeti vsaj 400 (v korpusu Gigafida pa vsaj 600) pojavitev. Na koncu je pomembno poudariti, da je pri oblikovanju geslovníka vedno smiselno kombinirati različna merila, kjer predstavlja pogostnost samo eno od njih.

5.1.2 Analiza konkordančnega niza

Pri pridobivanju in analizi korpusnih podatkov smo želeli postopke čim bolj optimizirati: naš končni cilj je bil pridobiti leksikalnogramatične podatke iz korpusa čim bolj avtomatsko, hkrati pa v procesu ne izgubiti ničesar, kar bi bilo glede na zastavljeni namen LBS in njene uporabnike relevantno. Ta cilj je bilo mogoče doseči postopoma, zato je bil prvi korak v opisu dane leme analiza konkordančnega niza, na podlagi katere smo ugotavljali posamezne leksikalne enote in njihovo obnašanje v realnih besedilih. Nato smo postopoma z uporabo različnih orodij in s sprotnim vključevanjem izboljšav ter lastnosti, specifičnih za slovenščino, določali način organizacije podatkov v leksikalni bazi.

Analiza konkordančnega niza je potekala v več fazah. Najprej smo s pomočjo konkordančnika v orodju SkE izdelali naključni ali po določenih kriterijih filtrirani vzorec najmanj 150 in navadno ne več kot 300 konkordanc (več o možnostih filtriranja konkordančnega niza gl. v Gantar et al. 2009a: 19). Konkordance smo nato s posebnim programom⁷¹ prenesli v Excelovo datoteko. Število konkordanc v naključnem vzorcu ni bilo dokončno določeno, ampak se je spreminjalo glede na besedno vrsto leme in leksikografsko presojo o pomenski kompleksnosti. Ker pa je bil konkordančni niz izbran naključno, je bil lahko problematičen v smislu korpusnega šuma; za leksikalni opis leksikalne enote nas je namreč zanimala predvsem njena besedilna raba in ne razbesediljene realizacije, kot se kažejo npr. v televizijskih sporedih in drugih seznamih.

71 Program za pretvarjanje konkordanc v Excelovo datoteko je podrobneje opisan v Gantar et al. 2009a: 73.

V takih primerih smo število konkordanc ustrezno povečali oz. vzorec dodatno filtrirali.

Konkordančni vzorec, ki smo ga prenesli v Excelovo datoteko, je bil pripravljen za nadaljnjo obdelavo. Analiza je bila primarno namenjena pomenski členitvi in registriranju udeleženske zgradbe ter stavčnih vzorcev pri posameznih pomenuh. Z analizo posameznih konkordanc so postale očitne tudi podrobnosti skladienskega okolja preučevane leksikalne enote, kar nam je koristilo pri oblikovanju pomenske sheme, npr. pri določanju tipičnih semantičnih tipov na posameznem udeleženskem mestu (Slika 18). Poleg tega smo na ta način prepoznavali slovnično obnašanje leksikalne enote, npr. prevladujoče slovnične oblike kot je tretjeosebna raba ali množina, prevladujočo upovedovalno možnost (npr. zanikanje), besedilni tip, govorni položaj ter pragmatične in druge slovnično-pomenske lastnosti. Z analizo dovolj obsežnega konkordančnega niza je bilo mogoče registrirati tudi stalne zveze in frazeološke enote ter njihove leksikalnogramatične lastnosti.

še rasti . Zamenju je strah , nekaj ga je	stisnu	ven , tko ko gnoj iz mozola . Seveda ,	iztisniti snovlekočino iz česa - preneseno
vzamemo iz lonca in ju s pomočjo vilice	stisnemo	iz ovoja . Ker sta tako česen kot čebula	iztisniti snovlekočino iz česa v drugo stvar
na cedilu . Če pa sem ter tja v črpaiko	stisnete	kapljico silikonske masti , bo delovanje	iztisniti snovlekočino iz česa v drugo stvar
segrevajte marelično marmelado , v katero ste	stisnili	malo limoninega soka . Nato zmes pretlačite	iztisniti snovlekočino iz česa v drugo stvar
varuje pred prehladom . V ne pretopen čaj	stisnimo	sok limone . Prav tako prija topla limonada	iztisniti snovlekočino iz česa v drugo stvar
opozori , da ti je nekdo na rano piljnil ali	stisnil	šampon . Zadnji trik sem doživel	iztisniti snovlekočino iz česa v drugo stvar
prebavni cikel , tako da mu moram občasno	stisniti	v danko (z nastavkom za modolex) nekaj	iztisniti snovlekočino iz česa v drugo stvar
pustijo v posebnih sušilnicah , potem pa se	stisnejo	in ko kvasovke še ne predelajo vseh	izžeti kaj
jabolka je treba najprej umiti , razrezati in	stisniti	Jabolčni sok nato pustimo stati v nepredušno	izžeti kaj
slami sušili grozde in ga po nekaj mesecih	stisnili	. Letos so s sušenjem grozda poskusili	izžeti kaj
je začeti dan s svežim sadnim sokom .	stisnite	si limono , pomarančo , grenivko , melono	izžeti kaj

Slika 18: Analiza konkordančnega niza z vidika določanja semantičnih tipov in stavčnih vzorcev

Analiza konkordančnega niza, ki je potekala neodvisno od izdelave besednih skic v orodju SkE, je bila pomembna tudi zaradi dejstva, da so nekateri zelo splošni ali primarni pomeni besede lahko razpršeni in nimajo tipične besedilne okolice, zaradi česar se ne potrjujejo s konkretnimi kolokacijami in jih iz besednih skic v orodju SkE ni mogoče razbrati neposredno. Takih pomenov v leksikalni bazi nismo hoteli spregledati, zato smo jih obravnavali enako kot druge pomene, le da smo jih namesto s kolokacijami ponazarjali s korpusnimi zgledi (gl. tudi pogl. o nekolokabilnih pomenih), hkrati pa smo morali to dejstvo upoštevati pri avtomatskem luščenju podatkov neposredno iz korpusa ter temu ustrezno prilagoditi slovnico besednih skic.

5.1.3 Leksikalnogramatični profil leksikalne enote in orodje Sketch Engine

Orodje Sketch Engine⁷² smo uporabili kot osnovni pripomoček za pridobivanje in analizo leksikalnogramatičnih podatkov iz korpusa in za postopek avtomatizacije, ki ga bomo podrobneje opisali v nadaljevanju. Sprva je bilo orodje namenjeno le obdelavi angleškega jezika (Kilgarrieff in Rundell 2002), nato pa so ga začeli nadgrajevati tudi za korpusse drugih jezikov (Kilgarrieff et al. 2004), med drugim tudi za slovenščino (Krek in Kilgarrieff 2006). Orodje vsebuje več funkcij, ki omogočajo analizo posamezne leme v lematiziranem in oblikoslovno označenem korpusu. Poleg tega potrebuje orodje za ustrezno delovanje tudi pripravo slovnice besednih skic, tj. specifikacijo slovničnih relacij, na katerih temeljijo njegove bistvene funkcije.

Orodje SkE je v svojem razvoju doživelo številne nadgradnje in izboljšave, ki so vsake dve leti predstavljene na delavnicah SKEW,⁷³ pri čemer je za slovenščino pomembna predvsem nadgradnja za avtomatsko luščenje kolokacijskih ter drugih relevantnih leksikalnogramatičnih podatkov neposredno v slovarski vmesnik ter lokalizacija celotnega orodja za slovenščino. V začetni fazi izdelave LBS korpus Gigafida še ni bil vključen v orodje SkE, prav tako še ni bila na voljo lokalna različica za slovenščino, besedne skice pa smo uporabljali na starejši različici slovnice besednih skic, ki je podrobneje opisana v Gantar et al. 2009a: 46–64; prim. tudi Yerošina et al. 2009. V nadaljevanju zato pri pojasnjevanju posameznih postopkov s konkretnimi podatki in pri vključevanju slik, razen pri opisu avtomatizacije, uporabljamo starejšo različico orodja in njegovih funkcij, ki je delovalo na korpusu FidaPLUS.

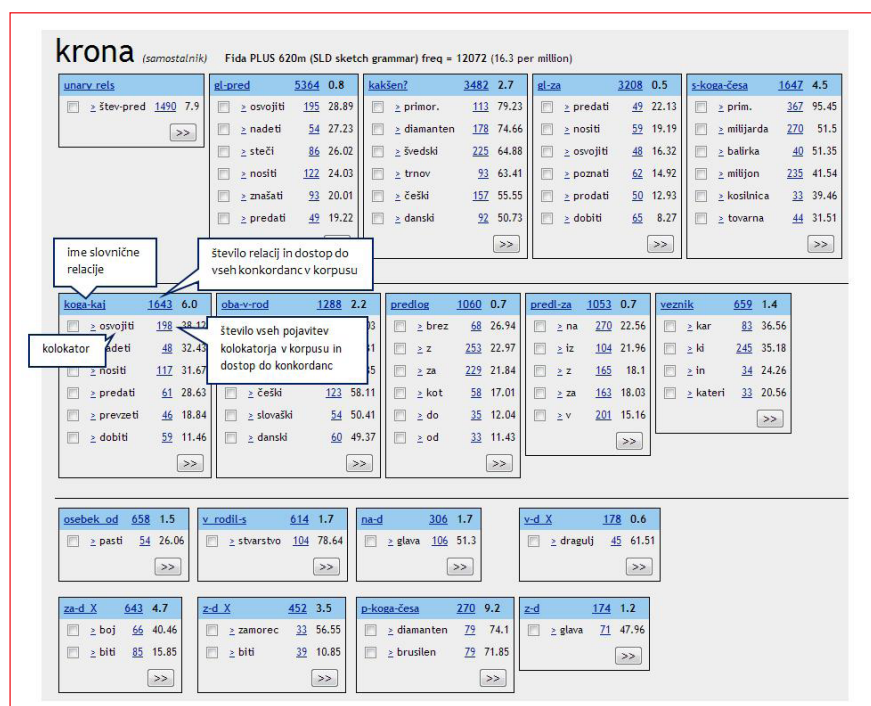
5.1.3.1 Besedne skice

Najpomembnejša funkcija SkE je izdelava besedne skice (ang. *Word Sketch*) za obravnavano besedo. Besedna skica nam pokaže, katere besede (t. i. kolokatorji) se tipično pojavljajo z besedo, ki jo analiziramo v njenem besedilnem okolju. Besedilno okolje pa je določeno s konkretno slovnično relacijo, npr. z zvezo glagola in samostalnika v tožilniku. Če je naprimer beseda, ki jo analiziramo, samostalnik *krona*, bomo pod slovnično relacijo »koga-kaj« našli tipične glagolske kolokatorje, ki se pojavljajo s samostalnikom *krona* v tožilniku, npr. *osvojiti*,

72 Gre za izdelek angleškega podjetja Lexical computing Ltd. Sketch Engine: <http://www.sketchengine.co.uk/>.

73 Več o posameznih delavnicah je mogoče najti na spletni strani: <http://www.sketchengine.co.uk/documentation>.

nadeti, nositi itd. Kot kaže izdelana besedna skica za samostalnik *krona* (Slika 19), dobimo podatke o tipičnih kolokatorjih zanj v 18 različnih slovničnih relacijah.⁷⁴ Besedna skica poleg kolokatorjev znotraj posamezne slovnične relacije podaja še informacije o pogostnosti celotne relacije ter posameznih kolokatorjev v izbranem korpusu ter njihovo statistično povezovalno vrednost – t. i. jakost (angl. *salience*). S klikom na relacijo si lahko ogledamo zapis sintakse konkretne slovnične relacije z regularnimi izrazi, s klikom na pogostnost posameznega kolokatorja pa se nam odpre ustrezeni konkordančni niz. Prazen kvadrateg pred posameznim kolokatorjem je namenjen klikosikografiji, tj. možnosti izbora in prenosa dobrih korpusnih zgledov za izbrane kolokatorje v slovarski vmesnik.



Slika 19: Besedna skica za samostalnik *krona*

Na podlagi podatkov iz besedne skice smo za posamezni pomen in ustrezno skladenjsko strukturo izločili pripadajoče relevantne kolokacije (Slika 20). Če za ponazoritev vzamemo slovnični relaciji »kakšen?» in »koga-kaj« pri samostalniku

⁷⁴ V izhodiščni različici smo upoštevali 32 slovničnih relacij. Podrobnejši opis je v Gantar et al. 2009a: 51–71.

krona, ki jima ustrezata skladijski strukturi pbz0 SBZ0 in gbz SBZ4, smo v slovarski vmesnik vnesli kolokacije, ki ustrezajo omenjenima skladijskima strukturama in posameznim predhodno ugotovljenim (pod)pomenom. Relevantnost posamezne kolokacije in korpusni šum smo preverjali z vpogledom v konordančni niz kolokatorja.

geslo

1.pomen
indikator: **denarna enota; valuta**
pomenska schema:
skladijske skupine ▢
a) struktura: pbz0 SBZ0
kolokacije ▢
■ [švedska, češka, danska] krona
zgledi ▢ (4)
↑ Back to Top

2.pomen
indikator: **naglavno okrasje**
pomenska schema:
skladijske skupine ▢
a) struktura: pbz0 SBZ0
kolokacije ▢
■ [trnova] krona
zgledi ▢ (1)
b) struktura: gbz SBZ4
kolokacije ▢
■ [nadeti, nadeti si, nositi, dobiti] krono
zgledi ▢ (1)
↑ Back to Top

3.pomen
indikator: **naslov prvaka**
pomenska schema:
3.1.
indikator: **pri športu**
pomenska schema:
skladijske skupine ▢
a) struktura: gbz SBZ4
kolokacije ▢
■ [osvojiti, nadeti si, prevzeti, dobiti] krono
zgledi ▢
↑ Back to Top
3.2.
indikator: **pri lepotehni tekmovanjih**
pomenska schema:
skladijske skupine ▢
a) struktura: gbz SBZ4
kolokacije ▢
■ [nositi, predati, dobiti] krono
zgledi ▢

kakšen? 3482 2.7

☐ primor.	113	79.23
☐ diamanten	178	74.66
☐ švedski	225	64.88
☐ trnov	93	63.41
☐ češki	157	55.55
☐ danski	92	50.73

>>

koga-kaj 1643 6.0

☐ osvojiti	198	38.12
☐ nadeti	48	32.43
☐ nositi	117	31.67
☐ predati	61	28.63
☐ prevzeti	46	18.84
☐ dobiti	59	11.46

>>

Slika 20: Prenos kolokacij iz besedne skice v slovarski program

Skupaj s kolokacijo, ki smo jo umestili pod ustrezní pomen ali podpomen in skladenjsko strukturo, smo v slovarski vmesnik prenesli tudi korpusne zglede. SkE nam za ta postopek ponuja funkcijo GDEX.

5.1.3.2 Izbor dobrih zgledov in klišikografija

GDEX (ang. *Good Dictionary Examples*) je sistem za avtomatsko ocenjevanje stavkov, ki določa njihovo primernost oz. neprimernost za kandidate za dober korpusni zgled. Funkcija deluje tako, da posamezne konkordance – tj. potencialne zglede – razvršča na lestvici od bolj do manj primernih na podlagi vnaprej določenih merljivih skladenjskih in leksikalnih kriterijev. Med najpomembnejšimi so: dolžina stavka (ustreznejši so stavki, ki vsebujejo od 10 do 25 besed) in dolžina ter pogostnost besed v stavku (ustreznejši so stavki, v katerih so uporabljene najbolj pogoste besede v jeziku, manj primerni pa so stavki, ki vsebujejo zaimke ali anaforične izraze, kot so npr. *ta, tisti* itd., zlasti če stojijo na začetku stavka). Višje so ovrednoteni stavki, v katerih se ciljna kolokacija pojavi v glavnem stavku, in stavki s t. i. razširjenimi kolokacijami, tj. kolokacijami, ki presegajo klasično dvodelno zgradbo, npr. pasti pod [določeno, magično, psihološko, kritično] mejo; prosto, novo, razpisano (delovno) mesto. Ustreznejši so tudi stavki, v katerih se najprej pojavi dovolj sobesedila, ki mu sledi kolokacija (tipično proti koncu stavka). Stavki se vrednotijo tudi glede na to, ali vsebujejo besede s črne liste. Nanjo sodijo zlasti številke in spletni naslovi. Poleg tega se pri vrednotenju stavkov lahko uporabljajo tudi druga merila, ki so na voljo za določen jezik, npr. oblikoskladenjski podatki o besedah, stopnja njihove večpomenskosti ipd.

V izhodišču je bil GDEX izdelan in preizkušen pri izdelavi spletne različice Macmillanovega angleškega slovarja za tujce (MED 2002, 2007), zaradi česar je bil izbor dobrih zgledov prilagojen specifikam angleškega jezika (Kilgariff et al. 2008: 426–427). Čeprav prvotna funkcija deluje tudi na drugih jezikih (oz. korpusih), je bila zaradi angleških jezikovnih specifik njena uporabnost za slovenščino omejena. Iz tega razloga je bila pri izdelavi LBS izdelana slovenska različica (Kosem et al. 2011), ki je občutno izboljšala izbor dobrih zgledov za slovenščino, hkrati pa je opozorila tudi na možnost izboljšave sistema za angleščino in druge jezike. GDEX je bil nato prilagojen še za postopek avtomatizacije pri izdelavi LBS (podrobneje v nadaljevanju), s čimer samo pospešili postopek izdelave gesla in optimizirali kakovost korpusnih zgledov, vključenih v slovarsko bazo.

V začetni fazi izdelave LBS je ročni postopek izbire zgledov zahteval natančno pregledovanje večjega števila konkordanc, med katerimi je leksikograf v skladu z določenimi merili izbral ustrezno število korpusnih zgledov in jih umestil v slovarski program pod ustrezen pomen in kolokacijo, stalno zvezo ter frazeološko enoto. Izkušnje so pokazale, da je poleg prenosa kolokacij v slovarski program postopek izbiranja zgledov izredno zamuden, zato se je možnost avtomatske oz. polavtomatske izbire korpusnih zgledov pokazala kot izredno dragocena.

V nadaljevanju izdelave LBS je postopek izbire korpusnih zgledov potekal v orodju SkE prek funkcije, ki smo jo delovno poimenovali kliksikografija (ang. *Tick-Box Lexicography*). Funkcija omogoča izbiro več kolokatorjev za določen pomen v izdelani besedni skici, ki v naslednjem koraku generirajo določeno število zgledov, ki jih je sistem GDEX najvišje ovrednotil. Število zgledov je mogoče poljubno določiti, med ponujenimi pa izbrati take, ki ustrezajo vnaprej določenim merilom.

Druga možnost kliksikografije, ki ne uporablja funkcije GDEX, je prenos zgledov prek konkordančnega niza, ki vsebuje na koncu vsake konkordance ikono za prenos izbranega zgleda neposredno v program za izdelavo slovarjev, kot prikazuje Slika 21.

Corpus: Fida PLUS 620m (SLD sketch grammar)
Hits: 26

Page 1 of 2 Go Next Last

Dnevnik, časopis	Zaradi političnih spletk proti grofu in predsem zaradi ljubečumja podpiše njegovo usmrtilitev, se odpove kroni in jo preda nečaku, škotskemu kralju.	
knjiga	Njegov sin, valižanski princ Edvard, je 1356 premagal Francoze v bitki pri Politiersu in utri pot sporazumu v Calaisu 1360, s katerim si je Edvard III. zagotovil fr. posesti, vendar se je odpovedal fr. kroni .	
knjiga	Tam se je eden odpovedal kroni in je to isto krono (rad ali nerad) predal drugemu vladarju.	
knjiga	Leta 1806 se je odpovedal sveti rimski kroni in postal cesar Franc I. Avstrije.	
Delo, časopis	Princ je premieru napisal pismo, v katerem ga je obvestil, da se bo z dekletom kljub prepovedi poročil, in se zato odpovedal kroni .	
revija	Ne bom se odpovedal jugoslovanski kroni , kroni, ki sem jo podedoval od svojega očeta, deda in praded.	
revija	Tam se je eden odpovedal kroni in je to isto krono (rad ali nerad) predal drugemu vladarju.	
časopis	Odpovedal se je ogrski kroni .	
Dnevnik, časopis	Poraženi kralj se je moral odpovedati angleški kroni .	
Delo, časopis	Kraljica Elizabeta II. je bila gotovo vesela, da se je izognila neprijetnosti, da bi morala avstralskim rugbyjašem čestitati na dan, ko bi se njihova domovina odpovedala kroni .	
Dnevnik, časopis	Nizozemski princ Johan Friso se je odpovedal kroni .	
Dnevnik, časopis	Nekateri avstralski republikanci so proti njej nastopili z demonstracijami, saj si že dolgo prizadevajo, da bi se njihova država odpovedala britanski kroni in postala republika.	
revija	» Ne bom se odpovedal jugoslovanski kroni , ki sem jo podedoval od svojega očeta, deda in praded.	
knjiga	Rousseau je bil jasnejši, ko je kralja konfrontiral s človekom: če se kralj odpove kroni , se dvigne v položaj človeka: il monte a t'etat d'homme.	
revija	odpovedal kroni zaradi ameriške ločenke Wallis Simpson za dvor nista prišla v poštev.	
Delo, časopis	Avstrija se je odpovedala španski kroni v zameno za oblast nad Belgijo, Luksemburgom, Milanom, Mantovo, Neapljem in Sardinijo.	
neoznačeno	Rousseau je bil jasnejši, ko je kralja konfrontiral s človekom: če se kralj odpove kroni , se dvigne v položaj človeka: il monte a t'etat d'homme.	

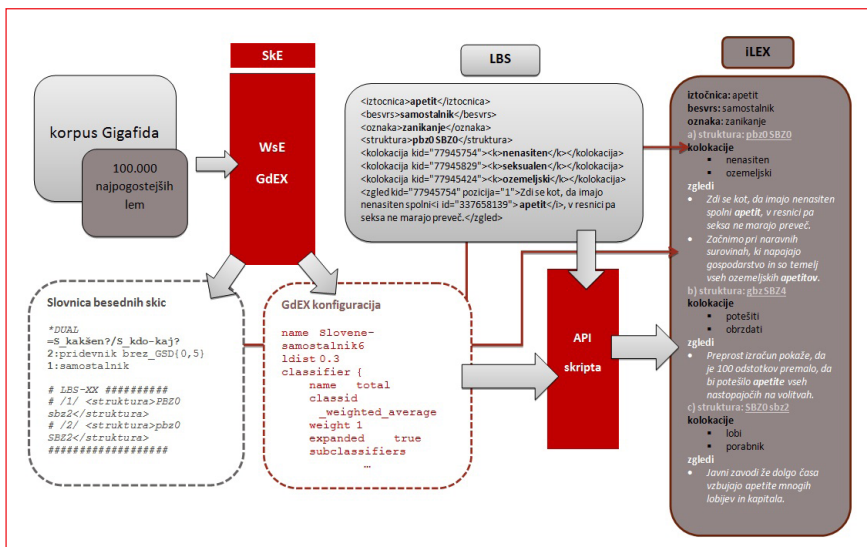
Slika 21: Prenos dobrih zgledov prek konkordanc v orodju SkE

5.1.4 Avtomatsko pridobivanje leksikalnogramatičnih podatkov iz korpusa

V zaključni fazi oblikovanja LBS smo del gesel izdelali s postopkom avtomatskega luščenja leksikalnih podatkov iz korpusa Gigafida prek orodja SkE neposredno v program za izdelavo slovarske baze iLex (Kosem et al. 2012; Kosem et al. 2013). Motivacijo za tako odločitev je predstavljalo dejstvo, da adekvaten opis jezika zahteva veliko časa, tudi zaradi vse večjih besedilnih korpusov, ki jih je potrebno analizirati. Pri gradnji LBS smo zato na podlagi izkušenj pri že izdelanih geslih predvideli, da bi se proces izdelave lahko precej skrajšal, če bi imel leksikograf možnost avtomatsko izluščiti relevantne podatke o besedi neposredno iz korpusa v slovarski vmesnik, jih pregledati, selekcionirati in po potrebi dopolniti. Dodatna analiza korpusa bi se ob izluščenih podatkih, ki bi bili leksikografsko relevantni, omejila na posamezne primere in na preverjanje aktualne rabe.

Zamisel smo začeli uresničevati ob zaključku aktivnosti, saj je bilo pomembno, da sta bila vrsta leksikalnogramatičnih podatkov in način njihove organizacije v LBS dokončno opredeljena s shemo DTD in da je bila zadostna količina gesel predhodno že izdelana. Le na ta način je bilo namreč mogoče predvideti optimalne rešitve, ki bi jih prinesel postopek avtomatizacije.

Glede na zgradbo LBS smo predvideli, da bi s postopkom avtomatizacije lahko pridobili podatke na ravni skladijskih struktur, pripadajočih kolokacij in relevantnih korpusnih zgledov. Poleg tega pa še podatke o tipičnem skladijskem ali besedilnem obnašanju leme v korpusu, kot je denimo sopojavljanje z lastnimi imeni ali količinskimi izrazi, izstopajoča možnost tretjeosebne rabe glagola ali nastopanje v *se*-glagolskih ali citatnih zgradbah. Prenos omenjenih podatkov naj bi potekal iz korpusa Gigafida prek orodja SkE oz. aplikacije besedne skice in v LBS registriranih skladijskih struktur z uporabo skripte API (ang. *Application Programming Interface*) v program iLex (Erlandsen 2004), v katerem smo gradili slovarsko bazo, kot prikazuje Slika 22.



Slika 22: Postopek avtomatskega pridobivanja podatkov iz korpusa GF v slovarski program

5.1.4.1 Slovnica besednih skic, različica ALLP 0.8

Za potrebe avtomatskega luščenja leksikalnih podatkov iz korpusa Gigafida je bila izdelana nova slovnica besednih skic (Krek 2012), ki izkorišča tudi nekatere elemente, ki so bili v orodje SkE dodani v novejših različicah. Med njimi je treba omeniti t. i. direktive (ang. *directives*) *CONSTRUCTION, *COLLOC in *SEPARATEPAGE. Prva omogoča prepoznavanje skladenjskih struktur brez kolokacij, kar je primerno predvsem za luščenje vezljivostnih vzorcev pri glagolih. Druga je namenjena izločanju elementov, ki v LBS spadajo v kategorijo skladenjskih zvez, denimo zvez predlog-samostalnik-predlog: *v primerjavi z, v odnosu do*, tretja pa je namenjena odpiranju relacij s tremi elementi (direktiva *TRINARY) na novi spletni strani, kar omogoča uvedbo natančnejših relacij s predlogi, npr. samostalnik-predlog-samostalnik, glagol-predlog-samostalnik, pridevnik-predlog-samostalnik itd., kjer lahko upoštevamo tudi sklon predloga, ki v prejšnji slovnici besednih skic ni bil upoštevan. Poleg tega smo uporabili tudi kombinacijo direktiv CONSTRUCTION+UNARY, ki je namenjena temu, da nas sistem z izpisom v posebnem stolpcu Constructions opozori, da se neka kombinacija pogojev v korpusu pojavlja nadpovprečno pogosto. S pomočjo te direktive je pri luščenju mogoče tudi avtomatsko generirati opozorila, ki bi jih v slovarjih pričakovali v t. i. slovnicih kvalifikatorjih ali oznakah, npr. pogosto zanikano, pogosto v 3. os. ednine

itd. Opisana različica slovnice besednih skic ima tako bistveno več relacij – vsega skupaj 103 – kot slovnica besednih skic, ki je bila uporabljena pri ročni izdelavi LBS, zato je namenjena zgolj avtomatskemu luščenju podatkov iz korpusa, saj je za človeškega uporabnika razmeroma težko berljiva.

5.1.4.2 GDEX za selekcionirano luščenje korpusnih zgledov

Obstoječa za slovenščino že prilagojena različica aplikacije GDEX (Kosem et al. 2011) ni bila ustrezna za potrebe avtomatizacije zaradi razlik v konceptu računalniško-leksikografskega dela. Ker smo z avtomatskim luščenjem želeli občutno skrajšati postopek ročnega polnjenja posameznih elementov geselske zgradbe, hkrati pa razbremeniti tudi postopek odstranjevanja nerelevantnih ali neustreznih podatkov, ki se v slovarsko orodje prenesejo zaradi korpusnega šuma, lematizacijskih in drugih bolj ali manj predvidljivih napak, je bil naš cilj izdelati konfiguracijo, pri kateri bi bili prvi trije ponujeni zgledi že dovolj dobri za ustrezno ponazoritev predhodno evidentiranih kolokacij.

Predvidevali smo, da bodo rezultati pri izboru kakovostnih zgledov pri posameznih besednih vrstah različni. Zato smo za vsako besedno vrsto, ki je zastopana v LBS, izdelali samostojno konfiguracijo. Pri določanju nastavitve smo analizirali zglede, ki so bili v LBS že ročno izbrani na podlagi meril dobrih korpusnih zgledov (Atkins in Rundell 2008: 458). Na ta način smo dobili izhodiščne statistične vrednosti za klasifikatorje za vsako besedno vrsto. Rezultate smo večkrat preverili na vzorčnem izboru lem in po potrebi prilagodili nastavitve. Postopek smo ponavljali, dokler nismo izoblikovali optimalne končne različice konfiguracije GDEX za postopek avtomatskega luščenja. Pomemben rezultat tega dela analize je bil dodaten nabor klasifikatorjev, ki jih prvotna različica GDEX ni vključevala. Zlasti npr. oblikovanje seznama prepovedanih besed ali zvez na začetku povedi in upoštevanje t. i. tretjega kolokatorja, ki upošteva merilo koligacijske tipičnosti določene kolokacije. Npr. pri kolokaciji *klavrn* + *podoba* klasifikator višje točkuje zglede s statistično pomembnim tretjim kolokatorjem *kazati*. Izbrana konfiguracija pa ponudi zglede, ki vsebujejo tipično širšo strukturo kolokabilne okolice: *kazati klavрно podobo česa*.

5.1.4.3 Izdelava skripte API

Prilagoditev slovnice besednih skic in konfiguracije GDEX sta bila predpogoja za pripravo skripte API, ki je zahtevala tudi posodobitev orodja SkE. Skripta API je napisana v programu Python in omogoča luščenje podatkov s povezavo na strežnik, kjer je nameščen SkE, z določitvijo ukaznih parametrov, kot so:

- korpus
- datoteka s seznamom lem
- datoteka s seznamom slovničnih relacij
- konfiguracija GDEX
- število zgledov na kolokator
- število kolokatorjev na slovnično relacijo
- minimalna frekvenca kolokatorja
- minimalna frekvenca slovnične relacije
- minimalna jakost kolokatorja
- minimalna jakost slovnične relacije

Za izdelavo API skripte je bilo potrebno pripraviti XML predlogo in jo poenotiti s shemo DTD v slovarski bazi, da smo lahko avtomatsko izluščene podatke uvozili v slovarski program. Pri elementih <kolokacija> in <zgled> smo v shemo DTD dodali identifikacijsko številko za kolokator, indeksno številko pojavnice pri elementu <zgled> ter zaporedno številko zgleda za vsak kolokator v GDEX-ovi razvrstitvi zgledov, kar je omogočilo njihovo identifikacijo v korpusu.

5.1.4.4 Rezultati in evalvacija

Pri vrednotenju rezultatov avtomatskega luščenja podatkov iz korpusa je treba upoštevati dvoje. Ročni pomensko-skladenjski opis je nastal na robustnejši besedni skici (manjše število gramatičnih relacij in manjša oblikoskladenjska razčlenjenost kolokatorjev) in na manj obsežnem korpusu (FidaPLUS tagset ver. 6.0). Avtomatska izdelava gesel pa je temeljila na bolj podrobno razčlenjeni besedni skici, ki je vključevala funkcije za detaljnješo analizo in bila implementirana v obsežnejši korpus Gigafida. Tako detajlno razčlenjena besedna skica za leksikografsko delo ni primerna in je bila že v izhodišču namenjena le avtomatski ekstrakciji. Pri primerjavi leksikografske kakovosti semi-ročno oz. na starejši različici besednih skic izdelanih gesel in gesel, izdelanih na podlagi avtomatsko izluščenih podatkov, je zato potrebno upoštevati, da smo pri določanju najoptimalnejših parametrov za prikaz relevantnih relacij, kolokatorjev in zgledov, ki so v slovarskem programu pripravljeni za nadaljnje leksikografsko delo, večkrat ročno testirali in prilagajali glede na ugotovitve. Osnovna ocena relevantnosti izluščenih podatkov v smislu, kaj v besedni skici relevantnega morda ne pride do leksikografa, je bila torej opravljena že v procesu določanja parametrov.⁷⁵

⁷⁵ Seznam 1000 lem, ki smo jih določili za avtomatsko luščenje podatkov, ni vseboval lem, ki so bile predhodno že ročno izdelane, saj smo z avtomatizacijo želeli izdelati predvideno število gesel v LBS, ne pa jih podvajati. Zato v tej fazi primerjave med gesli, izdelanimi na podlagi ročne analize besednih skic in na podlagi avtomatsko izluščenih podatkov, še ni bilo mogoče izvesti.

Ocena, ki je v tej fazi avtomatizacije še nismo opravili, jo pa vsekakor nameravamo v prihodnje, bo skušala ugotoviti, kolikšno je ujemanje med avtomatsko izluščenimi podatki in leksikografovimi izborom. V tem primeru bi bilo evalvacijo smiselno opraviti na obsežnejšem vzorcu istih lem in na podlagi enako obsežnega korpusa in iste različice besednih skic. Bolj uravnotežena izhodišča bi omogočila primerjavo števila avtomatsko izluščenih relacij in kolokatorjev z leksikografovimi izborom in priklic relevantnih kolokatorjev v izluščenih podatkih in pri ročno izdelanih geslih. Primerjava pomenske členitve in pomenskih opisov je sicer bolj relativna, saj je leksikografova odločitev vedno na nek način subjektivna, čeprav hkrati tudi usklajena s slovarskim konceptom. Bi pa taka evalvacija morala vključevati primerjavo pomenske razčlenjenosti in prisotnost/odsotnost posameznih pomenov, stalnih zvez in frazeologije. Zanimivo bi bilo tudi ugotoviti, ali leksikograf in računalnik izbereta iste zglede in če da, koliko in pri katerih skladenjskih strukturah, saj bi to posredno odgovorilo tudi na vprašanje, v katerih primerih se mora leksikograf vračati nazaj v korpus po podatke in na drugi strani, koliko časa mora nameniti čiščenju nerelevantnih podatkov. Pri nadgradnji avtomatizacije z vidika prepoznavanja in odstranjevanja nerelevantnih podatkov pa je smiselno razmišljati še o postprocesiranju in uporabi množičenja.

Kljub temu so se preliminarne ocene, ki temeljijo na primerjavi časa, potrebne za izdelavo srednje zahtevnega gesla na podlagi ročne analize besednih skic in časa, potrebne za izdelavo gesla na podlagi avtomatsko izluščenih podatkov, zelo obetavne. Leksikograf je na podlagi ročne analize izdelal v eni uri slabo četrtno gesla oz. povprečno 0,23 gesla/uro, medtem ko je bilo mogoče na podlagi izluščenih podatkov neposredno v slovarski program izdelati geslo v dveh urah oz. povprečno 0,5 gesla/uro. Seveda pa računalnik opravi le izbor in sortiranje podatkov, medtem ko v obeh primerih ostaja identifikacija pomenov, stalnih zvez, frazeoloških enot in ubesedenje pomenskih razlag, v domeni leksikografa.

5.2 Urejanje jezikovnih podatkov (programa za izdelavo podatkovne baze)

Pri izdelavi LBS smo uporabljali dva programa za izdelavo slovarjev. V obdobju snovanja in določanja vrste ter načina organiziranja leksikalnogramatičnih podatkov ter v obdobju izdelave glavnine gesel, ko je bilo v proces vključenih več avtorjev, smo uporabljali program Entry Editor, ki je del obsežnejšega slovarskega programskega paketa DPS: Dictionary Production System francoskega podjetja IDM,⁷⁶ v

⁷⁶ IDM DPS: <http://dps.cw.idm.fr/>.

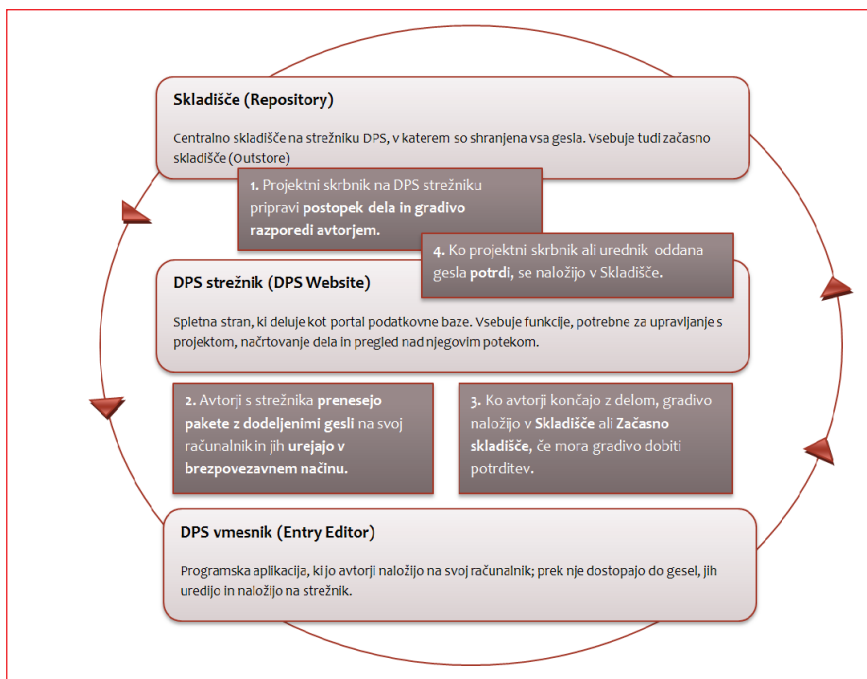
zaključni fazi izdelave LBS ter v fazi avtomatizacije postopkov pa program iLex⁷⁷ danskega podjetja EMP.

5.2.1 IDM Dictionary Production system

Dictionary Production System (DPS) francoskega podjetja IDM je program, zasnovan za specifične potrebe sestavljanja slovarjev in podobnih jezikovnih virov, zato so med ciljnim uporabniki leksikografi, uredniki, projektni vodje, založniške hiše in raziskovalne ustanove, ki upravljajo z jezikovnimi podatkovnimi bazami. Program je zasnovan tako, da avtorji dostopajo do podatkovne baze prek spleta in z njo delajo bodisi neposredno na strežniku ali pa vnaprej pripravljeno gradivo prenesejo s strežnika na svoj računalnik, ga uredijo in vrnejo v bazo, pri čemer za obdelavo uporabljajo programski vmesnik za delo v brezpovezavnem načinu. Program DPS omogoča tudi preprost nadzor nad gradivom – tako v tehničnem kot vsebinskem smislu, kar je bilo zlasti pomembno v fazi izdelave LBS, ko je hkrati delalo več leksikografov. Nadzor nad opravljenim delom vključuje določitev faz obdelave, uporabniških pravic, razdeljevanje gradiva po različnih kriterijih, primerjave stare in nove različice gradiva, preverjanje in preusmerjanje gradiva ipd. Poleg tega se med ključne prednosti programa uvršča možnost kompleksnih iskanj ter obdelave na celotnem slovarskem gradivu v formatu XML. Program DPS omogoča tudi pretvorbo v PDF ter hiter izvoz/uvoz podatkov.

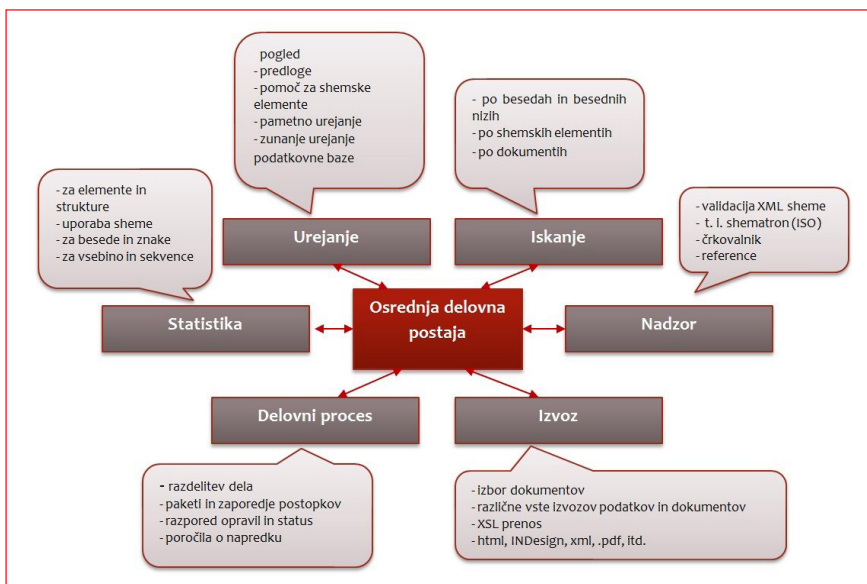
Kratika DPS se uporablja za celoten programski paket, ki ga sestavljata dve osnovni komponenti: strežnik DPS (*DPS Website*) in vmesnik DPS (*Entry Editor*). Potek dela je osredinjen na Skladišče (*Repository*), kjer je shranjena vsebina podatkovne baze. Strežnik DPS je spletna aplikacija, do katere avtorji dostopajo s standardnimi orodji za navigacijo po spletu. Na strežniku so orodja za vodenje projektov, administriranje in konfiguriranje sistema ter uporabniški dostop do podatkov. Vmesnik DPS pa je aplikacija za uporabo v okolju Windows, ki jo avtor namesti na svoj računalnik. Osnovni postopek dela je, da avtor s strežnika DPS na svoj računalnik prenese podatke, jih obdelava in vrne v Skladišče. Možno je tako delo neposredno s Skladiščem, ki je bolj prosto organizirano, ali pa s t. i. paketi (*batch*), ki omogočajo večji nadzor nad postopkom dela. Slika 23 prikazuje povezavo med različnimi komponentami programa DPS.

77 EMP: iLex: http://issuu.com/jens.erlandsen/docs/ilex_brochure_120dpi/2.



Slika 23: Program DPS: povezava med posameznimi deli programa in uporabniki

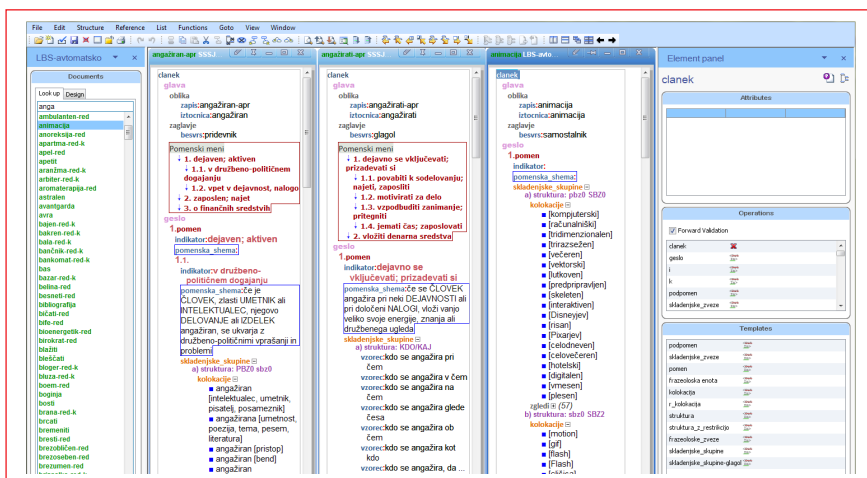
Osrednja komponenta celotnega slovarskega paketa za izdelavo gesel je programsko okno vmesnika DPS. Kot prikazuje Slika 24, je celotno programsko okno razdeljeno na štiri dele. Od leve proti desni so to: 1. okno (*Navigation Panel*), namenjeno dostopu do gesel v dodeljenem paketu ter za komuniciranje s strežnikom DPS, 2. okno (*Wysiwyg View*), namenjeno ogledu gesla brez XML elementov v izgledu za branje. Besedilo v tem oknu je možno tudi urejati. Enak izgled je predviden tudi za natis. 3. okno (*Tree View*) je namenjeno prikazu drevesne strukture gesla z XML elementi. To je glavno okno in je primarno namenjeno urejanju gesla. 4. okno (*Attributes and Annotations*) je namenjeno urejanju atributov pri XML elementih ter vstavljanju redaktorskih opomb k vsebini posameznega dela gesla.



Slika 25: Osnovne funkcije programskega paketa iLex

Osnovni namen orodja je integracija podatkovnih baz in uredniškega sistema, v katerem je v procesu urejanja mogoče posamezne funkcije uporabiti simultano za različne podatke v več dokumentih v eni sami ali več podatkovnih bazah. Posamezna opravila, kamor sodijo urejanje, iskanje, nadzor oz. validacija, statistične analize, spremljanje delovnega procesa in izvoz podatkov in dokumentov, so združena v osrednjem programskem vmesniku, ki je namenjen urejanju posameznih dokumentov v podatkovni bazi oz. bazah. Kot prikazuje Slika 26, je na levi strani t. i. projektni razdelek, ki omogoča dostop do vseh podatkovnih baz in dokumentov, npr. gesel znotraj posamezne baze, na desni strani pa lahko izbiramo med različnimi razdelki:⁷⁸ predstavitveni, urejevalni, iskalni, izvozni, statistični, validacijski, razdelek s predlogami, referenčni razdelek in razdelek z odložiščem.

⁷⁸ Opis, ki ga navajamo, se nanaša na različico programa 2.3 071. V nadgrajenih različicah so bile nekatere funkcionalnosti dodane ali izboljšane.



Slika 26: Osrednji vmesnik za urejanje podatkovne baze v programu iLex

Ena od pomembnejših odločitev za program iLex je, poleg sprejemljive cene za posamezno licenco, dejstvo, da podpira izdelavo podatkovne baze v formatu XML. Format XML ni samo prevladujoč standard na spletu, ampak se vse bolj uporablja tudi za izmenjavo podatkov, njihovo vzdrževanje, tiskanje in elektronsko založništvo. Zlasti zadnje zahteva dosledno ločevanje med vsebino in njeno prezentacijo: pomembno je namreč, da je mogoče isto vsebino uporabiti ne le na različnih medijskih platformah, npr. v tiskani obliki, na spletu ali mobilnem telefonu, pač pa tudi v različnih kontekstih in kombinacijah, ki jih predpostavlja namen končnega izdelka. Poleg tega format XML omogoča nadzor nad strukturiranostjo podatkov, tj. skladnost gesel z elementi sheme DTD, in kompatibilnost tudi v primeru, ko prihaja med dokumenti do večjih razlik.

**C. LEKSIKALNO-
GRAMATIČNI PODATKI
V LEKSIKALNI BAZI ZA
SLOVENŠČINO**

I. IZRAZNI NIVO

T. i. izrazni nivo se v zgradbi LBS nanaša na tiste podatke, ki se nanašajo na izrazne, lahko bi rekli oblikovne lastnosti besede, ki jo obravnavamo. Z vidika organizacije podatkov predstavlja izrazni nivo tisti del celotnega opisa besede v slovarski bazi, ki je neposredno povezan s korpusnimi podatki, tj. z lematizacijo (še prej tokenizacijo) in oblikoskladenjskim označevanjem. Vključuje torej podatke o lemi, njenih oblikah in variantah (podatki o oblikah in variantah se lahko sicer pripenjajo tudi na posameznje pomene) ter oblikoskladenjskih kategorijah, kot so besedna vrsta, sklon, spol, število, pa tudi občno- oz. lastnoimenskost.

Podatki, ki v shemi DTD pripadajo izraznemu nivoju, so podatki v <glavi> v elementih <oblika> in <zaglavje>. Oblika, ki je v korpusu prepoznana kot lema, je zapisana v elementu <iztočnica>, njena besednovrstna kategorija pa v elementu <besvrs>. Na lemo oz. iztočnico je vezan tudi podatek o njenih oblikah (s povezavo na Leksikon besednih oblik Sloleks) in korpusni frekvenci.

Na ravni iztočnice je bilo za slovarski opis treba premisliti o vrstah iztočnic, zapisu osnovne oblike in načinu beleženja in posredovanja slovničnih podatkov, zlasti v razmerju med izrazno podobo in vsebino leksikalne enote (homonimija in večpomenskost) in med vsebino ter njenim skladenjskim obnašanjem (besednovrstna konverzija).

6 Iztočnica

Odločitve, ki smo jih v zvezi z iztočnico kot naslovno enoto gesla sprejeli v LBS, izhajajo iz organizacije podatkov v slovarski bazi, ki obravnava iztočnico kot izrazno in ne kot pomensko enoto, ter iz organizacije, kot jo narekuje spletni medij. Zato ima iztočnica v geselski zgradbi v prvi vrsti povezovalno oz. organizacijsko vlogo. V spletni postavitvi dostop do slovarskih podatkov in njihova vsebina namreč nista povezana s poznavanjem makro- in mikrostrukturnih razmerij, ki jih uporablja tiskani slovar (npr. sistem abecedne razporeditve iztočnic, kazalčni sistem, sistem podgesel itd.), pač pa je pomembna predvsem opredelitev slovarskih enot in znotraj njih leksikalnih oz. pomenskih enot, tj. tistih, ki so nosilke pomena in na katere se v slovarju pripenjajo informacije, ki za uporabnika predstavljajo odgovor na potencialni jezikovni problem (Gantar 2011: 78). Vloga iztočnice v LBS je zato v okviru gesla združiti vse slovarske enote, ki vključujejo iztočnico kot možne različne nize črk (t. i. različnice), ki pripadajo isti osnovni obliki – lemi in imajo prekrivno oblikoslovno in naglasno paradigmo.

Podatki, ki so v LBS vezani neposredno na iztočnico, so v shemi DTD vsebovani v elementih <glava> in <zaglavje>. Znotraj elementa <glava> so vsebovani še podatki o zapisu iztočnice, ki so namenjeni iskanju po bazi, pogostnost njene pojavitve v korpusu ter sama iztočnica. Znotraj elementa <zaglavje> pa podatki o besedni vrsti leme, njenih pregibnih oblikah (prek povezave na Sloleks).

Način leksikografske predstavitve in izbor podatkov, vezanih na iztočnico, sta zahtevala vrsto pomembnih leksikografskih odločitev, npr. določitev tipov iztočnic, zapis iztočnice v osnovni obliki ter njena besednovrstna opredelitev. V zvezi s tem je bilo potrebno premisliti, kako reševati problem homonimije v odnosu do večpomenskosti, problem zapisovanja iztočnic glede na razmerje domače – prevzeto, glede na variantni zapis skupaj in narazen ter glede na besednovrstno opredelitev zlasti v luči kategorialno povezanih oblik (npr. pridevnik – deležnik) in besednovrstne konverzije.

6.1 Vrste iztočnic

V LBS smo obravnavali dve vrsti iztočnic: eno- in večbesedne. V primeru večbesednosti smo se na ravni iztočnice odločali na podlagi pomenskih in formalnih meril. To pomeni, da je večbesedna leksikalna enota lahko bodisi samostojna iztočnica bodisi samostojni segment gesla, v katerem ena od besed večbesedne enote nastopa kot iztočnica. Formalno lahko torej iztočnica združuje več tipov leksikalnih enot (pomene, podpomene in večbesedne leksikalne enote), lahko pa je tudi sama

večbesedna. Obstoj večbesedne leksikalne enote ne pomeni nujno tudi njene obravnave na ravni iztočnice (o tem odloča zlasti njena pogostnost in pomenska razvejavnost), pač pa jo, skladno z merili, ki jih bomo opisali v nadaljevanju, lahko uvrstimo na več ravni geselske zgradbe (hkrati), npr. na raven kolokacij in stalnih zvez.

6.1.1 Enobesedne iztočnice

Pri izdelavi LBS smo se osredotočali na enobesedne iztočnice. Kot enobesedne iztočnice smo obravnavali tudi glagole s prostim glagolskim morfemom *se/si*, kjer je morfem sestavni del glagola v vseh njegovih pomenih, npr. *odkrižati se*, *svitati se*, *zdeti se*. (Glede drugih rešitev gl. poglavje 7.1.1.2.1 Zapis glagolov s prostim glagolskim morfemom *se/si*.)

6.1.2 Večbesedne iztočnice

Razen v nekaj primerih, npr. *državni zbor*, *nivojski pouk*, *žiro račun*, ki so bili namenjeni poskusnim redakcijam stalnih zvez kot samostojnih iztočnic, v LBS na ravni iztočnice večbesednih leksikalnih enot nismo obravnavali. Tako tudi ne večbesednih enot tipa *al dente* ipd., ki jih ne razumemo kot klasične stalne zveze. V LBS smo se odločali za beleženje stalnih zvez v samostojnih razdelkih gesla, tj. v elementih <stalne_zveze> in <frazeološke_enote>. Medtem ko smo stalne zveze pod posebnimi pogoji lahko prekvalificirali v večbesedne iztočnice, smo frazeološke enote vedno obravnavali le znotraj gesla v samostojnem razdelku, neodvisno od pomenske členitve. Kot osnovni merili za premik stalne zveze na raven samostojne iztočnice smo upoštevali pogostost in leksikaliziranost, ki se kaže v pomenski samostojnosti zveze kot celote, smiselno pa bi bilo razmisliti tudi o pomenski razvejanosti zveze in vezanosti na splošni jezik.

6.1.3 Lastna imena

V LBS na ravni iztočnice nismo beležili lastnih imen, npr. imen bitij in zemljepisnih imen. Kot samostojnih iztočnic tudi nismo obravnavali stvarnih lastnih imen. Z vidika pomenske obvestilnosti bi bilo na ravni iztočnice smiselno obravnavati vrstna poimenovanja ter skrajšana ali poljudna poimenovanja skupnosti, društev, ustanov, listin ipd., ki so jezikovno in kulturno specifična. Ker gre pri vrstnih poimenovanjih pogosto za večbesedne enote, smo jih obravnavali bodisi pri stalnih zvezah (*državni zbor*, *slovenska matica*, *zdravstveni dom*) ali kot del kolokacijskega niza: slovenske [železnice, železarne]. Primere, kjer samostalni preide iz občnega v (neosebno) lastno ime, kot npr. pri iztočnici *otok* -> Otok – Velika Britanija, smo obravnavali kot samostojne pomene.

7 Zapis

Razumevanje iztočnice sledi v LBS spletni zasnovi slovarja in je zato v marsičem drugačno, kot ga poznamo iz klasičnih tiskanih slovarjev. Bistvena razlika izhaja iz dejstva, da izhodiščna vloga iztočnice ni (več) izrazno-pomenska, pač pa predvsem izrazna. Iztočnice torej ne enačimo z leksemom, pač pa le z njegovo izrazno podobo, katere osnovna naloga je združevati vse pomenske enote (pomene, stalne zveze, FE), ki jih je mogoče prepoznati pod istim izrazom. Tako izhodišče v marsičem spreminja predstavo o zgradbi gesla v smislu ustvarjanja razmerij med naslovno enoto, njenimi oblikami in leksikalnimi oz. pomenskimi enotami, ki jih zastopa. Neposredno se to kaže, kot bomo prikazali v nadaljevanju, pri zapisu osnovne oblike, pri obravnavi homonimije v odnosu do večpomenskosti, pri obravnavi skladijske konverzije, pa tudi pri obravnavi stalnih zvez in frazeoloških enot v zgradbi gesla.

7.1 Oblika

Iztočnico smo zapisovali v osnovni obliki, za vse druge oblike znotraj iztočniške oblikoslovne paradigme pa smo predvideli povezavo na Sloleks (Erjavec et al. 2008; Arhar 2009), ki trenutno vsebuje približno 100.000 leksikonskih enot.⁷⁹ Na ta način je mogoče dostopati do vseh oblik določene leksikalne enote oz. njenih oblikoslovnih lastnosti (Preglednica 5), vključno s podatki o občno- oz. lastnoimenskosti na ravni osnovne oblike, podatki o (besedotvorni) povezanosti leksikonskih enot, v primeru v jezikovni rabi izpričane oblikovne variantnosti (*datelj* – *dateljna/datlja*; *jokati* – *joka/joče*), pa tudi do podatkov o pogostnosti posameznih oblik ter o njihovi trenutni normativni vrednosti. V nadaljnjih izboljšavah leksikona je predviden tudi podatek o kolikostni in naglasni podobi posamezne oblike.

⁷⁹ Trenutna različica Sloleksa vsebuje podatke za enobesedne lekseme. Testno so vključene le tri večbesedne enote. Način njihovega vključevanja je opisan v Kazalniku (Erjavec et al. 2008: 52), realiziran pa ni bil.

Preglednica 5: Podatki za osnovne besedne vrste v LBO

	Vrsta	Spol	Živost	Sklon	Oseba	Število	Oblika	Stopnja	Vid
samo-stalnik	občno	m	živo	im, rod,		edn.			
	ime	ž	neživo	daj, tož,		dvoj.			
	lastno	sr		mest,		množ.			
	ime			orod					
glagol	glavni	m ⁸⁰			prva	edn.	sedanjik		dov.
	pomožni	ž			druga	dvoj.	deležnik		nedov.
		sr			tretja	množ.	nedoločnik		dvovoid.
							namenilnik velelnik prihodnjik pogojnik		
pri-devnik	splošni	m		im, rod,		edn.	nedoločno	osnovnik	
	deležniški	ž		daj, tož,		dvoj.	določno	primernik	
	svojilni	sr		mest, orod		množ.		presežnik	
prislov	splošni							nedolo-	
	deležje							čeno primernik presežnik	

Čeprav je podatek o pogostnosti osnovne in posameznih oblik razviden iz povezave na Sloleks, smo v strukturi geselskega članka predvideli tudi element <korpus>, ki je namenjen eksplicitnemu beleženju pogostnosti obravnavane leme, kot je izkazana v korpusu Gigafida. Številčni podatek je mogoče v geselsko strukturo vnesti avtomatsko, zato ga ročno nismo dodajali, pomembno pa je opozoriti na možnost različnih vizualizacij tega podatka v spletnem slovarju, npr. v obliki avtomatsko generiranih grafov, obarvanih likov, ki zastopajo določeno količinsko enoto, s čimer je mogoče pokazati tudi razmerje pogostnosti obravnavane besede do drugih bolj oz. manj pogostih besed v korpusu.

7.1.1 Osnovna oblika

Element <iztocnica>, ki v shemi DTD sledi elementoma <zapis> in <korpus>, je namenjen zapisu obravnavane leksikonske enote v osnovni obliki, razvrščanju podatkov, združenih pod konkretno iztočnico, in razvrščanju iztočnic v programu iLex, zato pri zapisu iztočnice v osnovni obliki nismo beležili dodatnih informacij, npr. roditeljske oblike ali variantnih zapisov. Pri iztočnici tudi nismo beležili izgovora, saj smo za te podatke predvideli samostojni element v geselski zgradbi, in

80 Pri glagolih je kategorija spola upoštevana samo pri deležnikih na -l.

sicer z namenom, da jih bo mogoče v spletnem slovarju prikazovati celostno, npr. v samostojnem zavijku (prim. Krek et al. 2013).

Merila za zapis besede v iztočnici v osnovni obliki smo podrobneje določili glede na besedno vrsto, pri tem pa smo sledili oblikoskladenjskim oznakam v korpusu Gigafida (Krek 2010).⁸¹ To je najprej potrebno zaradi možnosti pridobivanja podatkov iz korpusa za leksikografsko analizo⁸² in zaradi možnosti nadaljnjega urejanja in procesiranja podatkov.

7.1.1.1 Samostalnik

Osnovna oz. slovarska oblika, ki jo enačimo z lemo, je pri samostalniku imenovalnik ednine oz. imenovalnik množine, če samostalnik nima edninske oblike: *grablje*, *blačke*. V primeru, ko lahko govorimo o več možnih osnovnih oblikah, smo sledili odločitvam, ki so bile sprejete pri oblikoskladenjskem označevanju korpusa Gigafida. Posamezne izjeme smo reševali z dodatnimi merili, ki jih postavimo v nadaljevanju.

Enoštevilske samostalnike (Toporišič 2000: 271) smo navajali v njihovi edini številski obliki, najpogosteje v ednini, npr. *rukola*, *satén*, redkeje v množini, npr. *grablje*, *blačke*. Eksplicitno podatka o tem, da gre za enoštevilske samostalnike, nismo navajali, ga je pa mogoče razbrati posredno, in sicer iz (a) slovnične oznake *neštevno* neposredno za iztočnico, če je njena osnovna oblika samoedninska in če do s pomenom povezanih sprememb v števnosti ne prihaja pri nobenem od registriranih pomenov, npr. *divjačina*, *rukola*, (b) z oznako *neštevno* ob posameznem (pod)pomeni, kadar je sprememba kategorije števnosti vezana na pomenske spremembe, kot jih izkazujejo posamezni pomeni ali podpomeni, npr. *korespondenca*, *aplikacija* ipd., in (c) z oznako *množina* neposredno za iztočnico v primeru samomnožinske oblike, npr. *spodnjice*, *tangice*, ker prihaja dejansko do odstopanja od pričakovane edninske osnovne oblike.

Po enaki logiki, tj. na pomenski ravni z uvrstitvijo v samostojni (pod)pomen, smo obravnavali spremembo števila, ki je povezana s spremembo pomena, npr. *lisica* – žival; *lisice* – naprava. V teh primerih smo opozorilo, da gre za ustaljeno številsko obliko, npr. *navadno v množini* ali *v množini*, navajali v elementu <oznaka> neposredno za indikatorjem.

81 Sistem oblikoskladenjskega označevanja za slovenščino, ki je v razmerju korpus – leksikon – leksikalna baza usklajen, je bil razvit pri projektu JOS.

82 Za leksikografsko analizo je pomembno, da leksikografi poznajo seznam oblikoskladenjskih oznak in sprejete sklepe pri oblikoskladenjskem označevanju korpusa, saj lahko samo na ta način korpus ustrezno interpretirajo.

Pridevnikov, ki se v samostalniški vlogi pojavijo samo v določenih skladenjskih položajih – v mislih imamo predvsem *posamostaljene pridevnike* (Toporišič 2000: 255; posamostaljena pridevniška beseda) – nismo obravnavali na ravni samostojnih iztočnic z različnimi lemami. Taka rešitev namreč ni v skladu z lematizacijo v korpusu, iz katerega smo pridobivali podatke, saj bi v primeru strojne lematizacije pomenila predvidljivo slabe rezultate. Poleg tega gre za pomensko-skladenjsko problematiko, podprto z različnim besedilnim okoljem obeh rab, zato smo te primere obravnavali na ravni pomenov. Poleg tega se različni besednovrstno pogojeni osnovni obliki, npr. samostalniška *odvisni*, in pridevniška *odvisen*, redko rabita v celotni paradigmi in vseh številih, pač pa posamostaljene pridevnike zelo pogosto rabimo samo v množini za označevanje skupine referentov z določenimi lastnostmi (npr. *nezaposleni, dobri, slabi* itd.), imenovalnik množine pa ima enako obliko kot določna oblika pridevnika v ednini (Krek 2010: 47). Kot smo že zapisali, je izhodišče urejanja podatkov v LBS semantično, podatke na ravni skladnje pa premikamo v druge dele geselskega članka, medtem ko na ravni iztočnice ohranjamo oblikoskladenjsko prepoznavnost osnovne oblike. Način razporeditve podatkov in zapis osnovne oblike v iztočnici pri posamostaljenih pridevniki ponazarja iztočnica *odvisen* – pomen 3.1:

odvisen pridevnik

1 pogojen s čim

2 podrejen čemu

3 zasvojen

3.1 odvisnik; zasvojenec

v samostalniški rabi

navadno v množini

kdor uživa prepovedane droge ali se ukvarja z neko dejavnostjo do te mere, da škodi njegovemu zdravju ali socialnemu življenju

PBZ0<r>v samostalniški vlogi</r>od sbz2

odvisni od [drog, alkohola]

V primeru razlikovanja osnovne oblike na ravni spola, npr. *cigaret – cigareta, copat – copata*, smo se odločali za samostojni iztočnici. Kljub pretežni pomenski prekrivnosti namreč lahko prihaja tudi do določenih razlik na ravni kolokacij, stalnih zvez pa tudi do oblikovanja samostojnih pomenov, kot to npr. velja za žensko obliko samostalnika *copata*, ki se v pomenu 'neodločen, pomehkužen moški, zlasti v odnosu do ženske', uveljavlja le pri tej obliki.

7.1.1.2 Glagol

Osnovna oblika, ki smo jo pri glagolih navajali v elementu <iztocnica>, je nedoločnik. Oblik, kot so opisni deležnik na -l, namenilnik in velelnik, ki jih povezuje

glagolska lema, nismo obravnavali pri iztočnici, ampak smo nanje opozarjali v skladenjskih strukturah, ki so podlaga kolokacijam, npr. gbz Nam-gbz: [iti, odpraviti se, priti] iskat, lahko pa tudi pri samostojnem pomenu ali podpomenu z ustrezno slovnično oznako:

paziti *glagol*

4 previdno ravnati

4.4 kot opozorilo

navadno v velelniku

če zakličemo komu »Pazi!«, ga želimo opozoriti na nevarnost, ki mu grozi v danem trenutku

Druge oblike, ki se povezujejo z glagolom, npr. deležja, smo obravnavali kot samostojne iztočnice, in sicer v njihovi osnovni ali edini obliki z besednovrstno oznako prislov, npr. *kleče*. Deležnike na -n in -t, ter deležnike stanja na -l smo obravnavali kot pridevnike, npr. *akreditiran, oblegan, obrit, izumrl*, glagolnike pa kot samostalnike, npr. *arhiviranje, bučanje, deložacija*. V primeru izrazno prekrivne osnovne oblike, kjer prihaja pod vplivom skladenjske vloge do pomenskih sprememb, npr. *noseč* – deležniški pridevnik in *noseč* – deležijski prislov, smo izdelali dve samostojni iztočnici v skladu z oblikoslovnim leksikonom.

7.1.1.2.1 Zapis glagolov z glagolskim morfemom *se/si*

V zvezi z zapisom glagolske iztočnice v osnovni obliki je bilo treba določiti tudi merila za obravnavanje primerov z glagolskim morfemom *se/si*. Glede na njihovo pomensko vlogo smo glagole razdelili v tri skupine.

7.1.1.2.1.1 Glagoli z morfemom *se/si* v osnovni obliki

Glagole, ki se lahko pojavljajo z morfemom *se* ali *si*, smo na ravni iztočnice obravnavali najprej glede na obvezno prisotnost morfema v vseh glagolskih pomenih, npr. *daniti se, navzeti se, postiti se, šopiriti se, zaljubiti se*. V teh primerih smo morfem obravnavali kot sestavni del glagola, torej tudi njegove osnovne oblike. V primeru variantne oblike glagolskega morfema smo v iztočnici zapisali oba, npr. *oddahniti si/se*, pri čemer smo na prvem mestu navedli bolj pogostega. Podatek o frekvenci je razviden iz besedne skice v orodju SkE, kot kaže Slika 27.

oddahniti (glagol)		
Constructions		
☐ O_povratni_si	8225	117.8
☐ V_lahko_G	1544	33.7
☐ O_povratni_se	1528	3.0
☐ O_tretja_oseba	1264	0.6
>>		

Slika 27: Pogostost morfemov se in si ob glagolu oddahniti

7.1.1.2.1.2 Glagoli z morfemom se/si v katerem od svojih pomenov

Če glagol izraža povratnost dejanja na osebek samo v katerem od svojih pomenov, npr. *ubiti se* 'umreti' : *ubiti sebe (koga)* (Toporišič 2000: 357), je taka možnost razvidna bodisi iz pomenskega indikatorja bodisi iz ubeseditve v pomenski shemi, v osnovni obliki pa morfema nismo navajali:

angažirati glagol

dejavno se vključevati; prizadevati si

če se ČLOVEK angažira pri neki DEJAVNOSTI ali NALOGI, vloži vanjo veliko energije, znanja ali ugleda

7.1.1.2.1.3 Glagoli z morfemom se/si kot upovedovalno možnostjo

Če glagolski morphem *se/si* ni pomensko odločujoč in izraža zgolj možnost prehanja dejanja na osebek, npr. *umivati se(be)* – *umivati koga*, ter s tem samo katero od upovedovalnih možnosti, ga v osnovni obliki nismo navajali, ampak smo ga prikazali kot možno skladenjsko realizacijo v pomenski shemi z ločenimi udeleženskimi zgradbami ali pa v stavčnih vzorcih. Med skladenjskimi upovedovalnimi možnostmi (tj. tistimi, ki nimajo neposredne pomenske vrednosti), smo registrirali naslednje:

- prava⁸³ povratnost: masirati koga – masirati se(be); česati koga – česati se(be);
- neprava povratnost: kdo cepi koga – cepiti se;
- vzajemnost: obiskovati koga – obiskovati se; ljubiti koga – ljubiti se;
- splošni vršilec:⁸⁴ kazati – kazati se; razlikovati – razlikovati se;
- trpnost: kaj barva kaj – kaj se barva (od česa).

7.1.1.2.2 Zapis glagolov z glagolskim morfemom *ga* ali *jo*

Glagole, ki v katerem od svojih pomenov tipično pridobijo prosti glagolski morfem *ga/jo*, smo v iztočnici navajali brez tega morfema, npr. *cukati*. Tipično realizacijo ob ustreznem pomenu smo zapisali v element <ustaljena_oblika>, ustrezen zapis glagola s prostim morfemom pa je razviden tudi iz pomenske sheme, kolokacij in stavčnih vzorcev:

cukati *glagol*

6 piti

v zvezi z *ga*

v neformalnih situacijah

pogosto s šaljivim podtonom

rbz GBZ

[rad] *ga* *cukati*

če rečemo, da *ga* ČLOVEK *cuka* |ali| če ČLOVEK *cuka* PIJAČO, želimo na šaljiv način povedati, da pogosto pije alkoholne pijače ali da se opija

7.1.1.3 Pridevnik

Pridevnike smo v LBS zapisovali v imenovalniku ednine moškega spola, npr. *kisel*, *boboten*, *obupan*, *preroški* itd., kar je v skladu s stanjem v korpusih FidaPLUS in Gigafida. Tako zapisana pridevniška lema pokriva na ravni iztočnice še stopnjevanje oblike, tj. osnovnik, primernik, presežnik (*debel* – *debelejši* – *najdebelejši*), ne pa tudi elativa, npr. *prenasičen*, o čemer več v nadaljevanju.

Če ima pridevnik v imenovalniku ednine realnih več oblik, npr. *briljanten* – *briljantni*, *kisel* – *kisli*, *črn* – *črni*, *daljnji* – *daljen* (vendar samo *bližnji*), smo kot osnovno obliko v iztočnici določili nedoločno obliko oz. obliko, ki ustreza lematizaciji v korpusu Gigafida in leksikonu besednih oblik Sloleks, pomensko razlikovanje v smislu lastnost, snovnost, vrstnost itd., pa smo prikazali na ravni (pod)pomenov z indikatorjem *in/ali* s pomenskim opisom, npr. *kliničen* – bolezenski : hladen, brezoseben. V takih primerih pričakujemo tudi razlikovanje pri kolokacijah, npr.

83 Izraza »prava« in »neprava« povratnost v literaturi nista ustaljena in ju uporabljamo kot delovna termina zgolj v okvirih omenjene razprave in leksikografskih rešitev v LBS.

84 Splošnega vršilca razumemo kot eno od pojavnih oblik potrpenja (prim. Toporišič 2000: 502).

[preveč] kliničen : klinični [znaki, simptomi], in stalnih zvezah: *klinični center, klinični psiholog*.

Primerniških in **presežniških oblik** pridevnikov (Toporišič 2000: 329), npr. *debelejši, aktivnejši* ipd., v LBS nismo obravnavali kot samostojnih iztočnic, ampak smo na tipične realizacije opozarjali v strukturah v elementu <restrikcija>, sicer pa so ustrezne oblike vedno razvidne tudi pri kolokacijah:

izkušen pridevnik

2 izurjen in spreten

Kol-rbz PBZ0<r>primernik</r>

[precej, veliko] izkušenejši

PBZ0<r>presežnik</r>v sbz5

najizkušenejši v [moštvu]

najizkušenejši v [naši] vrsti

Elativno stopnjo (Toporišič 2000: 325), npr. *predebel, prenasičen*, smo obravnavali kot samostojno iztočnico, saj prehaja od dejanskega izražanja stopnje, ki bi jo lahko parafrazirali s 'preveč', do pomensko samostojnih tvorbo, npr. *prefinjen, prehlajen, prekajen* ipd.

V zapisu variantne pridevniške pare tipa *težek – težak, lahek – lahak, gorek – gorak, močen – močan, dolžen – dolžan*, smo obravnavali pod eno lemo (*težek, lahek, gorek, močen, dolžen*) v skladu s stanjem v leksikonu in korpusu. Bi bilo pa v teh primerih, tako kot smo se odločali pri samostalnikih (*zjedalec – zajedavec*) in prislovi (gl. v nadaljevanju), zaradi možnih pomenskih sprememb in kolokacijskih posebnosti smiselno razmisliti o samostojnih iztočnicah, vključno z informacijami o razmerju do standardne oblike.

7.1.1.4 Prislov

Glede na to da prislov obravnavamo kot nepregibno besedno vrsto (Toporišič 2000: 406) smo kot osnovno obliko obravnavali edino možno, npr. *aktivno, evidentno, natančno, tešče*.

Posebnost, ki jo moramo omeniti v zvezi z zapisovanjem prislovov v osnovni obliki, je njihovo stopnjevanje, kjer smo ravnali enako kot pri pridevnikih: poleg osnovniške združuje prislovna lema še primerniške in presežniške oblike, npr. *aktivno – aktivneje, enakomerno – enakomerneje*. Kadar prihaja pri primerniških in/ali presežniških oblikah do variantnosti, npr. *eksplicitno – eksplicitneje/eksplicitnejše – najeksplicitneje/najeksplicitnejše*, ne gre za samostojne iztočnice, pač pa je podatek dostopen prek leksikona. **Elativne oblike** prislovov, tako kot pri pridevnikih, smo opisali v samostojnih iztočnicah, npr. *prezgodaj*. Zaradi številnih različnih skladenjskih

vlog, npr. števniske, členkovne ipd., smo obravnavali nekatere presežniške oblike, npr. *največ*, *najmanj*, kot samostojne iztočnice.

V primeru sicer redkih variantnih parov, kot *posebej* – *posebaj*, smo se v skladu z leksikonom odločali za samostojni iztočnici.

7.1.2 Homonimne oblike

Kot homonime,⁸⁵ tj. kot samostojne leme, smo v LBS obravnavali samo tiste besede, ki so izrazno prekrivne znotraj istih oblikoskladenjskih kategorij, in sicer v celotni sklanjatveni, spregatveni ter naglasni paradigmi. Tu je treba najprej obravnavati oblike, ki jim ni mogoče pripisati enotnih kategorij, značilnih za določeno besedno vrsto, čeprav imajo osnovno obliko prekrivno, npr. *bor* – pridevnik (s pomenom 'majhen po količini; slab') : *bor* – samostalnik, *plesen* – samostalnik, *plesen* – pridevnik; *prst* – sam. ž. sp. in *prst* – sam. m. sp. V SSKJ gre – za razliko od SP in SSKJ2 – za samostojne iztočnice brez indeksne oznake, ki jo imajo sicer homonimi. V LBS smo se tako odločali tudi v primeru prekrivnih besednovrstnih kategorij, če prihaja do sprememb v celotni oblikoslovni in naglasni paradigmi, npr. *poročen* – *poročnega* in *poročen* – *poročnega*; *pasti* – *padem* in *pasti* – *pasem*, kot tudi samo v naglasni paradigmi, npr. *molíti* – *mólim* in *molíti* – *molím*; *častíti* – *častím* in *častíti* – *častím*; *téma* – *téme* in *tèma* – *tème* oz. *temè*. Vse druge na ravni leme prekrivne oblike smo obravnavali kot večpomenske, tj. znotraj enega gesla pod skupno iztočnico, kot denimo v primeru samostalnika *vila*:

vila samostalnik

1 kot odraz premožnosti, statusa lepa, razkošna hiša

2 pravljичno ali mitoško bitje

7.1.3 Besedotvorno povezane oblike

Osnovne oblike, ki izhajajo iz istega korenskega morfema – tj. besedotvorno povezane leme – smo obravnavali kot samostojne iztočnice. Pri tem smo ločili dve vrsti razmerij:

- manjšalnice oz. oblike, ki so podobne manjšalnicam
- variantni zapis oblike glede na besedotvorne lastnosti

Med besedotvorno povezanimi oblikami je treba posebej izpostaviti obravnavo manjšalnic oz. besed, ki po obliki delujejo kot manjšalnice, npr. *barvica*, *cvetica*, ki smo jih obravnavali kot samostojne iztočnice, npr. *hlačke*, *jopica*, *koščica*, *mačica*, *stolček*.

⁸⁵ Vidovič Muha (1997: 9) predlaga za oba tipa homonimov, tj. enaklopnice in enakozvočnice, izraz enakoizraznice.

Pomene kot 'majhne hlače', 'majhen stol', če obstajajo, smo obravnavali kot samostojne pomene. Če so taki pomeni osamljeni, npr. *trobentica* – 'majhna trobenta', ali v realni rabi niso izkazani, npr. *marelica* – 'majhna marela', jih v LBS nismo beležili.

Variantna zapisa samostalnikov na -lec/-vec, kjer obe obliki presežeta frekvenčni prag za uvrstitev v LBS, smo prav tako obravnavali kot samostojni iztočnici, saj se predvideva tako pomensko razlikovanje kot različnost besedilnega okolja in rabe:

zajedavec <i>samostalnik</i>	zajedalec <i>samostalnik</i>
1 parazit	1 parazit
2 kdor kaj neupravičeno jemlje ali dobiva	

Na morebitne normativne zadrege, ki izhajajo iz različnih zapisov, na ravni leme nismo opozarjali, ker predvidevamo za ta segment ločen opis v samostojnem elementu gesla oz. slovarskega zavihka.

V LBS 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*). Besedotvorni problem zapisa skupaj, narazen in z vezajem obravnavamo v nadaljevanju.

7.2 Variantnost

Različni zapisi leme, ki se izkazujejo v rabi, so lahko v LBS obravnavani kot samostojne iztočnice ali pa kot njen variantni zapis. Variantnost je lahko izkazana tudi pri izgovornih lastnostih iztočnice, kar prav tako lahko vpliva na njen iztočniški status. Variantnost, ki smo jo upoštevali v LBS, se nanaša predvsem na različne stopnje podomačenja oz. citatnosti prevzetih izrazov in na izkazane različne možne zapise (skupaj, z vezajem in narazen).

7.2.1 Domače – prevzeto

V primeru različnih zapisov leme glede na razmerje domače oz. podomačeno – prevzeto oz. pisno nepodomačeno, npr. *piercing* – *pirsing*, *file* – *fajl*; *džezovski* – *jaz-zovski* itd., smo se odločali za dve oz. lahko tudi več samostojnih iztočnic, če imajo posamezne oblike zadostno število pojavitev v korpusu in izpolnjujejo še druge pogoje.⁸⁶ Odločitev za izdelavo samostojnih iztočnic je povezana s predvidevanjem,

⁸⁶ Novejše tuje besede se navadno v jezik vključujejo postopoma in izkazujejo različne stopnje podomačenosti, npr. za postopek prebadanja telesa za nošenje nakita, sta se po podatkih korpusa Gigafida še najbolj uveljavila nepodomačen zapis *piercing* (670 pojavitev) in delno podomačen *pirsing* (218 pojavitev), redkeje pa *piercing* (4 pojavitve) in *pircing* (8 pojavitev).

da lahko v povezavi z načinom zapisa pričakujemo tudi različnost na domala vseh ravneh leksikalnega opisa, kot je npr. število pojavitev v korpusu, različno kolokabilno okolje, izbira registra, besedilnega tipa, stilne vrednosti, tvorjenje stalnih zvez, frazeologije ipd.

7.2.2 Skupaj – narazen

Problem zapisovanja nekaterih besed skupaj, z vezajem ali narazen smo glede na gradivne podatke in semantično usmerjenost LBS razdelili v dve skupini. V prvi so primeri, ki se v rabi zapisujejo na vse tri možne načine, in sicer ne glede na pomen-sko in normativno ustreznost zapisa, v drugi pa predložne zveze oz. sklopi tipa *na glas* oz. *naglas*, *na lepem* oz. *nalepem* ipd.

7.2.2.1 Skupaj in narazen

Zapise skupaj in narazen, npr. *videonadzor/video nadzor*, *rjavorumen/rjavo rumen*, *žiroračunračun/žiro račun*, smo upoštevajoč merila za določanje večbesednih iztočnic v LBS, obravnavali samo v primeru izkazane pogostnosti in splošne relevantnosti zveze, npr. *žiro račun* kot besednozvezna iztočnica in kot stalna zveza pri iztočnicah *žiro* in *račun*, ter *žiroračun* kot samostojna iztočnica. Zapise narazen smo obravnavali tudi kot kolokacije pod ustrezno skladijsko strukturo in z opozorilom, npr. pri iztočnici *video*: SBZ1<r>v pridevniški rabi</r> sbz0: video [posnetek, projekcija, igrica, spot].

7.2.2.2 Skupaj in z vezajem

Pri zapisu skupaj in z vezajem gre, kot določa obstoječi pravopisni standard (SP 2001: 58, § 503; 59, § 523), za dve pomensko in oblikovno različni besedi: zapis skupaj predvideva, da je en del zloženke pomensko odločujoč, drugi pa ga dopolnjuje, npr. *ekonomskopolitičen* – ekonomska politika, zapis z vezajem pa pomensko prirednost, npr. *ekonomsko-političen* – ekonomski in političen. V rabi zapis pogosto ne sledi pomenski vrednosti besede, kar pomeni, da se lahko oba zapisa uporabljata za različni pomenski vrednosti. Vzrok za to je verjetno v tem, da med pomensko pod- in prirednostjo ni očitnih razlik, če pa so, jih je mogoče registrirati predvsem na kolokacijski ravni in pri tvorjenju stalnih zvez.

V LBS smo torej na ravni samostojne iztočnice obravnavali zapis skupaj, če je lema presegla prag pogostnosti za uvrstitev v bazo. Besed, zapisanih z vezajem, na ravni iztočnic

nismo obravnavali,⁸⁷ ampak vedno le znotraj kolokacijskega niza skupaj z vezajem, npr.: [moralno-, pravno-] etičen, družbeno- [civilni, ekonomski, politični, kritični, socialni], kar ustreza tudi stanju v leksikonu in lematizaciji v korpusu. Pomenske vrednosti različnih zapisov smo registrirali glede na stanje v korpusu, kot kažeta spodnja primera:

rjavorumen <i>pridevnik</i>	rjav <i>pridevnik</i>
1 rumen z rjavimi odtenki; rjavorumena [barva, voda]	1 take barve kot lešnik [temno, svetlo, čokoladno, sivo ...] rjav
2 rjav in rumen	rjavo- [bež, oranžen, rumen, zelen ...]

Na ravni zapisa iztočnice njenega pravopisnega statusa nismo beležili, saj nas je zanimala predvsem v realnih besedilih izkazana pomenska raba in besedilno okolje, ki je s tem povezano, in sicer tudi tistih oblik, ki niso skladne z obstoječo kodifikacijo, če so v rabi dovolj zastopane.

7.2.2.3 Sklopi

Pomensko samostojne leksikalne enote, najpogostejše predložne zveze, ki se v rabi zapisujejo skupaj in/ali narazen, smo v LBS obravnavali na dveh mestih. Zapis skupaj kot samostojno iztočnico, npr. *odkod, naglas* (prislov), in kot večbesedno enoto: *na glas*. Leksikografski opis zveze in njena uvrstitev v zgradbo gesla pa sta bila odvisna od leksikografove presoje, ali zveza potrebuje pomenski opis:

glas <i>samostalnik</i>	naglas <i>prislov</i>
1 zvok	1 tako da se sliši
1.1 ko govorimo	[brati, prebrati] naglas
• na glas	[izgovoriti, izreči] naglas
[smejati se, zasmejati se, zajokati, jokati]	[povedati si, reči si] naglas
na glas	[povedati, reči, govoriti] naglas
[brati, prebrati] na glas	[razmišljati] naglas
[izreči, govoriti, razmišljati, povedati,	1.1 bolj glasno
spregovoriti] na glas	[zelo, bolj, preveč] naglas
• na glas	[povedati, govoriti, razmišljati] naglas
<i>javno</i>	1.2 javno
na glas [povedati]	naglas [povedati, govoriti, reči]
	naglas [priznati]

87 Ker zapis z vezajem v korpusni tokenizaciji ne predvideva samostojne leme, tudi ni bilo mogoče pridobiti podatkov v besedni skici, pač pa zgolj na ravni gramatične relacije.

8 Besedna vrsta

V LBS smo zapisovali besedno vrsto iztočnice v elementu <besedvrs> v zaglavju gesla, merila za zapis besedne vrste pa smo upoštevali tudi pri zapisovanju skladenjskih struktur.

Glede na izrazno vlogo iztočnice kot povezovalca pripadajočih leksikalnih enot smo pri besednovrstni opredelitvi sledili njeni izrazni (leksikonski) in ne vsebinski (pomenski) podobi. Tako je na ravni iztočnice za nas odločilna tista besednovrstna opredelitev, ki jo sugerira primarna skladenjska vloga leksikalne enote in iz nje izhajajoče slovnične kategorije. Besednovrstna opredelitev osnovnih kategorij (samostalnik, glagol, pridevnik, prislov) in podkategorij (deležnik, deležje, glagolnik) je na ta način usklajena z oblikoskladenjskimi oznakami v korpusu in omogoča izrabo besednih skic v orodju SkE ter nadaljnje procesiranje v bazi strukturiranih podatkov.

8.1 Osnovne besedne vrste

Kot rečeno, smo pri določanju besedne vrste sledili oznakam v korpusu Gigafida, predvsem z namenom, da bi med korpusi in orodji usklajene podatke s pridom uporabili tudi pri leksikografski analizi, seveda ob zavedanju, da bodo nekatere korpusne odločitve zahtevale drugačne opredelitve v bazi in slovarju. Ker smo v LBS obravnavali le leksikalnopomenske besede, smo med besednovrstnimi kategorijami upoštevali le samostalnike, glagole, pridevnike in prislove.

8.1.1 Samostalnik (in glagolnik)

Samostalnike smo v elementu <besvrs> opredeljevali z besednovrstno kategorijo *samostalnik*. Glagolnike smo v skladu z njihovimi obliko(slo)vnimi lastnostmi obravnavali kot samostalnike, lahko bi rekli samostalniške oblike glagolov, pri čemer je njihova glagolskost pomenske narave in se pogosto kaže na ravni dejanje – rezultat dejanja:

smrčanje *samostalnik*

1 hropeče dihanje med spanjem

1.1 hropeči glasovi

8.1.2 Glagol

Z besednovrstno oznako *glagol* smo označevali glagole, z nedoločniško obliko pa so zastopane tudi oblike za sedanjik, namenilnik, opisni deležnik na -l in velelnik.

8.1.3 Pridevnik (in deležnik)

Z besednovrstno oznako *pridevnik* smo označevali vse tri pomenske sklope pridevnikov, ki se lahko pojavljajo kot iztočnice, in sicer lastnostne, svojilne in vrstne, npr. *ambiciozen, mamin, gorski*. Razlikovanje med vrsto in lastnostjo smo utemeljevali na pomenski ravni. Posamostaljene pridevnike smo glede na njihovo pridevniško paradigmo obravnavali kot pridevnike. Deležnike, npr. *žgoč, zmrznjen, zapit, zrasel* ipd., smo opredeljevali kot pridevnike v samostojnih iztočnicah, podatek o tem, da gre za deležnike, pa je dostopen v Sloleksu.

8.1.4 Prislov (in deležje)

Lastnostne, okoliščinske (krajevne in časovne) ter količinske prislove kot npr. *burno, brezglavo; navzdol, drevi; mnogokrat* ipd., smo označevali kot prislove. Prav tako smo označevali prislovne oblike glagolskega izvora, tj. deležja na -č, -e, -aje in -ši, npr. *noseč, nehot, kleče*. Kot prislove smo označevali tudi besede kot *rad, všeč* ipd., ki so v Slovnici (Toporišič 2000: 412) opredeljeni kot povedkovnik. Obravnavanje osnovnih kategorij in podkategorij v LBS je opisano v Preglednici 6.

Preglednica 6: *Besednovrstne kategorije in podkategorije v LBS*

Besedna vrsta	Podkategorija	Obravnava v LBS	
		Iztočnica	<besvrs>
samostalnik		<i>čas, spomin</i>	
	glagolnik	<i>bučanje, smrčanje, gradnja</i>	samostalnik
glagol		<i>boleti, dihati</i>	glagol
	deležje	<i>miže, grede</i>	prislov
pridevnik	deležnik	<i>vroč, rojen, zapit, zrasel</i>	pridevnik
		<i>čajen, deževen, afriški</i>	
prislov		<i>dodobra, aktivno</i>	prislov

8.2 Konverzija besednih vrst

Funkcijske spremembe, tj. spremembe, ki nastanejo, ko beseda nastopa v sekundarnih skladijskih vlogah in s tem prevzame tudi skladijske in pomenske lastnosti, ki so za ta položaj značilne, se kažejo v slovenščini kot univerzalen slovnični pojav, vendar pa so hkrati pri posameznih besedah oz. njihovih pomenih, kot kaže raba, različno zastopane. V LBS smo zato besednovrstno konverzijo obravnavali, če se je odražala tako v pomenskih kot skladijskih lastnostih

besede in je bila za obravnavano besedo tipična, tj. ne zgolj sistemsko predvidljiva ali izkazana v osamljenih primerih. Če sta bila pogoja izpolnjena, smo konverzijo v strukturi gesla prikazali bodisi znotraj samostojnega pomena z indikatorjem ali pa v skladenjski strukturi na skladenjskem nivoju. Pri tem smo zabeležili naslednje možnosti:

- posamostaljeni pridevniki, npr. *odvisen, debel, pričakovan*
- (primarni) pridevniki v samostalniški rabi, npr. *latino, mini, afro*
- (primarni) samostalniki v pridevniški rabi, npr. *jazz, evro*
- samostalniki v povedni rabi, npr. *balzam, blagor*
- pridevniki in prislovi v povedni rabi, npr. *pomemben, potreben; naprodaj, zunaj*
- samostalniki v prislovni rabi, npr. *dno, konec, vrh* in
- prislovi v predložni in pridevniški rabi, npr. *mimo hiše, narobe svet*.

8.2.1 Posamostaljeni pridevniki

Posamostaljene pridevnike, v LBS torej z oznako *pridevnik*, ki v določenih položajih opravljajo vlogo samostalnika, smo obravnavali v samostojnem pomenu ali podpomenu in s slovnično oznako *v samostalniški rabi*, npr.:

rujen <i>pridevnik</i>	grajski <i>pridevnik</i>
1 rdeč	1 o gradu
rujno [lice]	grajsko [obzidje, poslopje, dvorišče]
1.1 o vinu	1.1 o prebivalcih gradu
rujno [vince, vino]	grajski [pisar, oskrbnik]
v samostalniški rabi	grajska [gospoda]
[kozarec, kozarček, kapljica ...] rujnega	2 plemiči in graščaki; gospoda
	v samostalniški rabi
	• Če pa bi si kdo rad omislil pravo grajsko življenje vsaj za nekaj dni, ni ovir, da ne bi tega časa res živel tako, kot se za »grajske« najbolj spodobí.

8.2.2 Pridevniki v samostalniški rabi

Besede kot *latino, afro, mini* ipd. smo obravnavali kot pridevnike, ki se, kadar nastopajo v vlogi prilastka, sklanjajo z ničto končnico – po tej lastnosti se razlikujejo od posamostaljenih pridevnikov, ki smo jih obravnavali v prejšnjem razdelku, in se približujejo samostalnikom v pridevniški rabi, ki jih obravnavamo v nadaljevanju. Pri oblikoskladenjskem označevanju korpusa, kot prikazuje Preglednica 7, prihaja pri tem tipu besed do največ nedoslednosti.

Preglednica 7: Besednovrstna opredelitev samostalnik/pridevnik v korpusu Gigafida, Sloleksu in LBS

Lema	Gigafida	Sloleks	LBS
latino	samostalnik	-	pridevnik
afro	samostalnik	-	pridevnik
evro	samostalnik	samostalnik	samostalnik
mini	samostalnik in pridevnik	pridevnik	pridevnik in samostalnik (mini–minija)
jugo	samostalnik	samostalnik	samostalnik
kaki	samostalnik	samostalnik	samostalnik (kaki–kakija)
jazz	samostalnik	samostalnik	samostalnik
jeans	samostalnik in pridevnik	-- ⁸⁸	samostalnik in pridevnik
žiro	samostalnik	samostalnik	pridevnik

Do odstopanj od korpusa prihaja v LBS zato, ker pri tovrstnih lemah v dejanski rabi običajno prevladuje bodisi samostalniška bodisi pridevniška skladenjska vloga. V primeru poenotenja oz. načelne odločitve za samostalniško ali pridevniško besednovrstno oznako bi prihajalo do situacij, kjer bi večina besedilnih realizacij lahko izkazovala prav nasprotno skladenjsko vlogo, kot bi bila zapisana ob iztočnici. V LBS smo se zato odločili, da bomo leme, pri katerih prevladuje pridevniška vloga, izhodiščno opredelili kot pridevnike, konverzijo v samostalnike pa označili na ravni pomena v elementu <oznaka>:

latino pridevnik

nesklonljivo

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, ki mi v roke potisne ključe in me usmeri proti veliki temni gmoti.*

⁸⁸ V Sloleksu najdemo podomačen zapis *džins*, označen s samostalnik.

8.2.3 Samostalniki v pridevniški rabi

Nekateri samostalniki, ki se v večini primerov skladenjsko in pomensko obnašajo kot samostalniki, npr. *jeans*, *kaki*, *jazz*, *evro*, se lahko v položaju pred samostalnikom obnašajo kot pridevniki, ki se sklanjajo z ničto končnico.⁸⁹ Taka vloga je lahko povezana s pomenskimi lastnostmi samostalnika kot v primeru *jeans* (1.2 o barvi), kjer smo ustrezno označili (pod)pomen, ali pa ostaja na ravni skladnje:

jeans *samostalnik*

1 tkanina

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]

Na pridevniško vlogo samostalniških besed, natančneje na izražanje količine, smo opozarjali pri skladenjskih strukturah z zapisom Kol-sbz: par [besed, stoletij, ur], rezina [ananasa, avokada], kos [potice, torte], kilometer [stran, naprej, nazaj] ipd.

8.2.4 Samostalniki, pridevniki in prislovi v povedni rabi (obravnavo povedkovnika)

Čeprav Slovenska slovnica (Toporišč 2000: 412) določa povedkovnik kot samostojno besedno vrsto, se v LBS zanjo nismo odločili. Gre namreč za izrazito skladenjsko kategorijo, saj se lahko na predvidenih skladenjskih pozicijah, tj. za pomožnim glagolom *biti*, pojavijo elementi iz tako rekoč vseh drugih besednih vrst, z vpeljavo te kategorije pri oblikoskladenjskem označevanju pa bi označevalnikom povzročili precej težav, zelo verjetno brez zadovoljivega rezultata (Krek 2010: 217/218). Take tipične skladenjske položaje smo zato obravnavali na pomenski in/ali skladenjski ravni.

Za ta položaj zelo tipične in hkrati heterogene besede, ki izkazujejo lastnosti različnih besednovrstnih kategorij, kot npr. *rad* in *všeč*, smo obravnavali v skladu z oblikoslovnimi oznakami v Sloleksu kot prislove.

⁸⁹ V SSKJ so tovrstni skladenjski položaji označeni z »neskl. pril.«, in sicer pod samostalniškimi iztočnicami: oznaka je torej veljala za besede, ki niso bile prepoznane kot pridevniki.

Besede kot *balzam*, *blagor*, *mraz*, *tema*, *groza*, *škoda* ipd., ki pogosto zasedajo povedkovniški položaj, so na ravni iztočnice opredeljene kot samostalniki, v strukturi pa je na povedno rabo opozorjeno v restrikciji, vendar samo v primerih, ko je v takem položaju beseda, ki je obravnavana kot iztočnica, sicer ne.

tema *samostalnik*

1 odsotnost svetlobe

Vez-gbz SBZ1<r>v povedni rabi</r>

[je] tema, je [trda, čista, popolna] tema

V strukturi geselskega članka smo na tipične povedkovniške vloge pri pridevniki-kih opozorili v slovničnih oznakah neposredno za indikatorjem pomena, npr. pri pridevniki *aktiven* in *kriv*, kjer je skladijska vloga razvidna tudi iz pomenske sheme, ali na ravni skladijskih struktur, ki so potrjene z ustreznimi leksikalnimi zapolnitvami v kolokacijah: [dan] je deževen, [primeri] so ilustrativni, [vprašanje] je logično, [mnogi] so skeptični; dialog je [potrben, vzpostavljen]; Vez-gbz PBZ1: [biti] pasji, [biti] tešč itd.

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

sbz1 Vez-gbz PBZ1

[član] je aktiven, [društvo] je aktivno, [sekcija] je aktivna

Vez-gbz PBZ1

[biti, ostati] aktiven

kriv *pridevnik*

1 odgovoren za kaj

1.1 za negativna dejanja

v povedni rabi

če je ČLOVEK kriv, je odgovoren za nastanek NEGATIVNE SITUACIJE

Vez-gbz PBZ1

[biti, čutiti se, počutiti se] kriv

8.2.5 Drugi primeri

Pri opredelitvi besedne vrste iztočnice in v skladijskih strukturah smo se pri nekaterih mejnih primerih, ki so navedeni v nadaljevanju, odločali dogovorno, pri

čemer smo načeloma sledili merilu besednovrstne označenosti na ravni leme, kot je izkazana v Sloleksu.

Samostalniki kot npr. *konec*, *vrh*, *dno* ipd., se v nekaterih položajih obnašajo kot prislovi, na drugi strani pa imamo nekatere izvirne prislove, ki se v določenih položajih rabijo kot predlogi ali pridevniki. Na konverzijo smo opozarjali pri strukturah:

<i>Čredi se približajo le koncem jeseni, v času parjenja.</i>	SBZ6<r>v prislovni rabi<r>sbz2
<i>Zaključne izpite bodo opravljali koncem naslednjega tedna.</i>	SBZ6<r>v prislovni rabi<r>sbz2
<i>Novinarskih vprašanj mu je bilo vrh glave.</i>	SBZ1<r>v prislovni rabi<r>sbz2
<i>Drugače je pa pri nas narobe svet.</i>	RBZ<r>v pridevniški rabi<r>sbz0

V nekaterih primerih, kjer prihaja tudi do manj tipičnih prenosov, npr. samostalnik – članek (*oni pa v jok – Jok, brate!*), smo prenos obravnavali bodisi znotraj samostojnega pomena z oznako, v kateri smo izpostavili, da nastopa beseda v vlogi besedilnega modifikatorja (*poudarno* ipd.), ali pa znotraj frazeološkega razdelka kot frazeološko enoto. Obravnavo po konverziji povezanih besednih vrst v LBS prikazuje Preglednica 8.

Preglednica 8: Obravnavanje po konverziji povezanih besednih vrst

Kategorija	Primer	Obravnavanje v LBS	
		na ravni iztočnice	v strukturi gesla
posamostaljeni pridevniki	<i>rujen, grajski, debel, pričakovan, odvisen, dvoumen</i>	pridevnik	<oznaka>v samostalniški rabi </oznaka> na ravni pomena
	<i>revolucionaren, trivialen, tipičen</i>		<restrikcija>v samostalniški rabi </restrikcija> na ravni skladenjske strukture
nesklonljivi pridevniki	<i>latino, mini, afro</i>	pridevnik	<oznaka>v samostalniški rabi </oznaka> na ravni pomena
pridevniki v povedni rabi	<i>pasji, aktiven, pomemben, potreben, odločilen</i>	pridevnik	<oznaka>v povedni rabi</oznaka> na ravni pomena
	<i>ambiciozen, poučen</i>		na ravni struktur tipa: Vez-gbz PBZ1 [biti, postati, ostajati] ambiciozen sbz1 Vez-gbz PBZ1 [načrt, cilj] je ambiciozen
samostalniki v povedni rabi	<i>vonj, barva</i>	samostalnik	<oznaka>v povedni rabi</oznaka> na ravni pomena
	<i>tema, škoda, balzam, blagor</i>		<restrikcija>v povedni rabi</restrikcija> na ravni skladenjske strukture tipa: Vez-gbz SBZ1<r>v povedni rabi</r> sbz1 Vez-Gbz SBZ1<r>v povedni rabi</r>
samostalniki v pridevniški rabi	<i>jazz, evro, jeans, celofan, jugo</i>	samostalnik	<oznaka>v pridevniški rabi </oznaka> na ravni pomena; <restrikcija>v pridevniški rabi </restrikcija> na ravni skladenjske strukture
samostalniki v prislovni rabi	<i>dno, vrh, konec</i>	samostalnik	<restrikcija>v predložni rabi </restrikcija> na ravni skladenjske strukture
prislovi v povedni rabi	<i>naprodaj, mimo, zunaj</i>	prislov	<oznaka>v povedni rabi</oznaka> na ravni pomena
			<restrikcija>v povedni rabi </restrikcija> na ravni skladenjske strukture
prislovi v predložni rabi	<i>mimo</i>	prislov	<restrikcija>v predložni rabi </restrikcija> na ravni skladenjske strukture

II. POMENSKI NIVO

LBS, kot smo jo zasnovali pri projektu SSJ, je v prvi vrsti namenjena slovarskemu tipu jezikovnih podatkov. Predvsem nas je torej zanimalo, kaj določena beseda oz. leksikalna enota pomeni, posledično pa tudi vse, kar izhaja iz njenih pomenskih lastnosti, tj. obnašanje v sobesedilu, stilistični potencial, omejitve v rabi ipd. Izhajajoč iz teoretičnih smernic, ki jih podrobneje opišemo v nadaljevanju, je pomenski nivo v LBS hierarhično nadrejen vsem drugim podatkom, ki jih beležimo v zvezi s posamezno besedo v iztočnici. To, za razliko od iztočniške ravni, kjer smo izhajali iz izrazne podobe besede, pomeni, da je na pomenskem nivoju izhodišče za beleženje jezikovnih podatkov vedno konkretni, na podlagi korpusne analize registrirani pomen besede.

Pomembno za razumevanje semantičnocentrične zasnove LBS (prim. Hanks 2013), je ločevanje med pomenskim potencialom, ki ga ima beseda v svojem realnem življenju, tj. zunaj slovarja ali slovarske baze, na eni strani, in leksikografskim opisom, tj. leksikografsko razlago, ki je metaslovarski konstrukt. V prvem primeru, tj. v realni sporočanjski situaciji, z razbiranjem pomenskih tendenc govornici navadno nimamo težav. Običajno si pomagamo z vedenjem, ki ga delimo s sogovornikom, in z vrsto drugih signalov, ki nam jih sugerirata (so)besedilo in situacija, če sporočanjski proces opišemo zelo poenostavljeno. V drugem primeru, tj. pri oblikovanju slovarskih razlag, pa zagovarjamo stališče, da besede same na sebi nimajo čistih pomenov, pač pa le tendence, da nekaj pomenijo (Hanks 2009: 11). Leksikografski opis pomena oz. pomensko razlago zato razumemo kot poskus približevanja pomenskih tendenc naslovniku, ali kot pravi Bolinger (1965: 572) in z njim 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 v sporočanjem procesu in znotraj slovarja/baze smo pri oblikovanju LBS upoštevali stališče, da pomenske tendence besed vplivajo na izbire v besedilnem okolju, kar se odraža v stavčnih vzorcih, skladenjskih strukturah in kolokacijah, hkrati pa tudi v omejevanju nekaterih slovničnih kategorij, kot so denimo (ne)števnost, število, glagolska oblika ipd. Ta povezanost se kaže tudi v zgradbi LBS, kjer je vsak pomen besede predstavljen z vidika vseh predvidenih informacij na posameznem nivoju, eksplicitno pa zlasti v zapisu pomenske sheme, ki predstavlja izhodišče za oblikovanje stavčnih razlag.

Elementi, ki v shemi DTD vsebujejo podatke o pomenu, so: pomenski meni, indikator (pri pomenu in podpomenu, stalni zvezi in frazeološki enoti), definicija ter oznaka. Izhodišče pomenskega opisa predstavlja indikator, zato se nanaša na vse leksikalne enote, ki jih v LBS razumemo kot samostojne pomenske enote, obravnavane v okviru določene eno- ali večbesedne izrazne enote (prim. Atkins in Rundell 2008: 162).

V prvem delu razdelka se bomo osredotočili na opis strategij pomenske členitve, kjer smo težili k podrobnemu pomenskemu členjenju, zato smo več pozornosti namenili večpomenskim besedam in ugotavljanju osnovnih pomenov in podpomenov ter razmerij med njimi. Dotaknili se bomo tudi primerov, kjer pomenov ni smiselno podrobneje členiti, ter strategij za čim bolj enotne rešitve pri pomenski členitvi tematsko in besednovrstno povezanih iztočnic. Drugi del je namenjen predstavitvi strategij pomenskega opisa, čeprav sta oba postopka povezana. Pomenski opis zahteva premisleke o ubeseditvi pomena in o uporabi strategij, ki omogočajo konsistentnost pomenskih opisov v celotni slovarski bazi. V tem razdelku podrobneje opišemo vlogo pomenskega menija v spletnem slovarju, pomenske indikatorje in posamezne tipe razlag. Obsežen del tega poglavja je namenjen opisu vloge, zgradbe in načina izdelave pomenskih shem po posameznih besednih vrstah. Pomenske sheme predstavljajo v LBS izhodišče za oblikovanje t. i. stavčnih razlag (ang. *full sentence definitions*), značilnih zlasti za slovarje, ki so namenjeni učenju tujega jezika. V slovarski opis, kot ga predvideva LBS, jih vključujemo zaradi možnosti združevanja pomenskih in skladenjskih informacij na način, ki je uporabniku razumljiv in hkrati učinkovit v smislu čim boljšega izkoristka pri uporabnikovem reševanju konkretnega leksikalnega problema.

Leksikografske strategije pri obravnavi stalnih zvez in frazeoloških enot prikazemo v samostojnih poglavjih, kjer se osredotočimo na vrste registriranih stalnih zvez, izpostavimo merila za njihovo identifikacijo ter opredelimo tipe leksikalnogramatičnih podatkov, ki jih nanje pripenjamo. V zvezi s stalnimi zvezami in frazeološkimi enotami opredelimo tudi sistem njihovega vključevanja v geselsko zgradbo in razmerja do drugih tipov večbesednih zvez, ki jih v LBS ne obravnavamo kot samostojne pomenske enote, zlasti torej do kolokacij, razširjenih kolokacij in skladenjskih zvez.

Posebej v okviru pomenskega nivoja opišemo sistem oznak in način njihove integracije v slovarsko bazo, vključno z načinom ubeseditve in možnostjo navezave na druge v bazi obstoječe pomenske in slovnične podatke.

9 Pomenska členitev besede

Pri pomenski členitvi besede, zastopane v iztočnici, težimo k beleženju najočitejših pomenskih tendenc. Predpostavljamo torej, da večina pomenske informacije konkretne besede izhaja iz njenega bodisi neposrednega bodisi širšega besedilnega okolja. Ključni element za pomensko razdvoumljanje nam tako predstavlja sobesedilo, ali kot pravi Hanks (2009: 11), imajo besede potencial, da ustvarjajo pomene, če jih postavimo v neko sobesedilo – pomeni so namreč povezani z določenimi nizi sintagmatskih vzorcev, ki jih je mogoče odkriti s podrobno korpusno analizo. Na tej podlagi izdelana pomenska členitev ni absolutna in kot taka tudi ne more biti identična v različnih slovarskih opisih, ampak je (vsakokratni) izdelek leksikografa in temelji na njegovi interpretaciji realnih rab, ki se ustvarjajo v prototipičnih vzorcih in njihovih alternacijah. Leksikograf pomenske tendence besed ubesedi v obliki pomenskega opisa, ki v primeru LBS vključuje kombinacijo pomenskega indikatorja in bodisi klasične slovarske razlage bodisi pomenske sheme.

Pri pomenskem členjenju smo v LBS sledili naslednjim osnovnim načelom:

- Besede v LBS pomensko členimo v dveh stopnjah, tj. na *osnovne pomene* (oz. na glavni pomen) in na *podpomene*. Podpomene predvidevamo samo pri iztočnici; stalne zveze in frazeološke enote, ki jih sicer obravnavamo kot leksikalne enote, imajo lahko več osnovnih pomenov, na podpomene pa jih ne členimo.
- O *glavnem pomenu* govorimo, ko ima beseda en sam pomen ali ko ima tak pomen samo podpomene; o *osnovnih pomenih* pa, kadar ima beseda več pomenov, ki so lahko členjeni na podpomene.
- Podpomeni so glede na glavni pomen ali osnovne pomene v hierarhičnem razmerju.
- Osnovne pomene razvrščamo na podlagi ocene pogostnosti kolokatorjev, ki sodijo k obravnavanemu pomenu. Na enak način, torej od bolj k manj pogostemu, razvrščamo tudi podpomene, vendar vedno znotraj nadrejenega osnovnega pomena. V primeru, da je kateri od podpomenov pogostejši od nadrejenega osnovnega pomena, se njegova aktualnost podreja pomenski in ne frekvenčni hierarhiji.
- Za določanje pomenov in podpomenov je odločilno stanje, kot ga izkazuje referenčni korpus FidaPLUS/Gigafida oz. analiza najmanj 150 konkordanc za obravnavano besedo.

- Pomenske lastnosti obravnavane besede smo skušali členiti čim bolj podrobno, pri čemer smo načeloma težili k večjemu številu osnovnih pomenov in manjšemu številu podpomenov.
- Hkrati smo sledili tudi beleženju tipičnosti – tj. prepoznavni količini analognih pomenov – in ne k individualnim posebnostim in enkratnim pomen-skim rabam.

9.1 Členjenje pomenov

Posamezni slovarji, celo taki, ki so namenjeni istemu uporabniku in imajo primerljiv obseg, različno podrobno členijo posamezne besede na pomene (prim. Krek 2010a). Absolutni odgovor na vprašanje, koliko pomenov ima beseda ali kako bi morali biti razvrščeni, ne obstaja (Moon 1987: 86). Pomeni večpomenskih besed so namreč lahko predstavljeni v različnih stopnjah razdrobljenosti. Kako podrobno bo konkretni slovar pomensko členil besede, je odvisno od njegovega namena, obsega, ciljnega uporabnika, specializiranosti itd. Ker je LBS namenjena različnim slovarjem oz. različnim uporabnikom, sledi tudi v pomenski členitvi temu cilju. Ena izmed pomembnih lastnosti, ki jo izpostavlja sodobna leksikografska literatura v zvezi s tem, je, da je v fazi oblikovanja slovarske baze smiselno pomene razdeliti čim bolj podrobno, saj je, kot utemeljmeta Atkins in Rundell (2008: 268), v fazi združevanja lažje združevati povezane leksikalne enote, kot pa obrniti postopek in poskušati razčleniti krovni pomen na manjše enote. S podrobno razčlenjenimi pomeni v slovarski bazi imamo namreč na voljo celoten spekter možnosti, iz katerih lahko izpeljemo več različnih slovarskih informacij, lahko uporabimo različne strategije vizualizacije ali prilagodimo količino podatkov različnim stopnjam zahtevnosti slovarskega uporabnika. Hkrati smo pri oblikovanju LBS upoštevali tudi nasprotno vodilo: členjenje pomenov mora slediti še obvladljivim in dovolj homogenim enotam, ki predstavljajo zadostno število analognih primerov v dejanski jezikovni rabi.

9.1.1 Enopomenske iztočnice

Čeprav je koncept enopomenskosti, če ga obravnavamo v luči Firthove izjave »vsaka beseda, ki je uporabljena v novem kontekstu, je nova beseda« (cit. po Béjoint 1988: 14) veliko kompleksnejši, kot ga na prvi pogled sugerira Crystalova (1985) definicija: »ki vsebuje en sam pomen«, velja, da je velik delež besed v jeziku enopomenskih,⁹⁰

⁹⁰ Jezikoslovci, ki so ugotavljali kvantitativno zastopnost enopomenskosti v angleščini in drugih jezikih, so prihajali do zelo različnih rezultatov, pri čemer je vzrok za diskrepanco pogosto različno pojmovanje in nedoločenost pojma enopomenskost (Béjoint 1988: 13). Kot še navajata Béjoint (prav tam: 22) in Moon (1987a: 176), leksikografi navadno upoštevajo drugačna merila za določanje besedne večpomenskosti kot teoretični lingvisti.

ali kot pravi Atkinsova (1992/93: 6): »Po letih korpusne leksikografije se nagibam k prepričanju, da so številne besede pravzaprav enopomenske in da uporabnikom ne delamo usluge s tem, ko z razčlenjevanjem na pomene trdimo drugače.«

Ob tem velja poudariti, da enopomenske besede navadno niso najpogostejše besede. Pogostnost besede je namreč v veliki meri povezana z njeno večpomenskostjo in s tem, kakšna je njena kolokabilna moč oz. sposobnost povezovanja z drugimi besedami v smislu tvorjenja večbesednih leksikalnih enot. V LBS je med 2.500 besedami z enim samim pomenom opredeljenih 865 besed, kar predstavlja dobro tretjino. Besednovrstno prevladujejo med enopomenskimi samostalniške besede, manj pa je enopomenskih glagolov in pridevnikov, vendar pa je tako razmerje najverjetneje povezano s procentualno zastopanostjo besednih vrst v LBS in ga ni mogoče generalizirati na besedišče v celoti. Načeloma stanje v LBS potrjuje nekatere že znane ugotovitve, namreč, da so enopomenske besede navadno specifične za določeno specializirano področje, da gre za besede, ki poimenujejo novo predmetnost in da hkrati (še) niso razvile prenesene rabe.

V LBS smo pri enopomenskih besedah upoštevali načelo, po katerem je osnovni pomen vedno tudi osrednji ali splošni in hkrati tudi najbolj dobessedni, saj se predvideva, da se metaforična raba in drugi pomenski odmiki bodisi niso razvili bodisi niso izkazani v tolikšni meri, da bi predstavljali ustaljen pomenski potencial slovenščine. Meja, kdaj posamezni pomen določene besede, zlasti ko gre za prenesene rabe, postane del besedišča, kljub referenčnim korpusom, ki omogočajo analizo obsežne količine primerov, ni vedno jasno določljiva. Čeprav so nam orodja, ki smo jih uporabljali za analizo korpusa, pri detektiranju pomenske razpršenosti v veliko pomoč, saj lahko ugotavljamo pomensko razpršenost glede na časovna obdobja in preverjamo ustaljenost v različnih besedilnih tipih in žanrih, je končna odločitev o tem, ali se določen pomenski odtenek že uporablja splošno ali pa je zgolj individualne oz. prehodne narave, v pristojnosti leksikografa. V LBS smo pri tem skušali slediti načelu pogostnosti oz. opaznemu številu analognih rab v daljšem časovnem obdobju in različnih virih.

9.1.2 Večpomenske Iztočnice

Za večpomenske iztočnice, izhajajoč iz dvostopenjskega hierarhičnega členjenja, v LBS velja, da imajo vsaj en glavni pomen s podpomeni ali pa imajo več osnovnih pomenov, ki so lahko nadalje členjeni s podpomeni. Sistema za določanje osnovnih pomenov in meril podpomenskosti ni mogoče določiti absolutno, saj ima vsaka

beseda svojo individualno pomensko zgradbo ali notranjo pomensko logiko, ki jo določa več parametrov. V LBS smo upoštevali predvsem naslednje:

- podrobnost pomenske členitve, in sicer na podlagi skladenjskih oz. pomensko-skladenjskih elementov večpomenskosti;
- hierarhičnost oz. upoštevanje medsebojne znotrajpomenske (ne)povezanosti, ki se kaže v ohranjanju/spreminjanju bolj ali manj očitnih pomenskih elementov;⁹¹
- pričakovana pomenska povezanost oz. soodvisnost – t. i. regularna ali sistemska večpomenskost (prim. Apresjan 1973, 1992 in 2002);
- izpričanost (in pogostnost) posameznega pomena glede na realno jezikovno rabo.

9.1.2.1 Kazalci večpomenskosti

V nadaljevanju predstavimo nekaj prepoznavnih slovničnih in vsebinskih meril večpomenskosti, ki smo jih upoštevali pri pomenski členitvi. Med formalnimi oz. slovničnimi pokazatelji izpostavljamo pomensko specifične slovnične lastnosti, določene s skladenjsko vlogo besed, na vsebinski ravni pa vlogo sobesedila pri prepoznavanju besedne večpomenskosti in nekatere vsebinske lastnosti – t. i. *pomenske elemente*, ki smo jih določali na podlagi korpusne analize.

9.1.2.1.1 Slovnični kazalci večpomenskosti

Eno od osnovnih oblikoskladenjskih meril večpomenskosti je besednovrstna opredelitev besede, določena z njeno tipično skladenjsko vlogo v stavku (prim. Vidovič Muha 2000: 30). Kot odraz večpomenskosti je zato treba upoštevati spremembe v skladenjski vlogi besede, npr. pri posamostaljenih pridevniki, kot so *debel*, *odvisen*, ki v skladu s samostalniško vlogo prevzamejo tudi pomenske lastnosti samostalnika: 'kdor ima preveč telesne maščobe'; 'odvisnik; zasvojenec' (prim. poglavji 8, Besedna vrsta in 8.2, Konverzija besednih vrst). Sicer pa se kot pomenskorazločevalne kažejo za posamezno besedno vrstno njene imanentne slovnične kategorije, ki prek potencialne stalnosti skladenjskih funkcij postajajo kategorialne (pomenske oz. slovarske) lastnosti besednih vrst (Vidovič Muha 2000: 32), kot prikazuje Preglednica 9:

91 Za izraz pomenski element (gl. tudi poglavje 9.1.2.1.2.3 Pomenski elementi) smo se odločili, da bi vzpostavili razliko do izraza pomenska komponenta, ki ga uporablja komponentna teorija, ter izraza pomenska sestavina, kot je določen znotraj kategorialnega slovarskega pomena (Vidovič Muha 2000: 30) in klasične slovarske razlage, saj ga ne uporabljamo v neposredni navezavi s katero od teh teorij.

Preglednica 9: Slovnčni kazalci večpomenskosti v LBS

Besedna vrsta	Slovníčnopomenska kategorija		Mesto obravnave v LBS	Primer
samo-stalnik	spol		iztočnica (povezava na leksikon besednih oblik)	<i>cigaret, cigareta, slikar, slikarka, čistilec, čistilka</i>
	podspol živosti in človeškosti		pomen	<i>avtoriteta, bas, čarovnik</i>
	števnost		pomen	<i>aplikacija, gradnja</i>
	število		pomen	<i>lisica</i>
glagol	intenčnost	neprehodnost	pomen, pomen-ska shema, stavčni vzorec	<i>strmoglaviti, angažirati, cepiti, demonstrirati</i>
		prehodnost		
	sprememba udeleženske zgradbe		pomen, pomen-ska shema, stavčni vzorec	
	vid		iztočnica (povezava na leksikon besednih oblik)	<i>obiskovati, obiskati, krvaveti, izkrvaveti</i>
pridevnik	vrsta	lastnost,	pomen, struktura, kolokacija, SZ	<i>lesen, briljanten</i>
		sestavinskost : vrsta : svojina		
	stopnjevanje elativ	intenzifikator	pomen, oznaka	<i>čeden, baje, slep, spoštljiv</i>
		iztočnica		<i>prenasičen, prenatrpan</i>
prislov		intenzifikator	pomen, oznaka	<i>bridko, gladko</i>

9.1.2.1.1 Samostalnik

Primer pomenske členitve na podlagi prehoda nežive entitete (oz. lastnosti, predmeta) v živo oz. človeško entiteto (nosilca te lastnosti) ponazarjata pomenski členitvi pri samostalniki *avtoriteta* in *čarovnik*.

avtoriteta *samostalnik*

1 vpliv; moč

1.1 vpliven, ugleden človek

avtoriteta za področje [komunikacij]

[biti, postati] avtoriteta

čarovnik *samostalnik*

- 1 pravljичno bitje
- 2 kdor kaj zelo dobro obvlada
- 3 računalniški program
 - čarovnik [se zažene, se prikaže, se pojavi, se odpre]
 - [zagnati, pognati, uporabiti, uporabljati, odpreti] čarovnika

Po pričakovanju lahko do takih pomenskih prenosov prihaja tudi na ravni stalne zveze ali frazeološke enote, npr. pri samostalniku *buča*:

buča *samostalnik*

- 1 sadež: plod
- 2 glava
- **bribtna buča**
pameten, bister, iznajdljiv človek

Različnost samostalniških pomenov se pogosto odraža v kategoriji števnosti, kot denimo v spodnjih primerih:

aplikacija *samostalnik*

- 1 računalniški program
 - [spletna, internetna] aplikacija
 - [programska, računalniška] aplikacija
 - 2 prenos teorije v prakso; uporaba
- neštevno
- aplikacija [znanja, metode]

moškost *samostalnik*

- 1 lastnost moških
- neštevno
- [atribut, ideal, simbol, kult, ikona, koncept] moškosti
- 1.1 moški spolni organ; penis
- števno
- navadno v ednini
- prijeti za [njegovo] moškost
- pogled na [njegovo, svojo] moškost

Omenjali smo tudi samostojne pomene z ustaljeno številsko obliko samostalnika:

lisica *samostalnik*

- 1 žival
 - 2 priprava za vklepanje
- v množini
- 2.1 o vozilih
 - [natakiniti, nadeti, odkleniti] lisice
 - 2.2 o ljudeh
 - [vkleniti] v lisice
 - [prikleniti] z lisicami

9.1.2.1.1.2 Glagol

Pri glagolih so lastnosti, ki opozarjajo na večpomenskost, zlasti intenčnost⁹² oz. vezljivostne lastnosti glagola,⁹³ ki se strukturno in pomensko realizirajo v stavčni povedi, in glagolski vid, v LBS izražen že na ravni osnovne oblike glagola v iztočnici, npr. *obiskovati* – *obiskati*, *odganjati* – *odgnati*; *krvaveti* – *izkrvaveti*. Za razliko od klasičnih slovarjev, je tudi SSKJ, kjer se podatek o glagolskem vidu uporablja kot besednovrstna opredelitev, v LBS do njega dostopamo prek povezave na Sloleks.

Tipična sprememba pomena, ki vključuje glagolsko intenčnost, je pridobitev glagolske prehodnosti kot npr. pri glagolih *angažirati* in *strmoglaviti*:

angažirati *glagol*

1 dejavno se vključevati; prizadevati si

če se ČLOVEK angažira pri neki DEJAVNOSTI ali pri določeni NALOGI, vložijo vanjo veliko svoje energije, znanja ali družbenega ugleda

- kdo se angažira pri čem
- kdo se angažira v čem (itd.)

1.1 povabiti k sodelovanju; najeti, zaposliti

če ČLOVEK ali USTANOVA angažira STROKOVNJAKA, ga zaradi njegovega znanja ali usposobljenosti povabi ali najame za sodelovanje pri izvajanju konkretne DEJAVNOSTI ali PROJEKTA

- kdo angažira koga
- kdo angažira koga za kaj
- kdo angažira koga pri čem
- kdo angažira koga v čem
- kdo angažira koga kot koga
- kdo angažira koga, da ...

strmoglaviti *glagol*

1 zrušiti se na zemljo

če LETALO strmoglavi, se z velike višine zruši na zemljo, pri čemer se uniči ali močno poškoduje

- kdo/kaj strmoglavi
- kdo/kaj strmoglavi nad čim (itd.)

92 O tem npr. Vidovič Muha (2000: 34): »Sposobnost glagola, da spreminja svoje stavčnovezljivostno vedenje lahko vpliva nazaj na njegovo pomenskosestavinsko zgradbo in s tem seveda na nastanek večpomenskosti.«

93 V SSKJ so taki pomeni lahko opredeljeni s slovničnimi kvalifikatorji, npr. kot: *prehodno* in *brezosebno*, vendar ne vedno ustrezno, npr. v razmerju do splošnega vršilca dejanja (prim. frazariti, govoriti, hoditi itd.).

1.1 zrušiti letalo

če ČLOVEK ali SKUPINA LJUDI strmoglaví LETALO, povzroči, da se zruši na zemljo oziroma na nek točno določen cilj, navadno v vojaških ali terorističnih napadih

- kdo strmoglaví kaj

9.1.2.1.1.3 Pridevnik

Pri pridevnikih se sposobnost stopnjevanja kot slovničnopomenska lastnost kaže v ločevanju generičnega pomena oz. vrstnosti, ki ga ni mogoče stopnjevati, od pomena lastnosti, kjer je ta možnost pogosto tipična. Če torej pri pridevniku (v iztočnici zastopanem z obliko na -en) prihaja do prehoda vrstnosti v lastnostnost in obratno, smo v LBS na pomensko razliko opozorili v samostojnem pomenu, pri čemer se pri lastnostnih pomenih na ravni kolokacij in struktur (rbz PBZ0) lahko stopnjevanje s prislovi izraža kot tipično:

briljanten *pridevnik*

1 ki zbujá občudovanje

1.1 o dosežkih

- briljantna [obramba, tehnika]
- briljanten [esej]
- [naravnost, absolutno] briljanten

1.2 o umetnikih ali znanstvenikih

- briljanten [pianist]
- briljanten [znanstvenik, matematik, fizik]
- [najbolj] briljanten

2 ki vsebuje briljante

- briljantni [prstan, nakit]

Večpomenskost na podlagi sposobnosti stopnjevanja – natančneje izguba te možnosti pri prehodu pridevnika v označevanje vrstnosti – je očitna na ravni stalnih zvez, ki jih v LBS obravnavamo kot samostojne leksikalne enote, npr. pri pridevniku *kisel*:

kisel *pridevnik*

1 takega okusa kot limona

1.1 o jedeh in živilih

- *kislo zelje*
- *kisla smetana*
- *kisla juha*
- *kisla kumarica*

Sposobnost stopnjevanja se pri nekaterih pridevnikih lahko pomensko osamosvoji in postane podobna prislovnemu pomenu stopnje: govorimo o samostojnih pridevniških pomenih oz. pridevnikih v vlogi intenzifikatorjev ali poudarjalcev (ang. *intensifier*), npr. pri pridevniki *bajen* in *čeden*:

bajen <i>pridevnik</i>	čeden <i>pridevnik</i>
1 zelo velik	1 ki lepo zgleda; lep;
izraža poudarek	privlačen
bajni [denarci]	1.1 o človeku
bajen [zaslužek, honorar, dobiček, donos, dobiček]	1.2 o stvareh
[naravnost] bajen	2 ki obstaja v veliki količini
2 mitičen; pravljichen	izraža poudarek
	čeden kupček [denarja]
	čedna vsota [denarja]
	čedni [denarci]

9.1.2.1.1.4 Prislov

Slovničnopomenske lastnosti, ki jih je mogoče pripisati prislovom (prim. Vidovič Muha 2000: 36), izhajajo iz njihove modifikacijske (zlasti vrstni, lastnostni in količinski prislovi) ali t. i. propozicijske (krajevni in časovni prislovi) stavčne vloge. V prvem primeru se prislovi razvrščajo ob glagole, pridevnike in prislove, v drugem pa določajo okoliščine glagolskega dejanja. Te lastnosti se navadno odražajo kot osnovne pomenske lastnosti oz. skupine prislovov, medtem ko njihova notranja pomenska členitev sledi razmerjem, kot so dobesedno – preneseno, splošno – specializirano itd., o čemer več v poglavjih, ki sledijo. Na večpomenskost kot posledico sprememb pri imanentnih slovničnih lastnostih prislovov je mogoče opozoriti pri osamosvajanju pomena stopnjevanja, npr. *bridko*, *gladko*, in pri prehodu v povedkovnik ali predlog, npr. *mimo*:

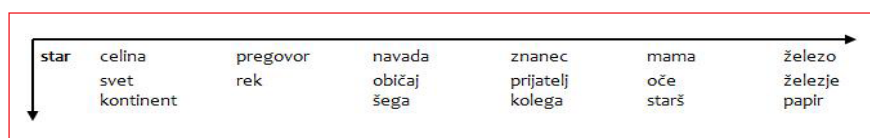
gladko <i>prislov</i>	bridko <i>prislov</i>
1 brez težav; z lahkoto	1 žalostno; nesrečno
izraža poudarek	2 močno; zelo
2 tako da ni več grudic ali izboklin	izraža poudarek
mimo <i>prislov</i>	
1 v smeri proti komu/čemu	
2 v smeri ob kom/čem	
RBZ<r>v predložni rabi</r>sbz2	
mimo [gola; koša]; mimo [odcepa]	
3 minulo; preteklo	
v povedni rabi	
4 nepremišljeno ali neprimerno	
v povedni rabi	

9.1.2.1.2 Sobesedilni in pomenski kazalci večpomenskosti

Vsebinskih kazalcev večpomenskosti, kamor vključujemo elemente besedilnega okolja obravnavane besede, zlasti torej kolokatorje ter udeležensko zgradbo pomena, posebej pa tudi prepoznavne vsebinske elemente posameznih besednih pomenov, ni mogoče obravnavati ločeno od njihovih slovničnih lastnosti, saj sta oba nivoja med seboj neločljivo povezana. Za samostojno obravnavo pomenskih kazalcev smo se odločili zaradi jasnejše opredelitve posameznih leksikografskih postopkov, ki se tičejo prepoznavanja pomenov večpomenske besede.

9.1.2.1.2.1 Kolokatorji

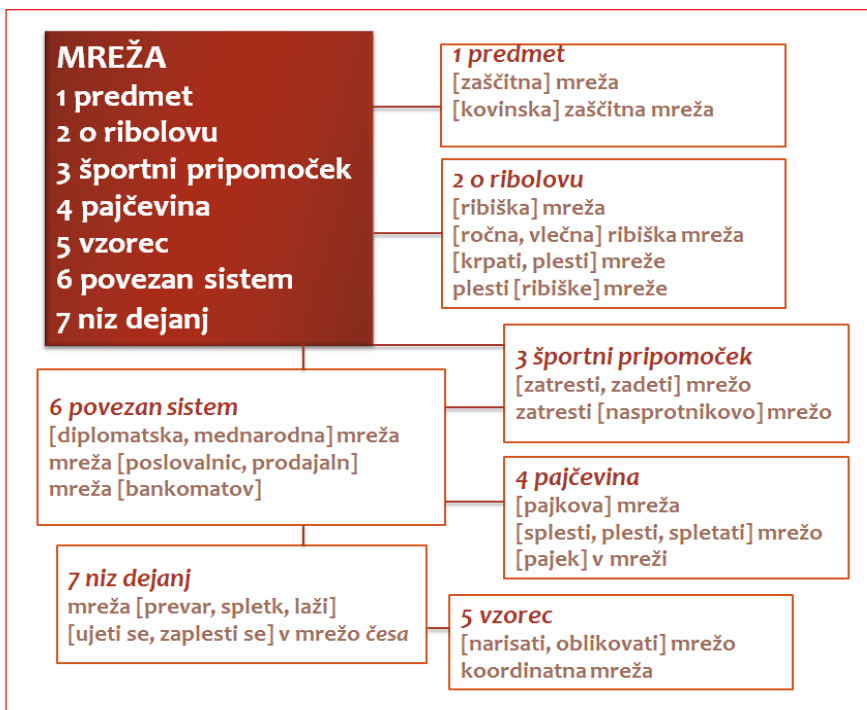
Če razumemo kolokacijo (frekventna sopoljavitev dveh ali več besed)⁹⁴ oz. kolokator (beseda, ki se glede na drugo besedo z njo sopoljavlja) kot leksikalno realizacijo situacijskega konteksta besede (Moon 1987: 92), lahko v njej prepoznamo najbolj tipičen preplet sintagmatskih in paradigmatskih razmerij, ki so ključna za ustvarjanje besedne večpomenskosti: na eni strani imamo torej besedno povezovalnost, ki predstavlja prvi pogoj za ustvarjanje pomenskih distinkcij, ki je na drugi strani omejena s pomenskimi izbirami, konkretnimi leksikalnimi zapolnitvami na paradigmatski ravni, kar ponazarja Slika 28 (Gantar 2007: 67):



Slika 28: *Sintagmatska in paradigmatska razmerja za pridevnik star*

Besedne skice v orodju SkE precej olajšajo prepoznavanje pomenskih lastnosti besede na podlagi povezovanja z drugimi besedami v okviru vnaprej določenih skladenjskih struktur, saj je pomenske tendence mogoče že na prvi pogled razbrati iz za različne pome ne tipičnih kolokatorjev. Slika 29 prikazuje odražanje večpomenskosti pri samostalniku *mreža* v konkretnih kolokatorjih.

⁹⁴ Diskusijo o obstoju več kot dvočlenskih kolokacij (prim. Siepmann 2005, 2006) puščamo ob strani. V LBS govorimo o večbesednih kolokacijah le v primeru kombinacij s predlogi in pri t. i. razširjenih kolokacijah. Več o tem v ustreznih poglavjih v nadaljevanju.



Slika 29 : Pomenska členitev samostalnika mreža na podlagi kolokatorjev

9.1.2.1.2.2 Udeleženska zgradba in udeleženci

V literaturi se pogosto razpravlja o tem, ali je vezljivost primarno skladijski ali pomenski pojav. Navadno prevladuje mnenje, da je vezljivost predvsem v pristojnosti skladijske, kar se opira na dejstvo, da se v številnih jezikih, kamor sodi tudi slovenščina, izraža z oblikoskladijskimi sredstvi – tj. skloni in skladijskimi razmerji (ujemanje, vezava, primik). Ta formalna sredstva realizacije vezljivosti pa ne bi smela zakrivati bistva stavčne organizacije elementov, namreč dejstva, da je sposobnost glagola povezovati se v besedilu z drugimi besedami primarno pomensko pogojena, tj. določena s pomenom povezujočih se besed (Pala in Ševeček 1997: 43). Skladijski vidik pri pojmovanju vezljivosti se osredotoča na določitev števila vezljivostnih mest in na tipologizacijo glagolov na tej osnovi. Leksikalni vidik skladijske realizacije vezljivostnih mest na površinski ravni ne izključuje, vendar pa vzroke za njihovo napoved in organizacijo znotraj stavka vidi v pomenskih lastnostih besede – vezljivost torej pojmuje kot leksikalno (pomensko; slovarsko) lastnost besed (zlasti glagolov), ki ima posledice v njihovi skladnji, in

ne obratno. Izhajajoč iz korpusne analize leksikalnega pomena, pojmujeemo tudi vezljivost ne zgolj kot skladijsko, pač pa kot leksikalno in skladijsko kategorijo hkrati (Čermáková 2009: 14) oz. kot sintagmatski pojav (Kopřivová 2006: 9), ki je utemeljen v kolokabilnosti⁹⁵ (Filipec in Čermák 1985) kot eni od temeljnih lastnosti leksikalnih enot (Čermák in Holub 1987: 10). Vezljivost⁹⁶ je torej določena v besedi kot njena imanentna lastnost, da veže nase druge besede, ki so nujne za njeno pomensko in skladijsko realizacijo (Filipec in Čermák 1985: 63). Kot taka vezljivost ne more biti vezana le na glagole, ampak tudi na nekatere pridevnike in samostalnike, kontekstualni (tudi sintagmatski) vidik pomenske določenosti pa ne izključuje niti prislovov.⁹⁷ Korpusno izhodišče z osredotočenostjo na kontekstualni vidik prepoznavanja pomena ne omogoča postavljanja ostrih mej med vezljivostjo in kolokabilnostjo, saj gre v obeh primerih za način besedne povezovalnosti (Čermáková 2009: 21), pri čemer je pri kolokabilnosti, kot jo definirata Filipec in Čermák (1985; Čermák 2010: 282), poudarek na leksikalnem vidiku besednega druženja,⁹⁸ vezljivost pa izpostavlja predvsem skladijski oz. oblikoskladijski vidik, znotraj katerega se formalno kategorizira povezovalnost leksema v odvisnosti do drugega leksema (npr. prek oblikoslovnih in skladijskih kategorij).

9.1.2.1.2.2.1 Glagol

Spremembe v vezljivostnem vzorcu glagola (tj. tako v kolokabilnosti kot njeni oblikoskladijski manifestaciji) so običajno tudi kazalec njegove večpomenskosti.⁹⁹ Kot kaže korpusna analiza stavčnih vzorcev, je pomenskoodločilno tako število in

95 Kolokabilnost razumemo kot sposobnost leksikalnih enot (tj. besed in večbesednih leksikalnih enot, kot so stalne zveze in frazeološke enote), da se povezujejo z drugimi leksikalnimi enotami v smiselne, slovnično pravilne in stilistično ustrezne izrazne enote.

96 Ker se v teoriji vezljivost pogosteje pojmuje kot skladijska, intencnost pa kot pomenska lastnost besed, bi bil izraz intenca na tem mestu nemara ustreznejši. Vendar pa dojemanje vezljivosti kot pomenskoskladijskega pojava, kar je tudi naše izhodišče, omogoča rabo izraza vezljivost tudi kot pomensko določenega pojma.

97 Po načelu skladijske relevance (Daneš in Hlavsa 1987: 65) lahko dobijo obvezno skladijsko mesto tudi glagolsko nevezljiva prislovna določila, npr. stavčnoočlensko (in tudi pomensko) obvezen prislov v: *Jana se je dobro odrezala* ali *Naloga se ni dobro iztekla*. (Povzeto po Žele 2001: 19.)

98 Leksemi se načeloma družijo na podlagi svojega pomena. Ta mora biti združljiv (kompatibilen) s pomenom drugega leksema. Čermák (2010: 282) kot primer navaja češki glagol psát (pisati) + dopis/knihu/poznámkú/vzkaz/ ... vendar ne *psát vodu/jidlo/rychlost/svetlo/ ... Kompatibilnost je torej pogojena s povezovalnostjo leksemov, ki pa je za vsak lekskem drugačna, določena z njegovo kolokabilnostjo, ki se formalno izraža v kolokacijah (prav tam: 283).

99 Npr. Vidovič Muha (2000: 34): »Sposobnost glagola, da spreminja svoje stavčnovozljivostno vedenje lahko vpliva nazaj na njegovo pomenskosestavinsko zgradbo in s tem seveda na nastanek večpomenskosti.«

skladenjska organizacija udeleženskih mest znotraj stavka kot tudi njihova leksikalna in oblikoskladenjska realizacija. Primer glagola *hvaliti* dobro ponazarja pomen-ske distinkcije, ki izhajajo iz različnih stavčnih vzorcev, kjer se udeleženska zgradba tako v skladenjskofunkcijskem, oblikoskladenjskem, udeleženskem in leksikalnem smislu realizira različno glede na posamezni pomen: *kdo hvali koga/kaj* : *kdo se hvali s čim pred kom*, ter v odvisnosti od leksikalnih zapolnitev, nakazanih z abstraktnimi semantičnimi tipi, kot so: ČLOVEK, USTANOVA, DEJANJA, DELO, na predvidenih mestih stavčne zgradbe:

hvaliti *glagol*

1 poudarjati dobre lastnosti

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, navadno pretirano poudarja svoje LASTNOSTI, SPOSOBNOSTI ali DOSEŽKE, predvsem z namenom, da bi ustvaril dober vtis pred drugim ČLOVEKOM

9.1.2.1.2.2.2 Samostalniki

Leksikalno izhodišče pri obravnavi vezljivosti, kot je sicer značilno za korpusno obravnavo problema (Kopřivová 2006; Čermáková 2009), ugotavlja, da je vezljivost tudi leksikalna lastnost nekaterih samostalnikov in pridevnikov ter jo je kot tako potrebno upoštevati tudi v slovarju. Vezljivi samostalniški pomeni so predvsem, vendar ne izključno (Čermáková 2009: 11) tisti, ki se obnašajo glagolsko. To se kaže na več načinov:

(a) V izražanju glagolskega pomena, zlasti delovanja ali dejanja, kar je mogoče preveriti z nadomestitvijo samostalnika z glagolom ali glagolsko zvezo v besedilu, pri čemer lahko zapis pomenske sheme sledi bodisi tipični skladenjski realizaciji – v spodnjem primeru z brezpredložnim določilom v rodilniku – ali s prikazom tipične realizacije v glagolski zvezi, npr.

komentar *samostalnik*

1 mnenje

1.1 izražanje mnenja; komentiranje

komentar DOGODKA, SITUACIJE ali TEME je izražanje mnenja o njegovi vsebini ali

če da ČLOVEK komentar na DOGODEK, SITUACIJO ali TEMO, izrazi o tem svoje mnenje

(b) V glagolniški obliki indikatorja, zlasti kadar glagolnik izraža dejanje (ne pa npr. rezultat dejanja):

gradnja *samostalnik*

1 izdelovanje objektov in predmetov

gradnja je izdelovanje STAVB, CEST in drugih OBJEKTOV ter večjih PREDMETOV iz različnih materialov

(c) Na površinski ravni je vezljivost kot pomenska lastnost samostalnika navadno izražena z desnimi določili, kar je razvidno tako v ubeseditvi pomenske sheme (gl. zgoraj iztočnici *komentar* in *gradnja*), kot iz skladenjskih struktur in pripadajočih kolokacij, npr.

adaptacija *samostalnik*

1 prenova stavbe ali prostora

SBZ0 sbz2

adaptacija [objekta, stavbe, šole, vrtca, hiše, hotela, zgradbe]

apetit *samostalnik*

1 občutek lakote

SBZ0 po sbz5

apetit po [hrani, mesu]

2 želja; potreba

SBZ0<r>v množini</r> po sbz5

apetiti po [ozemlju, zaslužku, nafti]

Tipično je desna določila mogoče pričakovati tudi pri samostalniških pomenih količine, npr. *debelina česa*, *morje česa*, *ščepec česa* ipd.

Samostalniška vezljivost je izpričana tudi v primeru ustaljenih skladenjskih nizov, ki niso nujno leksikalizirani v smislu nosilcev samostojnega pomena, zato jih v LBS ne obravnavamo kot leksikalne enote, pač pa kot skladenjske zveze, npr. *v odvisnosti od česa*; *v povezavi s čim*; *z ozirom na kaj* ipd.

9.1.2.1.2.2.3 Pridevnik

Leksikografski opis pridevnikov, izhajajoč iz kontekstualne obravnave vezljivosti, omogoča notranjo pomensko členitev večpomenskih pridevniških besed na podlagi izkazanih desnih predložnih ali brezpredložnih dopolnil,¹⁰⁰ npr. *alergičen na*

100 Kopřivová (2006: 3) izpostavlja naslednje tipe vezljivih pridevnikov: 1. z brezpredložnim sklonom, 2. s predlogom in ustreznim sklonom samostalnika/samostalniške zveze, 3. z nedoločnikom in 4. s stavkom, kar potrjujejo tudi slovenski primeri.

kaj,¹⁰¹ *odvisen od koga/česa, slep za kaj, opran česa* itd., ki se običajno realizirajo kot del povedka, tj. v vlogi povedkovnega določila, npr. *biti alergičen na kaj*, redkeje v prilastkovni funkciji: (*kdo*) *alergičen na kaj*. V strukturi pomenskega opisa v LBS se to lahko kaže že na ravni pomenskega indikatorja, sicer pa v pomenski shemi ter na ravni skladenjskih struktur in kolokacij:

izoliran <i>pridevnik</i> 1 ločen od česa; odmaknjen če je določen KRAJ izoliran od drugega KRAJA ali OBMOČJA, je od njega oddaljen ali z njim ni prometno, politično, kulturno ali kako drugače povezan PBZ0 od sbz2 izoliran od [sveta] 2 pridobljen iz česa izolirana SNOV je s posebnim postopkom pridobljena iz druge SNOVI ali ORGANIZMA PBZ0 iz sbz2 izoliran iz [rastlin, organizma]

9.1.2.1.2.2.4 Prislov

Kontekstualni vidik pri obravnavi prislovnih pomenov določa njihova t. i. ek-sogenega vloga v besedni zvezi. Ta namreč določa, da je jedro besedne zveze, za razliko od endogenih leksemov, zunaj njih samih, njihova vloga pa je posledično determinacijska. To velja poleg pridevnikov tudi za t. i. modifikacijske prislove, ob katerih je na podlagi sintagmatskih lastnosti razvrščanja ob glagole in pridev-nike mogoče določiti tudi semantično vrednost takega jedra. V pomenski shemi, ki je namenjena beleženju teh jeder v obliki semantičnih tipov, je to mogoče zapisati na naslednji način:

enosmerno <i>prislov</i> 1 od izhodišča proti cilju dejanje ali dogajanje, ki POTEKA enosmerno, se odvija zgolj v smeri od izhodišča proti cilju 2 o predstavah, sodbah; enostransko če človek VREDNOTI ali PREDSTAVLJA neko dejanje ali človeka enosmerno, ne upo-števa vseh dejstev in objektivnih meril

101 Kategorija človeškosti desnega določila, npr. *alergičen na kaj* : *alergičen na koga* je še nadalje lahko pokazatelj večpomenskosti.

ergonomsko *prislov*

udobno in enostavno za uporabo

ergonomsko OBLIKOVAN predmet, zlasti pohištvo ali naprava, človeku zagotavlja učinkovito, enostavno in udobno uporabo v delovnem okolju

Kot izkazuje pomenska shema prislova *mimo* spodaj, so z vidika sintagmatskih lastnosti zanimivi tudi udeleženci glagolskega dejanja, ki ga izraža jedrni glagol, prislov pa ga okoliščinsko določa. Vprašanje udeležencev se v tem smislu zastavlja tudi pri pretežno povedkovniških pomenih prislovov, kot izkazujeta 3. in 4. pomen. V takih primerih se zdi z vidika beleženja pomensko odločujočih delov udeleženske zgradbe za namene računalniške obdelave smiselno zapisati tudi udeležence v obliki semantičnih tipov:

mimo *prislov*

3 minulo; preteklo

ko bo ali če je nek DOGODEK mimo, je stvar preteklosti

4 nespametno; neustrezno; neprimerno
neformalno

če rečemo, da je neko DEJANJE, RAVNANJE ali IZJAVA mimo, se nam zdi, da je neustrezno ali neprimerno

9.1.2.1.2.3 Pomenski elementi

Med kazalce večpomenskosti štejemo tudi pomenske vsebine, ki jih v LBS pojmuje kot relevantne elemente pomena za določanje, povezovanje in razlikovanje posameznih pomenov in podpomenov. V slovenski leksikalni teoriji se je za »vsebinske lastnosti s pomensko razpoznavno vrednostjo« (Vidovič Muha 2000: 21) uveljavil izraz pomenske sestavine (prim. tudi Toporišič 1980, 1981; Müller 1980) ali semantične komponente (Orešnik 1972) po vzoru t. i. komponentne analize, katere namen je razgraditi pomen leksikalne enote na atomske koncepte (npr. Katz in Fodor 1963).

Pri prepoznavanju pomenskih razlik med posameznimi leksikalnimi enotami in znotraj leksikalne enote (kot tudi pri njenem pomenskem opisu) smo v LBS izhajali iz vsebinskih lastnosti, ki jih je bilo mogoče prepoznati na podlagi analize dejanske rabe konkretne leksikalne enote v realnem besedilu, vključno s tipičnimi upovedovalnimi možnostmi, kot je denimo zanikanje, in pragmatičnimi lastnostmi pomena, kot bomo prikazali v nadaljevanju na posameznih primerih za predstavnike v LBS upoštevanih besednih vrst.

9.1.2.1.2.3.1 Samostalnik

Vsebinske lastnosti večpomenskih samostalnikov je mogoče določati s podrobno analizo konkordančnega niza oz. na podlagi besedne skice v orodju SkE, kjer so elementi pomenske členitve zastopani s kolokatorji. Kot smo pokazali, je pomenske elemente nato mogoče ugotavljati glede na prepoznane vsebinske lastnosti, kot so npr.:

- entitete, na katere se pomen nanaša (človek, predmet, žival, deli telesa, snov);
- oblike in vizualni elementi predmetnosti, na katero se pomen nanaša (velik, ploščat)
- kognitivni, zaznavni elementi predmetnosti, na katero se pomen nanaša (čustva, lastnosti, dejanja)
- namembnost, ki jo pomen izraža (prehrana, zaščita, igra, dekoracija) oz.
- cilj dejavnosti/aktivnosti (zmanjševanje, razširjanje, preoblikovanje, izničenje, socialna interakcija, kulturno udejstvovanje), ki se povezujejo s pomenom
- področje, na katerem se pomen uporablja (šport, politika, umetnost)
- register, za katerega je pomen značilen (jezikovni status, družbene skupine)
- konotativne in pragmatične vsebine pomena (vrednotenje, učinkovanje; hierarhija govorcev)
- izkazane upovedovalne možnosti (zanikanje, trpna raba, druženje z lastnimi imeni) ipd., ki se opazno povezujejo s posamezno besedo/pomenom.

Pri pomenski analizi smo prepoznavali in upoštevali le tiste elemente, ki so se za konkretni pomen samostalnika kazali kot relevantni. Te pomenske elemente smo nato vključili v posamezne nivoje tristopenjskega (indikator, pomenska shema in/ali razlaga) pomenskega opisa. Nekaj primerov:

aranžma

Vsebinske lastnosti večpomenskosti			
entiteta	dejanje	namen	področje
turistična ponudba	pripraviti, organizirati	uporabnik	
cvetje, rastline	oblikovati, ustvarjati	priložnost, dogodek	
glasba	ustvarjati	izvajalec, vsebina	glasba
dokument	sklepati, dogovoriti		bančništvo

9.1.2.1.2.3.2 Glagol

Tudi pri glagolih se vsebinski elementi pomenske členitve izražajo v tipičnih kolokatorjih, glede na organizatorsko vlogo v stavčni povedi pa so pomenske spremembe povezane tudi s spremembami v številu predvidenih vezljivostnih mest, njihovih

skladenjskih realizacijah in leksikalnih zapolnitvah. Podobno kot pri samostalnikih smo tudi pri glagolih vsebinske lastnosti za pomensko razločevalnost določali na podlagi analize konkordanc in izdelane besedne skice. Z ugotavljanjem relevantnih pomenskih lastnosti in primerjavo njihove prisotnosti oz. odsotnosti smo pomensko razčlenili zlasti pomensko razvejane glagole:

razlesti se

Vsebinske lastnosti večpomenskosti			
entiteta	namen/dejanje/ posledica	dejanje	stavčni vzorec
usta, obraz, nasmeh	razpoloženje	narediti izraz na obrazu	kaj se razleze (komu) v (kaj)
prostor; tekočina plin, toplota, vonj	napolniti	premikati se	kaj se razleze kje (po, v čem)
prostor; ljudje	slišati se; prevzeti		kaj se razleze kje (po čem)
razpoloženje, zvok, občutek			
območje; ljudje, živali, rastline	iti, razširiti se, razraščati se		kaj se razleze kam/kje
človek, telo, fizični občutki (okus, bolečina, mravljinici)	širiti se		kaj se razleze (komu) kje (po čem)
človek, telo, snov (virus, serum)	razširiti se		kaj se razleze (komu) po čem
človek, telo	spreminjati obliko		kdo se razleze (kako, zakaj)
človek, negativna čustva (strah, tesnoba, nemoč)	širiti se; prevzemati		kaj se razleze komu kje (po čem)
čustvo (jok, stok), stanje	naraščati; prerasti		kaj se razleze v kaj
območje, negativna dejavnost; ljudje, ki jo opravljajo (kriminal, nedisciplina)	širiti se, zavzemati		kdo/kaj se razleze kam (na kaj)
konkretna ali abstraktna entiteta (mesto, podatki, dejavnost)	povečati se		kaj se razleze kako

9.1.2.1.2.3.3 Pridevnik

Pridevniki so glede na svojo prevladujočo atributivno vlogo, ki se kaže v pomenski odvisnosti od odnosnice in v tvorjenju stalnih zvez, pomensko razčlenljivi predvsem na ravni kolokabilnosti, ki pa se po pričakovanjih spreminja tudi glede na spremenjeno skladenjsko vlogo, npr. v vlogi povedkovnika.

lesen

Vsebinske lastnosti večpomenskosti			
entiteta	lastnost	področje	stavčna/besedilna vloga
predmet; orodje	izdelan iz lesa	obrt, gradbeništvo	atributivna
človek	tog, nespreten, okoren	nastopanje pred ljudmi	povedna

9.1.2.1.2.3.4 Prislov

Ugotavljanje večpomenskosti pri prislovih je pogojeno s primarnima skladenjskima funkcijama, tj. modifikacijsko in propozicijsko (Vidovič Muha 2000: 37). Določanje vsebinskih lastnosti pri prislovih ponazarja prislov *naravnost*, ki združuje tako okoliščinske (smer, cilj) kot načinovne oz. lastnostne pomene. Prislov je z vidika vsebinskih lastnosti pri določanju večpomenskosti zanimiv tudi zaradi izražanja poudarnosti:

naravnost

Vsebinske lastnosti večpomenskosti				
entiteta	namen	lastnost	besedilna vloga	dejanje
človek;	usmerjenost proti cilju, namenu	odločnost	poudarno	premikati se; hitro se premikati, biti usmerjen
človek, vozilo	prihod na cilj		poudarno	pripeljati, znati se, pasti
predmet, človek; del telesa	način premikanja, stanja; položaj	v ravni smeri, navpično		premikati se, stati
predmet, snov	neposredno prehajanje iz ene entitete v drugo, neposredna uporaba česa	neposrednost stičnost; odsotnost vmesnih elementov		uporabiti, položiti, nanesti
človek	neposreden premik k cilju	hitro, takoj, brez odlašanja, obotavljanja ali odstopanja od smeri	poudarno	iti, priti, odpraviti se
človek, pogled, oči	izražanje iskrenosti	odločno, pogumno	poudarno	gledati, zreti
človek	odkrito izražanje; neposredna komunikacija	neposrednost; odkritost, iskrenost		glagoli govorenja
predmeti, ljudje	poudarjanje lastnosti	skrajne pozitivne ali negativne lastnosti	poudarno	biti tak

9.1.2.2 Razmerja med pomeni

Pri pomenski členitvi smo poleg slovničnih in vsebinskih meril upoštevali tudi nekatera predvidljiva pomenska razmerja. Tem smo prilagodili tudi merila za beleženje pomenov znotraj baze, ki predvideva pomensko dvostopenjskost, kot se izraža v razmerju med glavnim pomenom oz. osnovnimi pomeni in podpomeni.

9.1.2.2.1 Glavni pomen in osnovni pomeni

V LBS govorimo o glavnem pomenu, ko ima beseda en sam pomen ali ko ima en pomen samo podpomene, o osnovnih pomenih pa, kadar ima beseda več pomenov, ki so lahko členjeni na podpomene.

Kadar ima glavni pomen več podpomenov, navadno nastopa kot združevalni oz. krovni pomen (gl. v nadaljevanju), sicer pa o glavnem pomenu govorimo predvsem pri enopomenskih besedah.

Osnovni pomeni predstavljajo osrednje ali splošne rabe besede, kot jih izkazuje referenčni korpus, vendar pa najbolj splošne ali najbolj dobesečne rabe pomena ni vedno lahko določiti. Glede na to, da smo celotni pomenski potencial besed v leksikografskem opisu omejili z razmerjem enostopenjske podrejenosti (lahko bi namreč uvedli tudi večstopenjsko členitev) in z razmerjem enakovrednosti med osnovnimi pomeni, se je kot najbolj zanesljivo pri pomenskem členjenju izkazalo čim bolj dosledno upoštevanje regularne jezikovne rabe v odnosu do njenih alternacij in posebnosti. Te tri nivoje, kot jih vidi Hanks (1994, 2004), je mogoče opazovati s postopki korpusnega vzorčenja, če orodju, ki nam služi za analizo, zagotovimo dovolj relevantnih podatkov.

Pri prepoznavanju osnovnih pomenov smo v LBS sledili naslednjim splošnim načelom:

(a) Vsaj en osnovni pomen je navadno tudi najbolj dobesečen, prim. *dihati* – 'vdi-havati in izdihavati zrak', vendar pa hkrati ta pomen ni nujno (ali prav ni) najbolj pogost.¹⁰² Glagol *dihati* se denimo pogosteje (ali vsaj enakovredno) rabi v pomenu 'sodelovati; živeti s čim', zlasti v interesni skupnosti ali v odnosu do okolja:

- *Vsi v ansamblu moramo dihati skupaj.*
- *Z rudnikom je svoje čase dihala celotna Idrija, danes pa za mesto pomeni le še ekološko breme.*
- *Zanima nas skupinsko delo na igrišču, vseh enajst igralcev mora dihati kot eden.*

¹⁰² Najbolj osnovni pomeni po pričakovanju tudi ne izkazujejo tipičnih kolokatorjev (npr. tek v pomenu 'hitro premikanje'), zato v nekaterih primerih iz besedne skice, ki temelji na tipičnosti besedilnega okolja, niso razvidni.

(b) Če ima beseda več osnovnih pomenov, smo v LBS sledili vodilu, da mora vsak osnovni pomen jasno izražati svoj pomenski doseg oziroma drugačno pomensko polje, kar naj bi se odražalo v pomenskem meniju:

cepiti <i>glagol</i>	voljno <i>prislov</i>	čaj <i>samostalnik</i>	hrpav <i>pridevnik</i>
1 prejeti ali vnesti cepivo	1 s pripravljenostjo; rade volje	1 pijača 2 rastlina	1 ki ni gladek 2 hripav
2 deliti na več delov	2 o testu in pecivu		
3 vnašati razdor	3 vojaško povelje		
4 izboljšati kakovost			

(c) Čeprav posamezni osnovni pomeni niso vedno povsem oddaljeni drug od drugega, smo jih beležili kot osnovne (tj. na isti ravni), kadar med njimi, sinhrono gledano, ni bilo mogoče vzpostaviti jasnega hierarhičnega razmerja, npr.

bazar *samostalnik*

- 1 orientalska tržnica
- 2 prireditve

9.1.2.2.1.1 Dobesedni in preneseni pomen

Najpogostejše razmerje med pomeni se ustvarja med dobesednim in figurativnimi ali prenesenimi (metafora, metonimija) pomenskimi rabami,¹⁰³ vendar pa stopnja asociativnosti med izhodiščnim in prenesenim pomenom ni vedno jasna, kot pogosto ni konsenza glede izhodiščnosti pomena samega (prim. Hanks 2004, 2006; Frath 2000: 50). Zato je razumljivo, da etimologi, pomenski analitiki, leksikografi in zlasti jezikovni uporabniki pogosto ne delijo enake predstave o naravi domnevne prenesene ali figurativne pomenske rabe (prim. Lyons 1977, cit. po Snoj 2010: 16). Zato tudi ne moremo postaviti absolutnega pravila, po katerem bi prenesene ali figurativne pomene¹⁰⁴ vedno obravnavali bodisi kot osnovne pomene bodisi kot podpomene, saj je pomenska struktura gesla specifična za vsako besedo posebej. V LBS se zato pri prenesenih in zlasti pri metaforičnih pomenih načeloma nismo odločali za vzpostavljjanje podpomenskih razmerij, ampak smo jih obravnavali kot enakovredne samostojne pomene:

103 V SSKJ je bil za preneseno rabo uporabljen kvalifikator pren. (preneseno), ki označuje podpomen, ki je vsebinsko blizu izhodiščnemu, konkretnemu pomenu, je pa »predstavno premaknjen v drugačno okolje« (interna navodila za izdelavo SSKJ).

104 Govorimo seveda o institucionalizirani, leksikalizirani metafori in ne o priložnostnih, ustvarjalnih figurativnih rabah.

buča <i>samostalnik</i>	hribček <i>samostalnik</i>	benigen <i>pridevnik</i>
1 rastlina	1 manjši hrib	1 o tvorbi v organizmu
2 glava	2 navadno v množini ali dvojini ženske obline	2 neškodljiv, nenevaren
izkašljati <i>glagol</i>	bobnati <i>glagol</i>	brezbarven <i>pridevnik</i>
1 izločiti s kašljanjem	1 igrati bobne	1 ki nima barve
2 izročiti denar	2 veliko govoriti	2 politično neopredeljen
3 razkriti		
4 povedati, kar koga teži		

Odločitev za obravnavanje potencialno povezanih pomenskih rab kot samostojnih pomenov omogoča tudi prikaz pomenskega razmerja, kadar raba očitno izkazuje prevlado figurativnega pomena, ki pa bi v hierarhičnem smislu moral biti podrejen. V razmerju izhodiščni (dobesedni) pomen : preneseni (motivirani) podpomen smo se odločali za prikaz razširjenosti v rabi pred izvornostjo.¹⁰⁵ Tovrstno rešitev prikazujejo spodnji primeri:

kovati <i>glagol</i>	čednost <i>samostalnik</i>	nečednost <i>samostalnik</i>
1 načrtovati; pripravljati	1 lepa človeška lastnost;	1 nemoralno, kaznivo
2 ustvarjati	vrлина; krepost	dejanje
3 izdelovati iz kovine	2 lepota	2 kar je neprijetno ali grdo

9.1.2.2.1.2 Konkretni in abstraktni pomen

Kot osnovni pomeni so v LBS obravnavani tudi pomeni v razmerju konkretno – abstraktno (prim. tudi fizično – psihično oz. telesno – duševno), pri čemer navadno oba ali več pomenov pripadata ločenima pomenskima poljema, npr.

¹⁰⁵ Trend osamosvajanja metaforičnih pomenov v sodobnih, zlasti na korpusu temelječih slovarjih, opuščanje označevanja metaforičnih pomenov z oznakami kot fig., pren. ipd. ter razvrščanje pomenov na podlagi pogostnosti in ne izhodiščnosti je v slovarski literaturi deležno tudi kritik, češ da gre za okrnitev obvestilnosti, zlasti s pedagoškega, pa tudi spoznavnega in pragmatičnega vidika (Moon 2004: 200). Na drugi strani je odsotnost metajezikovnih podatkov mogoče upravičiti s priročniško naravo slovarjev, kjer je obvestilnost o pogostnosti pomena pred njegovo izvornostjo za uporabnika prav tako pomembna, hkrati pa so metajezikovne oznake o vrsti pomenskega prenosa zanimive predvsem za jezikoslovne analize, za uporabnika pa navadno nimajo obvestilne vrednosti.

prostor <i>samostalnik</i>	blokirati <i>glagol</i>	boleti <i>glagol</i>
1 zaprto območje	1 ustaviti gibanje	1 o telesni bolečini
2 odprto območje	2 preprečiti uresničitev	2 o psihični bolečini
3 območje človekovega delovanja		

9.1.2.2.1.3 Splošni in specifični pomen

Kadar so osnovni pomeni v medsebojnem razmerju splošno – specifično, je specifična raba besede navadno vezana na konkretno področje (prim. *cepiti* – zdravstvo; vrtnarstvo; *bazičen* – kemija; *aplikacija* – računalništvo, *angažirati* – bančništvo) ali pa je določena s specifičnimi udeleženci, kot jih izkazujejo kolokatorji, prim. *dihati* (živo bitje : blago, čevlji, koža). Take pomenske rabe smo navadno prikazovali kot osnovne pomene, npr.

bazičen <i>pridevnik</i>	aplikacija <i>samostalnik</i>	angažirati <i>glagol</i>
1 o kemični sestavi; alkalen	1 računalniški program	1 dejavno se vključevati;
2 osnovni; temeljni	2 prenos teorije v prakso	prizadevati si
		2 vložiti denarna sredstva

9.1.2.2.2 Krovni pomen

V LBS smo kot tehnično rešitev pri organizaciji podpomenov uvedli možnost izpostavljanja pomena, ki služi zgolj kot krovna nadpomenka za vse podpomene, ki jih vsebuje. Gre za primere, kjer si podpomene sledijo v enakovrednem nizu, sodijo torej v okvir istega pomenskega polja, vendar nobenega izmed njih ni mogoče povzdigniti na raven glavnega pomena. V takih primerih je glavni pomen – kot krovni pomen – opredeljen zgolj z indikatorjem, drugi elementi gesla (tj. pomenska shema, stavčni vzorci, skladienske strukture, kolokacije in zgledi) pa niso navedeni:

vrečka <i>samostalnik</i>	pasti <i>glagol</i>	aktivno <i>prislov</i>
3 pripomoček	3 znižati se na lestvici	1 dejavno
3.1 v kuhinji	3.1 o količini, vrednosti	1.1 z veliko truda in sodelovanja
3.2 v medicini	3.2 o intenzivnosti	1.2 z namenom izboljšati
	3.3 o statusu, kakovosti	1.3 z vključevanjem različnih
	3.4 o socialnem položaju	dejavnosti

9.1.2.2.3 Podpomeni

Beseda ima lahko več osnovnih pomenov, ti pa imajo lahko svoje podpomenne. Za hierarhično ureditev pomenske členitve v LBS velja, da mora podpomen v odnosu do osnovnega pomena vedno ohranjati razvidno pomensko povezavo. V LBS so s podpomeni znotraj katerega od registriranih osnovnih pomenov zastopane podobnosti in razlike, kot se kažejo pri analizi vsebinskih lastnosti besede. Pri glagolih so to zlasti udeleženci in razmerja med njimi (*grmeti*), neposredna ali posredna udeleženosť katerega od udeležencev, vzrok ipd. Pri samostalnikih in pridevniki so to predvsem kolokatorji (*deblo, aktiven*). Te spremembe se odražajo tudi v pomenskem meniju:

grmeti <i>glagol</i>	aktiven <i>pridevnik</i>	deblo <i>samostalnik</i>
1 oddajati glasen zvok	1 dejaven	1 sistematični zapis izvora
1.1 o nevihti	1.1 ki veliko sodeluje in	2 del rastline
1.2 o predmetih	se redno udeležuje	3 del organa
1.3 o vozilih ali napravah	1.2 ki se poklicno ukvarja	
1.4 o glasbi	1.3 ki je usmerjen v	
1.5 o orožju	izboljšanje	
1.6 pri padanju	1.4 ki vključuje različne	
	dejavnosti	
	1.5 ki je kaj sposoben	
	opravljati	
	1.6 ki je v fazi delovanja	

Pri posameznih besednih vrstah se med pomeni ustvarjajo predvidljiva pomenska razmerja – t. i. sistemska polisemija (Apresjan 1973; 1992; 2002), ki pa se ne ureničujejo vedno absolutno. Razmerja, ki jih navajamo v nadaljevanju, predstavljajo ustaljene pomenske vzorce, kot smo jih registrirali pri analizi pisne rabe v korpusu.

(a) Razmerje med dejavnostjo/dejanjem in predmetom/pojavom, povezanim s to dejavnostjo, se kaže zlasti pri glagolnikih, iz tega izhajajoče pomenske spremembe pa so navadno povezane s spremembo slovnične kategorije števnosti kot v spodnjih primerih:

korespondenca <i>samostalnik</i>	bučanje <i>samostalnik</i>
1 dopisovanje	1 pojav ob delovanju narave ali naprav
neštevno	1.1 zvok
1.1 sporočila	

V primeru, da sta dejanje (ali dogajanje) in/ali rezultat dejanja (predmet, pojav) sama na sebi večpomenska, lahko pridobita v postopku pomenske členitve status samostojnih pomenov:

žur *samostalnik*

1 zabavno dogajanje

1.1 čas za zabavo

2 dogodek

1.1 zasebna zabava

1.2 javna prireditev

3 velika zmeda

Večpomenskost glagolnikov na ravni dejanje – rezultat dejanja oz. dejavnost – predmet dejavnosti je povezana tudi s skladejskimi lastnostmi glagolniške iztočnice, saj v pomenu dejanja glagolniki pogosto zahtevajo določila, podobno kot njihovi glagolski ustrezniki, npr. *bučanje česa* (*kaj buči*), *gradnja česa* (*graditi kaj*) ipd.

Pri pridevniki se na ravni podpomenov ustvarja zlasti razmerje med pojavom ali dejavnostjo in namenom oz. značilnostjo pojava:

dežen *pridevnik*

1 o dežju

1.1 namenjen za dež

baseballski *pridevnik*

1 o športu

1.1 namenjen igranju
baseballa

monsunski *pridevnik*

1 o vremenskem obdobju

1.1 značilen za to vremensko obdobje

(b) Na ravni pomen – podpomen smo v LBS obravnavali tudi pomenska razmerja med snovmi, rastlinami in živalmi ter njihovimi produkti v obliki jedi, začimb, izdelkov in med enotami, ki zadevajo količino izdelka ali snovi:

jetra *samostalnik*

1 organ

2 meso

2.1 jed

marmelada *samostalnik*

1 namaz

1.1 enota tega namaza

curry *samostalnik*

1 začimba

1.2 jed

riba *samostalnik*

1 žival

1.1 jed

Pridevniške iztočnice navadno izkazujejo vsaj en splošni pomen, pod katerega so uvrščeni kolokatorji, ki se na nanašajo na odnosnico v najbolj splošnem smislu, npr. *afriški* – o Afriki, *cvetlični* – o cvetlicah, v podpomenskih nizih pa je mogoče prepoznati nekatera regularna podpomenska razmerja.

Tipično razmerje sestavinskosti (Vidovič Muha 1978: 257) oz. vsebovanosti smo pri pridevniških pomenih reševali v razmerju osnovni pomen – podpomen, in sicer s t. i. nanašalnimi indikatorji na ravni osnovnega pomena in z opisnimi indikatorji, ki so bližji razlagam, npr. izdelan iz ..., ki vsebuje ... ipd., na ravni podpomena:

borovničev <i>pridevnik</i>	bambusov <i>pridevnik</i>	litijev <i>pridevnik</i>
1 o rastlini	1 o rastlini	1 o snovi
1.1 izdelan iz borovnic	1.1 o izdelkih	1.1 ki vsebuje litij

(c) Pomensko razmerje med osnovnim pomenom in podpomenom se ponekod izrazito ohranja tudi pri prenesenih pomenskih rabah ali pa obstaja znotraj posameznega osnovnega pomena več prenesenih rab, kar glede na pomensko strukturo gesla in predvideno hierarhično dvostopenjskost pomeni, da prenesenih pomenov ni mogoče vedno obravnavati kot samostojnih pomenov, ampak kot enakovredne podpomene, npr.:

izdati <i>glagol</i>	spati <i>glagol</i>	oblak <i>samostalnik</i>
4 razkriti	3 biti nedejaven	2 velika količina
4.1 skrivnost ali informacijo	3.1 o rastlinah	2.1 o delcih in plinih
4.2 dejansko stanje	3.2 o naravnih pojavih	2.2 zlasti o žuželkah
4.3 zlorabiti zaupanje		

Tendence podrejanja prenesenih rab na raven podpomenov so zlasti očitne pri metonimičnih samostalniških pomenih:

mošnjček <i>samostalnik</i>	kitara <i>samostalnik</i>	moškost <i>samostalnik</i>
1 vrečka	1 inštrument	1 lastnost moških
1.1 denarna sredstva	1.1 igranje tega inštrumenta	1.1 moški spolni organ; penis

Pri pridevniških in prislovih so prenesene rabe obravnavane kot podpomene samo v primeru, da je pomenska povezava z osnovnim pomenom očitna, npr.

zaspano <i>prislov</i>	čipkast <i>pridevnik</i>
1 v stanju utrujenosti ali zaspanosti	1 narejen iz čipk
1.1 brezvoljno; počasi	1.1 podoben čipkam
1.2 brez veliko dogajanja; nedinamično	

Poudarili smo že, da zgradba LBS predvideva podrejenost vseh podatkov posameznemu pomenu besede v iztočnici. To se kaže v tem, da je vsak pomen in podpomen

predstavljen z indikatorjem, v večini primerov pa še s pomensko shemo ali razlago, skladenjskimi strukturami ali vzorci, skladenjskimi zvezami in stalnimi zvezami. Poleg takih regularnih situacij pa prihaja tudi do primerov, kjer določena prepoznavna pomenska raba v svoji okolici ne izkazuje tipičnih kolokatorjev niti se ne realizira v prepoznavnih skladenjskih strukturah ali vzorcih, gre za t. i. *nekolokabilne pomen*. Taki pomeni zato tudi niso neposredno razvidni iz leksikalnogramatičnega profila besede, kot nam ga kažejo besedne skice. Navadno gre za zelo splošne, vendar redko tudi najbolj pogoste pomeni (npr. *tek*). V nekaterih primerih je za pomen prepoznavna le struktura (npr. *solata*), leksikalne zapolnitve, kot jih beležimo v kolokacijah, pa niso tipične:

tek *samostalnik*

1 hitro premikanje nog

- *Pri teku ne sodelujejo samo noge, temveč vse telo in razum.*
- *Slišim tek po stopnicah, po hodniku, zasopel vstopi Rok.*

solata *samostalnik*

1 rastlina

2 velika količina

solata je velika količina neurejenih ZAPISOV ali DOKUMENTOV

Pbz0 SBZ0

- *Količina papirnate solate, ki je je v državnem zboru že tako obilo, se je zaradi prispelih poročil povečala še za kakih deset višinskih centimetrov.*
- *Tisti, ki so si doslej sami omislili napis, da reklamne solate ne želijo prejemati, so brcnili v temo.*
- *Vso to solato bi moral vsaj še urediti, če že ne prebrati.*

9.1.3 Razvrstitev pomenov

V LBS nismo izhajali iz etimološke oz. zgodovinske razvrstitve pomenov, po kateri si pomeni sledijo glede na prvotni oz. najstarejši pomen besede, saj se pomeni skozi čas spreminjajo, hkrati pa so meje med homonimijo in večpomenskostjo večkrat nejasne, česar se je zavedala večina leksikografov že v času tiskanih slovarjev, ki so za ločevanje iztočnic na podlagi homonimije bolj primerni. Prav tako nismo kot ključne upoštevali razvrstitve, po kateri si pomeni sledijo glede na stopnjo prenesenosti oz. od izhodiščnega (dobesednega) k manj navadnemu, prenesenemu, abstraktnemu pomenu, saj obstajajo primeri, kjer se prenesena raba v sinhroni perspektivi prebije v ospredje ali pa se izhodiščni (dobesedni) pomen pomakne na obrobje ali celo popolnoma izgubi. Pri razvrstitvi pomenov smo težili k prikazovanju aktualnosti, pri čemer je treba povedati, da zanesljivih merljivih izhodišč za razvrščanje pomenov na podlagi njihove sinhrone rabe nimamo. Kazalec, ki ga je mogoče upoštevati, je število kolokatorjev, ki se pripenja na posamezni pomen besede, medtem ko je logika pomenskih razmerij v smislu nad- oz. podrejenosti in pomenska razvejanost besede

stvar leksikografskega koncepta. Čeprav izhajamo iz korpusnega gradiva, na podlagi katerega je mogoče ugotoviti številčno zastopanost marsikaterega jezikovnega pojava, je podatek o pogostnosti posameznega pomena, kot ga lahko pridobi leksikograf na podlagi pomenske analize, zelo pavšalen, saj je že sama pomenska členitev besede stvar leksikografske interpretacije, dogovorjene stopnje pomenske členitve in načina beleženja pomenskih podatkov. Razvrstitev pomenov pri večpomenskih iztočnicah je zato v LBS v prvi vrsti podrejena jasnosti in logiki pomenskega menija.

Pri razvrstitvi pomenov smo v LBS najprej upoštevali delitev na osnovne pomene in podpomene. Pri razvrstitvi osnovnih pomenov smo sledili pogostnosti, kar pomeni, da se lahko posamezni preneseni pomeni, če tako izkazuje raba, razvrščajo tudi pred dobesedne ali najbolj splošne pomene. Načelu pogostnosti nismo sledili le pri razvrstitvi stalnih zvez in frazeoloških enot, saj so te vedno zabeležene v samostojnih elementih, lahko pa si, če je znotraj razdelka enot več, sledijo po (relativni) pogostnosti.

Frekvenčna razvrstitev pomenov pomeni, da smo v LBS na prva mesta uvrstili tiste pomene, ki glede na analizo konkordanc in besednih skic predstavljajo najpogostejše rabe, kar se kaže zlasti v številu kolokatorjev, ki jih je mogoče pripisati posameznemu (pod)pomenu. Pri tem smo upoštevali možnost, da so isti kolokatorji, ki sicer izstopajo po frekveni, 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.

Preglednica 10 prikazuje število kolokatorjev v besedni skici glagola *brusiti* za gramatično relacijo: S_v_tožil, ki v LBS ustreza strukturi: GBZ sbz4 (tj. glagol + samostalni v tožilniku). Posamezne kolokatorje je mogoče razporediti v pomenske oz. tematske skupine, npr. deli telesa, predmeti, lastnosti ipd.

Preglednica 10: Pogostnost posameznega kolokatorja pri relaciji S_v_tožil za glagol brusiti

deli telesa	predmeti									lastnosti	
	orodje		orožje		kovine						
krempelj	82*	nož	145*	bajonet	14*	diamant	21	kolut	9	forma	40
krempeljč	7*	srp	6	meč	8*	kamnina	4	drsalka	5*	talent	15
krempeljček	4*	kosa	22*					robnik	4		
zobek	6	škarje	11*					valj	9		
zob	49*							pero	7*		
noht	8*										
rog	8										
peta	11*										
podplat	11*										

* vključuje FE ali preneseno rabo

Pri razporeditvi pomenov, kot je nakazano v nadaljevanju, pa je bilo ob pogostnosti kolokatorja treba upoštevati tudi dejstvo, da se isti kolokator lahko uporablja v konkretnem in prenesenem pomenu oz. da je nekatere metaforične rabe, ki izkazujejo omejeno kolokabilnost, smiselno predstaviti kot frazeološke enote:

brusiti *glagol*

♦ **brusiti jezike/brusiti si jezike**

opravljati; obrekovati

♦ **brusiti pete**

plesati ali veliko hoditi

Pri notranji razvrstitvi podpomenov smo upoštevali najprej logiko pomenske podrejenosti in s tem razumljivost pomenskega menija in šele nato tudi pogostnost. Sicer pa smo pri notranji razporeditvi podpomenov upoštevali še nekaj splošnih pravil:

- rezultat dejavnosti ali predmet je vedno podrejen dejavnosti:

korespondenca *samostalnik*

1 dopisovanje

1.1 sporočila

alkotest *samostalnik*

1 preverjanje alkoholiziranosti

1.1 naprava

- sadež, izdelek ali del rastline je vedno podrejen rastlini:

ingver *samostalnik*

1 rastlina

1.1 začimba

kanabis *samostalnik*

1 rastlina

1.1 mamilo

- jed je vedno podrejena živali/snovi ipd., iz katere je izdelana:

riba *samostalnik*

1 žival

1.1 jed

jezik *samostalnik*

1 organ

1.1 kot hrana

- del predmeta je vedno podrejen predmetu:

britvica *samostalnik*

1 pripomoček za britje

1.1 rezilo

- enota ali količina snovi je vedno podrejena snovi:

pivo *samostalnik*

1 alkoholna pijača

1.1 enota pijače

9.2 Združevanje pomenov

Čeprav je eno temeljnih načel, ki smo mu sledili pri izdelavi LBS, podrobna pomenska členitev, je leksikograf pogosto v dilemi, ali je določene pomenske rabe dejansko smiselno pomensko osamosvajati ali je, zlasti z vidika uporabnika, posamezne pomenske odtenke, kot jih navadno izkazuje živa jezikovna raba, bolje združiti v en sam (pod)pomen. Poleg tega se mora leksikograf zavedati, da namen slovarskih opisov ni in ne more biti zajetje celotne palete potencialnih sorodnih pomenskih rab in vseh podrobnosti, ki jih izkazuje posamezni pomen besede (gl. primer glagola *mleti* v nadaljevanju). Pomensko členjenje besed je namreč, in to ne glede na stopnjo podrobnosti, v primerjavi z naravnim sporočanjem nenaravno početje, katerega osnovni namen je organizacija in prikaz podatkov o pomenskih in leksikalnogramatičnih lastnostih besede v slovarju. Pri izdelavi LBS smo si, upoštevajoč omenjena dejstva, prizadevali prikazati čim bolj nedvoumno in prepričljivo pomensko situacijo. Zato je pomembno, da smo na eni strani jasno ločevali različne pomenske tendence in da smo hkrati določene pomenske odtenke zanemarili oz. združili pod en sam pomen ali podpomen.

9.2.1 Razpršena pomenska raba

Dober primer razpršene pomenske rabe predstavlja glagol *mleti*, pri katerem lahko na podlagi besedne skice s precejšnjo mero gotovosti izločimo najbolj splošen pomen, ki ima tudi prepoznavne kolokatorje:

drobiti na manjše delce

mleti [žito, pšenico, ajdo, rž, koruzo]; mleti [moko, zrnje]; mleti [kavo]

Poleg tega pa še relativno homogene prenesene rabe, kot:

počasi obravnavati ali reševati (zlasti v sodstvu)

[sodno, pravno] kolesje melje, [birokracija] melje; [počasi] mleti

potekati, izvajati kaj ne glede na okoliščine, navadno s silo (v zvezi z vplivnimi inštitucijami)

[mašinerija, stroj] melje, [tanki] meljejo

jesti, prebavljati

[trebuh] melje

premagovati nasprotnike (v športu)

mleti [tekmece, nasprotnike]

potekati, delovati, navadno nezavedno ali samodejno

[možgani] meljejo

Ob zgornjih relativno čistih pomenih, ki jih večinoma potrjujejo tudi kolokacije, se v realnih besedilih ustvarja še vrsta pomenskih rab, ki so bodisi od registriranih osnovnih in prenesenih precej oddaljene ali pa se jih v posameznih pomenskih lastnostih celo dotikajo. Za ponazoritev navajamo nekaj primerov iz korpusa Gigafida:

- *Vztrajno meljem strmino skozi pustinjo, posejano z nizkim bodikavim grmičevjem, šopi trave in bledikavimi majcenimi cvetovi omamno dišeče cvetlice.*
- *Nekateri se spet obnašajo, kot da se ni nič zgodilo in pridno meljejo marginalije, problemi pa se kopičijo.*
- *Zato sem poklapana in jezna nase. Dolgo meljem, ampak naredim pa zelo hitro.*
- *Turbodizel prijetno melje in je glede na značaj kot nalašč za mestne vožnje.*
- *Užitek je čofotati po njej – dno je namreč prekrito z mehko mivko, ki se prijetno melje med prsti.*
- *Težko se je upreti pritiskom, sugestijam, telefonskim odzivnikom, vse to vas melje iz dneva v dan.*

V situacijah, kot jih nakazujejo zgornji primeri, se zdi pomembno, da leksikograf ugotovi regularnost določenega pomena in ne skuša bodisi odpirati novih pomenov bodisi pomenskih opisov prilagajati tako, da zajamejo vse individualne pomenske rabe. Regularnost je, kot pravi Hanks (1994), nasproti ustvarjalnosti ena od prepoznavnih lastnosti jezikovne rabe, ki jo najbolj razkrivajo korpusni vzorci, zato je pomembno, da so leksikografske odločitve podprte s tistimi pomenskimi rabami, ki se izkazujejo kot najbolj tipične in ki imajo glede pomenskih tendenc, ki jih izražajo v besedilu, soglasje čim širšega kroga govorcev.

9.2.2 Sistemska polisemija – problem »harem«

Pri tipologiziranju samostalniške večpomenskosti je metonimija prepoznana kot eden najbolj tipičnih pomenskih izpeljav (Snoj 2010: 11), kamor sodijo denimo:¹⁰⁶ zamenjava celote z njenim delom (telefon – telefonska številka), dela s celoto (iti na pijačo – iti v bar), posloplje s prebivalci, uslužbenci ipd. (parlament – poslanci v

¹⁰⁶ Povzeto po Wikipediji: <http://sl.wikipedia.org/wiki/Metonimija>.

parlamentu), kraj s prebivalci (Ljubljana – prebivalci Ljubljane), prebivalci z državo (Nemci – prebivalci Nemčije), funkcija ali položaj z njegovim nosilcem (krona – kraljeva monarhija), avtor z delom (poslušati Mozarta – Mozartovo glasbo), predmet z dejavnostjo (vaditi kitaro – igranje kitare), orodje, predmet z njegovo funkcijo (slišati telefon – zvonjenje telefona), funkcija orodja z orodjem (vklopiti gretje – napravo, za gretje), znamka z izdelkom (jogi – vzmetnica), čas, obdobje s svojimi značilnostmi (renesansa – kulturne ipd. značilnosti renesanse), vsebina, npr. tekočina ipd. z embalažo (politi mleko – kozarec, embalažo z mlekom). Čeprav je metonimična pomenska raba v vsakdanjem sporočanju običajna, pa pri leksikografskem členjenju situacija navadno ni kristalno jasna. Kot že v primeru razpršene pomenske rabe, je tudi pri prepoznavanju metonimičnih pomenskih prenosov problem zlasti v tem, da želi slovar ustvariti črno-belo situacijo tudi tam, kjer so v realnem jeziku sivine.

Odločitve ali načela, ki smo jih v zvezi s tem sprejeli pri pomenskem členjenju besed v LBS, je mogoče ponazoriti s samostalnikom *harem*, ki zastopa problem, s katerim smo se pri pomenski analizi in členitvi pogosto srečali. Na eni strani namreč obstaja jasna možnost ločevanja dveh pomenskih rab, na primer 'harem kot skupnost vladarjevih žena, značilna za islamski dvor', in harem kot 'prostor, v katerem skupnost teh žensk biva'. Na drugi strani pa je realnih še vrsta pomenskih rab, kjer ločevanje med skupnostjo in prostorom za razumevanje besede v resnici ni relevantno. Preglednica 11 prikazuje poskus razvrstitve različnih rab samostalnika harem, kot se kažejo v zgledih, izbranih iz korpusa Gigafida, pod pomena 'skupnost' in 'prostor':

Preglednica 11: Ugotavljanje metonimičnih pomenov za besedo harem na podlagi korpusnih zgledov

Zgled	skupnost	prostor
1 <i>Vladarji so izvajali nasilne nabirke mladih žensk zato, da so jih pošiljali v hareme in podobne imperialne institucije.</i>	x	
2 <i>Kastiljski kralji so v času, ko so Arabcem še plačevali davek vsako leto, morali v hareme Seville poslati po 100 devic.</i>	x	x
3 <i>Haman, kraljev svetovalec, mu namigne, naj v svojem haremu zbere vse mlade device iz kraljestva in si med njimi izbere kraljico.</i>	x	x
4 <i>Esther s svojo preprostostjo, milino in dobroto povsem prevzame tako čuvaje harema kot kralja.</i>	x	x
5 <i>A glede na to, da si je Ismail zlagoma v harem natlačil od 350 do 500 žen in konkubin.</i>	x	x
6 <i>Takšno življenje je morda nekdaj obstajalo v starodavnih arabskih haremih.</i>	x	x

7	Indijski sultan Shah Jahan je imel v 17. stoletju sicer običajen harem številnih žena.	x	x
8	V harem pridemo skozi dvorišče evnuhov.		x
9	Po enotedenskem tavanju v puščavi se slučajno znajdem v haremu , polnem žensk!	x	x
10	Sultan je v harem prišel po zlatem hodniku, nastlanem s cekini		x
11	Na razglednici je bila odaliska, sužnja in priležnica v haremu .	x	x

Na podlagi stanja, kot ga prikazuje preglednica, kot tudi pri podobnih situacijah, ki se nanašajo na metonimične pomenske prenose, smo upoštevali naslednja načela leksikografske obravnave:

- Treba je ugotoviti, ali se metonimični prenos realizira v zadostni meri, da je smiselno uvesti nov (pod)pomen. Sistemska možnost, kot jo predvideva metonimija (pa tudi druge oblike regularne polisemije), sama na sebi ni zadostni razlog za ustvarjanje novega (pod)pomena.
- Na podlagi analize je treba ugotoviti, ali obstajajo povsem nedvoumni primeri izhodiščne pomenske rabe in zlasti, kateri primeri rabe so nedvoumno metonimični.
- Če prepoznamo metonimični pomen, ga moramo v odnosu do izhodiščnega potrditi s kolokacijami in korpusnimi zgledi, ki odločitev potrjujejo:

harem samostalnik

1 v muslimanskem svetu skupnost vladarjevih žena

[vladarjev, sultanov] harem; [muslimanski, islamski, orientalski, turški] harem

1.1 prostori, v katerih bivajo ženske na sultanovem dvoru

- *Sultan je v harem prišel po zlatem hodniku, nastlanem s cekini.*
- *V harem pridemo skozi dvorišče evnuhov.*

9.2.3 Individualna ali ustvarjalna raba

Kot samostojnih pomenov ali podpomenov nismo beležili individualnih ali ustvarjalnih pomenskih rab, ki jih govorci tvorimo priložnostno za konkretne sporočanje namene. Taki pomenski odtenki nimajo potrditev v velikem številu korpusnih zgledov in ne predstavljajo splošnih pomenskih asociacij, zato slovarsko niso zanimivi. Kljub temu pa smo bili pozorni na pomenske tendence, ki se lahko v jeziku oblikujejo postopoma. Mejni primer v tem smislu je npr. pomen 'onesposobiti' glagola *crkniti*:

crkniti *glagol*
3 onesposobiti

če ČLOVEK ali njegovo delovanje crkne ČLOVEKA ali NAPRAVO, ga onesposobi ali pokvari

- *Nato dobivaš cilje, ki pripeljejo do poskusa zmikavtenja najdragocenejšega predmeta: nekoga je treba olajšati ključa, crkniti alarm, ustvariti diverzijo.*
- *V eni od nalog je treba recimo crkniti stražarja, da kri z njegovega trupla odklene skrivnost.*

O uvrstitvi takih pomenskih rab smo se odločali tudi na podlag razpršenosti besedilnih virov ter različnega avtorstva, kar je mogoče razbrati iz metabesedilnih podatkov v korpusu.

10 Pomenski opis besede

Pomenski opisi, kot jih predvideva slovar, so neločljivo povezani s procesom pomenske členitve, zato sta v leksikografski obravnavi oba postopka združena, hkrati pa zahtevata ločeno opredelitev strategij, ki zagotavljajo čim bolj enotno obravnavo in zanesljivo slovarsko informacijo. Če želimo pomene med seboj hierarhizirati in ločevati, kot predvideva zgradba gesla v LBS, zahteva proces pomenskega členjenja tudi premislek o tem, kako določene pomenske tendence ubesediti in kateri so tisti pomenski elementi in situacije, na podlagi katerih se pomeni med seboj ločujejo.

Pri oblikovanju slovarskih razlag smo v LBS upoštevali naslednja splošna načela:

- Večina pomenske informacije konkretne besede izhaja iz njenega bodisi neposrednega bodisi širšega besedilnega okolja, vendar pa vsega, kar se nahaja vsakokrat v sobesedilu vsake konkretne besede, ni mogoče zajeti v njen pomenski opis, kot ga omogoča (ali bolje omejuje) slovar. Naloga leksikografa je zato predvsem ustvariti zadosten niz asociacij, ki bodo uporabniku pomagale povezati to, kar o svetu že ve, s tem, kar mu je še neznano (Bolinger 1965: 572).
- Slovarske razlage ne razumemo kot sredstva za popis pomenov, ki jih vsebuje določen jezik, ampak predvsem kot sredstvo za reševanje sporazumevalnih zadreg uporabnikov.
- Razlaga pomena besede v slovarju je vedno omejena na uporabo drugih besed, zato je smiselno uporabiti besedišče in določiti tipe slovarskih razlag, ki so za dekodiranje konkretnega pomena najprimernejše, in ki omogočajo zajetje tudi zunajjezikovnih sestavin pomena (tj. pragmatičnih, situacijskih ipd.), ki so v vsakdanjem sporočanju nujne za ustrezno razumevanje in rabo.
- Pri izbiri ustrezne slovarske razlage vedno upoštevamo dvojce: vsebino, tj. podatke, ki jih razlaga vsebuje, in obliko, tj. skladenjsko zgradbo, ki to vsebino posreduje uporabniku.
- Slovarske razlage so načeloma oblikovane tako, da ustrezajo procesu enkodiranja (ustrezni uporabi besede v besedilu). To predvideva vključitev podatkov o tipičnem kontekstu, v katerem se beseda/pomen uporablja, sociolingvističnih lastnosti, kot so žanr, regionalna, kulturološka ipd. vezanost, pragmatične in konotativne razsežnosti ipd., čeprav ni nujno, da vse te podatke vsebuje že pomenska razlaga.
- Pri oblikovanju pomenske razlage je treba upoštevati uporabnikovo (predhodno) jezikovno znanje (npr. šolski uporabnik, odrasli), jezikovno kompetenco (materni govorec, učenec tujega jezika;) ter (ne)razumevanje slovarskih konvencij.

Ker smo imeli pri snovanju LBS v mislih predvsem spletne slovarje, katerih pomembni lastnosti sta neomejenost s prostorom in različne možnosti prikazovanja podatkov, smo celotno pomensko informacijo razdelili na štiri segmente, ki so med seboj povezani, hkrati pa jih je mogoče brati tudi ločeno:

- **pomenski meni**, ki ga tvorijo **pomenski indikatorji**
- podatek v elementu **oznaka**
- **pomenska shema**, ki predstavlja pomen besede v obliki stavčne razlage z izpostavljenimi semantičnimi tipi na udeleženskih mestih in/ali
- **pomenska razlaga**.

10.1 Pomenski meni

Pomenski meni je značilen zlasti za elektronske in spletne slovarje ter slovarje, namenjene učenju tujega jezika (Ptasnik in Lew 2014). V tem smislu je nepogrešljiv del vsake slovarske baze, ki je osnova za slovarje, namenjene sodobnemu mediju in uporabniku. Pomenski meni (Slika 30) prikazuje pomensko zgradbo večpomenskega gesla in logična razmerja med pomeni in podpomeni.



Slika 30: Pomenski meni v programu iLex

Pomenski meni naj bi z razvrstitvijo pomenov ponudil tudi osnovno predstavbo o pogostnosti pomenov, vendar zgolj na ravni osnovnih pomenov, medtem ko podpomeni sledijo logiki pomenske podrejenosti in ne absolutni pogostnosti.

V spletnih slovarjih je klikljiv 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: vzbujati mora osnovno predstavbo o pomenski zgradbi gesla ter uporabnika napotiti na iskani pomen, ne

da bi moral preverjati posamezne pomenske razlage in iskati ustrezne zgleds rabe pri različnih pomenih. Pomenski meni ne sme biti zavajajoč v smislu premajhne medsebojne razločevalnosti ali prekrivnosti pomenov in podpomenov. Prav tako ne sme vzbujati napačnih predstav o pomenskih tendencah besede ali dajati napačnih pomenskih asociacij. Oblikovanje pomenskega menija je zato pomemben del slovarske informacije in zahteva uskladitev z informacijami na drugih nivojih slovarskega opisa.

Logičnost pomenskega menija ni povezana samo z ustrezno ubeseditvijo indikatorjev, pač pa tudi z ustrezno pomensko členitvijo iztočnice, zato je bilo treba najprej razmisliti o stopnji pomenske razdrobljenosti. Spodnji primer kaže možnost združitve podpomenov in razdelitev pomenske informacije med indikator in razlago:

skodelica <i>samostalnik</i>	→	skodelica <i>samostalnik</i>
1 majhna okrogla posoda ali njena vsebina		1 posodica; šalica
1.1 posoda za toplo pijačo, navadno z ročajem		<i>manjša, navadno okrogla</i>
1.2 posoda za hrano, navadno brez ročajev		<i>keramična posoda,</i>
		<i>namenjena shranjevanju in</i>
		<i>postrežbi tople hrane</i>

Pomenski meni smo skušali oblikovati tako, da indikatorji ne bi sugerirali prekrivnih pomenskih vsebin med različnimi pomeni:

kodirati <i>glagol</i>	→	kodirati <i>glagol</i>
1 zapisati s kodo		1 o besedilu; šifrirati
1.1 pretvoriti v računalniško obliko		2 zapisati v obliki kode
1.2 o možganih		2.1 o možganih
1.3 v genetiki		2.2 v genetiki
2 zavarovati s kodo		3 označiti s kodo
3 pripisati kodo		
4 označiti s kodo		

Ker t. i. krovni pomen ne vsebuje razlage in drugih elementov gesla, smo težili k temu, da je v pomenskem meniju že na prvi pogled jasno, da je krovni pomen samo združevalni pomen za več podrejenih podpomenov, npr.

čestitati <i>glagol</i>	glasovati <i>glagol</i>	ansambel <i>samostalnik</i>
1 izraziti lepe želje	1 izraziti mnenje	1 umetniška zasedba
1.1 ob pomembnem dosežku	1.1 na volitvah	1.1 glasbenikov
1.2 ob veselem dogodku	1.2 na tekmovanju	1.2 plesalcev ali igralcev

Zaradi tega je pomembno, da je indikator pri krovnem pomenu ubeseden tako, da ne vzbuja pomenske prekrivnosti s katerim od svojih podpomenov ali z drugimi pomeni znotraj gesla, kot prikazuje popravljena različica:

briti <i>glagol</i>	→	briti <i>glagol</i>
1 odstranjevati dlake		1 odstranjevati dlake
1.1 o moških brkih in bradi		1.1 o moških
1.2 o laseh in dlakah po telesu		2 močno pihati
1.3 odstranjevati dlake drugemu		
2 o vetru		

10.2 Pomenski indikator

Pomenski indikatorji so najpogostejše enobesedni ali kratki pomenski opisi leksikalnih enot v slovarjih, ki dajejo uporabniku osnovno informacijo o pomenu besede ali o pomenskem polju, na katerem se običajno uporablja. Pomenski indikatorji so bodisi namenjeni oblikovanju pomenskega menija, npr. v Macmillan English Dictionary for Advanced Learners (2007²),¹⁰⁷ spodaj na Sliki 31, bodisi predstavljajo izpostavljeni del širše definicije, kot npr. v Longman Dictionary of Contemporary English (2003⁵),¹⁰⁸ na Sliki 32.

fall - definition
★★★

VERB [INTRANSITIVE]  Pronunciation /fɔ:l/  Word Forms

Menu

- [move downwards](#)
- [become lower in amount](#)
- [belong to group/activity](#)
- [change to another state](#)
- [happen on particular day](#)
- [lose power/control](#)
- [be controlled by someone else](#)
- [be killed](#)
- [hang down](#)
- [slope downwards](#)
- [become quieter](#)
- [about light/shadow etc](#)
- [notice someone/something](#)
- [emphasize part of something](#)
- [+phrases](#)

Slika 31: Pomenski meni pri večpomenski iztočnici v MED.

¹⁰⁷ MED: <http://www.macmillandictionary.com/>.

¹⁰⁸ LDOCE: <http://www.longmandictionariesonline.com/>.

mem·o·ry plural memories

- 1 **ABILITY TO REMEMBER** [uncountable and countable] someone's ability to remember things, places, experiences etc

COLLOCATIONS ▲

have a good/excellent memory

have a bad/poor/terrible etc memory

have a memory like a sieve (=have a very bad memory)

have a long memory (=be able to remember things for a long time)

have a short memory (=only remember something for a short time, and soon forget it)

from memory (=using your memory, and not using notes or written instructions)

lose your memory (=lose your ability to remember things)

short-term memory (=your ability to remember things you have just seen, heard, or experienced)

long-term memory (=your ability to remember events that happened a long time ago)

remain/stay/be etched in your memory (=be remembered for a long time)

if my memory serves me correctly/right (=used to say that you are almost certain you have remembered correctly)

- ❖ My *memory* 's not as *good* as it once was.

memory for

- ❖ She has a *terrible memory* for names.

- ❖ Those of you with *long memories* will remember this song.

- ❖ The pianist played the whole piece *from memory*.

- ❖ I'm *speaking from memory*, but I believe it was last May.

- ❖ The first symptom of the disease is often *short-term memory* loss.

- ❖ The image has *remained in my memory* ever since.

- ❖ *If my memory serves me correctly*, he lived in Paris for a while.

- 2 **SOMETHING YOU REMEMBER** [countable usually plural] something that you remember from the past about a person, place, or experience

memory of

- ❖ She talked about her *memories of the war*.

happy/fond/bad etc memories

- ❖ He has lots of *happy memories of his stay in Japan*.

- ❖ My most *vivid memory* is not the accident itself but being in the ambulance.

- ❖ One of my earliest *childhood memories* is of my mother reading stories to me by the fire.

- ❖ Those old songs *bring back memories*.

- 3 **COMPUTER**

- a) [countable] the part of a computer where information can be stored:

- ❖ The data is stored in the *computer's memory*.

- b) [uncountable] the amount of space that can be used for storing information on a computer:

- ❖ 128 Mb of *memory*

- ❖ Personal computers now have much increased *memory capacity*.

Slika 32: Izpostavljeni deli slovarske razlage (ang. signposts) v LDOCE.

S pomenskimi indikatorji smo opisali pomensko zgradbo, tj. pomene in podpomene za vsako leksikalno enoto, ki jo beležimo v bazi, in sicer pri:

- posameznih pomenih in podpomenih
- stalnih zvezah in
- frazeoloških enotah

Za pomenski indikator, kot ga razumemo v LBS, velja, da na kratko in na razumljiv način določi pomenski obseg ali področje obravnavanega pomena, kot je razviden iz splošne rabe, pri večpomenskih besedah pa vzpostavlja še razliko glede na druge pomene ali podpomene.

10.2.1 Vrste pomenskih indikatorjev

Indikatorji, s katerimi najboljše dosegamo razumljive, kratke in pomenskorazločevalne pomenske opise, se pri posameznih besednih vrstah nekoliko razlikujejo, zato smo pri izbiri tipa indikatorja upoštevali predvsem:

- stopnjo pomenske razčlenjenosti besede
- razmerje med pomeni in podpomeni
- možnost razdelitve pomenske informacije na druge elemente pomenskega opisa

Čeprav je vrsta pomenskega indikatorja večinoma vezana na določeno besedno vrsto iztočnice, smo pri oblikovanju pomenskega menija pri večpomenskih iztočnicah uporabljali tudi različne tipe indikatorjev, s čimer smo se želeli izogniti uniformiranosti pomenskega menija. Menili smo, da je na ta način mogoče doseči večjo stopnjo berljivosti, fleksibilnosti in pomenske asociativnosti:

afriški *pridevnik*

1 o afriški celini

2 ki izhaja iz Afrike

2.1 o ljudeh

2.2 o rastlinah in živalih

3 značilen za Afriko

10.2.1.1 Sinonimni indikatorji

Sinonimni indikator se je pokazal kot najustreznejša enobesedna pomenska opredelitev, čeprav z enobesednim ustreznikom nikoli ni mogoče povsem zajeti konkretnega pomena, saj se sinonimni ustrezniki lahko v posameznih vsebinskih elementih, kot so denimo stopnja prevzetosti, konotativne in pragmatične vsebine, register ipd., ločujejo od pomenskih lastnosti razlaganega pomena besede. Pogosto sta iztočnica in

sinonimni pomenski indikator sicer v razmeroma prekrivnem pomenskem razmerju celo pri več pomenih, v določenih pomenih pa ne izkazuje skupnih pomenskih lastnosti. Zato smo pri oblikovanju LBS upoštevali pravilo, da se mora sinonimni indikator v čim večji meri (glede na kolokacijske, stilne, zvrstne, pogostnostne idr. lastnosti) približati pomenu besede v iztočnici. Stopnjo sinonimnosti je mogoče najbolje preveriti s primerjavo besedilnega okolja iztočnice in indikatorja za posamezni pomen, kar omogoča funkcija primerjalne skice v orodju SkE.

Sinonimni indikatorji najbolje opredeljujejo pomene, za katere obstajajo domači ustrezniki, npr. *imigrant* – priseljenec, *titula* – naziv, *abortus* – splav, *transparenten* – prozoren. Manj pa so primerni, kadar je situacija obrnjena, saj predvidevamo, da bo tak indikator sam potreboval razlago, npr. *adolescenca* – puberteta. Hkrati je pomembno, da sinonimni indikatorji ne zajamejo pomenskega polja preširoko ali preozko, zato smo upoštevali tudi možnost porazdelitve pomenske informacije med indikator, razlago in oznako. Upoštevajoč enobesednost, so sinonimni indikatorji primernejši za pomensko bolj razvejane besede, daljši opisni indikatorji pa za pomensko manj razčlenjene.

V indikatorju je lahko navedenih tudi več sinonimov, vendar gre v teh primerih za podatke, ki so primarno namenjeni ustvarjanju pomenskih povezav znotraj baze in z drugimi leksikalnimi bazami, prenos v slovar pa ni nujen:

žgoč <i>pridevnik</i> 3 intenziven; močan; začinjen	zapostavljati <i>glagol</i> 1 krivično obravnavati; prikrajšati; diskriminirati
temperamenten <i>pridevnik</i> 1 ki odkrito izraža čustva; strasten; vročekrven	čednost <i>samostalnik</i> 1 lepa človeška lastnost; vrlina; krepost

Pridevniške iztočnice smo, kot rečeno, zapisovali v osnovni obliki z obrazilom -en, in sicer ne glede na lastnostni ali vrstni pomen. Pri sinonimnem pridevniškem indikatorju pa upoštevamo pomensko razlikovalnost tudi v obliki indikatorja, konkretno pri distribuciji obrazila -en/-ni. To pomeni, da se pri pridevnikih, ki vključujejo tako lastnostni kot vrstni pomen, odločamo za pomensko ustrezno obliko sinonimnega indikatorja, ki zato ni vedno skladen z osnovno obliko pridevniške iztočnice:

evidenten <i>pridevnik</i> 1 očiten 2 najverjetnejši	rumen <i>pridevnik</i> 3 o tisku in publikacijah; tabloidni
---	---

10.2.1.2 Opisni indikatorji

Opisni indikatorji se približujejo kratkim pomenskim razlagam, zato smo jih v LBS navajali predvsem pri enopomenskih ali pomensko manj razčlenjenih besedah:

klikniti *glagol*

1 aktivirati s pomočjo računalniške miške

empiričen *pridevnik*

1 ki temelji na analizi dejanskega stanja

natančno *prislov*

1 z upoštevanjem vseh dejstev ali merljivih vrednosti

Pri večpomenskih besedah smo z opisnimi indikatorji opisovali glavni ali osnovni pomen oz. smo pomensko informacijo razdelili med elemente <indikator>, <definicija> in <pomenska_shema>:

patriarhalen *pridevnik*

1 značilen za moški princip organiziranja družbe

patriarhalna DRUŽBA, SISTEM ali VREDNOTE temeljijo na ureditvi ali mišljenju, ki se podreja moči in vplivu moških

Opisne indikatorje smo uporabljali tudi takrat, ko ustreznega sinonima ni bilo mogoče navesti, potencialna neposredna nadpomenka pa ni ustrezno zajela osnovnega pomenskega polja pomena:

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

10.2.1.3 Nanašalni indikatorji

Nanašalne indikatorje tipa *o čem* smo uporabljali za opredelitev najbolj splošnega pomena pri pridevniki in prislovih, npr. *afriški* – o afriki, *športni* – o športu in na ravni podpomenov za ločevanje kolokatorskih nizov:

alkoholen *pridevnik*

1 o pijači

1.1 o odvisnosti in zlorabi

1.2 o proizvodnji in porabi

2 o kemični spojini

2.1 o kemičnem procesu

2.2 o pripravkih in pripomočkih

balonarski *pridevnik*

1 o zračnem plovilu

1.1 o športnih prireditvah

1.2 o družabnih prireditvah

1.3 o organizaciji

1.4 o skupini ljudi

10.2.1.4 Indikatorji za vzpostavljanje pomenskih mrež – neposredne nadpomenke

Ena od nalog indikatorjev, ki opredeljujejo pomen samostalniške besede, je vzpostavljati pomenske mreže, zato smo v takih primerih izbrali najprimernejšo neposredno nadpomenko razlagane besede, npr. *angina* – bolezen; *arnika* – rastlina; *arheologija* – znanstvena veda, v razlagi ali pomenski shemi pa smo pomen natančneje opredelili:

biopsija *samostalnik*

medicinski postopek

medicinski postopek, pri katerem se iz določenega organa odvzamejo celice ali tkivo za mikroskopski pregled, na podlagi katerega se ugotovi morebitna prisotnost bolezni

Namen tovrstnih indikatorjev je omogočiti pomensko povezavo besed oz. njihovih pomenov prek sorodnih indikatorjev in hkrati vzpostaviti možnost povezave z obstoječimi leksikalnimi in ontološkimi bazami.

10.2.2 Oblikovanje pomenskih indikatorjev

Pri ubeseditvi pomena ali podpomena s pomenskim indikatorjem smo upoštevali tri osnovna načela:

- razumljivost
- kratkost
- pomenska razločevalnost

10.2.2.1 Razumljivost

Razumljivost indikatorja pomeni v prvi vrsti preprostost v načinu ubeseditve, ki ni rigidna in značilna za abstraktno metaslovarsko dikcijo, ampak se približuje vsakdanjemu sporočanju. Razumljivost smo skušali doseči z upoštevanjem načel, ki jih opišemo v nadaljevanju.

Asociativnost, ki jo v odnosu do pomena vzpostavlja indikator, pomaga uporabniku določiti pomensko polje, v katerem prepozna svoj pomenski problem. Asociativnost je mogoče doseči z izbiro ustrezne sopomenke, ki zagotavlja tudi primerno kratkost:

omrežiti *glagol*

- 1 očarati; zapeljati
 - 1.1 navdušiti; prepričati
- 2 onesposobiti
- 3 razširiti se; zavzeti
- 4 povezati v omrežje

Sinonimni, opisni in ontološki indikatorji (izjema so nanašalni) navadno izražajo tudi ustrezno besedno vrsto iztočnice. Spodnji primer predstavlja najprej izhodiščno nato pa redigirano rešitev v LBS:

obetati *glagol*

- 1 prihodnost, napoved
 - 1.1 verjetnost
 - 1.2 lastnost
 - 1.3 obljuba, napoved
 - 1.4 pričakovanje

→

obetati *glagol*

- 1 napovedovati se; kazati se
 - 1.1 vzbujati pričakovanja
- 2 napovedovati; predvidevati
 - 2.1 pričakovati; želeli si

Glagolski in deležniški pridevniški indikatorji naj bi poleg besedne vrste izražali tudi ustrezen glagolski vid iztočnice, vendar samo v primerih, kadar je dovršnost oz. nedovršnost pomensko predvidljiva oz. dejansko izražena, npr.

nakupovati *glagol*

- 1 pridobivati za plačilo

pobirati *glagol*

- 3 postavljati pokonci

obljubljen *pridevnik*

- 1 zagotovljen

izvožen *pridevnik*

- 1 prodan v tujino

Pri dvovidskih glagolih smo se odločali za nevtralnejšo, navadno nedovršno obliko glagolskega indikatorja, npr. *anketirati* – pridobivati informacije; *diskriminirati* – postavljati v slabši položaj, vendar tudi: *aktivirati* – povzročiti delovanje naprave, oz. za obliko indikatorja, ki je sugerirala pomensko ustrežnejši opis, npr.

angažirati *glagol*

- 1 dejavno se vključevati; prizadevati si
 - 1.1 povabiti k sodelovanju; zaposliti
 - 1.2 motivirati za delo
 - 1.3 vzbuditi zanimanje; pritegniti
 - 1.4 jemati čas
- 2 vložiti denarna sredstva

Glagolske prehodnosti v indikatorjih nismo izražali eksplicitno, npr. z navedbo določil v obliki poljubnostnega zaimka (gl. pogl. o opuščanju določil v nadaljevanju). V primeru pomenskoodločilne prehodnosti smo izpostavili samostojni pomen, v indikatorju pa smo uporabili ustrezno obliko glagola ali konkretni predmet (*strmoglaviti*), ki napoveduje tipično določilo. Pomensko-skladenjska podoba s povratno-svojilnim *si*, je razvidna iz pomenske sheme (*briti*):

strmoglaviti *glagol*

- 1 zrušiti se na zemljo
 - 1.1 zrušiti letalo

briti *glagol*

- 1 odstranjevati dlake

če si ČLOVEK brije DLAKE ali brije DLAKE drugemu ČLOVEKU, jih odstranjuje, navadno s PRIPOMOČKOM

V primerih, kjer glagol izraža povratnost dejanja na delovalnika samo v katerem od svojih pomenov in torej prosti morfem ni izražen na ravni iztočnice, je taka pomenska lastnost vedno razvidna iz pomenskega indikatorja in iz ubeseditve v pomenski shemi:

asimilirati *glagol*

- 1 vključiti se v večino

če se PRISELJENEC ali MANJŠINA asimilira, prevzame navade, kulturo in jezik LJUDI in OKOLJA, v katerem živi

10.2.2.2 Kratkost

Načelo kratkosti pri oblikovanju pomenskih indikatorjev ima v LBS dvojni namen: s kratkimi in jasnimi indikatorji smo se želeli izogniti kopičenju informacij in zapletenim skladenjskim konstrukcijam, saj velja, da daljši opisi otežujejo procesiranje vsebine, hkrati pa so namenjeni oblikovanju pomenskega menija, ki mora

biti pregleden, kar je zlasti pomembno pri pomensko obsežnejših iztočnicah. V ta namen smo se izogibali naštevanju in navajanju vezljivostnih mest v obliki zaimkov ter vzpostavljanju notranje pomenske krožnosti. Za ponazoritev navajamo ponekod izhodiščno in na to ustreznejšo rešitev:

(a) Daljši opisi in naštevanje pomenskih elementov v indikatorju.

pecelj <i>samostalnik</i> 1 del rastline, s katerim je list, cvet ali plod povezan z vejo	→	pecelj <i>samostalnik</i> 1 del rastline
cveteti <i>glagol</i> 3 biti srečen, zdrav, vesel	→	cveteti <i>glagol</i> 3 izgledati srečen

Naštevane indikatorjev, ko ti niso neposredni sinonimi, pač pa dodatno pojasnjujejo pomen, je bolj ustrezno pri pomensko manj razvejanih zgradbah:

grmeti <i>glagol</i> 1 oddajati glasen zvok 2 glasno izražati mnenje, nestrinjanje	gospodinjiti <i>glagol</i> 1 voditi in opravljati gospodinjska dela 2 ne zaposliti se
drama <i>samostalnik</i> 1 gledališko, filmsko ali literarno delo 2 napeto, razburljivo stanje	ležerno <i>prislov</i> 1 malomarno, podcenjujoče 2 brezskrbno, sproščeno

(b) Navajanje predmetnih določil:

blokirati <i>glagol</i> 2 ustaviti postopek ali nadaljevanje česa	→	blokirati <i>glagol</i> 2 preprečiti uresničitev 2.1 o postopku, dejavnosti 2.2 o delovanju snovi v telesu
---	---	--

Prav tako v indikatorjih načeloma nismo zapisovali predložnih predmetnih določil, zapisanih v obliki poljubnostnih zaimkov, npr.

komentirati <i>glagol</i> 1 povedati svoje mnenje o čem 2 poročati o čem 3 natančneje pojasniti kaj	→	komentirati <i>glagol</i> 1 izraziti mnenje 2 sproti poročati 3 natančneje pojasniti
---	---	--

(c) Sledili smo tudi načelu, da morajo indikatorji pomenski obseg besede pojasniti do te mere, da ni treba dodatno iskati podatkov v slovarju. To je pogosto takrat, ko določena beseda sama na sebi nima leksikalnega pomena, ampak se nanj le nanaša. Da bi se izognili krožnosti, smo v takih primerih upoštevali možnost povezav znotraj baze, npr. prek indikatorjev in semantičnih tipov v pomenski shemi:

karizmatičen *pridevnik*

1 o značaju

karizmatičen ČLOVEK ali ZNAČAJ ima lastnosti, s katerimi pozitivno vpliva na druge ljudi in jih privlači, tako da ga cenijo in spoštujejo

karizma *samostalnik*

1 značajska lastnost

značajska lastnost človeka, ki pozitivno vpliva na druge ljudi ter vzbuja občudovanje in spoštovanje

Optimalno informacijo, ki izhaja iz navajanja vseh potrebnih opisov na enem mestu oz. s pomočjo povezav znotraj baze, je mogoče doseči le, če v bazi dejansko obstajajo pomenski opisi pomensko povezanih iztočnic¹⁰⁹ ali stalnih zvez:

toplogreden *pridevnik*

1 o učinku tople grede

topla greda

ekologija

proces, pri katerem se toplotno sevanje, ki prihaja v ozračje z Zemlje, vrača nazaj in povzroča višjo temperaturo, kot bi bila, če bi Zemljino površje ogrevalo le sonce

Ker pomeni izogibanje krožnosti hkrati tudi navajanje več informacij na enem mestu, smo se pri razdelitvi pomenskega opisa med indikator in pomensko shemo/razlago odločali v skladu s funkcionalnostjo pomenskega menija in upoštevali uskladitev informacij med pomensko povezanimi iztočnicami.

Pri moško-ženskih parih, kot so: *etnolog* – *etnologinja* ipd., in pri iztočnicah za žensko obliko: *babica*, *sestra*, *deklica*, smo v indikatorju uporabili ustrezno obliko nadpomenke, npr. *etnologinja* – strokovnjakinja za etnologijo; *slikarka* – likovna umetnica, podrobnejši pomenski opis pa smo prenesli v razlago:

¹⁰⁹ Npr. rešitev: *marinirati* – pustiti stati v marinadi, kjer iztočnica *marinada* ni izkazana, je manj primerna.

babica *samostalnik*

2 poklic

strokovno usposobljena ženska, ki skrbi za zdravje nosečnice ter vodi in nadzoruje porod

Enako smo postopali tudi, kadar bi bila neposredna nadpomenka sama na sebi preohlapna:

čistilka *samostalnik*

1 ženska, ki čisti

ženska, ki se ukvarja s čiščenjem prostorov ali s kemičnim čiščenjem tkanin in oblačil

Pri živalskih parih, npr. *tiger – tigrica, lisjak – lisica*, smo pomene, ki označujejo spol, v indikatorju označevali z nadpomenko žival ali samec oz. samica, pomenske podrobnosti pa smo izpostavili v razlagi:

lisica *samostalnik*

1 žival

žival z rdečkastorjavim kožuhom in košatim repom

lisjak *samostalnik*

1 zviti, prebrisan moški

2 žival

samec živali z rdečkastorjavim kožuhom in košatim repom

10.2.2.3 Razločevalnost

Kot že večkrat omenjeno, je temeljna naloga indikatorjev oblikovati smiselni in pregleden pomenski meni obravnavane besede, iz katerega je mogoče razmeroma hitro ugotoviti njeno pomensko razvejanost ter razmerja med pomeni. Upoštevanje načel, ki smo jih navedli, nam pri tem vsekakor lahko pomaga, vendar pa je v končni fazi vedno leksikografova presoja tista, ki odloča o izbiri indikatorja, njegovi ubeseditvi in končni podobi pomenskega menija. Na tem mestu se zdi pomembno poudariti, da je treba pri izbiri indikatorjev vedno upoštevati razmerje do drugih indikatorjev oz. do pomenskega menija kot celote. Pri iskanju najboljše možnosti in na podlagi analiz več različnih pomenskih menijev smo zato pri izdelavi LBS upoštevali še dodatna navodila.

Izogibali smo se rabi večpomenskih besed v vlogi indikatorja. Ker pa tega načela ni mogoče upoštevati absolutno, smo težili k temu, da smo večpomenski indikator uporabljali v pomenu, ki predstavlja čim bolj osrednjo rabo te besede. Na podlagi

primerjalnih skic v orodju SkE smo lahko ugotovili, da se beseda *območje* tipično uporablja za opisovanje prostorskih pomenov, beseda *področje* pa za opisovanje dejavnosti:

domena <i>samostalnik</i> 1 področje dejavnosti, zanimanja	specializiran <i>pridevnik</i> 1 usmerjen na določeno področje delovanja
mokrišče <i>samostalnik</i> močvirnato območje	evakuirati <i>glagol</i> izprazniti stavbo ali območje zaradi nevarnosti

Izogibali smo se indikatorjem, ki nastopajo v katerem od svojih obrobni pomenov ali v za dani pomen netipični rabi, ker bi lahko na ta način vzpostavili neustrezno asociacijo:

brusiti <i>glagol</i> 2 odnašati površino	→	brusiti <i>glagol</i> 2 obdelovati hrapavo površino
zacviliti <i>glagol</i> 1 izpustiti visok glas		zacviliti <i>glagol</i> 1 oglasiti se z visokim glasom

Izogibali smo se rabi besednih oblik v vlogi indikatorja in besed, ki so redko zastopane v običajni rabi:

koketirati <i>glagol</i> 2 želeli delovati všečno	→	koketirati <i>glagol</i> 2 izražati naklonjenost; spogledovati se
klikniti <i>glagol</i> 2 proizvesti zvok	→	klikniti <i>glagol</i> 2 ustvariti zvok

Izogibali smo se indikatorjem v nikalni obliki:

civilist <i>samostalnik</i> 1 nečlan vojske ali policije	→	civilist <i>samostalnik</i> 1 kdor ni član vojske ali policije
--	---	--

Kot izjemo smo obravnavali indikatorje pri iztočnicah, ki vsebujejo nikalne pomenske vsebine, obenem pa smo, če je bilo mogoče, navedli tudi nezanikani sinonimni indikator:

odreči *glagol*

1 ne ugoditi prošnji; zavrniti

2.1 ne priznati; zanikati

3 ne želeli imeti; odpovedati se

V vlogi indikatorjev nismo uporabljali besed, ki izkazujejo v korpusu relativno malo pojavitev, ali besed, za katere je mogoče predvidevati, da so, zlasti z vidika šolskega uporabnika, težje razumljive in da tudi same potrebujejo razlago:

cveteti *glagol*

1.1 namnožitev alg na površini

→

cveteti *glagol*

1 pognati cvetove

1.1 o algah

gnati *glagol*

biti osnovna motivacija

→

gnati *glagol*

povzročati zanimanje; motivirati

smešen *pridevnik*

1 komičen

2 absurden

→

smešen *pridevnik*

1 ki vzbuja smeh; komičen

2 nepremišljen; nerazumljiv; čuden

olje *samostalnik*

1 viskozna tekočina

→

olje *samostalnik*

1 gosta prozorna tekočina

Izogibali smo se indikatorjem v *frazeološkem* ali *metaforičnem pomenu*, ki sami rabijo pojasnitev. K navajanju indikatorjev v frazeološkem pomenu nas navadno napeljujejo pomenske rabe, ki izražajo konotacijske ali pragmatične vsebine. V takih primerih smo tovrstno pomensko informacijo prenesli v oznako ali pa smo jo vključili v pomensko shemo oz. razlago, kot kažejo spodnji primeri:

izpljuniti *glagol*

1 spraviti na plano izdelek

→

izpljuniti *glagol*

1 o napravi; izvreči

če NAPRAVA izpljune POTRDILO
ali IZDELEK, ga izvrže iz posebne
za to izdelane odprtine

klikniti *glagol*

4 osebnostno se ujeti

→

klikniti *glagol*

4 postati si všeč

če rečemo, da med DVEMA
ČLOVEKOMA klikne, želimo
povedati, da se dobro ujameta ali
da si postajata všeč

leteti <i>glagol</i> vreti na plano	→	leteti <i>glagol</i> pojavljati se v veliki količini
lotiti se <i>glagol</i> 1 narediti prve korake	→	lotiti se <i>glagol</i> 1 začeti

10.3 Pomenska shema

Pomenska shema je v strukturi LBS tisti element geselskega članka, ki vsebuje skladenjski opis konkretnega pomena besede. Kot taka predstavlja ključni del pomensko-skladenjskega opisa leksikalne enote, ki izhaja iz predpostavke, da pomena besede ni mogoče ločiti od njegove skladenjske realizacije. Ker zapis pomenske sheme nosi v sebi ključne informacije o realizaciji pomena v okviru skladenjsko zaključene zgradbe, kot je prosti stavek¹¹⁰ ali (podredna) besedna zveza, predstavlja izhodišče za (a) oblikovanje pomenske razlage v obliki stavčne razlage, ki vključuje vse ključne skladenjsko-pomenske elemente posameznega pomena, tj. udeležence in okoliščine kot tudi konotativne in pragmatične pomenske elemente, ter (b) za strojno procesiranje in avtomatsko prepoznavanje pomenov oz. za semantično označevanje korpusov.

Pri zapisovanju pomenske sheme, kot jo razumemo v LBS, ločimo glagolske od samostalniških, pridevniških in prislovnih pomenov. Pri glagolskih pomenih je zapis pomenske sheme povezan z razumevanjem glagolske vezljivosti kot sposobnosti leksikalnih enot, da vežejo nase druge izraze, s čimer gradijo stavčno strukturo in povezujejo ravnino leksikalnih enot in njihovih pomenov z ravnino stavčne skladnje (Svozilová et al. 2005). O zapisu vezljivostnega vzorca v pomenski shemi je zato mogoče govoriti tudi pri nekaterih izglagolskih samostalnikih in pridevnikih, ki predvidevajo udeležence na predvidenih vezljivostnih mestih. Pri pridevniških in prislovnih pomenih, ki se realizirajo predvsem v modifikacijski pomenskopskladenjski vlogi, npr. *čeden fant/dekle*, *slabo/dobro se odrezati*, smo v pomenski shemi označevali tudi pomensko relevantna jedra.

Teoretično temelji zapis pomenskih shem v LBS na smernicah projektov FrameNet in Corpus Pattern Analysis, ki smo jih opisali v prvem delu monografije. Logika pomenskega opisa, kot jo predstavlja FrameNet, temelji na ugotavljanju potrebnih (jedrnih in nejedrnih) shemskih elementov za evociranje konkretnega

¹¹⁰ Vidik pomenske povezovalnosti (Čermák 2010: 282) predvideva izraženost pomensko kompatibilnih besed tudi zunaj stavčne povedi obravnavanega glagola, kot bomo prikazali v nadaljevanju. To koncept vezljivosti, ki navadno pomensko kompatibilnost in kolokabilnost oblikoskladenjsko kategorizira le znotraj stavka, širi na raven besedila. V zvezi s tem prim. tudi Žele (2010; 2012: 251).

pomena besede, ki jih je nato na podlagi skupnih oz. razlikovalnih elementov mogoče združevati v skupine, t. i. okvirje oz. sheme. Tako npr. shema *Staranje* (ang. *Aging*) vključuje glagola *postarati se*, *starati se* ter pridevnik *ostarel* (age, mature in aging). Obvezni ali jedrni element te skupine je element entiteta, ki je podvržena staranju, nejedrni elementi pa so lahko še: okoliščine, način, mesto oz. lokacija, rezultat in čas. Če skušamo pravilo zelo poenostavljeno prenesti na slovenski zgled, bi npr. v povedi: *pretirano je kadila, njen obraz se je hipoma postaral*, zvezo »njen obraz« označili za *entiteto*, »hipoma« za *način*, »pretirano je kadila« pa za *okoliščine*, v katerih je entiteta podvržena procesu staranja.

Drugo teoretično izhodišče predstavlja projekt Corpus Pattern Analysis oz. analiza korpusnih vzorcev Patricka Hanksa, katere namen je povezati pomen besede z njeno rabo na podlagi analize stavčnih vzorcev in kolokacij, v katerih se beseda pojavlja v korpusu. V slovarju Pattern Dictionary of English Verbs¹¹¹ je za vsak glagol določeno število osnovnih vzorcev, ki ustrezajo registriranim pomenom, ter vzorci, ki predstavljajo ustvarjalno rabo ali t. i. eksploatacije. V Hanksovi analizi korpusnih vzorcev tako npr. glagol *adore* (oboževati; častiti) predvideva dva osnovna vzorca, ki izkazujejo tudi možnost eksploatacije:

- 1 66 % **[[Human 1]] adore [[Human 2 | Deity | {Animal = Pet}]]**
 [[Human 1]] has extremely strong positive emotions with regard to
 [[Human 2 | Deity |
 {Animal = Pet}]] exploit.
- 2 34 % **[[Human]] adore [[Entity | Eventuality]]**
 [[Human]] enjoys [[Entity | Eventuality]] very much exploit.

Tak način označevanja pomenskih lastnosti besed služi predvsem v jezikovnotehnološke namene ter za empirične jezikoslovne analize. Obema projektoma, kot tudi oblikovanju pomenskih shem v LBS, je skupno, da izhajajo iz kontekstualnega načina določanja pomena oz. iz spoznanja, da pomeni izhajajo iz stavčnih in besednozveznih vzorcev, v katerih se besede tipično pojavljajo.

10.3.1 Oblikovanje in zapis pomenske sheme

Oblikovanje pomenske sheme je najprej povezano s pomensko členitvijo besede. Kot bomo pokazali v nadaljevanju, so spremembe v udeleženski zgradbi in semantičnih tipih na udeleženskih mestih največkrat neposredno povezane s pomenskimi spremembami, ki jih beseda aktivira v sobesedilu. Zato poteka postopek

¹¹¹ Opis korpusnih vzorcev in dostop do konkordanc, ki ustrezajo posameznemu vzorcu, je za trenutno 877 angleških glagolov dostopen na spletni povezavi: <http://www.pdev.org.uk/#browse?q=f=C>.

oblikovanja pomenske sheme skupaj z analizo konkordančnega niza oz. z analizo stavčnih vzorcev v besedni skici. Beležili smo torej tipične udeležence in okoliščine in jih povezovali v stavčne vzorce, značilne za konkretni pomen, kot prikazuje Slika 33 za glagol *kloniti*.

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 Mihijenkem .	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 zaslišanjem 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 33: Analiza konkordančnega niza z vidika udeležencev

Pomensko shemo smo oblikovali v treh korakih tako, da smo za vsak registrirani pomen glagola, samostalnika, pridevnika in prislova ugotovili:

- za realizacijo konkretnega pomena v korpusu izkazane udeležence in okoliščine, ki se realizirajo pretežno v besednozveznem in stavčnem vzorcu pomena
- semantične tipe izkazanih udeležencev in okoliščin
- pomensko relevantne situacije konkretnega pomena oz. t. i. pomenski scenarij.

Pri zapisu udeleženske zgradbe pomena v pomenskin shemi smo v prvem koraku zapisali tisti vzorec, ki je skladenjsko tipičen, hkrati pa predvideva vse udeležence, torej tudi tiste, ki se v dejanskih realizacijah lahko opuščajo oz. se ne izražajo.

Beleženje udeležencev je v prvem koraku zahtevalo ugotovitev njihovega števila ter realizacijo na oblikoskladenjski ravni, ki se kaže v razporeditvi in sklonskih oblikah določil in dopolnil. Prvi korak v oblikovanju pomenske sheme bi za enega od pomenov glagola *sesti*, samostalnika *nota*, pridevnika *pozoren* in prislova *vestno* zapisali takole:

sesti ustrezati	nota značilnost	pozoren ustrežljiv	vestno v skladu s pričakovanji
KAJ sede KOMU	KAJ daje noto ČEMU	biti pozoren DO KOGA	vestno GLAGOL

V drugem koraku smo beležili semantične tipe udeležencev na podlagi izkazane kolokacijske slike in pomenske členitve. Semantični tipi namreč predstavljajo abstraktne pomenske oznake tipičnih leksikalnih zapolnitev udeleženskih mest, ki jih na kolokacijski ravni potrjujejo kolokacije. Semantične tipe smo v pomenski shemi zapisovali z velikimi črkami, namenjeni pa so nadaljnjemu procesiranju in možnostim avtomatskega pomenskega razdvoumljanja:

sesti ustrezati
DEJAVNOST, STANJE ali LASTNOST česa sede ČLOVEKU
nota značilnost
LASTNOSTI česa dajejo noto IZDELKU, KRAJU ali DEJANJEM
pozoren ustrežljiv
ČLOVEK je pozoren do drugega ČLOVEKA
vestno v skladu s pričakovanji
vestno IZPOLNJEVATI

V tretjem koraku smo povezali skladenjsko realizacijo pomena, v kateri so registrirani semantični tipi na predvidenih udeleženskih mestih, z okoliščinami in lastnostmi sporočanje situacije v stavčno razlago. Pravimo, da registrirani skladenjski in leksikalni podobi pomena dodamo še pomenski scenarij:

sesti ustrezati
če rečemo, da neka DEJAVNOST, STANJE ali LASTNOST česa ČLOVEKU sede, s tem mislimo, da mu ugaja ali ustreza
nota značilnost
če LASTNOSTI česa dajejo IZDELKU, KRAJU ali DEJANJEM svojo noto, se v njem izražajo in ga delajo posebnega
pozoren ustrežljiv
če je ČLOVEK <i>pozoren</i> do drugega ČLOVEKA, je do njega ustrežljiv in mu izkazuje naklonjenost

vestno

v skladu s pričakovanji

če človek vestno IZPOLNJUJE pričakovanja in OPRAVLJA dolžnosti, dela to tako, kot se od njega pričakuje ali je predpisano

10.3.1.1 Udeleženci in okoliščine pomena

Upoštevajoč nalogi, ki ju ima pomenska shema, je pri obravnavi posameznih besednih vrst treba opozoriti na ločevanje zapisa (a) udeleženske zgradbe na ravni stavka (natančneje enostavčne povedi), kar sodi v okvir vezljivosti, ki se v literaturi pripisuje primarno glagolom, z vidika sintagmatskih pristopov pri leksikoloških in leksikografskih analizah (Kopřivova 2006; Čermáková 2009; Čermák 2010 et al.) pa tudi nekaterim samostalniškim in pridevniškim pomenom, in zapisa (b) kolokabilnosti pri tistih besednih vrstah (pridevniki in nekateri prislovi) oz. posameznih njihovih pomenih, ki se realizirajo primarno v modifikacijski oz. determinacijski vlogi.

Stavčni vzorec, ki ga vsebuje pomenska shema, naj bi praviloma vključeval vse pomensko relevantne udeležence oz. prikazoval polno udeležensko zgradbo danega pomena, ki jo je mogoče ugotoviti iz večjega števila realnih besedilnih rab, pri čemer smo se zavedali, da se ta ne uresničuje nujno le znotraj enostavčne povedi, pač pa lahko tudi v okviru širšega besedila. Pomensko prisotnost udeležencev smo zato v pomenski shemi prikazovali tudi takrat, ko se ta bodisi oblikoskladenjsko ne izraža oz. se opušča, in takrat, ko se izraža v besedilnem okolju, tj. znotraj stavčnega vzorca, ki ne pripada glagolu, ki ga pomensko analiziramo. Udeleženskih vlog tipa: vršilec, prizadeto, prejemnik, na posameznih mestih udeleženske zgradbe v pomenski shemi nismo beležili.

Zaradi terminološke jasnosti in sledenja tradiciji govorimo v primeru stavčne oz. povedkove vezljivosti o udeleženski zgradbi, udeležencih, določilih in dopolnilih, pri skupini pridevnikov in prislovov, ki pomensko določajo svoja jedra, pa izraza udeleženec in udeleženska zgradba uporabljamo tudi v pomensko širšem, torej ne zgolj vezljivostnem smislu, saj z njima označujemo vse pomensko in skladenjsko relevantne elemente v besedilnem okolju obravnavanega pomena. V primeru modifikacijske in determinacijske vloge pridevnikov in prislovov govorimo za razliko od vezljivosti o kolokabilnosti ter o kolokatorjih kot tistih elementih, s katerimi obravnavana beseda za določeni pomen kolocira na ravni podredne besedne zveze. Pridevnikom oz. večini pridevniškim pomenom so to, kar so glagolom udeleženci, samostalniška ali prislovna jedra, pri prislovih pa glagolska ali pridevniška jedra, kar ponazarja Preglednica 12.

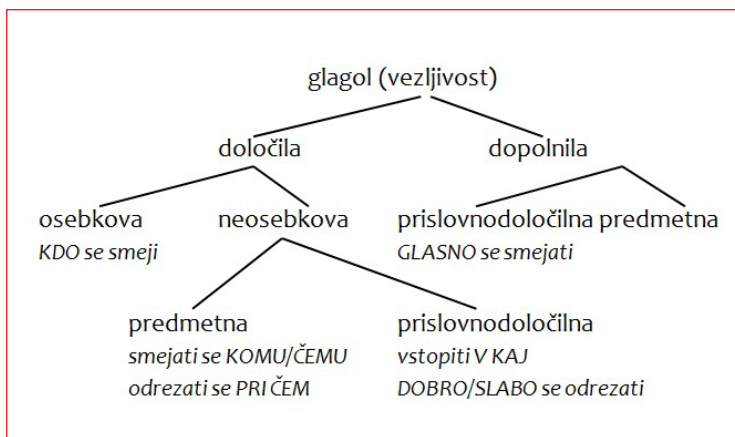
Preglednica 12: Udeleženci in kolokatorji v pomenski shemi z vidika besedne vrste

Vezljivost		Kolokabilnost
BESEDNA VRSTA	UDELEŽENCI	
	določila	dopolnila
glagol	KDO vzame KOMU KAJ	pasti KAM KAKO se odrezati
samostalnik	KAJ daje noto ČEMU poplava ČESA	odhod KAM
pridevnik	KDO je alergičen NA KOGA poln ČESA	čeden MOŠKI, ŽENSKA
prislov		vestno IZPOLNJE- VATI

10.3.1.1.1 Glagol

Glede na organizacijsko vlogo glagola v stavku smo se za zapis pomenske sheme odločali pri vsakem pomenu in podpomenu (razen pri krovnem pomenu). Pri zapisovanju glagolskih pomenskih shem smo uporabljali stavčno zgradbo (najpogostejše v obliki *če*-stavka), ki omogoča navedbo vseh udeležencev in okoliščin, njihovih semantičnih tipov in pomenskega scenarija. Tako oblikovane pomenske sheme predstavljajo izhodišče za oblikovanje stavčnih razlag, zapis udeležencev in okoliščin z velikimi črkami pa je, kot rečeno, namenjen avtomatskemu prepoznavanju udeleženske zgradbe pomena oz. nadaljnji računalniški obdelavi.

Pri tem je pomembno opozoriti, da z vidika stavčnega vzorca, znotraj katerega prepoznavamo pomensko intenčno polje glagola, obravnavamo kot pomensko relevantna tako (številčno omejena) vezljivostna določila kot tudi skladijsko nevezavne udeležence, ki so (običajno številčno neomejena) nevezljivostna dopolnila (prim. Žele 2003: 35), kot prikazuje Slika 34.



Slika 34: Glagolska določila in dopolnila glede na stavčne člene

Tako izhodišče narekuje skladenjski zapis pomena, v katerem so navedeni vsi pomensko relevantni elementi, ne glede na skladenjsko realizacijo (vezljivost : družljivost), saj se pomenska usmerjevalna ali opredelitvena zmožnost besede v procesu upovedovanja izraža tako z vezljivostnimi kot z družljivostnimi skladenjskimi sredstvi. Hkrati s številom udeležencev smo v pomenski shemi določili tudi njihove sklonske oblike in tipično razporeditev v stavčni zgradbi, npr.

pasti *glagol*

1 premikati se od zgoraj proti navzdol

če PREDMET zaradi sile težnosti pade na nek PROSTOR ali POVRŠINO, prileti tja iz zraka ali z višine

1.1 izgubiti; izpustiti

če ČLOVEKU pade PREDMET iz OBLAČILA, PREDMETA ali DELA TELESA na nek PROSTOR ali POVRŠINO, ga izgubi ali izpusti

Pri ugotavljanju in zapisovanju udeležencev in okoliščin predvidevamo predvsem dve tipični situaciji:

- različne možnosti opuščanja udeležencev in okoliščin oz. njihovo (ne)izražnost in
- spremembe v udeleženski zgradbi pomena, ki ne vplivajo na pomen besede v iztočnici.

10.3.1.1.1 (Ne)izraženost udeležencev

Zapis udeleženske zgradbe, ki predstavlja prvi korak v postopku oblikovanja pomenske sheme, zahteva prepoznavanje vseh pomensko nujnih udeležencev in okoliščin za konkretni pomen glagola. S pojmom vezljivosti je v literaturi navadno povezano tudi vprašanje obligatornosti in fakultativnosti, in sicer tako s skladenjskofunkcijskega kot sporočanjanskega oz. besedilnega vidika.¹¹² Glede na omenjene vidike so ti elementi (tj. določila in dopolnila) lahko pomensko (ne)nujna, skladenjsko (ne)obvezna in izrazno (ne)opustljiva (Žele 2001: 18). Izhajajoč iz analize realnih jezikovnih podatkov, pa je določitev meje med obveznimi in neobveznimi določili ter dopolnili pogosto težko določljiva. Pomenska obligatornost se tako v primeru nekaterih glagolskih pomenov tiče tudi prislovnih izrazov oz. izrazov v prislovni vlogi, npr. znati se *kje*, obnašati se *kako* ipd., ki se tradicionalno dojemajo kot fakultativni udeleženci.

Z vidika izraženosti pomensko relevantnih elementov smo pri analizi konkordančnega niza lahko ugotovili, da so pomensko obvezni elementi posameznega pomena v besedilnih realizacijah lahko eksplicitno izraženi bodisi v okviru stavčnega vzorca, ki ga narekuje glagol, bodisi zunaj njega ali pa so neizraženi.

V primerih, kjer so vsi udeleženci (in okoliščine), ki jih predvideva konkretni glagolski pomen, vključno z vršilcem na osebkovem mestu, izraženi znotraj stavčnega vzorca obravnavanega glagola, je zapis pomenske sheme razmeroma preprost, npr.

stisniti *glagol*
pridobiti tekočino

če ČLOVEK s PRIPOMOČKOM ali prsti stisne SADEŽ, RASTLINO ali SEME, pridobi iz njih užitno TEKOČINO

- Iz GROZDJA so na DOMAČI STISKALNICI stisnili 70 litrov LEDENEGA VINA.
- Mladi vršički imajo največ v sebi, ali pa če stisnemo SOK iz KOPRIVE s STISKALNICO.

Poleg običajnih realizacij udeležencev v okviru stavčne zgradbe glagolskega pomena so v nekaterih primerih pomensko obvezni udeleženci izraženi tudi v besedilnem okolju, tj. zunaj stavčnega vzorca obravnavanega glagola, kot prikazujejo spodnji zgledi:

¹¹² Žele (2001: 70; 2003: 23/24) govori v prvem primeru o skladenjsko oz. stavčnovzorčno blokiranih kategorijah, v drugem primeru pa o pragmatično oz. besedilno blokiranih kategorijah.

- *JABOLKA je treba najprej umiti, razrezati in stisniti. JABOLČNI SOK nato pustimo stati.*
- *Lani smo pridelali nekaj sončnic in SEME stisnili v oljarni. OLJE smo branili v velikanski steklenici v svetli sobi in letos ugotovili, da se je pokvarilo.*
- *Z rastlin odstranimo korenine in od bledega dela stebela, tik nad koreninami narežemo na 5–10 mm ODRESKE. Iz TEH v STISKALNICI ZA ČESEN stisnemo RASTLINSKI SOK.*

Kot sestavne elemente pomenske sheme smo obravnavali tudi pomenske elemente, ki se v besedilnem okolju realizirajo na način, ki ne predstavlja vezljivo-stnega vzorca glagola, katerega pomen obravnavamo, pač pa se pojavljajo znotraj stavčnega vzorca kot del udeleženske zgradbe katerega drugega glagola, kot ponazarjajo zgledi:

- *Labko pa JABOLKO razpolovite in stisnete SOK na OŽEMALNIKU ZA LIMONO.*
- *Če vam delajo cele JAGODE zaradi svoje lupine težave, potem stisnite SOK in ga po malem pijte.*

Če se taki elementi vsaj v nekaterih besedilnih realizacijah pojavljajo tudi kot sestavni deli udeleženske zgradbe obravnavanega glagolskega pomena, jih v pomenski shemi obravnavamo kot pomensko nujne. Tako ravnamo tudi v primerih, kjer so udeleženci zastopani z anaforičnim zaimkom ali stavčnim prilastkom:

- *Vse, kar mora uporabnik narediti, je, da potisne OŽEMALEC v LIMONO in nato iz nje stisne SOK.*
- *Po napovedi, naj bi trta letos kljub virusni okužbi obrodila vsaj deset kilogramov GROZDJA, ki ga bodo gorički vinogradniki stisnili v MOŠT in donegovali v vino.*

Kompleksnejše so situacije, kjer sicer pomensko nujni udeleženci bodisi običajno ali pa nikoli niso izraženi. Mejne so situacije, kjer je pomensko obvezni element izpuščen, ker je kljub njegovi odsotnosti zadoščeno pomenski obvestilnosti sporočila: kakršnokoli izpuščanje, izrazno urejanje, preurejanje ter preoblikovanje povedja in udeležencev namreč poteka v procesu upovedovanja v mejah ohranjanja istega sporočila (Hanks 2010; Žele 2010: 118).

Najbolj običajen način opuščanja udeležencev so različne stopnje deagentnosti, kjer je vršilec zastopan z osebno glagolsko obliko, zato smo delovalnike v pomenski shemi vedno zapisovali z velikimi črkami, razen v primeru brezosebnih glagolskih pomenov, npr. *bliskati se*. V primerih, kjer obstaja splošno vedenje o nujni prisotnosti udeleženca, smo v LBS z velikimi črkami zapisovali le tiste, ki so bili v večini besedilnih realizacij tudi dejansko izraženi, npr.

cepiti *glagol*

4 izboljšati kakovost

4.3 o rastlinah

ČLOVEK cepi RASTLINO z drugo RASTLINO z namenom, da bi izboljšal kakovost sorte

- *Tu nimamo več kaj reševati, razen če na novo cepimo žlahtno sorto.*
- *Po gozdu je iskal divjake jablan in brušek ter jih cepil s cepiči, ki jih je dobil po pošti.*

sesti *glagol*

1.1 prenesti težo telesa na zadnjico

če ČLOVEK ali ŽIVAL sede na neko MESTO ali PREDMET, namenjen sedenju, se nanj namesti tako, da prenese težo telesa na zadnjico

- *Povabil jih je, naj sedejo, in jim ponudil pijačo.*
- *Sedel je na posteljo in mi pomignil, naj sedem zraven njega.*

Izpuščeni udeleženec je lahko tudi sestavni, inherentni del glagola, kot npr. *pop*er v *popoprati*. Navadno so taki udeleženci izraženi, če je izpostavljena njihova lastnost ali vrsta, zato jih obravnavamo kot obvezne elemente pomenske sheme:

poprati *glagol*

1 dodati poper

če ČLOVEK popra JED ali ŽIVILO, mu doda POPER

- *Kuhan brokoli dobro odcedimo, solimo in popramo ter primešamo sir.*
- *Zelenjavo solite in poprate na koncu.*
- *Ko zavre, juho solimo in popramo z zmletim belim poprom.*

V nekaterih primerih so pomensko logični udeleženci izraženi v besedilnem okolju na način, ki ne predstavlja vezljivostnega vzorca konkretnega glagola. Pri glagolu *sprati* spodaj je snov v obliki umazanije ali ostankov hrane, ki jo speremo s predmeta/oblačila ali živila sicer pomensko obvezni element, vendar običajno ni eksplicitno prisotna v stavčnem vzorcu obravnavanega glagola: kdo spere kaj (= predmet ipd.) s čim in ne: kdo spere kaj (= umazanijo ipd.) s česa (= predmeta ipd.) s čim, kot prikazujeta besedilna zgleda v (1). Lahko pa se, kot je ponazorjeno v (2), pomensko obvezni udeleženec izrazi v širšem besedilnem okolju oziroma kot del stavčnega vzorca katerega drugega glagola, ki se sicer pomensko navezuje na glagol, ki ga obravnavamo v iztočnici. Z zapisom tovrstnih udeležencev v pomenski shemi z malimi črkami smo v LBS ohranjali razliko med strukturno možnimi, a hkrati za posamezni pomen realno neizraženimi udeleženci.

sprati *glagol*

1 umiti z vodo; splakniti

če ČLOVEK spere PREDMET, OBLAČILO ali ŽIVILO s TEKOČO VODO ali VODNIM CURKOM, z njega odstrani umazanijo ali delce kake snovi

(1)

- *Oluščeno SEME, ki ga kupimo v trgovini, le speremo z VODO. (in ne: umazanijo z oluščene semena)*
- *Izdatno poškrbimo notranjost PEČICE in jo speremo s toplo VODO. (in ne: z nje speremo umazanijo)*

(2)

- *Po kopanju je KOPALKE treba sprati s sladko VODO, saj sol uničuje tkanino.*
- *ZVON sperejo z vodnim CURKOM, da odstranijo ostanke gline.*

Ob zgoraj izpostavljenih je potrebno omeniti še razmeroma pogoste situacije izražanja glagolskih pomenov, ki predvidevajo vsaj tri potencialne udeležence, ki jih realno zapolnjujejo različni semantični tipi, kjer je lahko kateri od udeležencev bodisi izražen ali neizražen (*stisniti*) ali pa sploh nikoli izražen (*sprati*) oz. pridobi drugačno udeležensko vlogo, npr. *stisniti* TEKOČINO iz SADJA oz. stisniti SADJE (ne nujno izraženo: v TEKOČINO) ali sprati SNOV (ne nujno izraženo: s PREDMETA) in sprati PREDMET.

Odsotne pomensko logične udeležence, ki se navadno zaradi sporočilne redundančnosti ne izražajo v obliki glagolskih določil ali dopolnil, smo v pomenski shemi obravnavali kot del pomenskega scenarija, kamor sodi npr. opis »s prsti« v pomenski shemi za obravnavani pomen glagola stisniti (gl. v nadaljevanju tudi glagol *olupiti*), čeprav nekateri redki besedilni zgledi to tudi eksplicitno izražajo: *Korenike nastrgamo in nato s prsti stisnemo sok.*

10.3.1.1.1.1 Pripomočki in orodja

Zlasti pogosto so v besedilnih realizacijah neizraženi pripomočki ali orodja. V takih primerih smo pripomočke navajali kot obvezne udeležence z velikimi črkami, če so se realizirali vsaj v nekaterih besedilnih zgledih. Na ravni semantičnih tipov smo se tu odločali bodisi za splošnejšo oznako (PRIPOMOČEK, NAPRAVA, ORODJE), kadar je šlo za več različnih orodij ali pripomočkov, ali pa smo navedli konkretne pripomočke, kadar je bila raba izkazano omejena:

česati *glagol*

1 urejati lase ali dlako

če ČLOVEK SEBI, drugemu ČLOVEKU ali ŽIVALI češe LASE ali DLAKO, jih ureja z GLAVNIKOM ali s KRTAČO

- *Las si ne češite s krtačo, temveč z glavnikom.*
- *Saj se vse ženske češejo, kadar gredo na stranišče.*

1.1 o vlaknih

če ČLOVEK češe ŽIVALSKA ali RASTLINSKA VLAKNA, jih gladi s pomočjo ORODJA ali STROJA

- *Manjšo količino runa je polagal na pokrov z žeblički in ga česal z leseno pripravo, mikalnikom, na katerem so bili tudi žeblički.*
- *Gostje, člani Turističnega društva iz Šalamencev, so česali perje in pripovedovali smešne prigode.*

barvati *glagol*

1 spreminjati ali dodajati barvo

1.1 o predmetih

če ČLOVEK barva POVRŠINO PREDMETA ali DELA TELESA, nanj s PRIPOMOČKOM nanaša BARVO v določenem odtenku

- *Večje površine barvamo s ploščatim čopičem, podrobnosti pa s tankim.*
- *Fantje plemena Surma v Etiopiji si barvajo telo, da bi tako zapeljali dekleta.*

V primerih, ko namesto orodja ali pripomočka (ali skupaj s pripomočkom) nastopa del človeškega telesa, zlasti roka, prsti, dlan ipd., smo ločili dve možni situaciji:

Del telesa ali pripomočka smo zapisali v obliki semantičnega tipa, kar mora biti razvidno iz vsaj nekaterih realizacij, npr.

masirati *glagol*

če ČLOVEK masira SEBE ali drugega ČLOVEKA, z ROKAMI ali s PRIPOMOČKOM pritiska in gnete telo, navadno zato, da bi zmanjšal bolečino ali nudil ugodje

- *Moškemu zelo godi, če ga partnerica v predigri malo masira, prekrije s poljubi, mu šepče kaj na uho.*
- *Podplate masirajte s preostalimi prsti z nežnimi gibi od prstov k peti.*
- *Za poživitev krvnega obtoka in boljše počutje si telo med kopanjem masiram z masažno ščetko.*

mečkati *glagol*

če ČLOVEK mečka PREDMET ali MATERIAL, ga z ROKAMI stiska tako, da postane zguban ali stisnjen

- *Iz žepa je potegnil še en robček, ki pa je bil že ves raztrgan, saj ga je prej ves čas mečkal s prsti.*
- *Veselo je pogledala skozi okno, Sandro pa je živčno mečkal neke papirje.*

Zapis pripomočka/dela telesa z malimi črkami napoveduje pomensko nujno vključenost dela telesa ali pripomočka, ki pa se v realnih zgledih običajno ne izraža, zato smo ta element obravnavali kot del pomenskega scenarija:

olupiti *glagol*

1 odstraniti vrhno plast

če ČLOVEK olupi SADJE, ZELENJAVO ali vrhnji DEL ŽIVILA, mu s prsti ali s PRIPOMOČKOM odstrani lupino ali vrhno plast

- *Limoni in pomaranči skrbno operete in z nožem olupite le zunanjo lupino.*
- *Gladko, bleščečo se in sluzavo lepljivo kožico zlahka olupimo s klobuka.*

bobnati *glagol*

2 igrati bobne

če ČLOVEK bobna, navadno v skupini glasbenikov igra na inštrument tako, da nanj udarja s pripomočkom ali rokami

- *Jeb Bishop igra pozavno in kitaro, alt saksofonist je Dave Rempis, Kent Kessler basira, bobna pa Tim Mulvenna.*
- *Cecil Durkin je bil narkič, pretepač z noži in jetniški posiljevalec, je pa tudi bobnal v nekaj dobrih jazz bendih.*

10.3.1.1.1.2 Stavčni udeleženci in okoliščine

Nekateri pomeni za svojo realizacijo potrebujejo udeležence ali okoliščine, ki se tipično realizirajo v obliki (odvisnega) stavka ali zloženega stavčnega člena, posledično pa je semantični tip, ki ga zastopajo, težje opredeljiv z eno samo besedo ali besedno zvezo. Kot izkazujejo primeri, ne gre vedno za udeležence obravnavanega glagola, pač pa za kontekstualne pomenske elemente, ki so za obravnavani glagol pomensko odločilni, skladijsko pa pripadajo kateremu drugemu glagolu, kot smo opozorili že v poglavju o (ne)izraženih udeležencih. V pomenski shemi smo take udeležence in okoliščine obravnavali kot pomensko nujne, zato smo jih v pomenski shemi navajali bodisi (1) z velikimi črkami, če jih je bilo mogoče razbrati iz besedilnih zgledov, v nasprotnem primeru pa (2) z malimi črkami kot del pomenskega scenarija:

(1)

bučati *glagol*

1 oddajati glasen zvok

1.1 o množici ljudi

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.*
- *Mall of America je bil podoben ogromnemu, elegantnemu, štirinadstropnemu blodnjaku za podgane s širokimi hodniki, ki so bučali od mrgolenja ljudi, ki so brezglavo tekali iz ene trgovine v drugo.*

razkačiti *glagol*

če ČLOVEK, njegova DEJANJA ali BESEDE razkači drugega ČLOVEKA, ga zelo močno razjezi

- *Lepo sem ji povedal, da zasluži – kljub bolniški – bolje kot jaz in to jo je do konca razkačilo.*
- *Neodgovoren prijatelj vas bo pošteno razkačil.*

izkašljati *glagol*

če ČLOVEK ali USTANOVA izkašlja DENAR |ali| 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

- *Ko je pred kratkim prepričal britansko vlado, da je izkašljala 280 milijonov funtov za bolj zdrave šolske malice, je bil verjetno še sam osupel nad sabo.*
- *5 milijonov odpravnine, ki jih je moral ob ločitvi izplačati prvi ženi Barbari, in 5 milijonov enkratne preživnine za otroka, ki jih je izkašljaj Angeli.*

(2)

bučati *glagol*

1 oddajati glasen zvok

1.4 o glasbi

če GLASBA buči, se jo sliši zelo na glas

- *Glasba, ki je bučala iz gostilne, se je razlegala po vsej vasi in kot jutranji zvon k maši narod klicala k popivanju, ki je sledilo.*
- *Za vozili z nameščenimi zvočniki, iz katerih v isti maniri kot pri govorih buči divja glasba, se vijejo procesije.*

improvizirati *glagol*

2 izdelati brez primernega orodja ali materiala

če ČLOVEK improvizira neko DEJAVNOST ali OPRAVILO, ga opravlja brez vnaprejšnje priprave, navadno zato, ker tako zahtevajo nepredvidljive okoliščine

- *Pri vsaki intervenciji pa gasilci improvizirajo, saj ni nobeno reševanje podobno drugemu.*
- *Pri tem bodo morali improvizirati, saj bo vroče poletje, tako da se bo mladina s Kapitolskega griča spuščala po suhem, na nekakšnih vozičkih.*

izkrivljati *glagol*

1 spreminjati podatke

če ČLOVEK izkrivlja PODATKE ali DEJSTVA, jih namenoma prireja ali navaja neresnično stanje, navadno zato, ker hoče kaj prikriti ali doseči korist zase

- *V sindikatu so ogorčeni, ker nekateri sindikalni liderji v javnosti izkrivljajo podatke o plačah carinikov, da bi na ta način prizadevanja za višje plače v svoji dejavnosti prikazali kot upravičena in legitimna.*

pozdraviti *glagol*

1 izreči ali izraziti dobrodošlico, pozdrav

1.4 prijetno presenetiti

če ČLOVEKA ob prihodu v nek kraj pozdravi PREDMET, mu pogled nanj daje občutek, da je dobrodošel

- *Tik za mejo pri Lazaretu pozdravi popotnike tabla z napisom - Dobrodošli v deželi refoška!*
- *Po vsej Mehiki so papeža pozdravili ogromni panoji, na katerih je pisalo Mehika, večno verna!*

Pogosto je kateri od udeležencev glagola, ki ga obravnavamo v iztočnici, v stavku izražen s kazalnimi zaimki, ki referirajo na vsebine, izražene v predhodnem besedilu. V takih primerih smo skušali s semantičnim tipom zajeti pomenske lastnosti

vsebin na enak način kot v primeru udeležencev, ki se realizirajo znotraj glagolove stavčne zgradbe:

razkačiti *glagol*

1 močno razjeziti

če ČLOVEK, njegova DEJANJA ali BESEDE razkači drugega ČLOVEKA, ga zelo močno razjezi

- *To me je pa do kraja razkačilo in sem ji rekla, da mu v postelji noben šarm ne bo pomagal, če ne more nič spraviti skupaj.*

Če udeleženca v smislu ustreznega zastopnika semantičnega tipa ni bilo mogoče določiti, ker je besedilna raba izkazovala pomensko razpršene entitete, smo semantični tip zapisali v obliki poljubnostnega ali nedoločnega zaimka KAJ, NEKAJ (*čemeti*) oz. smo iz odvisnega stavka oblikovali »udeleženca« v obliki bolj konkretnega semantičnega tipa (*rotiti*), npr.

čemeti *glagol*

2 biti prisotno

če rečemo, da KAJ čemi v ČEM ali v ČLOVEKU, se tam prikrito nahaja, sčasoma pa lahko postane očitno, zlasti če je negativno

- *Film Obsedno stanje svari pred latentnim fašizmom, ki kot nekaj parazitski alien čemi v nedrjih demokracije.*
- *Verjetno v meni čemi preveč predsodkov, ki zapirajo kaj vem kakšne čakre, da se moja duša ne more osvoboditi.*

rotiti *glagol*

prositi

če ČLOVEK roti drugega ČLOVEKA, ga prosi, naj mu pomaga iz STISKE ali naj stori nekaj, kar je zanj zelo pomembno

- *»Prosim,« me je rotil, »daj mi priložnost, da spoznam svojega sina ...«*
- *Že kot mlado dekle so jo nekateri rotili, češ ne hodi v samostan, imaš preveč energije, strasti in ognja v očeh.*
- *Vse svoje prijatelje je rotil, naj nikomur ne izdajo njegove skrivnosti.*

Nasprotno pa so pogosto izražene pomensko obvezne krajevne, časovne in druge okoliščine konkretnega pomena, kot je razvidno iz spodnjih primerov:

čemeti *glagol*

1 biti kje; nahajati se

1.1 o človeku

če ČLOVEK čemi v nekem PROSTORU ali ob neki DEJAVNOSTI, se tam nahaja ali kaj počne, navadno že dlje časa in brez vidnega zanimanja

- *Živo sem si ga predstavljala, kako bo cele dneve sam zamišljeno čemel v svoji sobi in žalostno strmel skozi okno.*
- *Potem je zagledal, da v celici na ležišču v kotu čemi še nekdo, ves zgrbljen in v starih, porvaljanih cunjab.*

zaplavati *glagol*

1 začeti se premikati po vodi

ko ČLOVEK zaplava, se začne premikati po površini vode v določeno SMER

- »Dobro razmisli,« mu je odvrnila ribica in zaplaval po vodi.
- Nekajkrat sem globoko vdihnil iz jeklenke in ponovno na vso moč zaplaval proti toku.

V primeru, da take okoliščine niso bile določljive s konkretnimi semantičnimi tipi, smo uporabili ustrezen splošnejši zaimek:

hiteti *glagol*

1 hitro delati

ČLOVEK hiti, ko se mu KAM mudi ali ko mora KAM hitro priti

- Prvi mrak je ležal nad vasjo, ko sem hitela proti domu.
- Opazili smo tudi žensko v krznenem plašču, ki je objokana hitela proti hotelu.

10.3.1.1.1.3 Pomensko neobvezni udeleženci

Pomensko relevantnost izkazanih udeležencev ni vedno lahko določiti, še posebno, kadar se pomenski elementi izražajo v skladijsko neobveznih vlogah, v katerih se pogosto znajdejo prislovna določila, zlasti tista, ki določajo krajevne in časovne okoliščine konkretnega glagolskega pomena. Gre za situacije, kjer so udeleženci ali okoliščine v dejanski rabi tipično izraženi, vendar se za določen pomen ne zdijo odločilni. V takih primerih jih v pomenski shemi nismo navajali kot obvezne elemente in jih nismo zapisovali kot semantične tipe, pač pa smo na skladijskem nivoju registrirali stavčne vzorce, ki take okoliščine izražajo, in jih beležili pod opredelitvijo: LOKACIJA, SMER, ČAS, NAČIN in VZROK, pod strukturo rzb GBZ (prislov + glagol) ali v skladijskih zvezah. Tipičen primer je glagol *diplomirati*:

diplomirati *glagol*

pridobiti diplomu

če ČLOVEK diplomira, zaključi ŠTUDIJO in napiše diplomu

diplomirati v čem

diplomirati na čem

diplomirati v čem na čem

diplomirati v [Ljubljani]

diplomirati na [fakulteti, akademiji, univerzi]

- Po zaključenih študijskih letih, je leta 1971 diplomiral na Akademiji za likovno umetnost v Ljubljani.
- Leto po izidu pesniške zbirke je diplomiral na ljubljanski pravni fakulteti.

rzb GBZ

[lani, letos] diplomirati

diplomirati leta [x]

Pogosto izražene okoliščine in udeležence, ki za dani pomen niso pomensko odločujoči, smo v pomenski shemi navajali za izrazom »navadno«, vendar jih v tem primeru nismo zapisovali z velikimi črkami. Konkretna realizacije teh udeležencev so razvidne iz stavčnih vzorcev na skladenjskem nivoju in iz korpusnih zgledov:

iskati *glagol*

2 skušati pridobiti

2.3 z računalnikom

če ČLOVEK ali elektronska NAPRAVA išče PODATKE, ki so na voljo v elektronski obliki, jih pridobiva iz obsežnih podatkovnih baz navadno po določenih merilih

iskati po čem

- *Z oglaševanjem na iskalnikih si oglaševalec izbere besede in besedne zveze ter z njimi cilja na uporabnike, ki iščejo po teh izrazih.*
- *Išči po: naslovu, igralcih, režiserju, zvrsti, letu, državi ...*
- *Na strani za iskanje ljudi lahko uporabnike iščete po imenu ali po telefonski številki.*

10.3.1.1.1.4 Inherentni udeleženci

Kot smo že omenjali, se nekateri udeleženci v pomenski shemi pojavljajo kot inherentni deli glagola, ki je naša iztočnica, in so lahko z njim tudi morfemsko prekrivni, npr. *soliti* – sol; *pomokati* – moka. Te elemente je navadno mogoče prepoznati kot pomensko obvezne za konkretni pomen glagola, v dejanskih realizacijah pa se ne izražajo vedno eksplicitno. Kot v drugih podobnih primerih smo tudi tu glede zapisa postopali v skladu z izraženostjo v dejanskih realizacijah, in sicer:

- (a) Če v realizacijah t. i. inherentni udeleženci običajno niso izraženi, jih tudi v pomenski shemi nismo navajali kot semantične tipe, npr.

previti *glagol*

2 zamenjati plenico

če ČLOVEK previje OTROKA, mu zamenja plenico

- *Oče zgodaj vstane, previje dojenčka in je pol ure z otrokom, medtem ko mamica še spi.*
- *V prvih mesecih naj bi otroka previli pred vsakim hranjenjem, tudi nočnim.*

dihati *glagol*

1 vdihavati in izdihavati zrak

če ČLOVEK ali ŽIVAL diha, vdihava in izdihava zrak

- *Otrok se zbudi in težko diha.*
- *Mačka hitreje prede, kot diha, torej predenje ni sinhronizirano z dihanjem.*

zamižati *glagol*

če ČLOVEK zamiži, zapre oči

- *Za trenutek zamižite in postanite pozorni na svoje dihanje.*
- *Elizabeth je trdno zamižala in si z rokama zakrila obraz.*

(b) Inherentni udeleženci so razvidni tudi v dejanskih realizacijah (čeprav ne vedno), zato smo jih v pomenski shemi navajali kot semantične tipe, npr.

previti *glagol*

1 obvezati

če ČLOVEK previje drugemu ČLOVEKU ali ŽIVALI DEL poškodovanega ali ranjenega TELESA, mu ga obveže s POVOJEM

- *Po končanem odvzemu medicinska sestra sname prevezo, vbodno mesto pa previje s sterilnim povojem.*
- *Okolico ugriza očistite, kožo pa dezinficirajte in jo prekrijte s sterilno gazo ter po potrebi previjte s povojem.*
- *Imel sem srečo, da mi je Harry takoj pritekel na pomoč in mi previl rane.*

pomokati *glagol*

potresti z moko

če ČLOVEK pomoka ŽIVILO, ga potrese z MOKO

- *Nato jih narežemo na koščke, ki jih pomokamo s preostalo moko, in potisnemo v z maslom premazane modelčke.*
- *Pomokamo z ajdovo moko in zalijemo z juho.*
- *Testo pomokamo in pokrijemo ter pustimo na toplem, da vzbaja.*

10.3.1.1.2 Spremembe v udeleženski zgradbi pomena (uporaba skladijskega |ali|)

Osnovni nalogi pri zapisovanju pomenske sheme, ki predvideva beleženje udeleženske zgradbe pomena oz. njegove tipične skladijske realizacije sta (1) ugotovitev najbolj tipične realizacije, saj ta ni vedno neposredno razvidna iz kolokacijske slike pomena, in (2) dejstvo, da se določen pomen realizira v več enakovredno relevantnih skladijskih realizacijah. S tem je povezan tudi problem oblikovanja stavčne razlage, ki naj bi iz pomenske sheme izhajala, saj, kot pravi Rundell (2006: 332), ena stavčna razlaga pogosto ne more izraziti vseh slovničnih in stavčnih vzorcev, v katerih se pomen besede lahko realizira.

V času snovanja LBS in postavljanja temeljev stavčnih razlag v pomenskih shemah je analiza gradiva pokazala, da so problematične predvsem različne skladijske upovedovalne možnosti, ki izhajajo iz (a) skladijskih zgradb, ki nastajajo ob rabi glagolskega morfema *se* oz. *si*, (b) različne razporeditve in števila udeležencev, ki nimajo neposrednega vpliva na pomen – gre torej za problem, ki je povezan z neizražanjem

udeležencev zaradi njihove logične ali sobesedilne samoumevnosti (*stisniti pest – stisniti roko v pest*) ter (c) možnost različnih predložnih realizacij (npr. *dvomiti o čem/v kaj* ipd.).

Čeprav so spremembe v udeleženski zgradbi pomena (tako v številu udeležencev kot v oblikoskladenjski realizaciji in v zastopanosti semantičnih tipov) navadno znanilec pomenskih sprememb, so izkazani tudi primeri zgolj skladijsko pogojenih različnih realizacij, ko spremembe v udeleženski zgradbi ne vplivajo bistveno na pomen glagola, ki ga obravnavamo. V ta namen smo v LBS tako z vidika avtomatske prepoznavnosti dveh ali več prototipičnih vzorcev v skladijski zgradbi pomena kot z vidika ločevanja dveh ali več udeleženskih zgradb v zapisu pomenske sheme predvideli t. i. skladijski |ali| znotraj navpičnic, ki se razlikuje od pomenskega »ali«, s katerim lahko uvajamo različne pomenske odtenke znotraj pomenske sheme. Strategije pri obravnavi nekaterih tipičnih stavčnih alternacij podrobneje opišemo v nadaljevanju.

10.3.1.1.1.3 Strukture s prostim glagolskim morfemom *se* oz. *si*

Največ različnih možnosti alterniranja pomensko obveznih udeležencev predvidevajo udeleženske zgradbe za pomene glagolov, ki vključujejo prosti morfem *se* ali *si*. V nadaljevanju ločimo pet osnovnih tipov: (1) glagole, ki brez morfema *se/si* ne nastopajo, zato ga zapisujemo že na ravni iztočnice (*navzeti se, zaljubiti se*), (2) »pravo« povratnost (*osvežiti si; česati se, masirati se*), (3) t. i. »nepravo« povratnost (*cepiti se, smučati se*), (4) vzajemnost (*kregati se, obiskovati se*) in (5) splošnega vršilca (*išče se*). Strukture s prostima morfemoma ločujemo od trpne rabe s *se*, ki jo sicer prav tako obravnavamo kot samostojno udeležensko zgradbo v pomenski shemi, ločeni s skladijskim |ali|.

10.3.1.1.1.3.1 Povratni glagoli

Zapis pomenske sheme z vidika skladijskih alternacij pri glagolih, ki se brez prostega morfema ne pojavljajo, sama na sebi ni problematična. Prosti morfem *se* lahko izraža v sinonimnem indikatorju, npr. *daniti se* – delati se dan; *svitati se*, pri katerem od pomenov, sicer pa vedno na ravni stavčnih vzorcev:

navzeti se *glagol*

1 pridobiti lastnost

če se neka SNOV, zlasti hrana ali obleka, navzame VLAGE, VONJA ali OKUSA, ga sprejme vase, tako da postane njen sestavni del ali se v njej zazna

kaj se navzame česa

kaj se navzame česa iz česa

(ČAS) navzeti se med čim

2 prevzeti miselnost, navade

če se ČLOVEK navzame IDEJ, NAVAD ali RAZPOLOŽENJA od drugega ČLOVEKA ali iz OKOLJA, jih izraža kot del lastnega prepričanja ali značaja

kdo se navzame česa

kdo se navzame česa od koga

kdo se navzame česa pri kom

kdo se navzame česa po kom

10.3.1.1.3.2 Prava povratnost

Situacija, ko se udeleženec izraža kot neposredni predmet, npr. *česati koga*, hkrati pa se dejanje vrača nazaj na osebek, npr. *česati se* – govorimo o pravi povratnosti¹¹³ – se pri nekaterih glagolih kaže kot jezikovnosistemska možnost (*ustreliti* – *ustreliti se*, *nazivati* – *nazivati se* itd.). Ker gre za skladenjsko alternacijo, pri kateri do (ključnih) pomenskih sprememb ne prihaja, smo predvideli naslednje možnosti zapisa pomenske sheme:

Povratne udeleženske zgradbe s *se* smo beležili samo na ravni stavčnih vzorcev, v pomenski shemi pa povratnosti nismo posebej izpostavljali, npr.

ustreliti *glagol*

1 sprožiti strelna orožje

1.1 zadeti s strelnim orožjem

če ČLOVEK ustrelí drugega ČLOVEKA, ga zadene z nabojem iz OROŽJA, navadno z namenom da bi ga poškodoval ali ubil

kdo ustrelí koga/kaj

kdo se ustrelí

kdo ustrelí koga/kaj v kaj

kdo se ustrelí v kaj

Številčni podatek v orodju SkE pod direktivo Constructions: »O_tretja_oseba« nam lahko pomaga pri odločitvi o tem, kateri vzorec je za dani glagol bolj tipičen. Pri večpomenskih glagolih je treba biti pri interpretaciji tega podatka pozoren, saj se izkazana preferenca nanaša na obravnavano glagolsko lemo v celoti in ne na posamezni pomen glagola.

Če smo želeli opozoriti na dve (redkeje več) skladenjskih možnosti že v zapisu pomenske sheme, smo jih ločil s t. i. skladenjskim |ali|. Druga možnost je bila, da smo udeleženski zgradbi med seboj povezali. Spodnji primeri izkazujejo obe:

¹¹³ Samo v tem primeru gre pri *se* oz. *si* dejansko za povratnosvojilni zaimek.

briti *glagol*

1 odstranjevati dlake

1.1 o moških brkih in bradi

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

briti se

kdo se brije

kdo si brije kaj

briti si kaj

briti se s čim

1.2 o laseh in dlakah

če si ČLOVEK brije DLAKE |ali| če ČLOVEK brije DLAKE drugemu ČLOVEKU, jih odstranjuje s PRIPOMOČKOM

če ČLOVEK SEBI ali drugemu ČLOVEKU brije DLAKE, jih odstranjuje s PRIPOMOČKOM

kdo si brije kaj

briti si kaj

briti se po čem

kdo brije koga

odvaditi *glagol*

1 prisiliti k prenehanju

če se ČLOVEK z zavestnim trudom odvadi |ali| če ČLOVEK s trudom odvadi drugega ČLOVEKA ali ŽIVAL SLABE ali ŠKODLJIVE NAVADE, doseže, da tega ne počne več

če ČLOVEK z zavestnim trudom odvadi SEBE, drugega ČLOVEKA ali ŽIVAL SLABE ali ŠKODLJIVE NAVADE, doseže, da tega ne počne več

kdo se odvadi česa

kdo se odvadi od česa

kdo odvadi koga česa ...

10.3.1.1.3.3 Neprava povratnost

Pri izražanju t. i. neprave povratnosti, tj. v primerih, ko prosti glagolski morfem *se/si* ne pomeni nujno vračanja dejanja na osebek, npr. *smučati* – *smučati se*, *poslabšati* – *poslabšati se*, *ponavljati* – *ponavljati se*, je prisotnost morfema navadno pogojena z upoštevanjem najširšega stavčnega vzorca pomena, v danem primeru torej glede na to, ali je prisoten osebek, in še, ali njegova prisotnost kakorkoli vpliva na pomen glagola. V takih primerih smo načeloma navajali obe različici udeleženskih zgradb, ločenih z |ali| (*cepiti*), oz. smo se v pomenski shemi odločili za prikaz tiste udeleženske zgradbe, ki se v rabi pojavlja pogostejše (*smučati*), stavčni vzorec s prostim morfemom pa smo navedli na skladenjskem nivoju:

cepiti *glagol*

1 prejeti ali vnesti cepivo

če se ČLOVEK cepi, prejme CEPIVO proti določeni BOLEZNI |ali| če ZDRAVNIK cepi ČLOVEKA, mu v telo vnese CEPIVO proti določeni BOLEZNI

kdo se cepi
kdo se cepi proti čemu
kdo cepi koga/kaj

smučati *glagol*

1 voziti se s smučmi

1.1 na snegu

če ČLOVEK smuča, se s smučmi večkrat spusti navzdol po zasneženem in za to urejenem OBMOČJU

kdo smuča
kdo se smuča

10.3.1.1.3.4 Vzajemnost

Kadar udeleženske zgradbe s prostim morfemom *se/si* izražajo vzajemnost dogajanja, smo v pomenski shemi navedli obe skladenjski realizaciji, tj. tako z enim kot z dvema udeležencema, ter ju ločili z |ali|:

dogovarjati se *glagol*

1 izmenjavati mnenja

če se ČLOVEK dogovarja z drugim ČLOVEKOM |ali| če se LJUDJE ali USTANOVE dogovarjajo med sabo, si izmenjujejo mnenja

kdo<r>navadno dvojina ali množina</r> se dogovarja
kdo se dogovarja s kom

kregati se *glagol*

1 jezno izražati nestrinjanje

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

kdo<r>navadno dvojina ali množina</r> se krega zaradi koga/česa
kdo se krega s kom glede česa
kdo se krega s kom o čem

10.3.1.1.3.5 Splošni vršilec in trpna raba

Prosti morfem služi tudi za umikanje vršilca, ki je sicer nujni del pomenske celote glagola, in sicer tako, da ga pomensko posploši in skladenjsko opusti. V takih

primerih gre navadno za izrabljanje stilističnih možnosti, zato smo bili pozorni, ali taka posplošitev še ustreza okvirom konkretne pomenske sheme ali gre za samostojni pomen. V nasprotnem primeru, če torej umik vršilca ni izkazoval relevantnih pomenskih specifik, pomenske sheme nismo obremenjevali z razdelitvijo na več delov, razen če se ta možnost za dani glagolski pomen ni izkazala kot tipična (npr. *barvati*), ampak smo to možnost navedli v stavčnih vzorcih, npr.

arhivirati *glagol*

1 shraniti

1.1 v ustanovi

če ČLOVEK arhivira DOKUMENTE, pisno, slikovno ali zvočno GRADIVO, jih shrani v prostor ali mapo ter uredi tako, da lahko išče po njih

kdo arhivira
kdo arhivira kaj
kaj se arhivira

domnevat *glagol*

1 vnaprej sklepati; predvidevati

če ČLOVEK domneva, da se bo KAJ zgodilo ali da je KAJ resnično, o tem vnaprej sklepa ali predvideva na podlagi okoliščin ali izkušenj

kaj se domneva
kdo domneva kaj

barvati *glagol*

2 pridobivati barvo; obarvati se

2.2 o naravnih pojavih; odsevani

če se VODA ali NEBO barvata |ali| če SONCE ali svetlobni POJAV barva VODO ali NEBO, odseva v njiju rdeče barvne ODTENKE

kaj se barva
kaj se barva v kaj
kaj barva kaj
kaj barva kaj v kaj

10.3.1.1.4 Dodajanje oz. odvzemanje udeležencev (stisniti pest – stisniti roko v pest)

V nekaterih primerih različne skladenjske realizacije stavčnega vzorca zahtevajo prerazporeditev udeležencev, kar se na površinski ravni kaže tudi v različnih sklon-skih realizacijah (*roditi*) oz. se število udeležencev poveča, s tem pa nastane tudi potreba po izpostavitvi dodatnih semantičnih tipov v pomenski shemi. V primerih, ki jih navajamo v nadaljevanju, gre za glagolske pomene, ki predvidevajo vsaj tri udeležence, od katerih je lahko kateri neizražen ali pa nastopa v dveh udeleženskih

vlogah, ki se vsaj delno prekrivata (Žele 2011: 51), npr. *dotočiti vino v kozarec* – *dotočiti vino* – *dotočiti kozarec*; *sprati madež z obleke* – *sprati madež* – *sprati obleko*. V primerih, kjer prihaja v besedilnih realizacijah do različne razporeditve ali drugačnega števila udeležencev, hkrati pa se pomen glagola ohranja, smo v pomenski shemi posamezne skladijske realizacije ločevali bodisi z |ali| (*cveteti*) ali pa smo jih združili v eno pomensko shemo (*iskati*). Odločitev smo pogojevali s stopnjo berljivosti in razumljivosti pomenske sheme:

cveteti <i>glagol</i> 2 prinašati dobiček; uspevati če DEJAVNOST cveti ali če DEJAVNOST cveti ČLOVEKU, ORGANIZACIJI ali DRŽAVI, se dobro razvija in prinaša dobiček kaj cveti kaj cveti komu/čemu
roditi <i>glagol</i> 1 spraviti na svet 1.1 o materi in otroku če se OTROK rodi ŽENSKI ali če ŽENSKA rodi OTROKA, ga potisne iz maternice kdo rodi koga kdo rodi kdo se rodi
iskati <i>glagol</i> 2 skušati pridobiti 2.2 najboljšo možnost če ČLOVEK ali USTANOVA ZASE ali za drugega ČLOVEKA išče ČLOVEKA, PREDMET ali DEJAVNOST, skuša izbrati možnost, ki najbolj ustreza njegovim potrebam ali zahtevam kdo išče koga/kaj kdo išče koga/kaj komu/čemu kdo išče koga/kaj za koga/kaj

Analiza gradiva je pokazala, da prihaja do spreminjanja udeleženske zgradbe zlasti v naslednjih primerih:

človek – sredstvo

napihnniti *glagol*

1 napolniti z zrakom

če ČLOVEK napihne PREDMET z ZRAKOM |ali| če ZRAK napihne PREDMET, ga z njim napolni tako, da se raztegne in spremeni obliko

kdo/kaj napihne kaj
kdo napihne kaj s čim
kaj se napihne

fotokopirati *glagol*

narediti papirnato kopijo

če ČLOVEK ali NAPRAVA |ali| če ČLOVEK z NAPRAVO fotokopira LIST PAPIRJA ali DOKUMENT, naredi papirnato kopijo

kaj fotokopira
kdo fotokopira kaj
kdo si fotokopira kaj
kdo fotokopira kaj s čim

strmoglaviti *glagol*

1 zrušiti se na zemljo

če LETALO strmoglavi |ali| če ČLOVEK strmoglavi z LETALOM, se z velike višine zruši na zemljo, pri čemer se uniči ali močno poškoduje

kaj strmoglavi
kdo strmoglavi s čim

človek – del telesa

masirati *glagol*

1 obdelovati telo

če ČLOVEK masira SEBE ali drugega ČLOVEKA |ali| če ČLOVEK masira drugemu ČLOVEKU ~~DEL~~ telesa, z ROKAMI ali s PRIPOMOČKOM pritiska in gnete TELO, zlasti mišice, navadno zato, da bi zmanjšal bolečino ali nudil ugodje

masirati se
kdo masira koga
kdo masira komu kaj

stisniti *glagol*

1 premakniti del telesa

1.1 oblikovati pest

če ČLOVEK stisne PEST |ali| če se ROKA stisne v PEST |ali| če ČLOVEK stisne ROKO v PEST, skrči prste na roki tako, da z njimi oblikuje pest

3 potisniti, uporabiti fizični pritisk 3.6 poškodovati
če NAPRAVA stisne ČLOVEKA ali stisne ČLOVEKU DEL TELESA, ga močno poškoduje
kaj stisne koga kaj stisne komu kaj

predmet – vsebina predmeta/prostor

zasejati <i>glagol</i> 1 položiti semena v zemljo
če ČLOVEK zaseje RASTLINE ali zaseje določeno POVRŠINO z RASTLINAMI, položi semena v zemljo, da bi v njej zrasle rastline
kdo zaseje kaj kdo zaseje kaj na čem kdo zaseje kaj s čim

Razdvoumljanje udeležencev z različnimi udeleženskimi vlogami v takih primerih omogočijo kolokacije: zasejati [travo, lan, pšenico] : zasejati [gredico, njivo].

10.3.1.1.1.5 Možnost različnih predložnih realizacij

Različne predložne realizacije, kjer se pomenske vloge udeležencev ne spreminjajo, smo beležili zgolj na ravni stavčnih vzorcev, npr.

prekipevati <i>glagol</i> 1 biti poln česa 1.1 o pozitivnih občutkih
če ČLOVEK ali MESTO prekipeva od pozitivnih OBČUTIJ ali ENERGIJE, ima tega v izobilju
kaj/kdo prekipeva od česa kaj prekipeva v čem
mučiti <i>glagol</i> 1 imeti neprijetne občutke 1.1 zaradi bolezni
če ČLOVEKA muči BOLEZEN ali BOLEČINA, ga boli del telesa ali ima zdravstvene težave
koga muči kaj kdo se muči s čim

V tem okviru smo obravnavali tudi možnost predložne ali brezpredložne realizacije, npr. *čakati vlak* – *čakati na vlak*; *paziti otroke* – *paziti na otroke* ipd.

Več fleksibilnosti pri oblikovanju pomenskih shem in posledično stavčnih razlag je bilo potrebne v primerih, kjer znotraj pomenske sheme različni glagoli zahtevajo različno predložno in lahko tudi sklonsko realizacijo udeležencev, npr.

blanširati <i>glagol</i> preliti z vrelo vodo; popariti
če ČLOVEK blanšira SADJE ali ZELENJAVO, ga za kratek čas potopi v ali prelije z VRELO VODO
vreti <i>glagol</i> 3 o ljudeh, mislih, dogajanju 3.1 množično prihajati
če LJUDJE vrejo iz PROSTORA, množično prihajajo iz njega ali če ljudje vrejo v PROSTOR, se tja množično odpravljajo

10.3.1.1.2 Samostalnik

Glede na zapis pomenske sheme ločimo pri samostalnikih dva podtipa samostalniških pomenov, in sicer (a) t. i. vezljive samostalnike oz. tiste samostalniške pomene, kjer je mogoče znotraj stavčnega vzorca prepoznati udeležence po analogiji z glagolskimi udeleženci, kamor sodijo zlasti pomeni samostalnikov, ki se uresničujejo skupaj s pomensko izpraznjenimi glagoli, ter (b) vezljive samostalnike oz. samostalniške pomene, ki znotraj besednozveznega okolja izkazujejo pomensko odločilne elemente, kot prikazuje Preglednica 13:

Preglednica 13: *Vrste samostalniške vezljivosti*

besednozvezna vezljivost	(iz)glagolski samostalniki	utaja ČESA, zarota zoper KOGA /proti KOMU, apel KOMU za KAJ odhod v KAJ/KAM, porod KJE
	neglagolski samostalniki	aluzija na KAJ, čas ČESA
	količinski samostalniki	mlakuža ČESA, skodelica ČESA
stavčna vezljivost	samostalnik + pomensko izpraznjen glagol	KAJ daje noto ČEMU, KAJ ima vonj po ČEM

V nadaljevanju se bomo posvetili posameznim tipom samostalniških pomenov, kjer predvidevamo zapis pomenske sheme, znotraj tega pa vrstam samostalniške

vezljivosti ter udeležencem in strategijam pri zapisovanju samostalniških pomenskih shem.

10.3.1.1.2.1 Besednozvezna vezljivost samostalniških pomenov

Kot prikazuje Preglednica 13, sodijo v prvo skupino samostalniki oz. samostalniški pomeni, ki predvidevajo pomenskorelevantna določila ali dopolnila. Z vidika obliko-skladenjskega izražanja vezljivosti ločimo brezpredložna in predložna samostalniška določila, npr. *adaptacija ČESA*, *alergija NA KAJ*, samostalniška vezljivost pa je lahko tudi veččlenska, npr. *aplikacija ČESA VNA KAJ*, *asimilacija KOGA V KAJ*. Ob izpostavljenih tipih je potrebno omeniti še določila stavčnega tipa, ki smo jih registrirali pri skladenjskih strukturah, npr. *ambicija, da ...*; *apel, naj ...*; *dejstvo, da ...* ipd.

10.3.1.1.2.1.1 (Iz)glagolski samostalniki

(Iz)glagolski samostalniki, npr. *argument*, *apel*, in glagolniki kot *aplikacija*, *asimilacija*, *adaptacija* ipd., ki v marsičem prevzemajo tudi glagolsko skladenjsko obnašanje, izražajo predvsem pomen dejanja oz. dogajanja in stanja: *ohladev česa – kaj se ohladi*, *ohlajanje česa – kaj se ohlaja*, *gradnja česa – graditi kaj*. Izpostavljanje le določenih glagolniških pomenov je pri tem pomembno, saj obstajajo tudi bolj »samostalniški« pomeni glagolnikov (zlasti kot rezultat glagolskega dejanja), kjer pomensko relevantnih določil ne pričakujemo, kot npr. spodaj pri podpomenih samostalnikov *gradnja*, *bučanje* in *smrčanje*:

gradnja *samostalnik*

1 izdelovanje objektov in predmetov

gradnja STAVB, CEST in drugih OBJEKTOV ter večjih PREDMETOV je njihovo izdelovanje iz različnih materialov

gradnja [avtoceste, prizidka, kanalizacije, hidroelektrarne, ladje, plovila, makete]

1.1 zgradba; stavba

bučanje *samostalnik*

1 oddajanje zvoka

bučanje NARAVNEGA POJAVA je oddajanje glasnega bobnečega zvoka ob njegovem delovanju ali premikanju

bučanje [valov, morja, vode, slapov, brzic, reke]

bučanje [vetra]

bučanje [stroja, orgel]

1.1 bobneč zvok

[zamolklo, mogočno, glasno] bučanje

smrčanje *samostalnik*

1 hropeče dihanje med spanjem

smrčanje je občasno hropeče dihanje ČLOVEKA med spanjem

1.1 zvok

[slišati, zaslišati] smrčanje

smrčanje [se sliši, moti]

10.3.1.1.2.2 Neglagolski in količinski samostalniki

Sintagmatsko izhodišče pri obravnavi besedne večpomenskosti ter formalizacija skladenjskih vzorcev za namene računalniškega procesiranja sta pri zapisovanju pomenskih shem narekovala tudi obravnavo: (a) samostalniških pomenov z desnimi določili, ki so v nekaterih primerih pretvorljivi v levi prilastek, npr. *čas počitnic* – *počitniški čas*, vendar ne vedno, npr. *čas cvetenja*, *čas nastanka*, *čas osamosvajanja* ipd., ter (b) samostalniških pomenov, ki določajo količino samostalniškega jedra. V takih primerih smo udeležence v pomenski shemi zapisovali le, če smo z njo dopolnili informacijo v pomenskem indikatorju, sicer so določila razvidna iz skladenjske strukture Kol-SBZ0 sbz2 (količinski samostalnik v imenovalniku in samostalnik v rodilniku) in pripadajočih kolokacij:

spomin *samostalnik*

1 pomnjenje česa iz preteklosti

spomin na DOGODKE in LJUDI je njihova podoba iz preteklosti, ki se je ohranila v človekovi zavesti

spomin na [otročstvo, mladost, preteklost, dogodek, doživetje]

mreža *samostalnik*

6 povezan sistem

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

mreža [šol, poslovalnic, trgovin, bankomatov]

čas *samostalnik*

1.5 glede na značilnosti obdobja

čas nekega DOGAJANJA, DOGODKOV je obdobje, ki ga zaznamujejo značilnosti tega dogajanja ali dogodkov

čas [osamosvajanja, ministrovanja, županovanja]

čas [parjenja, cvetenja]

čas [dopustov, počitnic]

poplava *samostalnik*

1 velika količina

Kol-SBZ0 sbz2

poplava [informacij, besed, ponudbe]

poplava [hormonov]

V sklopu besednozvezne samostalniške vezljivosti smo obravnavali tudi samostalnike s pomensko nujnimi družljivimi elementi, kot npr. *odhod kam – odhajati/oditi kam, porod doma – roditi doma/na domu, pogled naprej – gledati naprej/v prihodnost* ipd. Taki primeri nakazujejo, da bi bilo v sistemu slovarske baze tako z vidika skladijskih analiz kot z vidika strojnega procesiranja pomena smiselno tudi pri vezljivih samostalniških pomenih (taka možnost je za zdaj predvidena samo pri glagolih v elementu <zorec>) uvesti združevanje skladijskih struktur pod pomenske skupine, ki označujejo različne načine izražanja okoliščin, zlasti predložnih in brezpredložnih prislovnih določil, npr.

SMER		LOKACIJA	
sbz0 v sbz4	odhod v Ljubljano	sbz0 na sbz5	porod na domu
sbz0 do sbz2	pot do vrha	sbz0 v sbz5	korupcija v Sloveniji
sbz0 rbz	odhod domov	sbz0 rbz	porod doma

10.3.1.1.2.2.1 Razmerje med zapisom pomenske sheme in razlage

Pri zapisovanju besednozveznih udeležencev kot semantičnih tipov v primeru neglagolskih samostalnikov tipa *mreža, dres, aerodinamika* ipd. in glagolskih pomenov samostalnikov, npr. *komentar, nakup* ipd., smo se pri zapisu pomenske sheme v odnosu do pomenske razlage (v elementu <definicija>) odločali tako, da smo v pomenski shemi sledili primarno skladijskemu zapisu pomena, ki neposredno izraža njegovo vezljivost, pri zapisu razlage, namenjene uporabniku, pa smo se odločali za čim bolj razumljivo ubeseditev, ki pa ne vključuje nujno tudi tipične skladijske realizacije, kot prikazuje Preglednica 14:

Preglednica 14: *Razlike v zapisu pomenske sheme in razlage*

Pomenska shema	Razlaga
mreža USTANOV, POVEZAV ali NAPRAV je javni sistem ...	mreža je sistem ustanov, povezav ali naprav, ki s svojim obstojem ali delovanjem zagotavljajo storitve na določenem območju
spomin na DOGODKE in LJUDI je njihova podoba ...	spomin je pomnjenje dogodkov in ljudi iz preteklosti, ki so se ohranili v človekovi zavesti
aerodinamika PREDMETA je njegova učinkovitost ...	aerodinamika je učinkovitost predmeta, zlasti vozila, pri gibanju skozi zrak

Taka odločitev sledi splošnim načelom, ki veljajo za oblikovanje LBS oz. njeno nivojsko zgradbo: ker so pomenske sheme v prvi vrsti izhodišče za računalniško procesiranje, je upoštevanje skladenjskih realizacij v njih bolj pomembno kot doseganje občutka naravnega razlagalnega diskurza. Razlage imajo za razliko od pomenskih shem drugačno vlogo: podati želijo informacijo o pomenu in rabi besede, pri čemer je tipična skladenjska raba lahko v pomoč pri njenem oblikovanju, ni pa nujno neposredno izražena. Razlagalne strategije za posamezne besedne vrste so opisane v samostojnem poglavju.

10.3.1.1.2.3 Stavčna vezljivost samostalniških pomenov

V okviru t. i. stavčne samostalniške vezljivosti smo obravnavali tudi pomene samostalnikov, ki se v stavčni zgradbi uresničujejo v obliki glagolske zveze, sestavljene iz glagola z oslabljenim pomenom¹¹⁴ in (obravnavanega) samostalnika v tožilniku, npr. *kaj daje noto, barvo čemu, kaj ima vonj po čem*. V nekaterih primerih so take zveze pretvorljive v glagole, ki izražajo splošnega vršilca oz. je ta pomaknjen iz stavka, npr. *vonjati se, barvati se* ipd. Ker v vseh teh primerih govorimo po analogiji z glagolom o stavčni vezljivosti, so nas v pomenski shemi zanimali udeleženci tako na osebkovem mestu kot na mestu predmetnih določil:

nota *samostalnik*

2 značilnost

če LASTNOSTI česa dajejo IZDELKU, KRAJU ali DEJANJEM svojo noto, se v njih izražajo in jih delajo posebne

[dodati, dajati, dati, dodajati, pridati, vdihniti, vnesti, vdahniti] noto
[imeti, vsebovati, dobiti] noto

vonj *samostalnik*

1 občutek česa

če ima PROSTOR ali PREDMET vonj po LASTNOSTIH nekega ČASA ali DOGAJANJA, se to v njem na poseben način izraža ali občuti

vonj po [ljubezni]

¹¹⁴ Kot glagole z oslabljenim pomenom smo obravnavali glagole, ki potrebujejo dopolnilo za izražanje svoje pomenske vsebine (prim. Černelič-Kozlevčar 1975: 76), npr. glagol *ležati* v pomenu 'nahajati se' potrebuje prislovno določilo kraja, npr. *vas leži v dolini* (prav tam). Podobno kot nakazuje omenjena razprava, gre tudi v LBS za glagole, ki sicer ne nastopajo v pomožniških vlogah (tj. kot deli povedka), ampak imajo v drugih kontekstih polnopomensko vlogo, prim. tudi glagole premikanja, npr. *priti do prodaje, razhajanj, zamenjave* ... v pomenu 'zgoditi se' : *priti do vasi*.

barva *samostalnik*

6.1 značilnost; zanimivost

barva so DEJANJA in LASTNOSTI česa, ki dajejo ČEMU pozitivne značilnosti in prepoznavnost

Kadar predstavlja omenjeni način skladenjskega izražanja, tj. v kombinaciji s pomensko oslabljenim glagolom, samo eno od izraznih možnosti pomena samostalnika, smo zveze zapisovali v element skladenjska zveza:

posel *samostalnik*

3 delo, opravilo ali določena aktivnost; opravek

[nujen] posel

[lotiti se] posla

[opraviti] posel

ne (hoteti) imeti posla s kom/čim

imeti posla s kom/čim

10.3.1.1.3 Pridevnik

Zapis pomenske sheme pri pridevnikih je podobno kot pri samostalnikih povezan z glagolskostjo pomenov, zato jo zapisujemo zlasti pri tistih (a) pridevniških pomenih, ki se tipično aktivirajo v povedku in predvidevajo tudi predložna ali brezpredložna določila, hkrati pa z vidika računalniške procesljivosti predvidevamo – podobno kot pri besednozvezni samostalniški vezljivosti – zapis pomenske sheme tudi pri (b) pomenih pridevnikov, ki se tipično realizirajo v prilastku, pri čemer je samostalniška odnosnica glede na pomen obravnavanega pridevnika pomensko odločujoča. V prvem primeru govorimo o stavčni vezljivosti, v drugem pa o kolokabilnosti, kot prikazuje Preglednica 15:

Preglednica 15: Pridevniška vezljivost in kolokabilnost

pridevniška vezljivost	pridevniki v povedni rabi in (iz)glagolski pridevniški pomeni	(biti) odvisen od KOGA/ČESA, (biti) natančen pri ČEM, (biti) pozoren na KOGA/KAJ; (biti) podoben KOMU, (biti) neprilagojen ČEMU/NA KAJ, (biti) lačen ČESA, (biti) nasičen s ČIM
kolokabilnost	pridevniki (zlasti neglagolski)	priden OTROK, čeden MOŠKI/ŽENSKA, mačje POTEZE, borovničeva SLADICA ali NAPITEK, modna OBLAČILA nepremostljiva OVIRA, bakren PREDMET, bakrena BARVA, bazična SNOV/RAZTOPINA, bazična DEJAVNOST

V nadaljevanju bomo natančneje obravnavali načine zapisovanja udeležencev v pomenski shemi pri vezljivih pridevniških pomenih, nato pa na kratko še zapis pomenske sheme pri neglagolskih pridevniških pomenih, najpogostejše realiziranih v vlogi levih prilastkov.

10.3.1.1.3.1 Udeleženci pridevniške vezljivosti

Nosilci pridevniške vezljivosti, kot nakazuje Preglednica 15, so pridevniki, ki imajo pomensko in posledično oblikoskladenjsko veliko skupnega z glagoli, med drugim tudi to, da lahko nastopajo v povedku. Pri nekaterih pridevniških pomenih je ta možnost domala izključna, npr. *divji*: ČLOVEK je divji, če se obnaša agresivno in ni sposoben obvladati svoje jeze ali drugih negativnih čustev. Pri drugih pomenih istega pridevnika, kot npr. 'ki nima urejenega pravnega ali zakonskega statusa' ali 'ki deluje v neskladju z veljavnimi zakoni ali predpisi', pa se raba v povedku navadno ne pojavlja: *divje odlagališče*, *divji lovec*, *divja privatizacija*.

Vezljive pridevniške pomene prepoznamo po tem, da zahtevajo predložna ali nepredložna določila v svojem neposrednem besedilnem okolju. Z vidika oblikoskladenjskega izražanja vezljivosti lahko določila natančneje sklonsko opredelimo, sicer pa je sklonska pridevniška vezljivost lahko tudi večdelna, kar smo v LBS obravnavali kot ustajene skladenjske zveze (v tretjem stolpcu):

obešen čez KAJ ₄	lačen ČESA ₂	izdelan na podlagi česa
brezobziren do KOGA ₂	neprilagojen ČEMU ₃	centraliziran na ravni [države]
koncentriran na KAJ ₄	konkurenčen KOMU ₃	datiran v čas od ... do ...
ljubosumen na KOGA ₄		angažiran na področju česa
informiran o ČEM ₅		
gnan od ČESA ₂		
izoliran iz ČESA ₂		
angažiran pri ČEM ₅		
nasičen s ČIM ₆		
neomajen v ČEM ₅		
dovzeten za KAJ ₄		
izsušen zaradi ČESA ₄		

Ob omenjenih je treba izpostaviti tudi pridevniška predložna in brezpredložna določila v vlogi prislovnih določil, npr. *obešen pod strop/okoli česa*, *izvožen v tujino*, *gojen na prostem*, *usmerjen v prihodnost*, *obrnjen nazaj* itd., in stavčna določila, kot npr. *kriv, da ...*; *motiviran, da ...* ipd.

Pri pridevniških pomenih, ki se skladenjsko realizirajo predvsem v povedkovem določilu, je to razvidno iz zapisa pomenske sheme, pri čemer nas tako s pomenskega vidika kot z vidika strojne obdelave podatkov zanimajo udeleženci ter njihove zapolnitve s konkretnimi semantičnimi tipi v celotnem stavčnem vzorcu, torej tudi na osebkovem mestu:

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

natančen *pridevnik*

2.1 pri športu

TEKMOVALEC je natančen pri METU, STRELU ali PODAJI, če pri tem doseže zadetek

alergičen *pridevnik*

2 ki vzbuja odpor

če je ČLOVEK alergičen na drugega ČLOVEKA, na neko DOGAJANJE ali STANJE, ima do tega negativen odnos ali mu vzbuja odpor

Tipična realizacija pridevnika v vlogi prilastka se v pomenski shemi kaže zgolj v zapisu samostalniškega jedra kot semantičnega tipa:

hlapen *pridevnik*

1 ki prehaja v plinsko stanje

hlapna SNOV, zlasti TEKOČINA, lahko preide v plinsko stanje že pri sobni temperaturi

10.3.1.1.3.2 Kolokabilnost pridevniških pomenov

Ob zgoraj omenjenih pridevniških pomenih so z vidika beleženja semantičnih tipov na mestu samostalniških jeder zanimivi tudi neglagolski pridevniki oz. (njihova) samostalniška jedra, npr. *mačje* POTEZE, *modna* OBLAČILA itd. Tako kot pri vezljivih pridevniki (*odvisen* OD ČESA) nas tudi v takih primerih zanima semantični tip jedra, ki ga zapišemo v pomenski shemi. V primeru izkazane enakovredne rabe v povedku ali prilastku se o najustreznejšem zapisu odločamo tudi na ravni pomenske sheme:

obupan *pridevnik*

1 ki nima več upanja

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

ali

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

Tudi pri pridevniki smo zapis pomenske sheme ločili od razlage na način, da odraža pomenska shema primarno skladijsko realizacijo pomena, medtem ko razlaga v elementu <definicija> s tem ni neposredno obremenjena. Zapis pomenske sheme ji služi le kot pomoč pri razumevanju pomensko relevantnega besedilnega okolja, sicer pa je pri razlagi na prvem mestu način ubeseditve, ki je razumljiv in hkrati zaobseže vse tipične pomenske elemente realne jezikovne rabe.

10.3.1.1.4 Prislov

Zapisovanje pomenske sheme z vidika udeležencev je pri prislovih oz. pri posameznih prislovnih pomenih vezano na njihove pomenske in skladijske lastnosti. O vezljivosti kot sposobnosti napovedovanja udeleženskih mest, kot jo poznamo pri glagolih ter deloma pridevniki in samostalnikih in ki se formalno izraža s skladijskimi označevalci (zlasti z določili) in skladijskimi razmerji (zlasti z vezavo in primikom), pri prislovih glede na ustaljeno vezljivostno teorijo¹¹⁵ ni mogoče govoriti. Skladijska neizraženost pomenskega vpliva besed(il)nega okolja, zlasti torej glagolskega, pridevniškega in samostalniškega jedra na prislove, pa ne izključuje potrebe po pomenski opredelitvi tako jeder kot prislovov, ki se z njimi družijo. Kot smo izpostavili že pri določanju pomenskih lastnosti prislovov na podlagi njihovega tipičnega besedilnega okolja v poglavju 9.1.2.1.2.2 Udeleženska zgradba in udeleženci, je tudi pri oblikovanju pomenske sheme za posamezne prislovne pomene smiselno razmisliti o udeležencih, ki jih predvideva njihova pomenska raba. K takemu pristopu nas ne navaja le izpostavitev kontekstualne rabe prislovov pri prepoznavanju njihovega pomena, pač pa tudi možnost avtomatskega procesiranja in pomenskega razdvajanja, saj mora leksikograf v procesu analize empiričnih podatkov sprejeti odločitve, ki so ključne tako za zapis pomenske razlage v slovarju kot za zapis vzorca, ki ga uspešno prepozna računalnik.

Na primeru prislova *vestno* lahko ugotovimo, da se tipično povezuje s tremi pomenskimi skupinami besed (glagoli in pridevniki), ki sugerirajo tudi njegovo pomensko rabo. V prvi najboljšežnejši skupini so glagoli, ki označujejo izvrševanje oz. izpolnjevanje nalog v skladu s pričakovanji ali predpisi, kot denimo: *izpolnjevati*, *spolnjevati*, *opravljati*, *ravnati*; *slediti*, *upoštevati*, *spoštovati*, *držati se* in *poslušati*. Kot podskupino – v smislu pomenske bližine – lahko prepoznamo glagole, ki zaznamujejo opravila in dejavnosti, ki zahtevajo ponavljajočo se skrb, pazljivost

115 Vsi primeri izrazno formalizirane družljivosti (vezavna družljivost ipd.) kažejo, da gre za neodtujljivo povezan pomenskoskladijski pojav, zaradi katerega bi bilo smiselno bodisi razširiti pojem vezljivosti tudi na raven skladijsko neformaliziranega pomena in govoriti o vezljivosti tudi pri drugih besednih vrstah ali pa jo skladijsko omejiti zgolj na tiste pojave pomenskoskladijske vezljivosti, ki imajo primarno skladijskoformalno potrditev.

in natančnost. Tretjo, relativno samostojno skupino, predstavljajo glagoli, ki bi jih lahko označili s skupnim pomenskim imenovalcem *beležiti*. Gre za glagole, ki označujejo zapisovanje in evidentiranje podatkov na način, ki zahteva natančnost, doslednost, vztrajnost (ponavljanje) in tudi podrobnost:

vestno *prislov*

1 v skladu s pričakovanji, pravili

če človek **vestno** IZPOLNJUJE pričakovanja in OPRAVLJA dolžnosti, dela to tako, kot se od njega pričakuje ali je predpisano

vestno [izpolnjevati, izvrševati, opravljati, ravnati]

vestno [izpolnjen]

1.1 redno in skrbno

če človek **OPRAVLJA** neko delo **vestno**, ga opravlja redno in skrbno

vestno [zaklepati, ločevati, bdeti, skrbeti, pospravljati, odstranjevati]

vestno [pokošen, vzdrževan]

2 natančno

če človek **vestno** BELEŽI podatke, jih natančno in redno zapisuje

vestno [zapisovati, beležiti, popisovati, dokumentirati, popisati]

vestno [popisan, zabeležen]

Pri zapisu pomenov lastnostnih prislovov, realiziranih v povedkovem določilu (npr. *akutno*, *cenovno*, *burno* ipd.), smo v pomenski shemi po analogiji z glagoli, samostalniki in pridevniki tega tipa poleg pomensko relevantnih glagolov in pridevnikov zapisovali z velikimi črkami tudi udeležence na osebkovem mestu:

akutno *prislov*

1 zelo resno; težko

1.1 o boleznih

če je ČLOVEK akutno BOLAN ali če njegova BOLEZEN POTEKA akutno, se razvija zelo hitro in je lahko smrtno nevarna

cenovno *prislov*

glede na višino cene

če je nek IZDELEK ali STORITEV cenovno OPREDELJEN, je določen glede na višino cene

10.3.1.2 Semantični tipi

Teoretična podlaga za določanje semantičnih tipov izhaja iz leksikalnogramatičnega opisa leksike, ki je primarno namenjen avtomatskemu prepoznavanju

skladenjskih razmerij (tj. skladenjsko razčlenjevanje) in avtomatskemu pomensko-mu razdvoumljanju besed. Za semantično prepoznavnost skladenjskih elementov besedila oz. njegovih delov, ki poteka s pomočjo semantičnega (pa tudi oblikoslovnega in skladenjskega) označevanja korpusov, je potrebna določena semantična ontologija. V okviru že omenjenih evropskih jezikovnotehnoških projektov lahko na tem mestu izpostavimo semantično ontologijo, izdelano v projektu SIMPLE, špansko leksikalno bazo GRIAL s pomensko opredeljenimi udeleženskimi vlogami ter leksikalno bazo SenSem (ang. *sentence semantic*), ki vsebuje pomensko in skladenjsko označene stavke, projekt ELSNET, v okviru katerega so bila označena predikatnoargumentna razmerja v italijanskih stavkih, ter vsaj še češki projekt VALLEX, nivo semantičnega označevanje v danski leksikalni bazi STO in semantično ontologijo, ki jo predvideva WordNet.

Semantični tipi udeležencev na posameznih mestih argumentne zgradbe so v omenjenih projektih različno poimenovani. Pri projektu FrameNet se govori o t. i. shemskih elementih, kot so npr. Agent, Theme, Degree, Duration, Manner, Place, Purpose, Time, Situation itd. V projektu CPA je govora o t. i. implikaturah, kot so npr. Human, Animal, Institution, Activity, Plan, Concept, Artifact, Force, Vehicle ipd., pri čemer so posamezne implikature lahko tudi podrobneje določene, npr. Human Group = Military, Human = bad Guy, Human = Victim, Human = Male, Human = Female ipd. V češkem projektu Vallex so posamezne udeleženske vloge v glagolski shemi, ki je izdelana za vsak posamezni pomen posebej, označene s t. i. funktorji, kot npr. actor, addressee, patient, effect, origin, difference, obstacle ter intent (zadnji trije kot kvazivalenčna dopolnila; Panevová 2005), poleg tega pa še: accompaniment, aim, benefactive, cause, complement itd. Funktorji, kot jih predvideva Vallex, naj bi označevali t. i. globinske udeleženske vloge, s katerimi se izražajo različni tipi razmerij med glagoli in njihovimi dopolnili, razumljenimi v splošnem smislu. Funktorji so še nadalje ločeni na notranje udeležence (aktante) in proste modifikatorje, kar v grobem ustreza dihotomiji določilo – dopolnilo (Panevová 1974; Panevová 1994). Pri pomenskem označevanju korpusa GRIAL govorijo o pomenskem opisu udeleženskih vlog ter o prepoznavanju t. i. dogodkovnih tipov, kot je denimo dogodek, proces, stanje ipd. Pri WordNetu, ki je najširše uporabljana ontologija za semantično označevanje korpusov, so leksikalne enote, in sicer samostalniki, glagoli, pridevniki in prislovi znotraj določenega pomenskega polja, zastopanega s krovnimi pojmi (ang. *concepts*),¹¹⁶ urejeni v sinonimske nize – sinsete,

116 Znotraj izhodiščnega princetonškega WordNeta obstaja nekaj vrhnjih konceptov (s prvotnih 25 se je npr. število za samostalnike znižalo na 9; Gangemi et al. 2001), ki jih praviloma uporabljajo vsi nacionalni WordNeti. Gre za neke vrste sinsete, ki so jezikovno neodvisni in so (lahko) osnova za domenske WordNete, ki se razvijajo v zadnjem času.

ki so med seboj povezani z različnimi semantičnimi in leksikalnimi relacijami, kot so npr. hipo- in hipernimija ter meronimija: {ata, atek, ati, očka, tata} -> {oče} -> {starš} -> {neposredni prednik} -> {prednik} -> {sorodnik, sorodstvo, žlahta, žlah-tnik} (Fišer 2005: 18–20).

V LBS razumemo semantične tipe v pomenski shemi LBS kot abstraktne zastopnike leksikalnih zapolnitev na predvidenih udeleženskih mestih, ki se na kolokacijskem nivoju realizirajo v obliki kolokacij. Pri glagolih v zvezi s semantičnimi tipi govorimo o pomenskih zastopnikih na mestu določil in dopolnil, ki so besednovrstno najpogostejše izraženi s samostalniki in samostalniškimi besednimi zvezami, in sicer tudi v primerih, ko se sicer pomensko relevantna prislovna določila realizirajo v primičnem skladenjskem razmerju, npr: če se ČLOVEK rodi v določenem KRAJU in v določenih OKOLIŠČINAH, ga to določa in vpliva na njegovo življenje. V nekaterih redkih primerih pa tudi s prislovnim zaimkom, ki zastopa krajevne ali časovne prislove, npr. če ŽIVAL kuka OD KOD, opazuje svojo okolico, navadno skozi manjšo odprtino. Tudi pri samostalnikih in pridevniki so udeleženci v pomenski shemi zastopani s samostalniškimi semantičnimi tipi, pri prislovih pa z glagoli in s pridevniki, kot smo prikazali v prejšnjem poglavju.

Semantični tipi so namenjeni formalnemu prepoznavanju relevantnih pomensko-skladenjskih elementov v besedilnem okolju določene besede. Predvidevamo, da bo podatke, ki jih zagotavljajo semantični tipi v tipičnem skladenjskem vzorcu, mogoče uporabiti za pomensko procesiranje podatkov, in sicer tako za semantično označevanje učnih korpusov¹¹⁷ kot za avtomatsko razdvoumljanje besednih pomenov, nena-zadnje pa tudi za integriranje oz. povezovanje z obstoječimi semantičnimi bazami.

Ker je LBS primarno namenjena izdelavi slovarjev in s tem načinu predstavitve podatkov, ki so razumljivi za človeškega uporabnika, je tudi formalizacija semantičnih tipov podrejena izhodiščno slovarskemu in ne jezikovnotehnološkemu konceptu. Ker je v pomenski shemi zapisana samo ena od skladenjskih možnosti realizacije pomena, dejansko pa se pomen vedno lahko realizira na več možnih skladenjskih načinov, je tudi zapis semantičnih tipov vezan le na izbrano skladenjsko realizacijo, ki pa naj bi bila za obravnavani pomen besede čim bolj tipična.

V nadaljevanju bomo opisali semantične tipe, kot jih razumemo v LBS. Izpostavili bomo različne pomenske skupine in način njihovega zapisovanja ter probleme, na katere smo pri tem naleteli in izhajajo iz specifik slovenskega jezika.

117 Semantični bazi, ki sta večinoma rabljeni za semantično označevanje korpusov, sta WordNet in FrameNet. V slovenskem pilotnem FrameNetu je trenutno devet glagolov gibanja v 225 stavkih, označenih na oblikoslovni, skladenjski in semantični ravni (Može 2013), ki jih predvideva framenetovski sistem označevanja.

10.3.1.2.1 Povezanost semantičnih tipov z elementi LBS za namene strojnega procesiranja

V spodnjem zapisu pomenske sheme so semantični tipi pomensko relevantnih udeležencev za enega od pomenov glagola *zmagati* označeni z različnimi barvami in tako povezani s svojimi pomenskimi (in skladenjskimi) ustrezniki na ravni besedilnih realizacij:

zmagati *glagol*

1 doseči najboljši rezultat

1.1 v športu

če **ŠPORTNIK** ali **ŠPORTNA EKIPA** zmaga, doseže na **ŠPORTNEM TEKMOVANJU** v določenem **ŠPORTU** ali glede na določena **MERILA** v odnosu do drugih **ŠPORTNIKOV** ali **ŠPORTNE EKIPE** najboljši **ŠPORTNI REZULTAT**

- *Odločale bodo malenkosti, zmagala pa bo bržkone ekipa, ki bo imela tudi kanček več sreče.*
- *V Rušah so naši zmagali s 75:59 (34:25), v Lenartu pa s 67:51 (34:24).*
- *Obeta se zanimiv obračun, v katerem bo zmagal najboljši.*
- *V zgodovini svetovnih pokalov je izmed slovenskih smučarjev v veleslalomu zmagal le Boris Strel in to davnega leta 1981 v Cortini.*
- *Na 90-metrski skakalnici je med člani zmagal Luka Bardorfer (SSK Ilirija) s skokoma 96 in 95 metrov ...*
- *Selektor mladinske reprezentance Miran Kavaš je bil z začetkom priprav za oktobrsko SP zelo zadovoljen, saj je Slovenija zmagala med ekipami, majice pa so osvojili še Švab za najboljšega tujca ...*
- *Šele v zadnji minuti mi je uspel veliki met in tako sem zmagal v kategoriji, v kateri je nastopilo več kot 20 mojstrov juda.*
- *Tislju so prišle na uho njihove napovedi, da želijo zmagati v skupini C, kar je v skladu z islandskim neomajnim optimizmom.*
- *Ves čas se je z ostalimi menjal na čelu in bil najmočnejši tudi na cilju ter zaslužen zmagal nad Andrejem Plešnarjem iz Nove Gorice.*

S pomenskosladdenjskega vidika je treba opozoriti na enake semantične tipe pri udeležencih z različnimi udeleženskimi vlogami, ki jih v pomenski shemi nismo označevali, npr. semantični tip: ŠPORTNIK, ŠPORTNA EKIPA ima v osebkovem položaju vlogo aktanta, v položaju predložnega predmetnega določila: v odnosu do drugih ŠPORTNIKOV ali ŠPORTNE EKIPE, pa nastopa v vlogi okoliščin (npr. ovira). Za strojno prepoznavanje pomensko relevantnih razmerij je povezava med udeleženci, njihovimi semantičnimi tipi in udeleženskimi vlogami nujna, vendar pa je hkrati potrebno tudi upoštevati, da ob določeni ročni formalizaciji teh podatkov, ki predstavljajo učno množico, nato večino dela lahko opravi računalnik sam, tj. avtomatsko zgenerira vse možne kombinacije semantičnih tipov in njihovo opuščanje, kot je nakazano v nadaljevanju. Pri jezikovnotehnoški nadgradnji slovarske baze je bilo zato predvideno, da bo postopek povezave med semantičnimi tipi v pomenski shemi, kolokacijami in stavčnimi vzorci opravljen naknadno. V spodnjem primeru je različna vloga semantičnega tipa

opredeljena s številko v oglatih oklepajih, posamezne realizacije pa so navedene v vzorcih, ki jih je prek semantičnih tipov mogoče povezati tudi s kolokacijami:

če ŠPORTNIK ali ŠPORTNA EKIPA_[1] zmaga, doseže na ŠPORTNEM TEKMOVANJU_[2] v določenem ŠPORTU_[3] ali glede na določena MERILA_[4] v odnosu do drugih ŠPORTNIKOV ali ŠPORTNE EKIPE_[5] NAJBOLJŠI REZULTAT_[6]

- **kdo**_[1] zmaga
- **kdo**_[1] zmaga **na čem**_[2]
- **kdo**_[1] zmaga **v čem**_[2]
- **kdo**_[1] zmaga **v čem**_[3]
- **kdo**_[1] zmaga **v čem**_[4]
- **kdo**_[1] zmaga **med kom**_[4] (množina)
- **kdo**_[1] zmaga **med kom**_[5] (množina)
- **kdo**_[1] zmaga **pred kom**_[5]
- **kdo**_[1] zmaga **nad kom**_[5]
- **kdo**_[1] zmaga **s čim**_[6]

[1] , [5] – ŠPORTNIK, ŠPORTNA EKIPA: [velemojster, skakalec, smučar, domačin, domačinka, gostitelj, barca, španec, favorit, ekipa, KK, PGD, reprezentantka, posadka, kolesar]; [S_med-d_g6], [S_pred-d_g6]

[2] – ŠPORTNO TEKMOVANJE: [na_ tekma, na_turnir, na_dirka, na_tekmovanje, na_miting], [v_tekma]

[3] – ŠPORT

[4] – MERILA: [v_niz, v_etapa, v_disciplina, v_dvoboj, v_kategorija] [S_med-d_g6]

[6] – REZULTAT: [z_točka, z_gol, z_zadetek]

Na podlagi zapisa, v katerem so z navpičnicami ločeni posamezni semantični tipi,

kdo_[1] zmaga | **na čem**_[2] / **v čem**_[2] | **v čem**_[3] | **v čem**_[4] / **med kom**_[4] (množina) | **med kom**_[5] (množina) / **pred kom**_[5] / **nad kom**_[5] | **s čim**_[6],

s poševnicami pa različne skladenjske variante znotraj posameznega semantičnega tipa, lahko avtomatsko sprocesiramo potencialne kombinacije, ki jih lahko pričakujemo v realnih besedilih.

V primeru pomenskosladdenjskega opisa pomena glagola *zmagati* v pomenski shemi lahko zgeneriramo obsežen nabor potencialnih stavčnih vzorcev, ki vključujejo podatke o možnih skladdenjskih realizacijah in tipičnih leksikalnih zapolnitvah na predvidenih udeleženskih mestih. Spodaj prikazujemo samo potencialnih 50 takih vzorcev:

kdo_[1] zmaga
 kdo_[1] zmaga | na čem_[2]
 kdo_[1] zmaga | v čem_[2]
 kdo_[1] zmaga | v čem_[3]
 kdo_[1] zmaga | na čem_[2] | v čem_[3]
 kdo_[1] zmaga | v čem_[2] | v čem_[3]
 kdo_[1] zmaga | v čem_[4]
 kdo_[1] zmaga | med kom_[4] (množina)
 kdo_[1] zmaga | na čem_[2] | v čem_[4]
 kdo_[1] zmaga | na čem_[2] | med kom_[4] (množina)
 kdo_[1] zmaga | v čem_[2] | v čem_[4]
 kdo_[1] zmaga | v čem_[2] | med kom_[4] (množina)
 kdo_[1] zmaga | v čem_[3] | v čem_[4]
 kdo_[1] zmaga | v čem_[3] | med kom_[4] (množina)
 kdo_[1] zmaga | na čem_[2] | v čem_[3] | v čem_[4]
 kdo_[1] zmaga | na čem_[2] | v čem_[3] | med kom_[4] (množina)
 kdo_[1] zmaga | v čem_[2] | v čem_[3] | v čem_[4]
 kdo_[1] zmaga | v čem_[2] | v čem_[3] | med kom_[4] (množina)
 kdo_[1] zmaga | med kom_[5] (množina)
 kdo_[1] zmaga | pred kom_[5]
 kdo_[1] zmaga | nad kom_[5]
 kdo_[1] zmaga | na čem_[2] | med kom_[5] (množina)
 kdo_[1] zmaga | na čem_[2] | pred kom_[5]
 kdo_[1] zmaga | na čem_[2] | nad kom_[5]
 kdo_[1] zmaga | v čem_[2] | med kom_[5] (množina)
 kdo_[1] zmaga | v čem_[2] | pred kom_[5]
 kdo_[1] zmaga | v čem_[2] | nad kom_[5]
 kdo_[1] zmaga | v čem_[3] | med kom_[5] (množina)
 kdo_[1] zmaga | v čem_[3] | pred kom_[5]
 kdo_[1] zmaga | v čem_[3] | nad kom_[5]
 kdo_[1] zmaga | na čem_[2] | v čem_[3] | med kom_[5] (množina)
 kdo_[1] zmaga | na čem_[2] | v čem_[3] | pred kom_[5]
 kdo_[1] zmaga | na čem_[2] | v čem_[3] | nad kom_[5]
 kdo_[1] zmaga | v čem_[2] | v čem_[3] | med kom_[5] (množina)
 kdo_[1] zmaga | v čem_[2] | v čem_[3] | pred kom_[5]
 kdo_[1] zmaga | v čem_[2] | v čem_[3] | nad kom_[5]
 kdo_[1] zmaga | v čem_[4] | med kom_[5] (množina)
 kdo_[1] zmaga | v čem_[4] | pred kom_[5]
 kdo_[1] zmaga | v čem_[4] | nad kom_[5]
 kdo_[1] zmaga | med kom_[4] (množina) | med kom_[5] (množina)
 kdo_[1] zmaga | med kom_[4] (množina) | pred kom_[5]
 kdo_[1] zmaga | med kom_[4] (množina) | nad kom_[5]
 kdo_[1] zmaga | na čem_[2] | v čem_[4] | med kom_[5] (množina)
 kdo_[1] zmaga | na čem_[2] | v čem_[4] | pred kom_[5]

kdo _[1] zmaga	na čem _[2]	v čem _[4]	nad kom _[5]	
kdo _[1] zmaga	na čem _[2]	med kom _[4] (množina)	med kom _[5] (množina)	
kdo _[1] zmaga	na čem _[2]	med kom _[4] (množina)	pred kom _[5]	
kdo _[1] zmaga	na čem _[2]	med kom _[4] (množina)	nad kom _[5]	
kdo _[1] zmaga	v čem _[2]	v čem _[4]	med kom _[5] (množina)	
kdo _[1] zmaga	v čem _[2]	v čem _[4]	pred kom _[5]	

Končni rezultat je niz potencialnih stavčnih vzorcev, v katerih lahko pričakujemo realizacijo konkretnega pomena ob izpolnjenih leksikalnih pogojih in predvidenih opuščanjih, tako na ravni dela stavčnega vzorca kot v smislu variantnih leksikalnih, zlasti predložnih realizacij znotraj njih.

10.3.1.2.2 Vrste semantičnih tipov

V tem poglavju si bomo podrobneje pogledali pomensko vlogo semantičnih tipov v pomenski shemi, saj je njihov primarni namen pomensko opredeliti relevantne udeležence in okoliščine v tipični stavčni in/ali besednozvezni realizaciji pomena. Semantični tipi, ki se potrjujejo s konkretnimi kolokatorji na kolokacijskem nivoju LBS, so ob oblikoskladenjski realizaciji udeležencev nosilci pomenskih sprememb, kar ponazarja zapis pomenske sheme pri različnih pomenih istega glagola:

zbošti <i>glagol</i>
1 občutiti ali povzročiti ostro bolečino
če se ČLOVEK zbode z OSTRIM PREDMETOM v DEL TELESA, občuti ostro bolečino
1.1 o občutku zbadanja v telesu
če ČLOVEKA zbode kje v TELESU, občuti tam ostro bolečino
2 zbuditi pozornost, negativen odziv
če neka IZJAVA, BESEDILO ali določeno STANJE zbode ČLOVEKA, v njem zbudi pozornost, navadno negativen odziv ali neodobravanje
3 prizadeti ali biti prizadet
če neka IZJAVA zbode ČLOVEKA, ga prizadene ali če ČLOVEK zbode drugega ČLOVEKA, ga s svojo IZJAVO prizadene

Osnovno pravilo za določitev semantičnega tipa je skupna pomenska lastnost besed, ki se v določeni gramatični relaciji pojavlja na registriranem udeleženskem mestu. Zato se, za razliko od obveznih udeležencev in okoliščin, ki jih potrjujemo in preverjamo stavčnimi vzorci in z zgledi, semantični tipi potrjujejo predvsem s kolokacijami. V zgornjem primeru zastopajo semantične tipe naslednji kolokatorji:

OSTER PREDMET: igla, šivanka, trn, nož, bucika, injekcija
 TELO/DEL TELES: palec, prst, roka, zadnjica; prsi, križ, hrbtenica, želodec
 IZJAVA: opazka, dejstvo, navedba, govorica, trditev, očitek, pripomba, kritika, članek, komentar
 BESEDILO: zapis, tabla, napis
 STANJE: podobnost, odsotnost, ugotovitev

Določeni semantični tipi s kolokatorji niso zastopani, ker so v besedilnih realizacijah bodisi pomensko razpršeni in jih ni mogoče prepoznati v konkretnih kolokacijah bodisi so neizraženi ali zastopani z drugimi skladenjskimi sredstvi, npr. semantični tip ČLOVEK navadno vključuje osebna glagolska oblika.

V prvi fazi izdelave LBS smo semantične tipe zapisovali brez enotne taksonomije, saj gre za postopek, ki ga je potrebno reševati na poti od konkretnega kolokatorja k abstraktnemu pomenskemu zastopniku. V drugi fazi izdelave LBS smo zato posamezne semantične tipe poenotili in izdelali štiristopenjsko taksonomijo, ki jo na kratko predstavimo v nadaljevanju.

Taksonomijo sestavlja pet med seboj hierarhično povezanih nivojev. Hierarhično najvišje je t. i. (1) pomensko polje, ki vsebuje posamezne pomenske razrede, in sicer (1.1) pomenske tipe, (1.2) pomenske podtipe, (1.3) kolokatorske tipe in (1.4) kolokatorje.

Pomenska polja zastopajo vrhnje pomenske kategorije, ki služijo združevanju osnovnih pomenskih tipov. Ob zaključku LBS so bila taka pomenska polja tri: (1) živo bitje, (2) konkretna entiteta in (3) abstraktna entiteta. Oznak za pomenska polja v pomenskih shemah navadno nismo uporabljali za označevanje udeležencev. Načeloma najširšo pomensko opredelitev udeleženca zastopajo v pomenski shemi nedoločni zaimki, ki so v nekaterih primerih zamenljivi z oznako pomenskega polja, zastopajo pa zlasti udeležence v obliki odvisnega stavka, npr.

domnevat *glagol*

1 vnaprej sklepati, predvidevati

če ČLOVEK domneva, da se bo NEKAJ zgodilo ali da je NEKAJ resnično, o tem vnaprej sklepa ali predvideva na podlagi okoliščin ali izkušenj

- *Ljudje že dolgo domnevajo, da se staramo, ker naše celice posamič kopičijo mutacije.*
- *Mnogi domnevajo, da uživanje antibiotikov zmanjšuje tveganje srčnih obolenj.*
- *Opazovalci domnevajo, da bosta Gerharda Schröderja lokalna poraza dokončno spodbudila k odločnejšim reformam.*

Ker je za strojno procesiranje zapis z zaimkom v pomenski shemi manj primeren, hkrati pa je taka možnost predvidena za stavčne vzorce na skladenjski ravni, smo tudi v primerih, ko je abstraktni udeleženec zastopan tipično z odvisnim stavkom, težili k oblikovanju udeleženca v obliki bolj konkretnega semantičnega tipa, npr.

zažugati *glagol*
2 zagroziti

če PREDSTAVNIK ustanove ali USTANOVA zažuga drugemu ČLOVEKU ali USTANOVI, ji zagrozi z BESEDAMI ali DEJANJI

- *Premier je zato zažugal, da bo krščanske demokrate v tem primeru vrgel iz vladajoče koalicije.*
- *Rusija ima v rokavu pomemben adut, s katerim lahko tudi zažuga začasni iraški vladi.*

Pomenski tipi predstavljajo pomenske kategorije znotraj pomenskih polj in smo jih v pomenski shemi uporabljali za označevanje udeležencev, kadar besedilnih realizacij ni bilo mogoče natančneje opredeliti. Npr. pomensko polje živo bitje vključuje semantične tipe, kot jih prikazuje Preglednica 16:

Preglednica 16: Primeri pomenskih tipov v pomenski taksonomiji

Pomensko polje	Pomenski tip
(1) ŽIVO BITJE	
	ČLOVEK
	BITJE
	ŽIVAL
	RASTLINA
	TELO

Najpogosteje uporabljan semantični tip, ki na ravni udeležencev zastopa živo bitje, je ČLOVEK. Ta semantični tip smo uporabili v primerih, kjer je pomen glagola vezan na človeka kot živo bitje, ki je sposobno razumskega in čustvenega odzivanja, njegove natančnejše opredelitve pa za pomen glagola niso odločilne. V teh primerih se na mestu kolokatorjev navadno pojavljajo pomensko različni udeleženci z oznako človeško in živo ter lastna imena. Semantični tip ČLOVEK smo uporabili tudi v primerih, kjer se glagol nanaša tako na človeka kot na žival, vendar se glagolski pomen primarno osredotoča na človeka, kar pomeni, da se živali v dejanski rabi redko ali pa sploh ne izražajo. Taki so zlasti glagoli, ki označujejo fizična ali psihična stanja:

zaspati *glagol*

1 preiti v stanje počitka

ko ČLOVEK zaspi, preide v STANJE TELESNEGA in UMSKEGA POČITKA, navadno ponoči ali zaradi utrujenosti

krvaveti *glagol*

1 izgubljati kri

ČLOVEK krvavi, če mu iz RANE ali DELA TELESA teče kri

Za zapis obeh semantičnih tipov, tj. ČLOVEK in ŽIVAL, smo se odločali v primerih, kjer se ob nespremenjenem pomenu glagola oba izražata tudi v rabi:

mučiti *glagol*

2 povzročati neprijetne občutke

2.1 o bolečini

če ČLOVEK muči drugega ČLOVEKA ali ŽIVAL, mu povzroča BOLEČINO

mučiti [živali, mačke]

mučiti [zapornike, ujetnike, jetnike, nasprotnike]

pobožati *glagol*

1 pogladiti

če ČLOVEK poboža drugega ČLOVEKA ali ŽIVAL, ga z ROKO ljubeče pogladi po DELU TELESA

pobožati [psa, mačko, konja]

pobožati [otroka]

Seveda pa v isti pomenski shemi semantičnih tipov, kot so ČLOVEK, ŽIVAL (pa tudi RASTLINA), nismo združevali, če je zastopanost določenega semantičnega tipa vplivala na pomen ali na izražanje pragmatičnih in stilističnih odtenkov, kot npr. v spodnjem zgledu:

crkniti *glagol*

1 umreti

če rečemo, da ČLOVEK crkne, na zelo grob način povemo, da umre, da nam je za to vseeno ali da mu to privoščimo

- *Mojega očeta mi je bilo žal, ne rečem, moja stara pa labko crkne na licu mesta.*
- *Povem vam, da bi me možakar tam zunaj pustil crkniti od mraza in lakote, če mu ne bi dal denarja, ki ga je zahteval od mene.*

1.1 poginiti

če rečemo, da ŽIVAL crkne, na grob način povemo, da pogine

- *Te živali so pač večkrat takšne, da nepričakovano crknejo, ko jih udomačimo, ne da bi imele pravi vzrok.*
- *Neke bele gobe z rdečimi pikami naj bi bile strupene, ker je nekaj kozic crknilo, ko so jih požrle, pomešane med travo.*

1.2 oveneti

če rečemo, da RASTLINA crkne, na grob način povemo, da oveni, navadno zato, ker zanjo ne skrbimo ali ker nima ustreznih pogojev za uspevanje

- *... napolni z zemljo in v njo presadi kaktuse, ki pa obvezno (kdo ve, zakaj?) iz nerazumljivih razlogov crknejo čez kakšen teden.*
- *In nato se bočejo pogovarjati o tem, zakaj je omara premaknjena za deset centimetrov, zakaj ni pobrisan prah z nagačenega fazana, zakaj je »crknila« tista njihova lepa rožica, ki so vam jo zaupali, in druge takšne neumnosti.*

Na osebkovem mestu v udeleženski zgradbi glagolov, ki pomenijo kognitivne procese, se ob semantičnem tipu ČLOVEK po principu metonimije pojavlja zelo pogosto tudi USTANOVA. V primerih, kjer gre za dejanja in lastnosti, ki jih dejansko počnemo ljudje, npr. *dogovarjati se, odpustiti* ipd., običajna jezikovna raba pa na osebkovem mestu izpostavlja tudi ustanovo, kot v spodnjih primerih, smo se odločali načeloma v skladu s tipičnimi realizacijami, in sicer:

(a) v primeru metonimičnih rab in v zvezi s procesi, ki se nanašajo primarno na človeka oz. ljudi, smo se odločali samo za zapis semantičnega tipa ČLOVEK oz. LJUDJE ali pa smo izpostavili, da gre predvsem za LJUDI v ustanovi, npr.

odpustiti *glagol*

1 povzročiti, da kdo izgubi službo

če USTANOVA ali ODGOVORNI v ustanovi odpusti ZAPOSLENEGA iz službe, ta ostane brez zaposlitve in ne prejema več plačila

- *Šef ga lahko odpusti kadarkoli, kar ne bi bilo pri veliki brezposelnosti in šibki pravni varnosti zaposlenih nič posebnega.*
- *Nasploh so podjetja prepričana, da lahko odpustijo četrtno delavcev, ne da bi se zmanjšala raven proizvodnje.*

(b) Če je za pomen iztočnice relevantno, da na mestu delovalnika nastopa ustanova ali predvsem ustanova glede na svojo dejavnost, pravne, gospodarske ipd. pristojnosti in se to potrjuje tudi s kolokacijami in/ali besedilnimi zgledi, smo kot semantični tip izpostavili tudi ali samo ustanovo:

dogovarjati se *glagol*

1 izmenjavati mnenja

če se LJUDJE ali USTANOVA dogovarjajo z drugimi LJUDMI ali USTANOVO |ali|
 če se LJUDJE ali USTANOVE dogovarjajo med sabo, se pogovarjajo o tem, kako rešiti
 konkreten PROBLEM

[občina, vodstvo, predstavnik, podjetje, država] se dogovarja

potegovati se *glagol*

1 poskušati doseči

1.1 o funkciji ali koristi

če se ČLOVEK ali USTANOVA poteguje za določen STATUS, KORIST ali DOBI-
 ČEK, si med več možnimi kandidati prizadeva to doseči

- *Temnopolti Larry Richards se poteguje za sedež v amerškem senatu.*
- *Reiffeisen Zentralbank se je pred več kot letom dni že potegovala za nakup SKB banke, a je kasneje v tekmi, v kateri so zmagali Francozi, izpadla.*
- *Novomeška družba Portoval se je na razpisu potegovala za koncesijo za program kreiranja javnih del in jo tudi dobila.*

Pomenski podtipi določajo relativno samostojne pomenske skupine znotraj pomenskega tipa, npr. dele pojavnosti ali entitete: TELO: DEL TELESA; OBLAČILO: DEL OBLAČILA; needninsko pojavnost, npr. ČLOVEK – SKUPINA LJUDI, ter pomenske podtype na istem nivoju, npr. PREDMET: IZDELEK, PROIZVOD. Nekatere pomenske podtype znotraj hierarhično višjih pomenskih tipov in pomenskih polj prikazuje Preglednica 17.

Preglednica 17: Primeri semantičnih podtipov v pomenski taksonomiji

Pomensko polje	Pomenski tip	Pomenski podtip
ŽIVO BITJE		
	ČLOVEK	
		SKUPINA LJUDI
KONKRETNA ENTITETA		
	PRIPOMOČEK	
		ORODJE
		REKVIZIT
	NAPRAVA	
		DEL NAPRAVE
	LOKACIJA	
		OBMOČJE
		KRAJ
		PROSTOR
		MESTO
ABSTRAKтна ENTITETA		
	ENERGIJA	
		TOPLOTA
		SVETLOBA
		ZVOK

Pomensko polje	Pomenski tip	Pomenski podtip
	DOGODEK	
		IZKUŠNJA
		VZROK
		POJAV

Če poskušamo opredeliti skupne pomenske lastnosti za enotno označevanje semantičnih tipov, ki npr. v zgornji tabeli predstavljajo pomenske podtipе za pomenski tip LOKACIJA: OBMOČJE; KRAJ; PROSTOR, velja, da smo najsplošnejšo oznako, tj. LOKACIJA, uporabili v primerih, ko smo glede na besedilne realizacije skušali zajeti neko nedefinirano prostorsko entiteto, zlasti:

(a) nahajanje koga ali česa v neki točki ali na neki širši nedefinirani lokaciji, npr.

gnezdti *glagol*

2 biti kje udobno nameščen

če rečemo, da ČLOVEK gnezdi na določeni LOKACIJI, je tam udobno nameščen

- *Gnezdila sva na gričku in gledala čez nizko ogrado na gubajočo se gladino jezera ljudi, ki se je razlivalo pred odrom, na katerem je Kali z Martinika igral svoj martinški reggae.*
- *Edino razkošje, ki si ga občasno privoščim, je počitek na zaslanem ležišču, če na njem ne gnezdi kateri od potnikov ali njegovih pomočnikov.*

(b) namenjenost koga na različne, nedefinirane lokacije:

iskati *glagol*

5 pobrati; dvigniti; odpeljati

če gre ČLOVEK iskat ČLOVEKA ali PREDMET na neko LOKACIJO, ga od tam odpelje

- *Na policijo ga je prišla iskat mama, ker je oče montiral telefonske centrale nekje na drugem koncu sveta.*
- *Damirjeva sestra je povedala, da je mama odšla iskat brata v bolnišnico in da se še nista vrnila domov.*

Semantični tip KRAJ smo uporabljali za izražanje konkretnih lokacij, zlasti geografskih enot, ki predpostavljajo ali zastopajo konkretno zemljepisno ime, npr.

deževen *pridevnik*

1 o dežju

1.1 z veliko dežja

v deževnem OBDOBJU ali KRAJU veliko dežuje

deževen [kraj]

deževno [mesto]

- *V Posočju, ki ima sloves zelo deževnih krajev, ne pomnijo tako velike požarne ogroženosti.*
- *Naziv deževno mesto pa si je prislužilo čisto po krivici, saj so podatki, da je to mesto z največ dežja v ZDA, čisto napačni.*
- *Pa lep pozdrav izpod deževnega Pohorja!*

zaspan *pridevnik*

2 miren, neaktiven

v zaspanem KRAJU ali v zaspanem OBDOBJU se ne dogaja veliko

zaspano [mestece, mesto, gnezdo]

zaspano [podeželsko] mestoce

zaspana [vasica, vas]

zaspan [kraj]

Semantični tip PROSTOR smo tipično uporabljali za zaprte oz. notranje lokacije, lahko pa zastopa tudi bolj abstraktna področja in krajevne okoliščine, npr.

vlomiti *glagol*

na silo vstopiti; vdreti

če ČLOVEK vlomi v nek PROSTOR, vanj vstopi z uporabo sile in z namenom, da bo kaj ukradel

bivalen *pridevnik*

1 o prebivanju, stanovanju

1. 1 namenjen prebivanju, stanovanju

bivalni PROSTOR ali STAVBA je namenjen prebivanju

bivalni [prostor, vikend, objekt, vrt]

bivalni [zabojnik, kontejner, modul, čoln, avtomobil]

Nasprotno smo semantični tip OBMOČJE uporabljali za označevanje zunanjih, odprtih prostorov v naravi in za običajno manj definirane lokacije, kar se kaže tudi v tipičnih kolokacijah, npr.

smučati *glagol*

1 voziti se s smučmi

1.1 na snegu

če ČLOVEK smuča, se s smučmi večkrat spusti po zasneženem in za to urejenem OBMOČJU

smučati po [strmini, progi]

smučati po [označeni, strmi] progi

smučati na [smučišču, progi]

smučati z [vrha]

Kolokatorski tipi natančneje opredeljujejo pomenske tipe ali podtipe s konkretnjšimi predstavniki, ki pa so še vedno zastopniki večje skupine sorodnih primerkov, npr. ČLOVEK: MOŠKI, ŽENSKA, OTROK, ali s pomočjo prilastka, npr. BITJE: MITOLOŠKO BITJE, ŽIVO BITJE, BITJE z nadnaravno močjo; DELOVANJE: UMETNIŠKO DELOVANJE. Kolokatorske tipe lahko členimo na relativno samostojne pomenske skupine, npr. USTANOVA: PODJETJE, ORGANIZACIJA; DRŽAVNA USTANOVA; MEDIJI. Kot je razvidno iz Preglednice 18, kolokatorski podtipi niso nujno podrejeni neposredno pomenskemu podtipu, ampak lahko opredeljujejo skupino primerkov istega razreda tudi znotraj širšega pomenskega tipa:

Preglednica 18: Primeri kolokatorskih tipov v pomenski taksonomiji

Pomensko polje	Pomenski tip	Pomenski podtip	Kolokatorski tip
(1) ŽIVO BITJE	ČLOVEK		UMETNIK
			STROKOVNJAK
			ŠPORTNIK
			MOŠKI
			ŽENSKA
			OTROK
			PARTNER
		SKUPINA LJUDI	ŠPORTNA EKIPA
			MNOŽICA
BITJE			MITOLOŠKO BITJE
			BITJE z nadnaravno močjo
ŽIVAL			PTICA
			ŽUŽELKA
RASTLINA			SADJE
			SADEŽ, PLOD
			DREVO
			ZRAVILNA RASTLINA
			ZELENJAVA
TELO	DEL TELESA		
			ORGAN
			GLAVA, OBRAZ
			OKONČINE

Semantični tipi, ki jih v pomenski shemi zastopajo kolokatorski tipi, se na ravni kolokacij potrjujejo s konkretnimi besedami, kot je prikazano za pomen pridevnika *čeden*:

čeden *pridevnik*

1 ki lepo izgleda; lep; privlačen

čedna ŽENSKA ali MOŠKI se nam zdi privlačna, navadno zato, ker je lepa in urejena

čedna [punca, rjavolaska, ženska, dekle, svetlolaska, mladenka]

čeden [mladenič, fant, moški, dečko, temnolasec]

čeden [natakar, neznanec]

S *Kolokatorji* v vlogi semantičnih tipov smo označevali udeležence pri posameznih pomenih besede, kjer je realizacija pomena vezana na konkretno leksikalno zapolnitev, kot je prikazano v Preglednici 19.

Preglednica 19: Primeri kolokatorjev v pomenski taksonomiji

Pomensko polje	Pomenski tip	Pomenski podtip	Kolokatorski tip	Kolokator
ŽIVO BITJE	ČLOVEK		UMETNIK	PLESALEC, PEVEC
	STROKOVNJAK			ZDRAVNIK, ZNANSTVENIK
KONKRETNA ENTITETA	PRIPOMOČEK	ORODJE		
		REKVIZIT		
		ŠPORTNI REKVIZIT		
				ŽOGA, MREŽA
	NAPRAVA			FOTOAPARAT, TISKALNIK, RAČUNALNIK

Prikazana 5-stopenjska taksonomija vključuje semantične tipe, zastopane s samostalniki oz. samostalniškimi besednimi zvezami, kar velja za pomenskosladojenjske lastnosti glagolov, samostalnikov in pridevnikov, ne pa tudi za označevanje

udeležencev pri prislovnih pomenih, ki so navadno zastopani z glagoli ali pridevniki. Pri prislovnih pomenih je semantične tipe glagolov in pridevnikov mogoče razvrstiti v samostojne pomenske skupine, kot prikazuje Preglednica 20:

Preglednica 20: Pomenska tipologija glagolov in pridevnikov v vlogi semantičnih tipov pri prislovih

GLAGOLI USMERJENEGA MISEL-NEGA DELOVANJA	PRIČAKOVATI	DOLOČEN
	ODLOČITI	POVEZAN
	SPREJETI	
	DOSEČI	
	ODLOČATI	
	VREDNOTITI	
	PRESOJATI	
GLAGOLI IZRAŽANJA, GOVORJENJA	IZRAŽATI SE	POVEDAN
	IZRAZITI SE	IZGOVORJEN
	GOVORITI	
	POVEDATI	
	REČI	
	DOGOVARJATI SE	
	IMENOVATI	
načelno prislov		
če se človek ali organizacija načelno ODLOČA ali DOGOVARJA o neki ideji ali načrtu načelno, ju obravnava na splošno, brez podrobnosti		
GLAGOLI ČUSTVENE IN MISELNE DEJAVNOSTI	OBNAŠATI SE	NEZADOVOLJEN
	ODZVATI SE	
	REAGIRATI	
	MISLITI	
	ČUTITI	
intimno prislov		
če človek intimno nekaj MISLI ali ČUTI, tega ne deli z drugimi ali če se nekaj DOGAJA intimno, se dogaja v majhnem, zasebnem krogu ljudi		
GLAGOLI TELESNE DRŽE	DRŽATI SE	
	STATI	
pokončno prislov		
če SE človek DRŽI pokončno, ima hrbtenico poravnano in glavo vzravnano		

GLAGOLI DOGAJANJA IN POTEKA	POTEKATI	
	POJAVLJATI SE	
	ZGODITI SE	
	DOGAJATI SE	
	TRAJATI	
avtomatično prislov		
če se nekaj ZGODI avtomatično, izhaja iz predhodnega stanja ali vnaprej določenih pravil		
GLAGOLI DELOVANJA	DELATI	IZDELAN
	NAREDITI	OBLIKOVAN
	POČETI	UREJEN
	OPRAVITI	OBDELAN
	OPRAVLJATI	
	STORITI	
	OBDELOVATI	
	IZVESTI	
	DELOVATI	
	IZPOLNJEVATI	
	UREDITI	
	RAZPOREDITI	
lično prislov		
če je predmet lično IZDELAN ali OBLIKOVAN, je narejen z občutkom za lepoto in urejenost		
GLAGOLI SPREMINJANJA	POVEČATI SE	DRUGAČEN
	SPREMENITI	
čezmerno prislov		
če SE nekaj čezmerno POVEČA ali SPREMENI, preseže normalno, priporočljivo ali zdravo mejo		
GLAGOLI PREMIKANJA	PREMAKNITI	
	PREMIKATI	
nerodno prislov		
človek SE nerodno PREMIKA ali OBNAŠA, kadar je v zadregi		
GLAGOLI ZAPISOVANJA	ZAPISATI	ZAPISAN
	BELEŽITI	
izrecno prislov		
če je kaj izrecno POVEDANO ali ZAPISANO, je izraženo na zelo jasen in neposreden način		

V posameznih primerih so semantični tipi glagolov in pridevnikov lahko zastopani tudi s konkretnjšimi glagoli ali pridevniki, kot spodaj za pomen prislova *tešče*. V tem smislu ustrezajo nivoju kolokatorskih tipov ali kolokatorjev, kot smo jih predstavili zgoraj:

tešče *prislov*

pred uživanjem hrane

če človek na tešče ZAUŽIJE hrano, pijačo ali zdravilo, to naredi še preden je tisti dan kaj pojedel ali spil

10.3.1.2.3 Zapis semantičnih tipov

Poleg pomenskih in skladijskih lastnosti semantičnih tipov je bilo pri oblikovanju pomenskih shem potrebno določiti tudi enotne strategije pri njihovem zapisu z velikimi črkami, saj je na ta način omogočena njihova avtomatska prepoznavnost. V nadaljevanju se posvetimo posameznim izraznim možnostim in problematičnim konstrukcijam, ki izhajajo predvsem iz jezikovnosistemskih lastnosti slovenskega jezika.

V primerih, kjer je za ustrezen opis pomena besede v pomenski shemi potrebno udeležence podrobneje določiti, smo predvideli naslednje možnosti:

(a) konkretizacija semantičnega tipa s stavčnim prilastkom

pozdraviti *glagol*

1 izreči pozdrav

1.1 imeti kratek govor

če ČLOVEK, ki je uradni predstavnik neke ustanove, pozdravi SKUPINO LJUDI na srečanju ali prireditvi, ima v imenu te ustanove kratek pozdravni govor

(b) konkretizacija semantičnega tipa s predložno zvezo

odpustiti *glagol*

1 povzročiti, da kdo izgubi službo

če USTANOVA ali ODGOVORNI v ustanovi odpusti ZAPOSLENEGA iz službe, povzroči, da ostane brez zaposlitve in ne prejema več plačila

2 dati dovoljenje za odhod iz ustanove

2.1 o vojaški službi

če ODGOVORNI odpusti ČLOVEKA na služenju vojaškega roka z VOJAŠKEGA USPO-SABLJANJA, uradno dovoli, da služenje prekine, navadno zaradi zdravstvenih razlogov

10.3.1.2.3.1 Večbesedni semantični tipi

Pri zapisovanju večbesednih semantičnih tipov smo upoštevali načelno pravilo, po katerem smo z velikimi črkami zapisali jedrni del besedne zveze v vlogi semantičnega tipa oz. celotno besedno zvezo, če se je pokazalo, da je za pomen odločujoča kot celota.

(a) Zapis levega prilastka

V večbesednem semantičnem tipu smo levi prilastek zapisovali z velikimi črkami samo v primeru pomenske relevantnosti (*ožigosan*, *blanširati*). V primeru, da je bila razširitev semantičnega tipa z levim prilastkom namenjena zgolj berljivosti pomenske sheme, se za zapis z velikimi črkami nismo odločali (*mečkati*):

blanširati *glagol*

preliti z vrelo vodo; popariti

če ČLOVEK blanšira SADJE ali ZELENJAVO, ga za kratek čas potopi ali prelije z VRELO VODO

ožigosan *pridevnik*

1 negativno opredeljen

ČLOVEK je ožigosan, če ga drug ČLOVEK, SKUPINA LJUDI ali OKOLJE, v katerem živi, proglasi za nosilca NEGATIVNIH LASTNOSTI, navadno neupravičeno, zaradi osebnih ali političnih interesov

mečkati *glagol*

2 božati in objemati

če se ČLOVEK mečka z drugim ČLOVEKOM, se z njim objema, boža in poljublja

(b) Zapis desnega prilastka

Desni prilastek smo zapisovali z velikimi črkami, če predstavlja pomensko relevantni del semantičnega tipa (*boleti*, *doživeti*) in vedno takrat, kadar je jedro zaimek (*obetati*). Desni prilastek v obliki predložne zveze smo, kot rečeno, zapisovali z malimi črkami (*kriti*):

boleti *glagol*

1 o telesni bolečini

1.1 čutiti bolečino

če ČLOVEKA ali ŽIVAL boli DEL TELESA, čuti v njem bolečino

doživeti *glagol*

3 dočakati določeno starost

če ČLOVEK doživi določeno ŠTEVILO LET ali nek DOGODEK, potem živi toliko let in dočaka ta dogodek

obetati *glagol*

2.1 pričakovati; želeli si

če si ČLOVEK od ČESA obeta KAJ DOBREGA, zlasti USPEH ali DOBIČEK, pričakuje, da se bodo njegova pričakovanja uresničila

kriti *glagol*

2 izpolnjevati potrebe

2.1 po surovinah, energiji

če PODJETJE ali USTANOVA krije POTREBE po energiji, surovinah ali storitvah, jih zagotavlja v zadostni meri glede na povpraševanje

(c) Zapis stalne zveze

Stalne besedne zveze v vlogi semantičnih tipov smo zapisovali z velikimi črkami:

odplakniti *glagol*

1 odstraniti z vodo

če VODA ali NARAVNI POJAV odplakne LJUDI, PREDMETE, ZEMLJO ali NASELJA, jih odnese stran ali spere s površin

10.3.1.2.3.2 Ujemanje

Pri zapisovanju udeleženskih mest s semantičnimi tipi naletimo na problem ujemanja med zloženim osebkom in povedkom ter na problem navezovanja med zloženo nanašalnico¹¹⁸ in anaforičnim zaimkom. Z vidika zapisa semantičnih tipov v pomenski shemi je bilo treba posebej obravnavati skladišne zgradbe, ki zahtevajo:

- (1) izbiro ustreznega števila pri glagolu v povedku, kadar se mora ta ujemati s priredno ločno sestavljenim OSEBKOM,¹¹⁹ npr.

če IDEOLOGIJA ali PREPRIČANJA **razdvaja/razdvajajo** ljudi, jih **deli/delijo** na dve ali več ideološko nasprotujočih si skupin

¹¹⁸ Nanašalnico (prim. Korošec 1985: 14) razumemo v smislu izhodišča, zastopanega s samostalniško besedno zvezo, na katero se nanaša anafora.

¹¹⁹ V prikazanih primerih uporabljamo zapis z velikimi črkami samo pri tistih delih pomenske sheme, ki ponazarjajo nanašalnico, ki je predmet stavčnega ali anaforičnega ujemanja.

- (2) določitev oblike **pridevnika** v prilastku (spol in število) v priredno ločno zloženem OSEBKU ter oblike **veznega glagola** in **pridevnika** v povedkovem določilu npr.

mršav/mršava/mršavi ČLOVEK, DEL TELESA ali **ŽIVAL je/so**
zelo **suh/suha/suhi** in nezdravega videza

(v obeh primerih gre za stavčno ujemanje), in

- (3) izbiro spola in števila **anaforičnega zaimka**, kadar se mora ta ujemati z ločno priredno zloženo NANAŠALNICO (tj. anaforično ujemanje), npr.

če človek ali žival kolovrati po nekem PODROČJU ali STAVBI, se po **njem/njej/njima** premika, pogosto na nenavaden način in brez določenega namena.

Za povezovanje priredno zloženih delov v zapisu semantičnih tipov, ki kažejo na možnost izbire in se hkrati vsebinsko ne izključujejo, smo uporabljali veznik *ali*,¹²⁰ saj omogoča izbiro in s tem tudi možnost odločitve le za enega od elementov zloženega semantičnega tipa, po katerem naj bi se zgledovala (tj. se z njim ujemala) tudi oblika glagola in pridevnika v povedkovem določilu. V primeru številsko različnega večdelnega zapisa semantičnega tipa, kot v: ČLOVEK, USTANOVA ali LJUDJE, bi torej pričakovali ujemanje povedka samo z enim delom, ne pa z obema/vsemi, npr. če se ČLOVEK, USTANOVA ali LJUDJE loti problema, ga začne reševati. Taka možnost pa ni povsem skladna s predlagano v Slovenski slovnici. Kot navaja Toporišič (2000: 609), zahtevata ločno in priredno zloženi osebki needninski povedek, npr. *Tja bova šla ali ti ali jaz*.¹²¹ Med izjemami za priredno zloženi osebki, kjer pa ni primerov z veznikom *ali*, Toporišič (2000: 610) navaja možnost ravnanja povedka po tistem delu zloženega osebka, ki je povedku najbližji: *Na mizi je bila pogača in mleko* (poleg običajnega: *Na mizi sta bila pogača in mleko*). S tem soglaša tudi večina raziskav tega problema (Bajec 1955/56: 13; Marušič et al. 2007; Toporišič 2000: 610), kjer pa se kot pomembni dejavniki pri odločanju o tem, s katerim od elementov zloženega osebka naj se ujema povedek, poleg bližine in položaja pred ali za povedkom navajajo še nevtralnost moškega spola, človeškost, množinskost samostalnikov, njihova (ne)števnost, pomenske lastnosti ipd.

Ker so pomenske sheme izhodišča za stavčne razlage in je zato njihova berljivost ter razumljivost za uporabnika pomembna, smo želeli različne možnosti realizacij problematičnih zgradb ob upoštevanju specifičnosti veznika *ali* preveriti

¹²⁰ Ne gre torej za zveze tipa: *vse ali nič*, *denar ali življenje* (Toporišič 2000: 605).

¹²¹ Za razliko od izključevalnega: *Tja boš šel ali ti ali jaz* (Toporišič 2000: 609).

pri potencialnih uporabnikih. V ta namen je bila izdelana anketa (prim. Polončič Ruparčič 2012), ki je vključevala več primerov znotraj posameznega problematičnega tipa. Analiza rezultatov, ki jo na kratko povzamemo v nadaljevanju, je bila izhodišče za reševanje te problematike pri oblikovanju pomenskih shem in stavčnih razlag v LBS.

10.3.1.2.3.2.1 Stavčno ujemanje

Problem stavčnega ujemanja se je pri oblikovanju pomenskih shem pokazal predvsem pri dveh tipih skladenjskih zgradb: (1) pri zapisu pomenskih shem glagolskih pomenov s priredno ločno sestavljenim osebkom, ki zahteva ustrezno število glagola v povedku, in (2) pri zapisu pomenskih shem pridevniških pomenov, kjer je bilo treba določiti spol in število prilastka v priredno ločno zloženem osebku ter število veznega glagola in število ter spol pridevnika v povedkovem določilu, ki se z osebkom ujema.

10.3.1.2.3.2.1.1 Ujemanje med zloženim osebkom in povedkom

Izhajajoč iz zgradbe pomenskih shem v obliki *če*-stavka, obstaja, kot je razvidno iz spodnjih primerov, nekaj podtipov, ki smo jih obravnavali v ločenih skupinah. V prvi skupini so primeri, kjer se morata z zloženim osebkom ujemati oba glagola, in sicer v glavnem in odvisnem stavku:

- če se USTANOVA ali SKUPINA LJUDI **loti** problema, ga **začne** postopoma in premišljeno reševati
- če IDEOLOGIJA ali PREPRIČANJA **razdvajajo** ljudi, jih **delijo** na dve ali več ideološko nasprotnojučih si skupin.

V drugi skupini so primeri, kjer se ugotavlja ustrezna oblika veznega glagola in pridevnika v povedkovem določilu glede na zloženi osebek, npr.

- če **je** ČLOVEK, RASTLINA ali PREDMET **neprilagojen** razmeram v okolju, **nima** za bivanje ali delovanje v njem ustreznih lastnosti ali pogojev
- če **je** STORITEV ali IZDELEK cenovno **opredeljen**, je **določen** glede na višino cene
- če **je** TELO ali **so** MOŽGANI **omamljeni** od substance, je **njihovo** delovanje spremenjeno.

V tretji skupini pa so primeri ujemanja povedka z zloženim osebkom v glavnem stavku in z anaforičnim zaimkom v odvisnem:

- če ČLOVEK, KRAJ ali UMETNIŠKO DELO **uroči** človeka, ga **njegova** lepota povsem prevzame, tako da ni sposoben jasno razmišljati
- če ČLOVEK, SKUPINA LJUDI ali STVAR **doživi**, da **nanjo** nepričakovano vpliva nek dogodek, ima to lahko negativne posledice.

V odgovorih, ki so jih navedli anketiranci,¹²² konkurirata predvsem množinska in edninska oblika glagola, v primeru dvodelnega semantičnega tipa pa prihaja tudi do seštevanja, torej dvojine: če ČLOVEK ali ŽIVAL sedeta, se namestita ... Izbira množinske oblike prevladuje predvsem takrat, ko je semantični tip več-delen, če je kateri njegov del v množini ali se dojema kot množina (npr. SKUPINA LJUDI so namesto je). Pri oblikovanju pomenskih shem smo se odločali za množinsko obliko glagola, pridevnika v povedkovem določilu in anaforičnega zaimka le v primerih, ko je bil kateri od delov zloženega osebka v množini, sicer smo se pri ujemanju nanašali na njegov zadnji, povedku najbližji del, pri čemer smo množinski del zloženega osebka, če je bilo mogoče, navedli na koncu niza: če IDEOLOGIJA ali PREPRIČANJA razdvajajo ljudi, jih delijo ...; če je TELO ali so MOŽGANI omamljeni ..., je njihovo delovanje ... Enako smo se odločali tudi pri pridevniških pomenih, kjer smo spol pridevnika v povedkovem določilu prilagodili najbližjemu elementu zloženega osebka: če je ČLOVEK, RASTLINA ali PREDMET neprilagojen ...; če je IZDELEK ali STORITEV cenovno opredeljena, je določena ...

10.3.1.2.3.2.1.2 Ujemanje med prilastkom zloženega osebka in pridevnikom v povedkovem določilu

Zgradbe, v katerih se je bilo potrebno odločiti za ustrezno obliko prilastka v zloženem osebku in ustrezno obliko veznega glagola in pridevnika v povedkovem določilu, so se pokazale kot precej trd oreh, posledično pa so bile tudi odločitve anketirancev razmeroma neenotne. Pri ujemanju prilastka z jedrom zloženega osebka je bila izkazana težnja v prevladi prvega, tj. prilastku najbližjega samostalnika, npr. mršav ČLOVEK, DEL TELESA ali ŽIVAL, otečene OČI ali GRLO, kar je mogoče pripisati bližini samostalnika, saj je to merilo prevladujoče tudi pri ujemanju med zloženim osebkom in povedkom,¹²³ verjetno pa tudi pomenski vlogi veznika *ali*, kjer je druge samostalniške dele mogoče razumeti kot možnost izbire oz. zamenjavo, o čemer smo že pisali.

122 Na anketo je ob izločitvi neveljevanih odgovorov odgovorilo 98 ljudi, ki so bili razdeljeni v dve skupini: na jezikoslovce (71,4 %), in sicer na slovenske lektorje na tujih univerzah (18,3 %), druge sloveniste (25,5 %) in jezikoslovce drugih smeri (27,5 %). V drugi skupini so bili nejezikoslovci (28,5 %) z najmanj visokošolsko izobrazbo. Razdelitev je bila narejena glede na predvideno poznavanje jezikovne norme pri zapisovanju ujemanja.

123 Tako predlaga tudi spletna svetovalnica ŠUSS (dostopno na: <http://www2.arnes.si/~lmarus/suss/arhiv/suss-arhiv-000417.html>): »Zdi se, da je princip ujemanja v obeh primerih /tj. v primeru ujemanja glagola s sestavljenim osebkom in v primeru pridevniškega ujemanja z jedri zloženega osebka/ primerljiv. Gre za to, da se najbolj sprejemljivo vedno slišijo primeri, kjer se pridevnik ujema z najbližjim samostalnikom.«

Če so si bili vprašani pri tem delu skladenjskih zgradbe razmeroma enotni, si v drugem delu, kjer je bilo treba izbrati ustrezno obliko veznega glagola glede na zloženi povedek in ustrezno obliko pridevnika v povedkovem določilu, konkurirata dve merili, in sicer merilo seštevanja samostalniških elementov oz. merilo števila:

mršav ČLOVEK, DEL TELESA ali ŽIVAL **so suhi**
otečene OČI ali GRLO **so obarvani/obarvane**
ožigosana POŠILJKA ali DOKUMENT **sta opremljena**
čeden MOŠKI ali ŽENSKA **sta privlačna**

in merilo prvega samostalnika v osebku:

mršav ČLOVEK, DEL TELESA ali ŽIVAL **je suh**
ožigosana POŠILJKA ali DOKUMENT **je opremljena**
čeden MOŠKI ali ŽENSKA **je privlačen**

V LBS smo na podlagi analiz in čim večje enotnosti pri vseh odločitvah, ki se tičejo problematike skladenjskega ujemanja, prilastek zloženega osebka načeloma prilagajali prvemu samostalniku, torej: mršav ČLOVEK, DEL TELESA ali ŽIVAL; ožigosana POŠILJKA ali DOKUMENT itd., medtem ko smo se pri izbiri oblike veznega glagola in pridevnika (spol in število) v povedkovem določilu odločali za seštevanje oz. množino:

mršav ČLOVEK, DEL TELESA ali ŽIVAL **so suhi**
ožigosan DOKUMENT ali POŠILJKA **sta opremljena**

Hkrati smo pri posameznih ubeseditvah skušali prilagoditi semantične tipe skladenjski zgradbi, ki je omogočala sprejemljivejši zapis ujemanja. Tako smo npr. množinsko obliko samostalnika spremenili v edninsko: otečeno OKO ali GRLO, pri spolsko različnem osebku tipa MOŠKI ali ŽENSKA pa smo navadno uporabili dvojino: čedna ŽENSKA ali MOŠKI sta privlačna.

Pri ujemanju znotraj številsko in/ali spolsko različnega osebka smo v primeru, da je bila pomenska shema na ta način bolj berljiva, iztočnico skupaj s semantičnim tipom, ki zastopa v spolu in/ali številu različno samostalniško zvezo, ponovili, in sicer, če je ponovitev odpravila problem ujemanja med pridevnikom in samostalnikoma različnih slovničnih spolov ali števil (*ne-premostljivo*) oz. če so različni semantični tipi ustvarjali pomenske odtenke, ki so zahtevali uporabo različnih glagolov v razlagi (*aktiven*):

- **nepremostljiv KONFLIKT** ali **nepremostljive RAZLIKE nastopijo**, ko nasprotni strani ne najdeta soglasja ali skupne rešitve
- **aktiven ČLOVEK izvaja** ali **aktivna DEJAVNOST vključuje** več različnih dejavnosti

10.3.1.2.3.2.2 Anaforično ujemanje

Pri anaforičnem ujemanju, ki poteka med večjdrno nanašalnico in anaforičnim zaimkom (t. i. naveznikom), je smiselno najprej ločevati med dvo- in večdelno nanašalnico, saj se je pokazalo, da so v prvem primeru odgovori anketirancev bolj heterogeni kot v primeru več kot dvodelnih nanašalnic.

10.3.1.2.3.2.2.1 Dvodelna nanašalnica

sam. m. sp. edn. + sam. sr. sp. edn. = sam. m. sp. edn.

Zgradbe z dvodelnimi nanašalnici, ki so sestavljene iz samostalnikov moškega in srednjega spola v ednini, z vidika izbire ustrezne oblike anaforičnega zaimka niso problematične, saj oblika za moški spol zastopa tudi srednji spol (*ga*). Ker sta oba samostalnika v ednini, tudi ni potrebe po množinski obliki anafore (*jih*), vprašanje je le, ali se jedra seštevajo (*ju*), kot to denimo velja za ujemanje povedka z osebkom v: *Borut in Meta pridno študirata* (Toporišič 2000: 609). Anketiranci so se v večini primerov odločali za prevlado zaimka v moškem spolu in za ednino:

- če človek dokumentira DOGODEK ali DOGAJANJE, **ga** opiše ali posname z namenom, da za- beleži in ohrani bistvene podatke

sam. m. sp. edn. + sam. ž. sp. edn. = sam. najbližji povedku, po možnosti m. sp.

Kadar nanašalnico sestavljata samostalnik moškega in samostalnik ženskega spola in sta oba v ednini, so bili anketiranci v svojih odgovorih bolj negotovi, hkrati pa posamezne odločitve številčno niso posebej izstopale. Analiza je v splošnem pokazala, da so anketiranci pri obliki anaforičnega zaimka upoštevali najprej število, tj. needninsko obliko anafore (ČLOVEKU ali USTANOVI -> jima), nato moški spol (ČLOVEKA ali ŽIVAL -> ga) in nazadnje povedku najbližje jedro (DO-KUMENT ali FOTOGRAFIJO -> jo). Čeprav odločitev za needninsko obliko najboljše ustreza stanju v slovnici (Toporišič 2000: 609), se v LBS pri tovrstnih zgradbah nismo odločali zanjo predvsem zaradi že omenjene izbirne pomenske vloge veznika *ali*, pač pa smo se v večini primerov odločali za navezavo na samo-stalnik, ki je bil glagolu najbližji in – kjer to ni vplivalo na berljivost sheme kot celote – poskrbeli, da je bil na tem mestu samostalnik moškega spola:

- če predstavnik ustanove zažuga USTANOVI ali drugemu ČLOVEKU, **mu** zagrozi z besedami ali dejanji
- če človek skenira FOTOGRAFIJO ali DOKUMENT, **ga** s posebno napravo pretvori v digitalno obliko in shrani na računalniku
- če človek cuka ŽIVAL ali drugega ČLOVEKA, **ga** večkrat zapored na kratko povleče za del telesa ali oblačila

sam. sr. sp. edn. + sam. ž. sp. edn. = sam. najbližji povedku, po možnosti sr. sp.

Pri nanašalnici, sestavljeni iz samostalnika srednjega in samostalnika ženskega spola v ednini, merila moškega spola samo na prvi pogled ni mogoče upoštevati, dejansko pa ta prevladuje, saj je oblika anafore za moški in srednji spol enaka. Kljub temu analiza izbir, ki so jih opravili anketiranci, kaže prevlado kategorije spola šele na tretjem mestu, medtem ko so se najpogosteje odločali za prevlado kategorije števila (STALIŠČE ali TRDITEV -> ju) in bližine glagola v povedku (po PODROČJU ali STAVBI -> njej). Pri oblikovanju pomenskih shem smo se tudi v primeru kombinacije samostalnika srednjega in ženskega spola v ednini odločali podobno kot v prejšnjem primeru: anaforični zaimek smo navezali na samostalnik, ki je bil najbližji povedku in težili k temu, da je bil to samostalnik srednjega spola:

- če človek argumentira TRDITEV ali STALIŠČE, **ga** utemelji z razlago
- če človek improvizira DEJAVNOST ali OPRAVILO, **ga** opravlja brez vnaprejšnje priprave
- če človek kolovrati po STAVBI ali nekem PODROČJU, se po **njem** premika brez določenega namena

sam. m. sp./sr. sp./ž. sp. edn. + sam. m. sp./sr. sp./ž. sp. mn. = sam. m. sp. mn., po možnosti najbližji povedku

Kadar nanašalnico sestavljajo samostalniki različnih spolov, od katerih je eden v množini, so anketiranci preferirali izbiro množinske oblike anaforičnega zaimka. V LBS smo se prav tako odločili za prevlado needninske oblike, težili pa smo tudi k temu, da je bil samostalnik v množini najbližji povedku, če je to omogočala berljivost sheme kot celote:

- če človek arhivira zvočno, pisno ali likovno GRADIVO ali DOKUMENTE, **jih** shrani v prostor ali mapo
- če človek češe DLAKO ali LASE, **jih** ureja z glavnikom ali krtačo
- če človek uživa HRANO ali HRANILNE SNOVI, se z **njimi** prehranjuje

10.3.1.2.3.2.2 Več kot dvodelna nanašalnica

Kadar je večdelna nanašalnica sestavljena iz spolsko različnih elementov v ednini ali množini, je anaforični zaimek največkrat množinski, ki je tudi glede na spol nevtralen, torej moškega spola. Druga, vendar redkejša alternativa, je nanašanje na edninski samostalnik moškega oz. srednjega spola, še posebej, če je ta hkrati tudi najbližji glagolu v povedku. V LBS smo se v skladu s predhodnimi rešitvami

odločali za drugo možnost, tj. za edninsko obliko anaforičnega zaimka v moškem spolu, samostalniki moškega ali srednjega spola pa smo navedli na koncu niza, tako da je bil povedku najbližji. Če je bil tak, povedku najbližji samostalnik moškega spola v množini, smo temu prilagodili tudi anaforični zaimke (*absorbirati*):

- če človek hvali drugega ČLOVEKA, USTANOVO ali DELO, poudarja **njegove** dobre lastnosti, sposobnosti ali dosežke
- če človek skenira drugega ČLOVEKA, OKOLICO ali PREDMET, **ga** pozorno opazuje, da bi videl vse podrobnosti
- če človek olupí SADJE, ZELENJAVO ali vrhnji DEL ŽIVILA, **mu** odstrani lupino ali vrhnjo plast
- če človek popopra DOGODEK, IZJAVO, ali VZDUŠJE, **ga** s pripombo ali z dejanjem naredi bolj zanimivega in pikrega
- če snov absorbira ENERGIJO, SVETLOBO ali DELCE, **jih** vsrka vase.

10.3.1.3 Pomenski scenarij

Tako imenovani pomenski scenarij predstavlja poleg števila in razporeditve udeležencev v stavčnem vzorcu ter njihovih leksikalnih zapolnitev s semantičnimi tipi tretji korak v oblikovanju pomenske sheme in je primarno namenjen opisu pomena, torej njegovi razlagi. Potem ko smo registrirali in v obliki semantičnih tipov navedli vse udeležence, ki jih pomenske lastnosti neke besede zahtevajo v sobesedilu stavčnega (pa tudi zunaj stavčnega) okvira, je bilo treba opisati še sporočanje situacije, razmerja med udeleženci, okoliščine, namen, posledice, pozitivno ali negativno vrednotenje, konotacijo in pragmatične lastnosti.¹²⁴ V scenarijskem delu pomenske sheme smo torej opisali scenarij situacije, potrebne za realizacijo pomena besede, s čimer smo pojem pomenskega scenarija tesno povezali s pojmom slovarske razlage. V tem smislu je pomenski scenarij najbolj razlagalni del pomenske sheme in hkrati najbolj pragmatični del na podlagi pomenske sheme izdelane pomenske razlage.

10.4 Slovarske razlage

V LBS ločimo dva tipa razlag, in sicer:

(a) Razlage pri tistih besednih pomenih, ki ne potrebujejo pomenskega opisa z vidika njegove skladenjske vključenosti v sobesedilo. Gre zlasti za samostalniške in nekatere pridevniške pomene, ki nimajo glagolskega izvora in svojega pomena ne izražajo s pomočjo sporočanje vloge. V teh primerih smo se odločali bodisi za

¹²⁴ Na tem mestu je potrebno omeniti še pojem *semantična prozodija* (ang. semantic prosody; prim. Šorli 2011, 2012 in tam navedeno literaturo), ki se v korpusnem jezikoslovju uporablja za tisti del pomena, ki presega raven besede pa tudi njenega kolokacijskega okolja, hkrati pa se jasno loči tudi od konotativnega pomena. Kot taka naj bi pomenska prozodija izražala osebni odnos govorca ali pisca oz. njegovo vrednotenjsko in emotivno držo glede posameznega konteksta ali scenarija pa tudi glede (pričakovane ali dejanskega) rezultata, in sicer na diskurzni in ne zgolj ožje sobesedilni ravni.

opis pomena samo s pomenskim indikatorjem (*fen*) bodisi za t. i. klasično slovarsko razlago z uvrščevalno pomensko sestavino kot nadpomenko in zadostnim številom razločevalnih pomenskih sestavin znotraj nje, da je bil pomen besede ustrezno diferenciran glede na druge pomene npr. pomensko oz. tematsko sorodnih besed (*beluš*, *ingver*). Ta tip razlage smo v strukturi LBS zapisovali v element <definicija>.

fen *samostalnik*

1 sušilec za lase

2 topel in suh veter

beluš *samostalnik*

1 rastlina; špargelj

podolgovata zelena ali bela rastlina, katere steblo in vršički se uporabljajo kot zelenjava pri pripravi jedi in specialitet

ingver *samostalnik*

1 rastlina

čvrsta, svetlorjava korenina istoimenske rastline, ki se uporablja kot začimba ali živilo

(b) Drugi tip razlage je v LBS vključen v element <pomenska_shema>, pri čemer naj bi končne slovarske razlage na podlagi pomenskih shem izdelali naknadno za namene konkretnega slovarskega priročnika. Razlaga v pomenski shemi je usmerjena v beleženje najširšega skladišnega vzorca, ki je značilen za konkretni pomen razlagane besede, stavčna razlaga pa v razlagalne strategije ali tipe, ki omogočajo čim uspešnejše podajanje vsebinskih poudarkov pomena besede, vključno z njegovimi pragmatičnimi elementi. V nekaterih primerih je to v strukturi LBS nakazano tako, da ima iztočnica pri katerem od pomenov izkazani obe možnosti, tj. pomensko shemo in razlago. V teh primerih je za naš namen razlaga zapisana v ležečem tisku, pomenske sheme pa so navedene v temneje obarvanem delu gesla:

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

10.4.1 Klasične oz. analitične slovarske razlage

Glede na to da je informacija o tem, kaj določena beseda pomeni, osnovna informacija, ki jo pričakujemo v slovarju, je razumljivo, da je v leksikografski teoriji in praksi problemom pomenskega opisa in razlagalnemu jeziku namenjeno veliko pozornosti. V teoriji in praksi enojezične leksikografije sta se ob sicer številnih tipih in variantah (prim. Kosem 2006: 26) vzpostavila dva temeljna tipa razlaganja besednega pomena: klasična slovarska tudi aristotelovska razlaga in stavčna razlaga, ki ima več podtipov (prim. Lew in Dziemianko 2006).

Slovenska leksikografska teorija (prim. Vidovič Muha 2000) in praksa (SSKJ; SNB) je pri oblikovanju pomenskih razlag sledila tradiciji aristotelovske slovarske razlage, imenovane tudi klasična leksikografska ali analitična razlaga (Hanks 2005; Dziemianko in Lew 2013: 155). Gre za tip razlage, ki jo najpogosteje najdemo v splošnih enojezičnih slovarjih in je sestavljena iz dveh delov: nadpomenke (*genus proximum*) in ene ali več razločevalnih lastnosti (*differentia specifica*). Tak tip razlage, za katerega je značilno dosledno upoštevanje besednovrstnega ujemanja razlagane besede z razlagalno nadpomenko, kar v besedilu omogoča zamenljivost razlagane in razlagalne besede, je najbolj uporaben pri definiranju konkretnih samostalnikov in pri večini glagolskih pomenskih skupin (Atkins in Rundell 2008: 415). Konkretni samostalniki so namreč v mentalnem leksikonu organizirani kot leksikalne kategorije v medsebojnih hierarhičnih razmerjih nad- in podpomenskosti. Analitični tip razlage je na drugi strani veliko težje uporabiti za razlaganje nekaterih pomenskih skupin glagolov, pridevnikov, prislovov ter funkcijskih besed (Atkins and Rundell 2008: 415). Tak tip razlage je, kot rečeno, uporabljen tudi v SSKJ in ga je mogoče ponazoriti s spodnjim primerom:

spominjati -am nedov. 1. povzročati, da se komu kaj iz preteklosti znova pojavi v zavesti
genus proximum differentia specifica

V LBS smo klasične slovarske razlage uporabljali pri pomenih, kjer razpoznavanje udeležencev in njihovih semantičnih tipov ni bilo ključno za razumevanje pomena. V nasprotju s strukturalnim pomenoslovjem, kot se realizira v SSKJ, ki pri oblikovanju razlag izhaja iz skladenjske oz. stavčnoočlenske vloge besed, kjer imajo leksemi s slovarskim (pomenskosestavinskim) denotatom, npr. samostalniki, glagoli in nekateri prislovi, analitično slovarsko razlago, leksemi z denotatom govornega dejanja (členki, medmeti, krajevni in časovni prislovi ter vrstni pridevniki) pa stavčno razlago tipa »izraža nekaj«, npr. *baje* – izraža negotovost trditve (povzeto po SSKJ), smo v LBS izhajali iz teorije, po kateri se pomeni besed ustvarjajo v sobesedilu stavčne povedi, v kateri govorci določeno besedo

tipično rabimo. Pritegnitev sporočanjkega merila pri odločanju o tipu razlage nam je omogočila izbiro primerne tipa razlage glede na pomen in sporočanjko vlogo besede in ne primarno glede na njeno skladiščno vlogo. Opisi pomena, ki ne vsebujejo obveznih elementov pomenske sheme, kar pomeni, da iztočnica ni nujno sestavni del pomenskega opisa in da udeleženci niso izpostavljeni v obliki semantičnih tipov, so v LBS vezani predvsem na samostalniške pomene in pomene stalnih zvez, na primer:

imigrant *samostalnik*

priseljenc

kdor pride v drugo državo in se tam za stalno ali za daljši čas naseli

nacionalist *samostalnik*

1 privrženec nacionalizma

kdor zagovarja prepričanje, da je narod, ki mu pripada, več vreden od drugih narodov

aluminijasta folija

aluminij v obliki tankih lističev, navitih na rolo, namenjen uporabi v gospodinjstvu, zlasti za zavijanje in shranjevanje živil

bolniški list

zdravniško potrdilo o odsotnosti z dela zaradi bolezni ali nege bolnika

Ena od tez strukturalnega pomenoslovja je, da govorce pomensko konceptualiziramo svet na enak način, kot so zgrajene klasične slovarske razlage: »Pri poskusu določitve denotativnega pomena izhajamo torej iz slovarsko strukturirane leksike, ki je v veliki meri odslíkava dejanske predmetnostne strukturiranosti, gibajoče se predvsem znotraj splošnega pojmovnega sveta« (Vidovič Muha 2000: 51). Na drugi strani raziskave procesa pomenskega razdvoumljanja besed v realnih sporočanjkih situacijah (Sinclair in Coulthard 1975) ugotavljajo, da postopka strukturiranja pomena, ki je sicer povsem legitimen na ravni pomenske strukturalne analize, ni mogoče povezati s konceptualiziranjem sveta in tem, kako ljudje dejansko razbiramo pomene zlasti večpomenskih besed. Zato sta tudi postopek pomenskega dekodiranja v vsakdanjem jezikovnem sporočanju in postopek razdvoumljanja pomenov v slovarjih med seboj neprimerljiva (Atkins in Rundell 2008: 263). Osnovno izhodišče, ki ga upoštevamo pri pomenskem členjenju v LBS torej je, da obstaja bistvena razlika med slovarskimi pomeni (tj. oštevilčenimi pomeni, na katere so navadno razdeljene iztočnice v slovarjih) in pomeni, kot jih dojemamo govorce v vsakdanjem sporočanju. Osredotočenost na pomensko

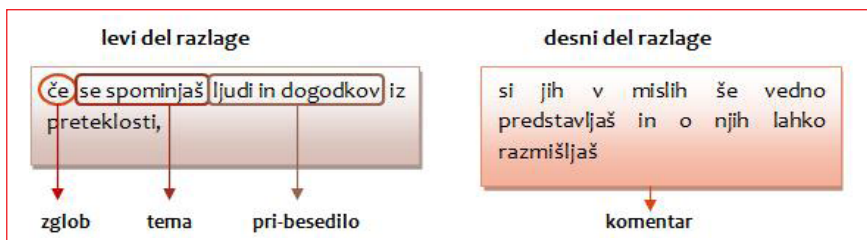
in sporočanjso izhodišče pri opisu besednega pomena v slovarju izhaja iz spoznanja, da se govorci pomenov besed učimo iz kontekstov, v katerih jih tipično slišimo, kjer pa ne gre samo za obstoj konkretne predmetnosti in razpoznavanje razmerij znotraj nje, ampak za prisotnost celotnega jezikovnega in zunajjezikovnega konteksta, v katerem se beseda razume in uporablja tako, kot se. Iz tega razloga in že omenjene potrebe po možnosti računalniškega procesiranja skladenjske realizacije pomena, smo želeli formulo stavčne razlage tipa *če*-stavek preizkusiti tudi na slovenščini.

10.4.2 Stavčne razlage

Za skladenjski zapis pomena, ki ga v LBS uvajamo zaradi možnosti strojnega procesiranja, je najprimernejša formula za oblikovanje razlag stavek. Kot smo prikazali zgoraj, pa stavčna oblika razlage, vključno s svojimi podtipi, ni enako primerna za vse skladenjskofunkcijske in pomenske razrede besed, zato smo v LBS stavčne razlage uporabljali predvsem pri glagolih oz. tistih pomenih besed, besednih zvez in frazeoloških enot, ki vključujejo udeležence in zahtevajo pojasnitev razmerij med udeleženci, njihov odnos do vsebine, okoliščin ipd. Naš namen je bil torej z razlago pojasniti predvsem sporočanjso vrednost besede v sobesedilu, manj pa – to informacijo smo namreč prestavili na druge segmente zgradbe LBS – na njeno slovnično, skladenjsko in skladenjskofunkcijsko vlogo.

Stavčna razlaga, ki jo je uveljavil leksikografski projekt Collins Cobuild (CCED; Sinclair 1987; Hanks 1987), je bila uporabljena predvsem v slovarjih, namenjenih učenju angleščine kot tujega/drugega jezika. Zanj v splošnem velja, da je primerna za šolske uporabnike, ki slovarje uporabljajo kot pripomočke pri učenju in usvajanju jezika. Levi del stavčne razlage podaja informacijo o rabi iztočnice, desni pa o njenem pomenu. Običajno je desni del bolj ekspliciten kot levi, med seboj pa sta povezana s t. i. zglobom (ang. *hinge*), ki je največkrat glagol, v nekaterih primerih tudi veznik (Krek 2004: 5). Bistvo stavčnih razlag je v tem, da sistematično vključujejo podatke o (glagolski) vezljivosti oz. udeleženskih vlogah s tipičnimi zapolnitvami, ki temeljijo na skladenjskopomenskih in kolokacijskih podatkih iz besedilnih korpusov. Zgradba stavčne razlage (Slika 35), kot smo jo uporabljali v LBS, je za glagol *spominjati* izdelana po vzoru razlag iz CCED:¹²⁵

125 Spletna različica CCED je dostopna na: <http://www.collinsdictionary.com/dictionary/english-cobuild-learners>.



Slika 35: Zgradba stavčne razlage tipa če-stavek

Razmišljanja o tipu slovarskih razlag, ki bi bile primerne za uporabnike, ki se jezika učijo, so sprožile ugotovitve, da obstoječi enojezični slovarji za angleščino ne zadovoljujejo tovrstnih uporabnikov oz. da ti ne uporabljajo celote vseh podatkov, ki jih vključujejo enojezični slovarji (Rundell 1988: 127). Snovalci Cobuildovih slovarjev za tujce so se zato lotili analize spontanega definicijskega diskurza v šolskih situacijah (Sinclair in Coulthard 1975) in rezultate v obliki podslovnice definicijskega jezika (Hanks 1987; Sinclair 1991: 123-137; Barnbrook 2002) prenesli v slovarsko okolje. Raziskave slovarskih uporabnikov in analize učinkovitosti izrabe obstoječih slovarjev z vidika tipa razlag, njihove oblike, vsebine in razlagalnega jezika so od pojava stavčnih razlag naprej ugotavljale tako njene prednosti kot pomanjkljivosti.

Stavčne razlage (npr. v CSD in MED) so navadno daljše od sinonimnih in analitičnih, vendar uporabniku bolj prijazne: v večini primerov leksikografom omogočajo opis pomena s preprostejšimi in lažje razumljivimi besedami, primernimi za šolskega uporabnika in/ali učenca jezika (Atkins in Rundell 2008: 38). Konverzijski obrazec *če*-stavka v šolskih slovarjih (npr. v CSD) je namreč veliko bolj v soglasju z naravno zvenečimi opisi pomena, saj odgovarja na uporabnikovo vprašanje »Kaj ta beseda pomeni?«. Glede na krajše razlagalne ustreznike v tradicionalnih slovarjih vsebujejo stavčne razlage posledično tudi več podatkov, in sicer v obliki, ki je uporabniku poznana iz običajnega sporočanja (prav tam: 329). Stavčne razlage z navajanjem tipičnih udeležencev, ki so pomembni za razumevanje določene besede,¹²⁶ hkrati na naraven način vključujejo tudi številne pragmatične informacije, ki se nanašajo na govorčev odnos do povedanega (Rundell 2006: 332).

Na drugi strani temeljni očitke stavčnim razlagam izhaja iz dejstva, da so navadno do dvakrat daljše kot ekvivalenti v tradicionalnih slovarjih,¹²⁷ hkrati pa

¹²⁶ V zvezi s pomenskosestavinskimi razlagami v SSKJ tudi Vidovič Muha (1999: 11) ugotavlja, da se na ta način npr. pri glagolih izpuščajo univerzalne pomenske sestavine, kot je usmerjenost glagolskega dejanja, realizirana v glagolski vezljivosti.

¹²⁷ To je imelo posledice predvsem za knjižno izdajo, saj je v večini primerov pomenilo zmanjšan obseg iztočnic.

tovrstne razlage predstavljajo učečim težavo pri procesiranju in razumevanju, saj se predpostavlja, da imajo omejeno jezikovno znanje (Rundell 2006: 328). S tem povezan je problem kopičenja anaforičnih zaimkov, za slovenščino, kot smo pokazali, pa tudi ujemanja med zloženim osebkom in povedkom ter izbira ustrezne anafore ali predloga, če v razlagi trčita skupaj dva glagola z različno predložno vezavo:

blanširati *glagol*

preliti z vrelo vodo

če človek blanšira SADJE ali ZELENJAVO, ga za kratek čas prelije z ali potopi v vrelo vodo

Stavčne razlage namreč pogosto vsebujejo stavke, v katerih razmerja med razlagano besedo ali katerim od udeležencev ter nanašalnico niso vedno nedvoumno izpeljana. Čeprav je prednost stavčnih razlag v tem, da vključujejo tipično skladenjsko okolje razlagane besede, Rundell (prav tam: 332) ugotavlja, da ena stavčna razlaga pogosto ne more izraziti vseh slovničnih vzorcev, v katerih se beseda tipično pojavlja.¹²⁸ Tak način definiranja prinaša tudi nove konvencije, ki jih učeči jezika pogosto niso vajeni, zato obstaja možnost, da bodo ostale nekatere informacije uporabniku skrite.

V slovenski leksikografski teoriji se na eni strani kaže težnja po prenovitvi obstoječega načina oblikovanja pomenskih razlag z vključitvijo stavčnega načina razlaganja pomena, in sicer z argumentoma, da je potrebno narediti razlage uporabniku bolj razumljive, saj sta bili poleg nedoslednosti prav nerazumljivost in togost lastnosti, ki so ju uporabniki v SSKJ najbolj kritizirali (Müller 2009: 19), ter z vidika uvajanja metodologije, ki omogoča vključitev situacijskih pomenskih lastnosti, zlasti torej pragmatike (Müller 2009: 24, 105; Kržišnik 2009: 42, 55) in tipičnih udeležencev pri posameznih pomenih. Po drugi strani pa je očitna tudi težnja po ohranjanju analitičnega pristopa k razlaganju besednih pomenov, podprta z argumenti slovenske leksikografske tradicije in strukturalističnega, tj. skladenjskopomenskega izhodišča pri obravnavi leksike v slovarju (Vidović Muha 2009: 24).

T.i. nanašalnih in slovničnih razlag tipa glagolnik od, manjšalnica od, ženska oblika od in nanašajoč se na, kjer je informacija o pomenu prenesena na motivirajočo besedo, v pomenski shemi ali razlagi nismo uporabljali, ker smo skušali zajeti pomen tudi pri tistih besedah, za katere naj bi veljalo, da nimajo samostojnega leksikalnega

¹²⁸ Kar je eden od razlogov, da smo v LBS vključili tudi zapis stavčnih vzorcev, o čemer več v samostojnem poglavju.

pomena.¹²⁹ Nasprotno smo se osredotočali na pomensko razlago povsod, kjer je bilo to mogoče. Razlike v opisu glagolnika in pridevnika v LBS ter SSKJ najbolje prikazuje spodnja primerjava:

SSKJ

bučanje -a s (â) *glagolnik od bučati*: mogočno bučanje morja; šum in bučanje viharja / bučanje čebel / bučanje orgel

áfriški -a -o prid. (â) *nanašajoč se na Afriko*: afriška celina / afriške države / publ. afriški vrh sestaneč šefov afriških držav "zool. afriški slon

LBS

bučanje *samostalnik*

1 pojav ob delovanju narave ali naprav

1.1 bobneč zvok

afriški *pridevnik*

1 o afriški celini

2 ki izhaja iz Afrike

2.1 o ljudeh

2.2 o rastlinah in živalih

3 značilen za Afriko

Tako izhodišče pri opisu pomenov je pomembno, saj je mogoče predvidevati, da uporabnik v slovarju išče predvsem pomensko informacijo – vsaj na ravni razlage. Pri besedah, ki so na pomenski ravni – tj. v pomenski razlagi – razložene funkcij-sko- ali pomenskosladijsko mora torej iskati pomensko informacijo na drugem mestu, pri t. i. motivirajoči besedi, pri čemer tako konceptualno izhodišče od uporabnika zahteva poznavanje vsaj tega, kateri od pomenov večpomenskih motivirajočih besed pomensko motivira pridevnik, kateri glagolnik ipd. Slovnicih oz. skladijskofunkcijske podatke je zato v slovarju smiselno posredovati v ločenem razdelku, s čimer se med drugim razbremeni tudi sicer obvestilno zelo kompleksno geselsko zaglavje.

¹²⁹ V zagovor k tipu razlag, kot so: manjšalnica od, glagolnik od ipd., je bilo podano mnenje, da tovrstne besede nimajo svojega pomena, ampak se z njimi posredno razlaga motivirajoča beseda, zaradi česar naj bi bilo v zvezi z njimi smiselno navesti le slovnicih oz. besednovrstno informacijo (Vidovič Muha 2009: 24).

10.4.2.1 Razlagalne strategije

V LBS smo pri oblikovanju pomenskih shem in razlag upoštevali strategije,¹³⁰ ki temeljijo na že omenjenih projektih FrameNet in Corpus Pattern Analysis ter na analizi stavčnih razlag, kot so bile zasnovane pri slovarskih projektih založniške hiše Collins Cobuild. Od omenjenih projektov smo prevzeli predvsem tiste strategije, ki so se v izbranih segmentih leksikografskega opisa v LBS izkazale za najbolj učinkovite. Hkrati pa seveda ni bilo mogoče v celoti in povsem nekritično prevzeti posameznih modelov, saj je bilo treba upoštevati tako specifične značilnosti in namembnost omenjenih projektov kot tudi slovenščine in izdelave LBS. Kot pomembno izhodišče in skupni imenovalc leksikografskih opisov omenjenih projektov in LBS je treba omeniti identifikacijo pomembnih in uporabnih dejstev na podlagi empiričnih korpusnih podatkov.

Bistvo stavčnih razlag, ki so jih zasnovali J. Sinclair in sodelavci, je v tem, da *če/ko*-stavek postavi razlago 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, v kateri je pomen iztočnice (definiendum) opisan s pomočjo drugega dela (definiens). Levi in desni del razlagalne strukture sta med seboj povezana bodisi z glagolom *biti* bodisi s kakim drugim glagolom, ali pa z veznikom tipa *če/ko* (gl. Sliko 37). Vloga levega dela razlage je v tem, da izpostavi tipične skladienske vzorce in besedilno okolje, v katerih se pojavlja iztočnica (kar je še posebej odločilno pri glagolu), desni, razlagalni del pa prispeva tisto, kar želimo (novega) povedati o pomenu konkretne iztočnice v prepoznanem besedilnem in skladienskem okolju. Pomembna značilnost tovrstnih razlag je povezana z odločitvijo, da je eksaktnost in točnost razlage drugotnega pomena, saj je njihov prvi in osnovni namen doseči uporabnikovo razumevanje, kar dobro ponazarja primer razlage samostalnika *asfalt* v SSKJ in Longmanovem slovarju¹³¹ za učenje angleščine :

SSKJ:

asfalt zmes bitumena in peska za asfaltiranje

Longman Dictionary of Contemporary English:

as-phalt črna lepljiva snov, ki se strdi, ko se posuši, in se navadno uporablja za prekrivanje cestnih površin¹³²

130 Poglavje temelji na analizah, ki jih je v okviru LBS izdelala Šorli (2014).

131 Spletna različica LDOCE je dostopna na: <http://www.ldoceonline.com/>.

132 **as-phalt** a black sticky substance that becomes hard when it dries, used for making the surface of roads.

Pri oblikovanju stavčnih razlag smo v LBS težili k zapisu, ki najbolje posreduje vsebino pomena, medtem ko je skladenjski realizaciji primarno namenjen zapis pomenske sheme, čeprav ni nujno, da se obe možnosti med seboj izključujeta. Ker gre predvsem za različne možnosti zapisa, smo zgradbo LBS organizirali tako, da sledi zapis pomenske sheme skladenjski realizaciji pomena (v levem okvirčku), zapis razlage pa lahko tipično skladenjsko realizacijo prestavi v desni del (za poševnico) razlagalne zgradbe (v desnem okvirčku):

če ČLOVEK ali USTANOVA **terja od** drugega ČLOVEKA, DRŽAVE ali USTANOVE, *da izpolni določeno OBVEZNOST ali se ravna po določenih PRAVILIH, /to od njega zahteva*

če ČLOVEK ali USTANOVA **terja** drugega ČLOVEKA, DRŽAVO ali USTANOVO, */od njega zahteva, da izpolni določeno OBVEZNOST ali se ravna po določenih PRAVILIH*

Pri oblikovanju pomenskih opisov smo na podlagi Cobuildove razlagalne taksonomije določili tipe razlag, ki smo jih uporabljali glede na razlikovalne lastnosti, ki se nahajajo praviloma v prvem, tj. levem delu razlage, pa tudi glede na besednovrstno pripadnost besede v iztočnici. Predstavljena taksonomija je namenjena lažjemu orientiranju med različnimi razlagalnimi strategijami, ki so nam na voljo, hkrati pa naj bi pripomogla tudi k poenotenju izbir v sorodnih situacijah.

Tip razlage brez če/ko-stavka, v kateri sta levi in desni razlagalni del povezana z 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 opisu samostalniških pomenov, pri čemer smo iztočnico iz razlage izločili (*branjevka*), v pomenski shemi pa smo jo ohranili (*ohladitev*):

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

Razlago, ki ne vključuje če/ko-stavka, je bilo mogoče uporabiti tudi pri večini pridevniških pomenov, ki se tipično realizirajo v vlogi prilastka:

renesančen *pridevnik*

1 ki ima značilnosti renesanse

renesančna UMETNOST ali LASTNOSTI so značilne za obdobje renesanse

smrdljiv *pridevnik*

1 ki ima neprijeten vonj

1.1 o snovi

smrdljiva SNOV ima zelo neprijeten vonj, navadno zaradi prisotnosti kemičnih sestavin ali zaradi procesa razkrajanja

Pri glagolskih pomenih smo ta tip razlage uporabljali redkeje, sicer pa predvsem za izražanje načinovnih ali vzročnih elementov pomena:

izkašljati *glagol*

1 izločiti s kašljem

1.1 odpraviti draženje v grlu

ČLOVEK se izkašlja tako, da večkrat zapovrstjo iztisne zrak iz dihal in s tem odpravi dražeč občutek v njih

Tip razlage s če/ko- stavkom, ki ustreza pogojnemu odvisniku, smo uporabljali predvsem za opisovanje glagolskih pomenov (*svitati se, siliti*) ter za opisovanje samostalniških in pridevniških pomenov, ki se tipično aktivirajo v povedni rabi (*ne-prilagojen, lahkomiseln*). Če/ko-stavek se lahko nahaja na začetku ali koncu stavčne zgradbe:

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

neprilagojen *pridevnik*

3 neprimeren za bivanje

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

lahkomiseln *pridevnik*

1 nepremišljen; neodgovoren

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

Tip razlage s če/ko-stavkom in metajezikovnim elementom, ki ima obliko t. i. projekcije (Barnbrook 2002: 152), smo uporabili pri opisu pomenov, kjer smo želeli izpostaviti okoliščine rabe ali odnos govorca do ubesedene vsebine (*cviliti*, *hrustljav*). Tipična formula za izražanje metajezikovnega komentarja je z glagolom *reči* v prvi osebi množine v odvisnem stavku: *če rečemo, da ...*, ter z ustreznim dopolnilom v glavnem stavku: *...menimo/se nam zdi/pomeni, da/želimo povedati, da ...* Če za pomen poudarek dejanja izrekanja ni bil nujen, smo razlago oblikovali ustrezno drugače (*lahkomiseln*):

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

lahkomiseln *pridevnik*

1 nepremišljen; neodgovoren

1.1 o ženski, lahek

ŽENSKA se zdi ljudem lahkomišeln, če se rada druží z moškimi ali pogosto menja partnerja

Pretvorba v tem tipu stavčne razlage vključuje poročani govor, ki spremeni obliko leksikografske enačbe v komentar o rabi namesto o notranjem pomenu iztočnice same. Prav v frazeologiji, kjer gre najpogosteje za metaforični pomen in vključevanje pragmatičnih elementov, je uporaba tovrstne razlage še posebej učinkovita (*čas*). Hanks (1987) npr. povezuje model razlage s projekcijo z definiranjem figurativnega ali ekstenzionalnega pomena (*crkniti*), kar je tudi osrednja cobuildovska praksa.

pustiti času čas

če rečemo pustimo času čas, sprejmemo trenutne razmere ali čustva, običajno neprijetna, in verjamemo, da bo sčasoma bolje

čas beži

če rečemo, da čas beži, imamo občutek, da hitro mineva

crkniti *glagol*

2 prenehati obstajati

2.1 o ustanovi

če rečemo, da USTANOVA ali PODJETJE crkne, želimo povedati, da ne obstaja več, ker ne prinaša dobička ali se je znašla v finančnih težavah

Obliko razlage s *če/ko*-stavkom smo uporabili tudi pri samostalniških pomenih, ki se uresničujejo v kombinaciji s pomensko izpraznjenim glagolom:

vonj *samostalnik*

1 zaznavna lastnost

1.2 občutek

če ima PROSTOR ali PREDMET vonj po LASTNOSTIH nekega ČASA ali DOGAJANJA, se to v njem na poseben način izraža ali občuti

Tip razlage z iztočnico v neimenovalniški obliki smo uporabili v primerih, ko je taka pretvorba omogočala bolj berljivo in posledično lažje procesljivo obliko razlage. Zanje smo se odločali zlasti pri pridevniških pomenih, redkeje pri samostalniških:

deževen *pridevnik*

1 o dežju

1.1 z veliko dežja

v deževnem OBDOBJU ali KRAJU veliko dežuje

prozoren *pridevnik*

1 o predmetih, snoveh in materialih

skozi prozoren PREDMET ali MATERIAL je mogoče videti

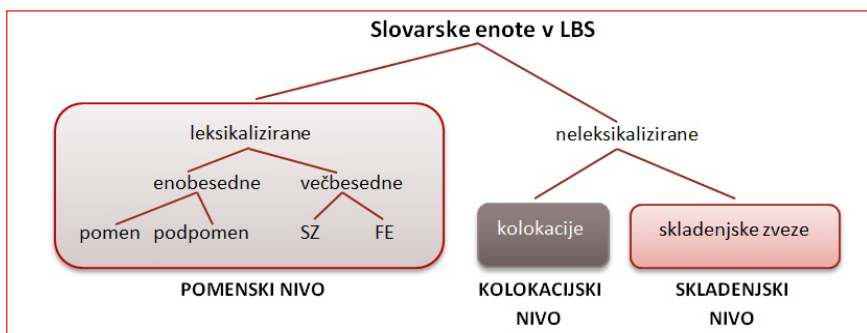
debelina *samostalnik*

1 lastnost predmeta

debelino ploščatega PREDMETA določa razdalja med njegovima nasprotnima ploskvama ali stranicama

11 Večbesedne leksikalne enote

V LBS kot leksikalne enote pojmujeemo posamezne pomene in podpomene ter stalne zveze (SZ) in frazeološke enote (FE) – vse tiste enote torej, ki so nosilke pomena, pri čemer so le-te na izrazni ravni lahko tudi več kot enobesedne. Tako izhodišče, ki v leksikografski teoriji in praksi ni novo (prim. Atkins in Rundell 2008: 162), nam omogoča izpeljavo slovarske zasnove, pri kateri se celotni nabor leksikalnogramatičnih podatkov podreja pomenski in ne izrazni vrednosti besede oz. zveze, hkrati pa nam omogoča delitev vseh slovarskih enot (tj. vseh nizov besed, na katere se navezujejo slovarske informacije) na leksikalizirane, ki so nosilke pomena in lahko eno- ali večbesedne, med zadnjimi stalne zveze in frazeološke enote, in na neleksikalizirane, ki niso nosilke pomena, izrazno pa so vedno večbesedne in izkazujejo zgolj strukturnoskladenjsko in statistično ustaljenost, tj. skladenjske zveze in kolokacije. V zgradbi geselskega članka v LBS so tako večbesedne leksikalne enote obravnavane na dveh nivojih: SZ in FE na pomenskem, skladenjske zveze in kolokacije pa na skladenjskem in kolokacijskem, kot prikazuje Slika 36:



Slika 36: *Slovarske enote v LBS*

Skladno s tem predvideva shema DTD za SZ in FE tudi bolj ali manj enako zgradbo in vrsto leksikalnogramatičnih podatkov kot za posamezne (pod)pomene pri enobesednih iztočnicah, medtem ko kolokacije in skladenjske zveze ne predvidevajo pomenskega opisa.

Leksikografske strategije pri obravnavi SZ in FE v LBS prikažemo v nadaljevanju v samostojnih poglavjih, kjer izpostavimo merila za njihovo prepoznavanje, opredelimo različne tipe ter se osredotočimo na vrste leksikalnogramatičnih podatkov, ki jih nanje pripenjamo. V zvezi s SZ in FE opredelimo tudi način njihovega vključevanja v zgradbo gesla in razmerja do drugih tipov večbesednih zvez, ki jih v LBS ne obravnavamo kot samostojnih pomenskih enot, zlasti torej do kolokacij, razširjenih kolokacij in skladenjskih zvez.

11.1 Stalne zveze

Kot ugotavljajo leksikološke analize (Cowie 1998: 9; Gantar 2007), predstavlja jo SZ obsežen del besedišča posameznega jezika. Zlasti produktivno naj bi bilo ustvarjanje večbesednih enot pri poimenovanju nove predmetnosti, hkrati pa je prepoznavanje SZ in njihov pomenski opis pomemben pri učenju tujega jezika.

11.1.1 Merila za določanje stalnih zvez

Za prepoznavanje SZ kot samostojnih leksikalnih enot smo upoštevali njihove pomenske, oblikovne in skladenjske lastnosti.

11.1.1.1 Pomenske lastnosti

V LBS smo kot SZ obravnavali večbesedne enote, katerih pomena ali sporočanje funkcije ni mogoče izpeljati iz vsote pomenov posameznih sestavin – to lastnost si delijo tudi s FE, hkrati pa za razliko od FE, njihov pomen kot celota ni metaforičen, npr. spalna srajca, glasbeni stolp, osnovna šola, lahko pa je nastal po metaforični poti, npr. *tiskarski skrat*, *rumeni tisk*, *ležeči policaj*, *medeni tedni* itd. Glede na njihovo vsaj delno pomensko neprozornost in posledično pomensko samostojnost smo jih obravnavali na enak način kot enobesedno leksiko.

SZ so lahko večpomenske: *topla greda* – 1 del vrta; 2 proces v zemeljskem ozračju, pri čemer lahko ob navadno terminološkem, npr. *zadruga* – promet sistem zadrge, SZ razvije tudi drugotni pomen: postopnost; izmeničnost postopka, ali pa ima neka zveza hkrati dobesedni (kolokacija) in lastni pomen (SZ), npr. *orjaški dežnik* – velik dežnik : goba; *poletni čas* – poletje : geografija časovni pas. Zlasti samostalniške SZ navadno poimenujejo konkretno predmetnost, npr. bančna kartica, kislja juha, lahko pa zastopajo tudi abstraktno predmetnost, npr. aktivno znanje jezika, fotografski spomin. Raba konkretnega pomena SZ je navadno omejena na določeno specializirano področje, npr. trojanski konj – računalništvo; kislja dež – ekologija. Poleg tega tujejezični ustrezniki SZ, kar sicer velja tudi za druge tipe večbesednih leksikalnih enot, pogosto niso dobesedni prevodi ali večbesedne enote, npr. *jedilna/prava/temna čokolada* – ang. *plain/cooking chocolate*; nem. *echte/reiner Schokolade*; *osnovna šola* – ang. *elementary/primary school*, nem. *Grundschule* ipd.

11.1.1.1 Oblikovne lastnosti

Strukturna ustaljenost SZ se kaže tako v njihovi skladenjski zgradbi, npr. *losjon po britju*, ne pa tudi (tj. v rabi neizkazano) *losjon za britje*, *zemlja krast* (in ne *krasti*

zemljo), in v pretvorbenih omejitvah, kjer denimo raba ne izkazuje zapisov kot *bri-vski losjon* ali *sejem bolh* (za *boljši sejem*), kot tudi v oblikoskladenjski ustaljenosti katere od sestavin SZ, npr. *krompir v oblicah*, ali zveze kot celote, npr. *vodne/vodene barve*, kemija *redke zemlje*, v neizkazovanju pregibnosti, stopnjevanja, npr. *stari ata*, ne pa tudi (*naj*)*starejši ata*, ipd. Na drugi strani lahko SZ podlegajo različnim predvsem sebi lastnim pretvorbam, npr. poenobesedenju, oz. imajo sinonimno kratično ime: *porodniški dopust* – *porodniška*; *upravna enota* – *UE*, *anonimni alkoholiki* – *AA* ipd.

Ob strukturni trdnosti izkazujejo SZ tudi visoko stopnjo variantnosti, ki pa ni prosta, ampak razmeroma natančno določljiva tako na ravni leksikalnih sestavin kot skladenjskih lastnosti. Pri tem se njihova strukturna in pomenska ustaljenost kaže v tem, da variantne sestavine pripadajo sorodnemu pomenskemu polju, npr. *debeli/krepki tisk*; *morilska čebela/čebela ubijalka*; *odzivni/reakcijski čas*, *kratkoročni/kratko-trajni spomin*; *glasbena skupina/zasedba*. Besednozvezna zgradba SZ omogoča tudi druge tipe variantnosti, zlasti na oblikoskladenjski ravni, npr. *toksični šok* – *sindrom toksičnega šoka*, *žvečilni gumi* – *žvečilna guma*, *tek štafet* – *štafetni tek* – *tek v štafeti*, *češnjevi paradiznik* – *paradiznik češnjevec* ipd.

SZ, kot tudi sicer leksikalne enote, pogosto izkazujejo lastno kolokabilnost: [*ponuditi*] *krompir v oblicah*; [*pravilnik, odlok*] o obratovalnem času, [*skrajšan, nedeljski*] obratovalni čas, [*iti se, igrati*] *zemljo krast*, ali pa se razširjajo z notranjimi kolokatorji in se sestavljajo: *iskanje (skritega) zaklada* – družabna igra, *predstavnik/predstavnica za javnost/ za stike z javnostjo*, (*ponuditi*) (*pečen*) *krompir v oblicah*; [*štajerska*] *kisla juha*.

11.1.1.2 Glede na skladenjsko vlogo

Funkcijskoskladenjsko – tj. glede na stavčnočlensko vlogo – se SZ obnašajo enako kot proste besedne zveze. V LBS smo zabeležili predvsem samostalniške besedne zveze, ki se stavčnočlensko realizirajo kot osebki in predmeti, redkeje kot povedkova določila in povedkovi prilastki, ter pridevniške besedne zveze v vlogi prilastka in/ali povedkovega določila. SZ s strukturo samostalniške besedne zveze se pogosto vežejo z glagoli, kar smo zapisovali v obliki kolokacij, npr. [*sejati, posejati, posaditi*] v *toplo gredo*, vendar pa je ta lastnost, kot bomo izpostavili v nadaljevanju, vezana predvsem na FE s strukturo samostalniške besedne zveze.

11.1.2 Merila za uvrščanje stalnih zvez v LBS

Osnovni pogoj za prepoznavanje SZ je v LBS potreba po njihovem samostojnem pomenskem opisu. To pri analizi posamezne leme, ki nastopa kot iztočnica, pomeni tudi prepoznavanje pomenske povezanosti SZ kot celote s katerim

od registriranih pomenov, zato je treba pri tem vprašanju upoštevati, katere tipe podatkov bo imel uporabnik na voljo, če bo SZ predstavljena bodisi kot samostojna iztočnica bodisi kot del registriranega pomena ali kot samostojni del gesla, tj. od posameznih pomenov sestavin neodvisno. Izkušnje pri izdelavi LBS so pokazale, da je prepoznavanje pomenske povezanosti SZ kot celote s katerim od pomenov sestavin v nekaterih primerih očitno in nedvoumno pri drugih pa je taka povezava nejasno določljiva in prepuščena uporabnikovi presoji in prepoznavanju diahronih izhodišč. Zato smo za organizacijo SZ v okviru slovarske mikrostrukture predvideli dve možnosti: SZ je lahko samostojna iztočnica ali pa je njenemu opisu namenjen samostojni del gesla. V drugem primeru je SZ uvrščena bodisi pod posamezni (pod)pomen bodisi od registriranih pomenov neodvisno, s čimer pridobi v geselski zgradbi enak položaj kot FE. SZ kot samostojna iztočnica izkazuje glede na druge SZ v LBS nekoliko drugačno geselsko zgradbo, posledično pa tudi večjo količino leksikalnogramatičnih podatkov, kot prikazuje Slika 37 za iztočnico *nivojski pouk*:

oblika

zapis:nivojski pouk

iztočnica:nivojski pouk

zaglavje

besvrs:samostalniška zveza

Pomenski meni

1. način poučevanje

geslo

1.pomen

indikator:način poučevanja

oznaka^{podr.}:šolstvo in izobraževanje

pomenska shema:

definicija1:način poučevanja v devetletni osnovni šoli, ki za razliko od temeljnega pouka poteka na dveh ali več ravneh zahtevnosti

skladenjske skupine

a)

struktura: sbz0 SBZ2

kolokacije

[izvajanje, izvedba, organizacija, uvajanje, uvedba] nivojskega pouka

zgledi (12)

b)

struktura: gbz SBZ4

kolokacije

[izvajati, uvajati] nivojski pouk

zgledi (5)

c)

struktura: gbz SBZ3

kolokacije

[nasprotovati] nivojskemu pouku

zgledi (3)

d)

struktura: SBZ0 sbz2

kolokacije

nivojski pouk [slovenščine, angleščine]

zgledi (4)

e)

struktura: SBZ1 gbz

kolokacije

nivojski pouk [poteka, se izvaja]

zgledi (3)

f)

struktura: sbz0 pri SBZ3

kolokacije

[delo, ocenjevanje] pri nivojskem pouku

[ocenjevanje [znanja] pri nivojskem pouku]

zgledi (6)

g)

struktura: sbz0 za SBZ4

kolokacije

[učilnica, učitelj] za nivojski pouk

zgledi (3)

skladenjske zveze

skladenjska zveza

[temeljni] in nivojski pouk

Slika 37: *Stalna zveza kot samostojna iztočnica*

255

SZ, ki so v LBS obravnavane kot samostojne iztočnice, vsebujejo podatke o besedni vrsti zveze kot celote. Ta se v skladu z določanjem besedne vrste enobesednih lem v LBS ravna po oblikoskladenjskih oznakah jedrnega dela besedne zveze v korpusu Gigafida, npr. *nivojski pouk* – samostalnišnik oz. samostalniška besedna zveza; *funkcionalno nepismen* – pridevnik oz. pridevniška besedna zveza; *aktivno obvladati jezik* – glagol oz. glagolska besedna zveza. Posebnost predstavljajo podatki na skladenjskem in kolokacijskem nivoju, kjer je za osnovo vzeta SZ (zapisana z velikimi črkami), njeno kolokabilno okolje pa predstavlja drugi del strukture, kot tudi sicer pri enobesednih lemah:

SBZ0 sbz2

nivojski pouk [slovenščine, angleščine]

gbz SBZ4

[izvajati, uvajati] nivojski pouk

Načeloma velja, da lahko kot samostojna iztočnica nastopa vsaka SZ, ki presega prag uvrstitve v LBS, vendar pa smo SZ beležili predvsem v samostojnem razdelku geselskega članka pri enobesednih iztočnicah, in sicer zgolj pri tisti iztočnici, ki smo jo sprejeli v slovarsko obravnavo. Če sta v obravnavo prišli obe besedi, ki se pojavljata kot sestavini SZ, je mogoče pričakovati, da so pri obeh besedah v vlogi iztočnice v razdelku <stalne_zveze> obravnavane tudi iste SZ (npr. izt. *vino* – *arhivsko vino*; izt. *arhivski* – *arhivsko vino*). Odločitev za obravnavo SZ, ki ni nujno izpeljana pri vseh v slovarsko bazo sprejetih iztočnicah, je podprta z več razlogi. Prvi je ta, da je bil geslovnik za LBS omejen na 5000 najpogostejših polnopomenskih lem, kar pomeni, da SZ, pri katerih katera od sestavin ni presegla praga 600 pojavitev v korpusu, ni bilo mogoče zajeti ali pa se katera od sestavin ni uvrstila med iztočnice, ker zunaj konkretne zveze ne obstaja ali pa se v dejanskih besedilih realizira zelo redko (npr. *oblica* v *krompir v oblicah*). Drugi razlog je tehnične narave in izhaja iz organizacije podatkov, kot jo določa shema DTD v formatu XML, ki omogoča iskanje vseh v bazi registriranih SZ ali njihovih delov kot tudi njihovo medsebojno usklajevanje, povezovanje in povezovanje z drugimi elementi geselskega članka, npr. kolokacijami. Posledično bi bilo v spletnem slovarju, ki na taki bazi temelji, mogoče dostopati do podatkov o SZ neodvisno, tj. na ravni samostojne iztočnice, ali v povezavi s katero od iztočnic, znotraj katere je SZ opisana.¹³³

Pri uvrščanju SZ v zgradbo gesla predvidevamo dve možnosti: (a) v primeru nedvoumno izkazane pomenske povezave med pomenom zveze kot celote in katerim od pomenov besede v iztočnici smo zvezo navedli v okviru danega pomena

¹³³ Podrobneje je ta možnost predstavljena v Krek et al. (2013: 33).

ali podpomena; (b) v primeru, kjer take pomenske povezave na podlagi analize jezikovne rabe ni bilo mogoče ugotoviti, pa smo SZ navajali kot samostojno, od posameznih pomenov neodvisno enoto:

mreža *samostalnik*

1 preplet niti, žic

1.1 material

1.2 predmet

viseča mreža

naprava za ležanje iz trpežne tkanine in vrvi, ki se napne med dve drevesi ali stojali

2 o ribolovu

vlečna mreža

stoječa mreža

3 športni pripomoček

4 pajčevina

5 vzorec

koordinatna mreža

geografija

geometrija

6 povezan sistem

7 niz dejanj

nevronska mreža

računalništvo

kristalna mreža

kemija

matna mreža

šah

Kot prikazuje umestitev SZ v zgradbo gesla, pridobijo pomensko podrejene SZ dodatno informacijo na ravni skupnega indikatorja, npr. *viseča mreža* – predmet, oz. z umestitvijo na določeno strokovno področje, npr. *vlečna mreža*, *stoječa mreža* – ribolov.

11.1.3 Opis stalnih zvez v LBS

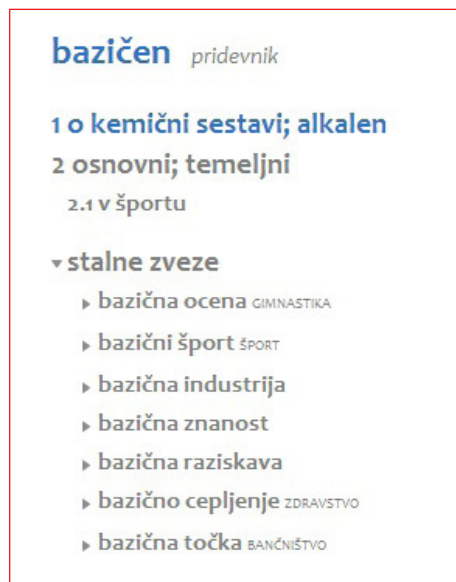
Kot izhaja iz sheme DTD, smo v LBS beležili v zvezi s posamezno SZ naslednje leksikalnogramatične podatke:

- identifikacija SZ (v odnosu do kolokacij, samostojnih in frazeoloških pomenov)
- pomenska členitev (eno- : večpomenske SZ)

- pomenski opis in ali/specifično področje rabe (indikator, razlaga; oznaka)
- oblike rabe (variantnost; raba v različnih skladijskih vlogah)
- tipično besedilno okolje (kolokabilnost SZ)
- zgledi rabe

11.1.3.1 Glede na pomen

V zgradbi geselskega članka, namenjenega beleženju SZ, so za pomenski opis predvideni elementi: <indikator>, <oznaka> in <definicija>, kar velja tudi za eno-besedne iztočnice, nimajo pa večpomenske SZ predvidenega pomenskega menija, čeprav je mogoče, kot smo predvideli v poskusnem prikazu LBS na spletu (Slika 38), na koncu pomenskega menija navesti vse SZ (in FE), ki pripadajo besedi, obravnavani v iztočnici, oz. njenemu posameznemu pomenu.



Slika 38: Seznam stalnih zvez v pomenskem meniju v demo spletni postavitvi LBS

Pri SZ posamezni elementi pomenskega opisa v geselski zgradbi niso nujno zapolnjeni. Ker s pomenskimi indikatorji pri SZ ne oblikujemo pomenskega menija, tudi opredelitev SZ z indikatorjem ni vedno nujna: SZ je tako lahko opredeljena zgolj z indikatorjem (*notranji glas*) ali z razlago (*globalna vas*) ali pa se oba podatka medsebojno dopolnjujeta (*namišljeni bolnik*):

notranji glas

intuicija

globalna vas

če poimenujemo svet globalna vas, ga vidimo kot okolje, v katerem so s pomočjo sodobnih komunikacijskih tehnologij vse informacije dostopne vsakomur

namišljeni bolnik

hipohonder

kdor je nenehno zaskrbljen glede svojega zdravja in misli, da je bolan, čeprav dejansko ni

Terminološka specifičnost SZ pogosto zahteva razlago, ki je v skladu s pojmovanjem zveze na terminološkem področju, in s tem konsenz stroke. V takih primerih smo SZ v bazi zgolj zabeležili in jo umestili na ustrezno strokovno področje s pomočjo oznake, razlage – če ustreznega pomena na podlagi splošne rabe, kot se izkazuje v referenčnem korpusu, ni bilo mogoče določiti – pa nismo zapisovali. Nasprotno smo se za razlago odločali v primerih, kjer sicer področno specifična SZ izkazuje poljudno razumevanje, kot se izkazuje v splošni rabi:

topla greda

ekologija

proces, pri katerem se toplotno sevanje, ki prihaja v ozračje z Zemlje, vrača nazaj in povzroča višjo temperaturo, kot bi bila, če bi Zemljino površje ogrevalo le sonce

Pri pomenski opredelitvi SZ z indikatorjem smo sledili splošnim načelom uporabe indikatorjev v LBS. Med različnimi tipi indikatorjev so za opredelitev SZ prišli v poštev zlasti t. i. ontološki indikatorji oz. neposredne nadpomenke, npr. za opredelitev živali in rastlin: *afriški slon* – žival, *afriški som* – riba, *afriška vijolica* – okrasna rastlina, ter sinonimni in opisni indikatorji, npr. *afriški kontinent* – Afrika, *velikonočno jajce* – pirh; *pisanica*; *Afriški rog* – Somalski polotok.

Poseben tip indikatorja, ki je v LBS vezan zgolj na SZ, je t. i. terminološki indikator, s katerim smo opredelili področje rabe določene SZ ter s tem nakazali bodisi potrebo po natančnejši terminološki definiciji bodisi smo podatek v indikatorju dopolnili s splošnejšo razlago v ustreznem elementu, kot smo prikazali pri SZ *topla greda* zgoraj.

V nekaterih primerih so SZ lahko tudi večpomenske ali pa imajo poleg konkretne-ga stalnozveznega še metaforični tj. frazeološki pomen, kar jih uvršča med FE. V primeru večpomenskosti smo SZ navedli v ločenih nadelementih <stalna_zveza>,

kar nam je omogočilo samostojni pomenski opis ter navedbo oznak, besedilnega okolja in zgledov pri vsakem od pomenov:

aktivno oglje

zdravilo

farmacija

- *V lekarni poišči aktivno oglje in lineks – zdravilo, ki ga uživaš, če dobiš močno drisko.*
- *Aktivno oglje veže v prebavnih povzročitelje driske in pline, ki so vzrok neugodja, napihnjenosti in prebavnih krčev.*

aktivno oglje

kemija

- *Obetavna metoda nadzora nad dimnimi izpusti je dodajanje aktivnega oglja v jaške kanale za odvod dima.*
- *Obdelava prezračenega vina z aktivnim ogljem je brez pomena, kajti oglje pospešuje oksidacijo.*

Pomenske lastnosti SZ je mogoče opredeliti tudi z vidika kategorij v pomenski taksonomiji, ki smo jo uporabili za določanje semantičnih tipov. Tudi pri SZ lahko namreč govorimo o izhodiščnih semantičnih poljih, ki jih na prvi stopnji določajo živo bitje ter konkretne in abstraktne entitete. SZ so znotraj semantičnega tipa ČLOVEK vezane predvsem na poimenovanja izmišljenih, pravljичnih in mitoloških bitij, kot npr. *zobna miška, krilati konj, morska deklica, dedek/Dedek Mraz*, ter na poimenovanja živali: *kraljevi/kraljevski pingvin, morski pes*; rastlin: *pegasta kukavica, pegasta kolobarnica; pasji zob, ptičje mleko, božična zvezda*, dele telesa: *periferni živec, rumena pega, ušesna školjka, mali možgani, mlečni zob, debelo črevo* ter imena (zlasti t. i. ljudska) za bolezni rastlin in bitij, npr. *pikasta pegavost, sladkorna bolezen, manična depresija, kurje oko, oslovski kašelj* itd. V semantično domeno konkretnih entitet sodijo tudi SZ, ki poimenujejo jedi in izdelke, npr. *francoska solata, pasja radost, pohorska omleta, trajno mleko, makaronovo meso, mesni sir, turški med, jajce na oko, čokoladni poljubček, vroča čokolada*; oblačila ali njihove dele: *spalna srajca, ruski ovratnik, kopalni plašč, večerna obleka, hlačno krilo, vroče hlačke*; pohištvo: *nočna omarica, lepotilna mizica*, ter ustanove, društva in stranke, kadar gre za vrstna poimenovanja, npr. *nacionalna stranka, društvo anonimnih alkoholikov, ambulanta za nujno medicinsko pomoč, glasbena šola*; tip dogodkov, prireditve ipd.: *tiskovna konferenca, maturantski ples, modna revija*; poklicne nazive: *medicinska sestra, lesni tehnik, častni doktor*. Pri posameznih iztočnicah smo zabeležili tudi konkretne zgodovinsko pomembne dogodke, sklepe, gibanja, procese, npr. *bencinski tolar, beg možganov, Bostonska čajanka, črni toreč, črni ponedeljek, pomlad narodov, zedinjena Slovenija, praznike: binkoštni ponedeljek, cvetna nedelja, kresna noč, debeli četrtek, dan žena*, običaje: *plezanje na mlaj, sekanje pirhov, skok čez kožo*; religiozne izraze in parafraze: *Izraelovi sinovi, božji sin, brezmadežno spočetje, drevo spoznanja, nauk o dveh kraljestvih* ipd. S SZ

pogosto označujemo tudi predmete in dejanja, ki imajo simbolično ali dogovorno vrednost, npr. *bela zastava*, *črna zastava*, *rumeni trak*; *lipov list*, *rdeča zvezda*, *Davida zvezda*, ter parafraze, in sicer tako za občna imena: *kraljica športov*, *kralj živali*, kot za konkretna stvarna, npr. *najboljši sosed*, geografska: *mesto pod Pohorjem*, *knežje mesto*, *dežela tulipanov*, *rdeči otok*; *zibelka renesanse*; *južni tečaj*, *afriški rog*; *Svilena pot*; *Veliki pes*, *Mlečna cesta*, ali osebna imena: *Sveti oče*, *Ocean modrosti* ipd.

Z vidika izražanja konotativnih pomenskih lastnosti so posebnost SZ, ki poimenujejo konkretno predmetnost, hkrati pa tako na izrazni kot pomenski ravni niso povsem nevtralne in tako zabrisujejo ostro mejo med SZ in FE, npr. *štirinožni prijatelj*, *tiskarska rasa*, *ljubljska srajca*, *oko postave*, *varuh/mož postave*, *mestni očetje*, *nedeljski voznik*, *mlado meso*, *mamina punčka*, *mamin fantek/sinček*, *ledena kraljica*, *zakonski jarem* ipd.

11.1.3.2 Glede na zgradbo

SZ so, strukturno gledano, zgrajene enako kot proste besedne zveze, zato načeloma zanje veljajo enake strukturne možnosti, kot smo jih zabeležili pri kolokacijah in opisali v poglavju Skladenjske strukture na skladenjskem nivoju. V LBS smo za beleženje skladenjske zgradbe SZ z vidika besednovrstne pripadnosti posameznih sestavin sicer predvideli samostojni element <struktura>, vendar pa se je v procesu izdelave LBS pokazalo, da je zaradi variantnosti in prekrivnosti z regularnimi strukturami beleženje nepotrebno. Podatek o besednovrstni zastopanosti sestavin SZ je sicer bolj kot za splošnega uporabnika zanimiv za računalniško procesiranje večbesedne leksike in za jezikoslovne analize. V nadaljevanju je prikazanih nekaj najbolj tipičnih strukturnih možnosti s primeri.

Med dvodelnimi strukturami tudi pri SZ prevladujejo samostalniške zveze s pridevnikom v vlogi levega prilastka in zveze s samostalniki:

Sam0+Sam2	gledališče absurda
Sam0+Sam1	avtomobil bomba
Sam1+sam0	čili paprika, balkan žur
Sam0+Sam0	čebela delavka
Sam4+Glag	zemljo krast

Med večdelnimi pa tako nepredložne, npr. *aktivna volilna pravica*, *aktivno znanje jezika*, *knjiga pohval in pritožb*, kot predložne, npr. *minister brez listnice*, *čokolada v prahu*, *študij ob delu*, *davek na dodano vrednost* itd.

SZ so v osnovni obliki zapisane v elementu <zveza>, pri čemer besednozvezni del geselske zgradbe predvideva v zaporednih elementih <zveza> možnost beleženja več oblikoslovno različnih SZ, ki se pomensko ne razlikujejo, npr. *princeskini krofi*, *princes krofi*; *biotska raznovrstnost/raznolikost*, *biotska pestrost*. Govorimo o t. i. variantnih oblikah rabe, ki so sicer značilne tudi za FE z glagolskimi sestavinami. Tipično besedilno okolje SZ smo beležili, tako kot pri enobesednih iztočnicah, v elementu <kolokacije>:

topla greda	čajna žlička
vrtnarstvo	čajna žlička [soli, sladkorja, medu]
[prenosna] topla greda	[zvrhana] čajna žlička
[sejati, posejati, posaditi] v toplo gredo	
[gojiti] v toplih gredah	

11.1.3.3 Glede na skladenjsko vlogo

SZ, razen v primeru, ko nastopajo kot iztočnica, npr. *nivojski pouk* – samostalniška zveza, nismo opredeljevali glede na besedno vrsto, in sicer iz dveh razlogov: Najprej zato, ker je skladenjska vloga, ki odloča o besedni vrsti zveze kot celote zaradi besednozvezne zgradbe bolj podvržena različnim pojavnim oblikam, npr. leksikalnim in skladenjskim variantam (Gantar 2007: 229), ki so v večji ali manjši meri tudi pomensko pogojene (*beraška palica* – *spraviti na beraško palico* – *biti na beraški palici*). Ker so take konverzije posebej tipične v povezavi z glagolskimi sestavinami, smo beležili SZ le pri samostalnikih, pridevniki in prislovih, pri glagolih pa smo pomensko samostojnejše dele beležili kot skladenjske zveze. In drugič zato, ker predvidevamo, da splošnemu uporabniku zadostuje informacija o pomenu in tipičnem besedilnem okolju SZ, ki jo potrjujejo tudi korpusni zgledi. Posamezne skladenjske kategorije SZ je v slovarski bazi seveda mogoče v vsakem trenutku izluščiti glede na različna merila, npr. SZ s samostalniškimi, pridevniškimi ali prislovnimi sestavinami, ter jih analizirati tudi z vidika skladenjske vloge, indikatorjev, glede na področje rabe ipd.

11.1.4 Stalne zveze v odnosu do drugih večbesednih enot

Obstoj različnih tipov večbesednih enot, ki segajo od kolokacij, fraznih glagolov, frazeoloških enot do idiomov, in različnih teoretičnih pristopov, ki skušajo pojav klasificirati ter opredeliti z nedvoumnimi merili, jasno kaže, da meje niso strogo začrtane, zlasti ne znotraj obeh skrajnosti, ki ju predstavljajo proste besedne zveze na eni strani in idiomatične oz. frazeološke enote na drugi. Zato je, še zlasti v primeru priročnikov, kjer ne želimo pogojevati razumevanja in dostopnosti do iskane

informacije s poznavanjem teoretičnih pristopov, njihova opredelitev v slovarski bazi in slovarju dogovorne narave. Ali kot pravita Rundell in Atkins (2011: 244):

Frazeologija je stopenjska lastnost jezika. Večbesedne zveze vključujejo širok razpon jezikovnih pojavov, od ustaljenih pomensko neprozornih idiomov do popolnoma odprtih pomenskih kombinacij /.../, pri čemer meje med njimi niso jasno začrtane. Zato je vprašanje, kam jih /znotraj slovarske baze/ vključiti, največkrat prepuščeno urednikovi presoji.

Ker so različni tipi večbesednih (leksikalnih) enot eden najtežje kategorialno ulovljivih elementov jezika in ker so meje med posameznimi večbesednimi enotami dostikrat zabrisane, smo se pri organizaciji z njimi povezanih podatkov v LBS odločili za slovarski opis, kjer lahko posamezne enote zamenjujejo položaj na lestvici izkazane pomenske trdnosti, posledično pa lahko (hkrati) nastopajo v različnih delih geselske zgradbe in predvidevajo različno količino ter vrsto leksikalnogramatičnih podatkov, pri čemer se ti med seboj ne izključujejo, ampak se doponjujejo. Predvideli smo naslednje razmerne možnosti:

- razmerje med SZ in samostojnim pomenom besede
- razmerje med SZ in kolokacijo
- razmerje med SZ in frazeološko enoto

11.1.4.1 Stalna zveza – samostojni pomen

Merilo za določanje SZ v odnosu do samostojnega pomena, registriranega pri kateri od iztočnic, smo povezali z ugotavljanjem kolokabilne trdnosti zveze kot celote. V primeru, da je t. i. kolokatna paradigma (Čermák 1985: 173) katere od obravnavanih besed v besedni zvezi omejena zgolj na sopojavljanje z drugo besedo oz. besedami, ki pripadajo istemu pomenskemu polju, npr. stari [kontinent, celi-na], sicer pa ne vstopa v širše pomenske kombinacije, ki bi kazale na pomensko neodvisnost sestavine glede na zvezo kot celoto, obravnavamo zvezo kot stalno, različne pomensko sorodne povezave, če se izkazujejo, pa kot njene variante. Na drugi strani relativno razpršenost kolokatorjev ob besedi, navadno z metaforičnim pomenom, npr. *slep* – brez izhoda, *rumen* – o tisku in publikacijah, obravnavamo kot samostojni pomen, tipične sopojavitve pa kot kolokacije, npr.

rumen *pridevnik*

3 o tisku in publikacijah; tabloidni

rumeni [tisk, medij]

rumene [strani]

rumena *mrzlica*

slep *pridevnik*

3 brez izhoda

slepa [ulica, dolina]

slepi [tir, rokav, rov]

slepi *potnik*

11.1.4.2 Stalna zveza – frazeološka enota

Eden od očitnih problemov pri določanju SZ je prehajanje med registriranimi (pod)pomeni, zlasti metaforičnimi, in FE. Samostalnik *vreča* se denimo pogosto uporablja kot metonimija za količino denarja oz. kot metafora za vir finančnih sredstev in v tem pomenu kolocira s pridevniki kot: [proračunska, državna, finančna] vreča; [polniti, napolniti, odvezati] vrečo, hkrati pa je zelo frekventna tudi zveza *vreča brez dna* s tipičnim lastnim kolokacijskim okoljem. V takih primerih smo se odločili za prikaz SZ pri samostojnem pomenu in hkrati kot FE z lastnim kolokabilnim okoljem:

vreča *samostalnik*

4 vir finančnih sredstev

[proračunska, državna, finančna] vreča

vreča brez dna | kaj ni vreča brez dna

metati denar v vrečo brez dna

metanje denarja v vrečo brez dna

[polniti, napolniti] vrečo brez dna

Podobno situacijo imamo tudi v primerih, kjer neka po metaforični poti nastala SZ s konkretnim denotatom razvije še dodatni metaforični pomen. Tudi v takih primerih smo zvezi opisali v dveh ločenih elementih, in sicer kot SZ in kot FE, seveda z različnima pomenskima opisoma:

slep *pridevnik*

slepo črevo

medicina

del prebavnega trakta, ki v prebavnem procesu nima posebne vloge in ga je potrebno odstraniti, če se vname

slepo črevo

če se neka dejavnost ali območje dojema kot slepo črevo, se zdi da je neperspektivno ali odmaknjeno od dogajanja

11.1.4.3 Stalna zveza – kolokacija

V primerih, kjer obstaja razvidna pomenska povezava med pomenom sestavne SZ in pomenom zveze kot celote, npr. *šola* – vzgojno-izobraževalni proces; *osnovna šola* – vzgojno-izobraževalni proces, ki traja 8 oz. 9 let in je obvezen. Kjer beseda *šola* izkazuje v danem pomenu obsežen kolokacijski niz, posamezni

kolokatorji pa določajo tip oz. vrsto vzgojno-izobraževalnega procesa: [osnovna, srednja, višja, visoka, poklicna] šola, smo navedli kolokacije pod ustrezno skladenjsko strukturo, nato pa še v razdelku SZ s podrobnejšo pomensko opredelitvijo:

šola *samostalnik*

1 o izobraževanju

1.4 proces

[osnovna, srednja, višja, visoka, poklicna] šola

osnovna šola

vzgojno-izobraževalni proces, ki traja 8 oz. 9 let in je obvezen

srednja šola

vzgojno-izobraževalni proces s poudarkom na strokovno usmerjeni izobrazbi, ki traja 4 leta in sledi obveznemu osnovnemu šolanju

visoka šola

izobraževanje na fakultetni ravni, ki traja 2 leti

11.2 Frazeološke enote

Da se FE dojema kot poseben, od preostalega besedišča neodvisen del leksike, dokazujejo ne samo samostojne jezikoslovne vede, ki jih preučujejo (tj. frazeologija, idiomatika; frazeografija), pač pa tudi potreba po samostojnem (frazeološki slovarji) ali od preostalega besedišča ločenem opisu (frazeološko gnezdo) v slovarskih priročnikih. Glede na svoje prepoznavne lastnosti, kot so večbesednost, pomenska in oblikovna nerazstavljivost, ustaljenost ter hkrati variantnost ter metaforičnost zveze kot celote, predstavljajo FE tisti segment jezika, ki prepoznavno izstopa glede na preostalo leksiko, hkrati pa je pri določanju prepoznavnih kategorij, kot so pomenska neprozornost, (ne)ustaljenost zgradbe, skladenjskofunkcijska pretočnost ipd., izrazito spremenljiv in neulovljiv.

V LBS smo za opis FE predvideli samostojni del geselskega članka, v katerem je FE prek katere od sestavin zgolj izrazno povezana z lemo v iztočnici. Ta del gesla vsebuje podatke o:

- osnovni obliki FE (oz. njeno identifikacijo v odnosu do SZ in samostojnih pomenov besede v iztočnici)
- pomenski členitvi (eno- in večpomenske FE)
- pomenu in stilnih ter pragmatičnih posebnosti (indikator, razlaga, oznaka)
- tipičnem besedilnem okolju
- tipičnih oblikah rabe (leksikalne in skladenjske variante)
- besedilni realizaciji v obliki korpusnih zgledov.

FE nismo opredeljevali glede na njihovo besednovrstno pripadnost in tudi ne glede na njihovo notranjo strukturiranost, smo pa več pozornosti zato namenili opisu različnih pojavnih oblik oz. t. i. variantnim oblikam rabe, ki jim FE glede na svojo besednozvezno ali stavčno zgradbo (in skladijsko vlogo) tipično podlegajo.

V nadaljevanju opišemo merila, ki smo jih v LBS upoštevali pri določanju FE, ter posamezne tipe, zlasti v strukturnem ter pomenskem smislu. Nato opišemo način njihovega umeščanja v zgradbo LBS, zlasti v odnosu do metaforičnih pomenov katere od sestavin, ki je obravnavana v iztočnici.

11.2.1 Merila za določanje frazeoloških enot

Področje frazeologije je znotraj jezikoslovnih raziskav usmerjeno na tiste zveze besed, ki glede na sestavne elemente izkazujejo samostojni, od sestavnih elementov neodvisni – tj. *frazeološki pomen*. O FE torej govorimo predvsem takrat, ko se ob večbesednosti in ustaljenosti upošteva tudi idiomatičnost kot tista lastnost, ki določa, da nekatere enote glede na slovnična in pomenska razmerja sestavnih elementov niso jasno razstavljive (Čermák 1985: 167, 172). Ob tem se, vsaj v smislu pojmovanja frazeologije v ožjem smislu, poleg večbesednosti za FE zahteva tudi metaforičnost ali prenesenost pomena, ki jo izkazuje zveza kot celota, kar ji daje glede na nefrazeološka in enobesedna poimenovanja ekspresivno oziroma zaznamovano noto.¹³⁴ Prav zadnje iz okvira FE izloča vsa večbesedna terminološka poimenovanja, in sicer ne glede na sicer izkazano pomensko nerazstavljivost ali izvorno metaforičnost, npr. *črna skrinjica*; *sončni veter*. Kot že omenjeno, uvrščamo ta tip večbesednih leksikalnih enot na raven SZ.

Med FE, ki ustrezajo zgornjim merilom, smo pri izdelavi LBS izpostavili naslednje osnovne tipe:

- FE z zgradbo nestavčne besedne zveze, npr. *dvorezni meč*, *rešilna bilka*, *trojanski konj*, *princ na belem konju*, *občutiti kaj na lastni koži*, *pljuvati v lastno skledo* ter
- FE s strukturo stavka, npr. *zob časa načne koga/kaj*, *mrak pade na oči komu* ipd.
- FE s strukturo povedi, tako stavčne kot nestavčne, kamor smo šteli (a) pregovore in (iz)reke, (b) izraze s pretežno pragmatično vlogo, ki se kaže v našanju na vsebino sporočane, vrednotenje udeležencev, okoliščin, splošnih nazorov in prepričanj, vezanih na katerega od udeležencev v sporočanski situaciji, ter (c) ustaljene konverzacijske formule in izraze, kot so pozdravi, opozorila, kletvice ipd., ki so sicer lahko tudi enobesedni.

¹³⁴ Ekspresivnost je definirana kot sposobnost izražanja subjektivnih razmerij do predmetnosti (Vidovič Muha 2000: 97), v literaturi pa se v zvezi s FE navaja tudi konotativnost, slikovitost in metaforičnost.

- V manjši meri smo v frazeološkem razdelku beležili tudi dele besedila, kot so otroške izštevance: *en kovač konja kuje (koliko žebljev potrebuje), muca je popapala kaj* ipd., ki se uporabljajo kot pragmatični komentarji na vsebino sporočanega ali udeleženca v sporočanju.

11.2.1.1 Glede na zgradbo

Zgradbeno so FE bolj zapletene kot SZ, ki so prekrivne s skladenjskimi strukturami, ki veljajo za proste besedne zveze. FE imajo namreč, kot smo izpostavili zgoraj, vsaj naslednje strukturne pojavne oblike: nestavčna podredna in priredna besedna zveza ter stavčna in nestavčna poved.

V prvi skupini gre za FE, ki so v svoji osnovni obliki prekrivne s samostalniškimi, pridevniškimi, glagolskimi in prislovnimi besednimi zvezami, npr. *nočni ptič, vonj po denarju, črv ljubosumja; blažen med ženami, sit kot boben, srečen, da bi objel ves svet; brusiti pete, dobiti jih po buči; od časa do časa, gor ali dol, (biti) ob pravem času na pravem mestu, v en glas* ipd. Sem sodijo tudi priredne zveze, kot npr. *bog in/i batina, jasno in glasno, iz mesa in krvi* ipd.

FE, ki so pomensko odvisne od vsebine širšega sobesedila, imajo navadno zgradbo stavčne ali nestavčne povedi, npr. *odrasti že (enkrat)!, daj/dajte (no) mir!*, vendar ne izključno, npr. *korajža velja!* Njihovo sobesedilnost in pogosto vezanost na govorjeni diskurz najbolje pojasnjujejo spodnji zgledi:

- *Ali ste res tako neobveščeni, da ne rečem neumni, da mislite, da je v življenju vse zastoj? Odrastite že enkrat in ozavestite se, da se v življenju vse plača.*
- *Daj mir, no. Če ti puščaš zapiske po celem stanovanju, lahko tudi on kaj pusti tukaj.*
- *»Korajža velja!« je v svojem nagovoru med drugim še dejal papež Janez Pavel II.*

Med FE z zgradbo stavka so najočitnejši pregovori in izreki tipa, *podarjenemu konju se ne gleda v zobe; beseda ni konj; kdor dolgo kašlja, dolgo živi; dobro blago se samo hvali; čas beži; obleka naredi človeka*, ipd., ki podobno kot zgoraj omenjene FE predstavljajo komentarje k vsebini, ki je izražena v sobesedilu.

V ta strukturno-pomenski okvir moramo šteti tudi t. i. pragmatične FE, kot denimo: *naj sosedu krava crkne, bog (naj) odpusti komu, to/kaj je (čisti) absurd* ipd., ki so izrazito vrednotenjske ali pa vzpostavljajo razmerje do vsebine povedanega ali udeležencev sporočanja ter so prav tako pogosto vezane na govorjeni diskurz. Tudi zato je v zvezi z njimi smiselno uporabiti stavčni tip razlage z metajezikovnim komentarjem *»če rečemo«*, npr.

kaj/to je čisti absurd | absurd je, da ...

če rečemo, da je neko dejanje ali dogodek čisti absurd, želimo izraziti svoje nestrinjanje, neodobravanje ali ogorčenost nad njegovo nesmiselnostjo

naj sosedu krava crkne

če rečemo, naj sosedu krava crkne, želimo, da se komu zgodi kaj neprijetnega, zlasti zato, ker smo nevoščljivi

Poseben tip stavčnih predstavljajo FE, ki vključujejo prosto a hkrati obvezno vezljivostno mesto, in sicer (a) v vlogi logičnega osebka, npr. *glava boli koga* (zaradi česa); *sekira pade v med komu* – t. i. nepravi glagolski frazemi (Krzlišnik 1994: 162 op. 25; 1996: 51, 63; Gantar 2007: 106), bodisi (b) na mestu osebka, kadar je to pogojeno s kategorijo neživosti oz. nečloveškosti: *kaj vzame/jemlje komu dah, kaj je/ni mačji kašelj* ipd.

11.2.1.2 Glede na skladenjsko vlogo

Ko izpostavljamo skladenjski vidik FE, imamo v mislih njihovo stavčnočlensko in širše besedilno vlogo, ne ukvarjamo pa se s skladenjskim razmerjem znotraj zveze, kot tudi ne z besedno vrsto sestavin, saj se ugotavlja da:

- prihaja znotraj FE v sicer predvidljivem sistemu jezika do nepredvidljivih in slovnično nepričakovanih kombinacij, npr. pričakovano: (kdo) iti + kam/od kod/kako ... proti slovnično nepričakovano: gre za + to/kaj ..., in da
- besednovrstna opredelitev FE, kot naj bi jo določala njena notranja besednovrstna zgradba po načelu: samostalniška zveza opravlja v stavku samostalniške stavčnočlenske vloge, ne odloča (nujno) tudi o besedni vrsti zveze kot celote.

Tipičen primer so rodilniške samostalniške zveze kot: *zdrave pameti, bistre glave, zrelih let* v pridevniških skladenjskih vlogah in predložne samostalniške zveze v vlogi prislovnih določil, npr. *na štiri oči, na čelu česa* itd. Ob tem, da zlasti FE v modifikacijski vlogi, tj. v vlogi členkov ali prislovov, niso nujno oblikovane kot členkovne, pridevniške ali prislovne besedne zveze, je treba znotraj frazeološke problematike in načina slovarske predstavitve opozoriti še na očitno skladenjsko večfunkcijskost, ki je povezana s pomenskimi odtenki in izrabo stilističnega potenciala (Gantar 2007: 115). Npr. zgradbeno in pomensko razmeroma preprosta FE *trda butica* se večinoma realizira v pomenu: 'neumna glava', pri čemer je sestavino *glava* in tudi zvezo kot celoto prej mogoče razumeti kot metonimijo za človekove miselne sposobnosti, ne pa toliko kot metonimijo za človeka samega, kar kažejo tudi primeri, kot: *To je nato še večkrat ponovila, da bi mu vbila v trdo betico*. V tipični kombinaciji z glagoloma *imeti* in *biti*: *težko ti je pomagat, če maš trdo betico*, gre

za nekoliko drugačno pomensko vrednost, tj. 'biti neumen, trdoglav', funkcijsko-skladenjsko pa za povedkovo določilo. Pomen, ki bi ga pričakovali po metonimiji 'neumen človek', ki predpostavlja osebkove in predmetne vloge, pa se v besedilih ne realizira.

Nekatere FE z zgradbo samostalniške besedne zveze se tipično realizirajo v prislovni vlogi, npr. *polomljena angleščina – v polomljeni angleščini* – 's slabim znanjem angleščine'. Hkrati je zanje značilno, da opravljajo v stavku predvidene skladenjske vloge samostalniških zvez, npr: *že vse življenje jo tepe njen dolgi jezik; reforma pomeni kislo jabolko kulturne politike*. Kot take se tipično vežejo z določenimi glagoli oz. pomenskimi skupinami glagolov, npr. *oprijeti se, okleniti se rešilne bilke; iskati, ponuditi rešilno bilko; ugrizniti, zagristi v kislo jabolko*.

Zlasti FE z zgradbo glagolske zveze se skladenjsko lahko realizirajo kot povedki, kar je razvidno iz njihovih enobesednih ustreznikov: *zarežati kot britev* – prizadeti, še pogostejše pa gre za realizacijo vezavnega ali primičnega razmerja, npr. *briti norce iz koga* – norčevati se iz koga.¹³⁵ Na drugi strani je mogoče pričakovati, da bodo FE z zgradbo nestavčne glagolske zveze, npr. *delati iz muhe slona*, nastopale tudi ali predvsem v besedilni funkciji, kar ima navadno posledice tudi za njihov pomenski opis (Gantar 2007: 110). Če denimo pomen FE *delati iz muhe slona* razložimo le z 'zelo pretiravati' (tako kot SSKJ), ki ustreza skladenjski vlogi povedka, tvegamo izpustitev pomembne informacije, ki se nanaša na vrednotenje predmetnosti, ki jo tisti, ki so pri vrednotenju udeleženi, različno dojemajo: če torej *nekdo1* reče, da *nekdo2 dela iz muhe slona*, potem (*nekdo1*) meni, da *nekdo2* pripisuje nečemu večjo pomembnost, kot sam (*nekdo1*) meni, da je smiselno, ustrezno. Vsebinski elementi oz. predmet vrednotenja so v takih primerih praviloma deli širšega sobesedila in predstavljajo v povezavi s FE njegovo semantično koherenco, kot kažejo korpusni zgledi:

- *Menim, da veliko ljudi zaradi nove valute dela iz muhe slona, vse, kar je pomembno, je to, da znamo računati.*
- *V službi bo nekaj zamud, toda nikar ne delajte iz muhe slona.*
- *»Kakšen prstan?« »Prstan, ki ste ji ga podarili, ko sta počitnikovala ob jezeru Balaton. Hočete reči, da je imela Ela halucinacije?« »Poslušajte, ne delajte iz muhe slona, prav! Ja, res je, bila sva pri njeni babici na Balatonu. Res je tudi, da sva bila vse noči skupaj. In res je tudi, da sem ji podaril prstan. Pa kaj?*

Skladenjskofunkcijska pretočnost FE predstavlja glede na to, da je neposredno povezana tudi z izražanjem pomenskih odtenkov, svojevrsten izziv pri določanju osnovne oblike FE v povezavi z večjo ali manjšo tipičnostjo njenega besedilnega

¹³⁵ Podrobno je na podlagi korpusne analize pri stavčnih in t. i. pragmatičnih FE prikazana skladenjska večfunkcijskost v Gantar (2007: 107–114).

okolja, kar je pomembno tudi z vidika strojnega procesiranja in možnosti avtomatskega prepoznavanja v korpusu, hkrati pa zahteva tudi premislek o najustreznejši uvrstitvi FE v zgradbo gesla. Rešitve, ki smo jih v zvezi z osnovno obliko, variantami in t. i. variantnimi oblikami rabe ter s tem povezanim pomenskim opisom sprejeli pri oblikovanju LBS, podrobneje opišemo v nadaljevanju.

11.2.2 Opis frazeoloških enot

V LBS FE vedno navajamo v samostojnem razdelku geselskega članka, pri čemer predvidevamo slovarski opis v celotni vertikali, torej z vsemi pripadajočimi elementi geselske zgradbe, kot predvideva shema DTD: FE v osnovni obliki, poleg tega pa še variante in variantne oblike rabe, ki zadevajo oblikoskladenjske spremembe FE kot celote ter njene različne upovedovalne možnosti, ki so pogosto tudi pomensko pogojene. Nadalje njeno kolokabilno okolje, če je tipično, ter pomenski opis v obliki indikatorja, ki ga lahko dopolnjuje oznaka in/ali pomenska razlaga. Vsi podatki so potrjeni s korpusnimi zgledi rabe:

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 dejstva o kaki neprijetni stvari ali dogodkih*

- *Nalijmo si čistega 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 Sloveniji ni.*

11.2.2.1 Osnovna oblika, variante in oblikoskladenjske pretvorbe

Tipična značilnost FE je, kljub pomenski in strukturni trdnosti, da izkazujejo veliko stopnjo variantnosti, in sicer tako pri posameznih sestavinah (leksikalne variante) kot pri besednih oblikah (oblikoslovne variante) in predvidenih prostih skladenjskih mestih: *luč na koncu tunela/predora; stopicati/capljati na mestu, stisniti pest/pesti; biti trn v peti komu/koga; biti trn v čigavi peti*. Te vrste variante smo v LBS navajali znotraj ene same FE za poševnico.

Nekoliko drugače smo obravnavali zapis FE v primeru variantnih oblik rabe in oblikoskladenjskih pretvorbenih možnosti. V prvem primeru gre za variantnost besednih oblik, ki so posledica oblikoslovnega prilagajanja sestavin znotraj konkretne skladenjske kombinacije. Tipično katera od sestavin FE, npr. variantni glagol, 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 oblikoskladenjskih pretvorbenih možnostih (prim. Kržišnik 1987, 1990) pa govorimo v primerih, ko gre za:

- različne skladenjske oz. stavčnoočlenske funkcije FE, npr. *zdrava pamet – biti zdrave pameti – pri zdravi pameti – (narediti kaj) po zdravi pameti*;
- različne realizacijske možnosti,¹³⁶ ki jih omogoča zgradba FE, zlasti posamostaljenje, npr. *prati možgane – pranje možganov, stopicati na mestu – stopicanje na mestu*, in potrpnjenje oz. prenos na splošnega vršilca dejanja, npr. *dvigati prah – prah se dviga*;
- 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*;
- 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*;
- fakultativne sestavine FE: *pasti kot (zrela) hruška in pasti komu (v naročje) kot (zrela) hruška*;
- predvidljive skladenjske pretvorbe, npr. *kaj/breme pade na čigava rame – na – kaj/breme pade na ramena koga* ipd.

V LBS smo take primere beležili zaporedoma znotraj istega nadelementa v elementu <enota> (v primerih spodaj za pokončnico), če ni prihajalo do bistvenih pomenskih sprememb. Pretvorbene možnosti in vseh vrst variante smo seveda beležili le, če so se kazale kot tipične in za posamezno FE relevantne:¹³⁷

rešilna bilka | videti v čem rešilno bilko | ponuditi komu rešilno bilko | oprijeti se česa kot rešilne bilke | oprijeti se/okleniti se edine/zadnje/vsake rešilne bilke

skakati/letati s cveta na cvet | letati od cveta do cveta

11.2.2.2 Tipično besedilno okolje frazeološke enote

Ob variantnih oblikah rabe smo relativno spremenljivo besedilno okolje FE beležili v elementu kolokacije, npr.

¹³⁶ Če se katera od pretvorb izkaže kot edina možna, npr. *kopja se lomijo* (*lomiti kopja), smo navajali samo to obliko.

¹³⁷ Ustvarjalnih pretvorb tipa, *Tako bodo mandat končali približno enako, kot so ga začeli – s trnom v peti*, nismo vključevali v slovarsko bazo.

ostrina britve | na/po ostrini britve

[hoditi, stopati] po ostrini britve

[hoja] po ostrini britve

[življenje, preživetje] na ostrini britve

v korak s časom | v koraku s časom

[stopati, iti, hoditi] v korak s časom

[biti, ostati] v koraku s časom

Pri oblikovanju osnovne oblike FE v razmerju do variantnih oblik rabe in elementov besedilnega okolja je treba posebej izpostaviti prosta oz. leksikalno ali katedralno predvidljiva vezljivostna mesta pri FE z glagolsko sestavino, in sicer tako na osebkovem mestu kot na mestu predmetnih dolpolnil. Taka mesta so lahko obvezna ali pa fakultativna – v tem primeru smo jih navajali v oklepaju:

spraviti (koga) na kolena

ne seči komu (niti) do kolen

jemati kaj z levo roko

pognati kje (močne/globoke) korenine

kdo¹³⁸ ne more iz lastne/svoje kože**11.2.2.3 Pomenski opis**

Za pomenski opis FE sta v zgradbi gesla predvidena elementa <indikator> in <definicija>, ki se lahko dopolnjujeta.

Pomenski indikator ima načeloma enako vlogo kot pri posameznih pomenih, tj. na kratko opredeli pomen FE, le da so glede na kompleksnost in pomensko razpršenost indikatorji v primeru FE najpogostejše enobesedni nefrazeološki sinonimni ustrezniki, npr. *stisniti zobe* – potrpeti, *stisniti pas* – varčevati, oz. vzpostavljajo razliko med pomeni večpomenske FE: *dol pasti* – 1 biti presenečen, začuden, 2 – biti utrujen, 3 – omagati, izgubiti zavest. Poleg sinonimnih indikatorjev, ki smo jih pogosto kombinirali z razlago, smo za pomenski opis FE uporabljali tudi daljše opisne indikatorje, zlasti v primerih, ko pragmatična pomenska komponenta ni bila očitna in smo nanjo opozarjali le z zgledi rabe:

138 Imenovalniško obliko poljubnostnega zaimka na mestu osebkva smo navajali le pred glagoli, kot so moči ipd., kjer bi nedoločnik predstavljal preveč abstrakten in s tem manj prepoznaven zapis FE.

dvorezen meč

tvegano dejanje

če ocenimo, da je neko dejanje dvorezni meč, menimo, da ima lahko tako dobre kot slabe posledice

(na široko) odpreti/odpirati denarnico

biti pripravljen veliko plačati ali porabiti veliko denarja

- *Visoka šolnina ga ne ovira, saj sta mama in oče pripravljena na široko odpreti denarnico.*
- *Dokler ste pripravljeni na široko odpreti svojo denarnico, bo vaša poroka kot iz škatlice.*

globoko/globlje seči v denarnico

plačati ali porabiti več denarja

- *Za darilo, ki ga je izbral, bo moral globoko seči v denarnico.*
- *Cene nekaterih izdelkov so se dvignile celo za polovico. Globlje bo treba seči v denarnico predvsem za marelice, jagode, papriko in paradižnik.*

Zaradi pomenske razpršenosti in vključenosti pragmatičnih elementov v pomen FE, smo več poudarka pri pomenskem opisu FE namenili stavčnim razlagam, največkrat v obliki *če*-stavka, redkeje v obliki *izraža, da* Tak tip razlage je omogočil navedbo vseh za pomen potrebnih udeležencev. Vključitev metajezikovnega komentarja pa je omogočila prisotnost zunanjega opazovalca ali neudeleženca v sporočanjski situaciji, ki v ospredje potisne njegovo vrednotenje:

utopiti/utapljati kaj v alkoholu | utopiti se/utapljati se v alkoholu

če rečemo, da se kdo utaplja v alkoholu ali da utaplja v alkoholu svoje skrbi in težave, želimo povedati, da se omamlja z alkoholnimi pijačami, da bi na težave pozabil

pljuvati v lastno skledo

če rečemo, da kdo pljuva v lastno skledo, menimo, da s slabonamernim dejanjem, uperjenim proti drugemu človeku ali ustanovi, škoduje predvsem samemu sebi

Opredelitev pomena pri FE, ki so pomensko odvisne od širšega sobesedila in imajo navadno zgradbo nestavčne povedi, npr. *odrasti že (enkrat)!*; *daj/dajte (no) mir!*, ter pri pragmatičnih FE, npr. *to/kaj je (čisti) absurd*, smo uporabili tudi tip razlage: *izraža, da/kaj ... ali kot ...*, npr. *če me spomin ne vara* – izraža omilitev, dvom v trditev; *fant moj!* – kot izraz presenečenja.

Posebnost v pomenskem opisu so FE, ki jih spremlja konkretna telesna gesta, obrazna grimasa, premik telesa ipd., na kar smo opozorili v indikatorju, pomenski opis pa smo navedli v razlagi:

pokazati komu dvignjen/iztegnjen (srednji) prst

kot vulgarna gesta

- Najstarejša hči Rita je bila zelo nadarjena, toda ko jo prišla v puberteto, je očetu pokazala čvrsto iztegnjen srednji prst in jo pobilala s teniškim čudakom.
- Ko je izstopil, se je nesramno režal in mi z iztegnjenim prstom kazal, da me je pač zafrknil.

stisniti pest/pesti

kot gesta

izraža jezo, kljubovalnost ali pripravljenost na pretep

- Markiz je prebledel od jeze in stisnil pesti.
- Ko je kapitan ukazal spremembo smeri, da bi se skrili za kak otok, je Cezar stisnil pesti. Rimljan vendar ne bi smel bežati.
- Ko sem videl, kako ženejo Italijani naše ljudi, sem stisnil pesti: nikdar!

stisniti pest/pesti

kot gesta

izraža veselje nad zmago, uspehom

- Jacob je spustil slušalko, triumfalno stisnil pest in vzklil: "Pogoltnil je!"
- Karol Kučera je po četrtfinalni zmagi na Grand Slamu v Melbournu zmagoslavno stisnil pest, Pete Sampras pa je razočaran odšel z igrišča.
- Srečko Katanec je stisnil pesti, skočil, zavpil in ganjen potočil nekaj solz ob prvi uvrstitvi Slovenske izbrane vrste na svetovno prvenstvo.

Od tega tipa ločimo FE, kjer predstavlja gesta ali telesni premik zgolj metonimijo za določeno aktivnost ali dejanje, npr. *stisniti koga v kot* – spraviti v brezizhoden položaj, *stakniti glave* – sestati se in skupaj razmisliti, *ne migniti (niti) s prstom* – ničesar ne narediti, *oblizniti si (vse) prste* – biti zadovoljen z rezultatom ali uspehom.

Kot smo že omenili, se FE pogosto uporabljajo pomensko zelo razpršeno (večpomenske FE). Če si v konkretnih besedilnih kontekstih ogledamo rabo posamezne FE, lahko ugotovimo, da izraža različne pomenske odtenke, ki so povezani tako z možnostjo različnih skladijskih rab kot z vsakokratnim kontekstom in govorčevo namero. Za ponazoritev si pogledajmo nekaj korpusnih primerov, ki vsebujejo FE *zaviti/zavijati v vato*:

- Tipično za razvujane kapitaliste, ki živijo zaviti v mehko vato blaginje in nočejo videti tistega, kar bi bilo lahko preveč za njihovo srce!
- Občutljive ljudi lahko z resnico prizadenete. Zato bodite bolj diplomatski in resnico zavijte v vato, manj bo bolela.
- To povezujem s tem, da sta me v otroštvu starša precej zavijala v vato, ker sta me hotela obvarovati pred neprijaznim zunanjim svetom.
- Vojno so mi, obenem z mlajšimi brati, sicer skrbno zavijali v vato in jo poskušali narediti nevidno ali vsaj blažiti.
- S 1. januarjem 2003 domači proizvajalci ne bodo več »zaviti v vato«, zato Golob pričakuje zanje precejšen šok.
- Ne zavijajte se v čustveno vato. Prav radi opazite vse, kar vas neprijetno tiši v okolju.
- Oba zavoda ustvarjata presežek prihodkov nad odhodki, kar, ne da bi ovijali v vato, pomeni dobiček.

- *Kdor laže, ta pač laže, kaj bi se sprenevedali in zadevo zavijali v vato!*
- *Mednarodni terorizem je huda stvar, o kateri se ljudem v tej v vato zaviti Sloveniji ne more niti sanjati.*

Večbesednost in metaforičnost govorcem omogočata tudi izrabo FE v ustvarjalne namene, ki za leksikografski opis, upoštevajoč njihovo enkratnost in individualnost, niso zanimive, lahko pa ustvarijo nov trend frazeološkega pomena, ki ga je v slovarski bazi smiselno evidentirati. Spodnji korpusni zgledi kažejo, da se pregovor *kdor drugemu jamo koplje, sam vanjo pade* redko realizira v svoji izvorni obliki:

- *Menimo da je dobil kar je iskal. Kdor drugemu jamo koplje sam pade noter.*

Na drugi strani je mogoče evidentirati rabe, ki so dovolj ustaljene in predstavljajo frazeološke oblike, ki jih je smiselno evidentirati in opisati vsaj na ravni slovarske baze:

drug drugemu jamo kopati

- *Slovenski politični eliti druga drugi kopljeta jami in po ljudski modrosti bosta sami vanju popadali-*
- *Ko bi v parlamentu poslanci drug drugemu kopali jamo, bi imeli svetovno posebnost.*
- *Takšni smo pač Slovenci: drug drugemu kopljemo jamo, čeprav od tega nimamo koristi, le škodo.*
- *Klerikalci in liberalci so se, skratka, med seboj grizli, drug na drugega pljuvali, drug drugemu kopali jamo, čeprav bi obojim moral biti glavni cilj boj za Zedinjeno Slovenijo.*
- *Ko si naši vrli politični nasprotniki v parlamentu drug drugemu kopljejo jamo, se lahko zgodi, da bodo pokopali ves narod.*

pasti v jamo, ki si jo je kdo sam/sebi izkopal | pasti v jamo, ki jo je kdo kopal drugemu

- *A naposled so Dunajčani padli v jamo, ki so jo sami skopali, saj so kar naprej grešili na začetnem udarcu.*
- *IJ je sam padel v jamo, ki jo je drugim kopal ...*
- *Toda politika je Ljubljancanom obljubila nov stadion, zdaj pa so padli v jamo, ki so jo skopali sami sebi.*
- *Kapitalizem je padel v jamo, ki si jo je skopal sam.*
- *Ruska pravoslavna cerkev bo seveda prejel ali slej sama padla v jamo, ki jo zdaj tako vneto koplje katolikom.*
- *Morda bo vendarle upošteval nasvet uglednih članov, ki opozarjajo na škodljive posledice za stranko. Sicer bo padel v jamo, ki jo je kopal drugim.*
- *Najhujše pa je, da so Palestinci sami padli v jamo, ki so jo kopali drugim.*

Ob pomenski razpršenosti je treba opozoriti tudi na samostojne pomene, ki jih nakazujejo posamezne strukturno razmeroma prekrivne FE. V zgradbi frazeološkega razdelka smo različne pomene ene same FE zapisovali kot ločene frazeološke iztočnice, torej v samostojnem frazeološkem razdelku. Odločitev je podprta s predvidevanjem, da lahko tudi večpomenske FE izkazujejo lastne variante, svoje tipično besedilno okolje in pretvorbene možnosti:

zaviti koga v vato | zaviti se v vato

z neodobravanjem če rečemo, da zavijamo koga, zlasti otroke, v vato, želimo povedati, da jih pretirano ščitimo ali jim skušamo pribrani neprijetne izkušnje

zavijati/ovijati kaj v vato

z neodobravanjem če rečemo, da besede ali dejanja zavijamo v vato, jih želimo omiliti in prikazati manj problematične, kot v resnici so

biti zavit v vato | živeti v vati

z neodobravanjem če rečemo, da je kdo zavit v vato, mislimo, da živi preveč brezskrbno in se ne zaveda težav realnega življenja

11.2.3 Merila za uvrščanje frazeoloških enot v LBS

Pri uvrščanju FE v zgradbo gesla oz. znotraj nadelementa <frazeološke_ zveze> smo upoštevali predvsem zgradbo FE oz. določitev njene osnovne oblike, variant, variantnih oblik v razmerju do njenega kolokabilnega okolja in njene pomenske lastnosti, ki so z obliko in obsegom FE neposredno povezane.

11.2.3.1 Glede na zgradbo

FE, ki vsebuje besedo iz iztočnice, smo navajali v frazeološkem razdelku pri vseh predmetnopomenskih besedah, ki jih vsebuje, npr. FE *princ na belem konju* smo navajali pri iztočnicah *princ*, *konj* in *bel*, če so bile del geslovnika. Pri variantnih sestavinah FE ali pri kolokabilnih elementih, npr. *čakati na (svojega) princa na belem konju*, *sanjati o princu na belem konju* ipd., smo navedli FE pri glagolskih sestavinah (tj. pri *čakati* in *sanjati*) le, če je prišla zveza na statistično površje v leksikalnogramatičnem profilu, ki ga izkazujejo besedne skice, in seveda, če je bila lema sprejeta v obravnavo kot del geslovnika. V fazi izdelovanja LBS istih FE pod različnimi iztočnicami nismo usklajevali, zato lahko prihaja pri posameznih FE do deloma različnih informacij oz. do različne količine informacij. To je v nekaterih primerih pričakovano glede na različno izhodišče:

brusiti glagol

brusiti (si) jezike

veliko govoriti o aktualni temi

- *No, fofušljivci iz krogov turističnih delavcev že brusijo jezike, da jim v Portorožu uspeva, ker nudijo erotično tajsko masažo, vendar pa ta s tradicionalno nima prav nič skupnega.*
- *Čevljarji ne osrečujejo le ljudi s premajhnimi ali prevelikimi stopali, temveč morajo občasno zakriti tudi neskladje med bodočima zakoncema, zaradi katerega bi si zlobneži brusili jezike.*

jezik *samostalnik*

brusiti (si) jezike (o čem)

veliko razpravljati o čem, navadno kritizirati ali opravljati

- *Akadska srenja se muza, zlobneži pa veselo brusijo jezike.*
- *A kdo bi ga obsojal, dokler teoretiki zarot brusijo jezike o bistveno močnejših klanib, kakršna je recimo družina Bush?*

cerkven *pridevnik*

biti reven kot cerkvena miš

biti zelo reven

- *Res so revni kot cerkvene miši, toda pred njihovim domačim pragom jih ne bo nihče poniževal.*
- *Obtoženec je reven ko cerkvena miš, niti toliko denarja nima, da bi si plačal odvetnika.*

miš *samostalnik*

reven kot cerkvena miš

zelo reven

- *Revni so kot cerkvene miši, še najosnovnejših potrebščin si ne morejo privoščiti.*
- *Prepričan, da mu v Sovjetski zvezi ne bo uspelo, je zapustil dom, družino in reven kot cerkvena miš se je znašel sredi zahodnega sveta.*

Uvrščanje FE pod sestavine fakultativnih elementov ali sestavin besedilnega okolja, ki jih ne navajamo v osnovni obliki, je podrejeno statistični analizi take sestavine kot iztočnice v besedni skici, saj izhajamo iz parametrov, ki določajo različno intenziteto pojavljanja določene besede z drugo glede na iztočniško izhodišče. To pomeni, da npr. FE *ostrina britve* | *na/po ostrini britve*, ki se, gledano z vidika samostalnika *britev* ali *ostrina*, tipično pojavlja z glagoli *hoditi*, *stopati* ter samostalniki *hoja*, *življenje*, *preživetje*, ki smo jih navedli v elementu kolokacija, ne bo nujno navedena tudi pri glagolih *hoditi* in *stopati*. Nasprotno je FE lahko navedena tudi pri iztočnici, ki predstavlja njeno kolokabilno okolje, če je kolokabilna povezava dovolj izstopajoča in če nastopa kot iztočnica. Posamezne rešitve prikazujejo spodnji primeri:

čas *samostalnik*

v korak s časom | v koraku s časom

[stopati, iti, hoditi] v korak s časom

[biti, ostati] v koraku s časom

korak *samostalnik*

v korak s čim | v koraku s čim

[iti, hoditi, stopiti, morati, hoteti] v korak s [časom]

[biti, ostati, razvijati se] v koraku s [časom]

stopati *glagol*

stopati v koraku/v korak s čim

stopati v koraku z [modo]

stopati v korak s [časom]

Taka rešitev je v skladu z beleženjem ustaljenega dela FE in relativno spremenljivostjo njenega besedilnega okolja, ki glede na obravnavano besedo v iztočnici podaja informacijo v skladu s pomenskim izhodiščem, ki predstavlja fokus opazovanja FE v besedilu. Z vidika uporabnika, ki dostopa do informacij, strukturiranih v slovarski bazi, prek enotnega poizvedovalnega okenca na spletu, se podatki lahko prikazujejo glede na ključne besede v FE (tj. *korak*, *čas*) ali glede na elemente besedilnega okolja (tj. *stopati*, *iti*, *hoditi*; *moda* itd.).

11.2.3.2 Glede na pomen

V LBS FE pomensko nismo vključevali pod posamezne pomene besed, ki so zastopani s katero od sestavin FE v iztočnici. FE so z iztočnico povezane zgolj izrazno, kot smo prikazali v prejšnjem poglavju. Čeprav sicer povezave med posameznim pomenom sestavine in FE kot celote obstajajo in so v nekaterih primerih očitne, se za pomensko podrejanje nismo odločali, ker izkušnje pri izdelavi LBS kot tudi navedbe v literaturi kažejo, da temeljijo take pomenske povezave na osebnih jezikovnih in kulturnih izkušnjah vsakega posameznega govorca (Cowie 1998: 215). Podrejanje FE pod posamezne pomene iztočnice bi pomenilo ugotavljanje njihove pomenske motiviranosti, o čemer pravi Čermák (2001: 11) naslednje:

Kot je na eni strani očitno, da obstaja določena pomenska povezava med pomenom sestavine in celostnim frazeološkim pomenom zveze, je na drugi strani jasno, da je status take pomenske udeležbe lahko različen in ga je težko natančno določiti, posledično pa tudi zanašanje na pomensko analizo pri pomenski opredelitvi FE in njihovi klasifikaciji na pomensko motivirane in nemotivirane ni zanesljivo.

Pomensko podrejanje FE glede na pomen njihovih sestavin bi bilo tako za hierarhizacijo FE znotraj slovarske baze pa tudi znotraj slovarskih priročnikov nezanesljivo in nujno nekonsistentno, posledično pa zavajajoče za uporabnika. S tem ko FE ne uvrščamo pod posamezne pomene besed v iztočnici, ohranjamo tudi razliko glede na SZ, ki so vsaj deloma pomensko motivirane in zato ponekod uvrščene v pomenski razdelek katere od svojih sestavin.

Podobno kot pri strukturni in skladijskofunkcijski opredelitvi tudi na ravni pomena FE niso vedno jasno določljive in jih ni mogoče ostro ločiti od drugih tipov večbesednih enot, zlasti v razmerju do kolokacij, uvrščenih pod katerega od metaforičnih ali prenesenih pomenov besede v iztočnici, do SZ, ki so nastale po metaforični poti, in do samostojnih prenesenih ali metaforičnih pomenov. V LBS smo se podobno kot pri SZ odločali za možnost obravnave FE v različnih elementih geselske zgradbe hkrati, pač glede na izkazano stopnjo metaforičnosti zveze kot celote oz. glede na razpršenost/omejenost besedilnega okolja pri izražanju konkretnega pomena besede ali zveze.

FE – samostojni pomen

Ker beležimo FE v samostojnem razdelku LBS in ker frazeologija poleg strukturne ustaljenosti navadno vključuje figurativne, zlasti metaforične in metonimične nasledke katerega od pomenov besede, ki jo obravnavamo v iztočnici, smo se v posameznih primerih morali odločiti bodisi za samostojne prenesene pomenne iztočnice v relativno ustaljenem besedilnem okolju bodisi za vključitev zveze v frazeološki razdelek. Npr. v zgledu: *Ves čas so jim govorili: Stisnite, bolje bo! In so stisnili, zdaj pa ni nič bolje*, gre za skrajšano varianto FE: *stisniti zobe* – 'potrpeti', je odločitev leksikografa za vključitev samostojnega pomena besede stisniti v odnosu do FE stisniti zobe povezana s tem, v kakšni meri je ustaljeno tudi besedilno okolje, v katerem se glagol stisniti pojavlja v prenesenem pomenu. Pri tem je seveda pomembno, da je raba takega samostojnega pomena čim bolj enotna, in relativno pogosta, kar pa pogosto ni realno, kot prikazujejo tudi spodnji zgledi, kjer prvi sugerira pomen 'potrpeti', druga dva pa 'potruditi se':

- *Malo boste pač morali stisniti, pa bo kmalu bolje.*
- *Za september mi bo ostalo še osem izpitov, zdaj pa moram stisniti, da bo dopust v Portorožu slajši.*
- *Dejstvo je, da pravi profesionalci znajo stisniti, ko gre zares.*

V danem primeru smo se zato odločili za FE *stisniti zobe* s pomenom 'potrpeti', ob zavedanju, da so nekatere rabe lahko tudi eliptične (prvi zgled zgoraj), hkrati pa smo glagolu *stisniti* pripisali tudi samostojni pomen 'potruditi se', ki navadno kolocira z glagoli *morati*, *znati* in *treba je*.

Vzemimo za zgled še en pomen glagola *stisniti*, ki se realizira v zvezah kot: [*srce*] *se stisne komu*; v [*prsib*, *želodcu*, *grlu*] *stisne koga* in pri [*duši*, *srcu*] *stisne koga*, pri čemer so posamezni deli zveze zamenljivi znotraj ožjega pomenskega polja, npr. *srce*, *prsi*, *želodec*, *grlo*. Ker pa se glagol *stisniti* v istem pomenu dovolj pogosto rabi tudi

v neustaljenem skladenjskem in leksikalnem okolju, npr. *nič se mi ni smilil, ampak vseeno me je malo stisnilo*, mu lahko ob FE pripišemo še samostojni preneseni pomen: 'občutiti tesnobo, stisko'.

kolokacija – stalna zveza – FE

Razmejitev FE od kolokacij temelji na prepoznavanju zveze besed kot samostojne leksikalne enote, kar je sicer pogoj tudi za določanje SZ. Pri razmejevanju SZ glede na FE je zato potreben še nadaljnji korak, tj. ugotavljanje metaforičnosti pomena zveze kot celote. Raba FE je glede na druge večbesedne leksikalne enote torej v besedilu vedno zaznamovana, ekspresivna in omogoča med drugim tudi vzpostavljanje razmerja med dobesednim in prenesenim pomenom. Če je torej raba neke zveze v besedilu izkazovala poleg dobesedne (kolokacija) še preneseno rabo, smo jo obravnavali tudi kot FE, obe zvezi, tj. dobesedno in preneseno, pa smo uvrstili v ločene elemente geselske zgradbe, kot jih predvideva slovarska baza:

alkoholni hlap – hlap

- *Za njenim hrbtom je stal moški, ki ji je za vrat glasno izdihoval alkoholne hlap.*
- *Rutinski pregled bi minil brez posebnosti, če ne bi zavohali alkoholnih hlapov.*
- *Termometer na klobuku kotla rabi za merjenje oziroma nadzor temperature alkoholnih hlapov.*

alkoholni hlap – alkohol; alkoholna pijača – SZ

- *Po dolenski deželi opojnosti alkoholnih hlapov ne poznajo le odrasli, če verjamete ali ne, celo osnovnošolci.*
- *Morda v temi in bržčas pod vplivom alkoholnih hlapov ni prepoznal »krivca«.*
- *Predhodno zaužiti alkoholni hlap so me počasi opogumljali.*
- *Alkoholni hlap so domačinoma popolnoma zameglili razum.*

stalna zveza – FE

Razlikovanje med SZ in FE temelji primarno na stopnji metaforičnosti oz. dobesednosti razumevanja pomena zveze kot celote, kar smo dokazovali oz. preverjali z realno rabo:

alkoholni hlap – ometičnost pod vplivom alkohola; alkoholiziranost tudi odvisnost od alkohola – FE:

- *28-letnik se je kar brez vozniskega izpita, z neregistriranim avtom pa še z alkoholnimi hlap pod kapo¹⁴⁰ odpravil na vožnjo po domači deželi.*
- *V alkoholnih hlapih se mu je prikazala žena ...*
- *Njegovo otroštvo je bilo zadnje čase le še neprijazen svet alkoholnih hlapov, vpitja in nasilja.*

139 Dobesednost pomena izkazuje tudi širše besedilno okolje, kjer se pojavljajo v zvezi s *hlapi* besede kot: *izdihovati, vohati in merjenje temperature*.

140 Gre dejansko za združitve dveh FE s sorodnim pomenom: *imeti ga pod kapo* in *alkoholni hlap*.

12 Oznake

Oznake (ang. *labels*) oz. kvalifikatorji in kvalifikatorska pojasnila (SSKJ, SNB) so različni tipi pojasnil v slovarjih, ki uporabniku sporočajo, kako je besedo, besedno zvezo, njeno varianto ali posamezni pomen treba razumeti/uporabiti v besedilu glede na to, katere slovnične omejitve 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.

Če želimo odgovoriti na vprašanje, kaj so oznake, moramo glede na razvoj leksikografije, uporabniške potrebe in sodobne možnosti analize realne jezikovne rabe, ki jo zagotavljajo korpusi, upoštevati, da so se tovrstne informacije premaknile tudi na druge elemente slovarske mikro- in makrostrukture, kot sta denimo pomenska členitev in pomenski opis pa tudi drugi, do sedaj za te namene manj uporabljeni elementi slovarja, kot so različna opozorila, poudarki, napotki ipd., ki jih omogoča spletni medij.

Učinkovitost informacij o rabi je tesno povezana z upoštevanjem tega, komu so oznake namenjene, in sicer tako v izbiri besedišča, kot v načinu ubeseditve in izbiri oznak. Med različnimi uporabniki je smiselno upoštevati predvsem razliko med splošnim in šolskim uporabnikom na eni strani in učencem slovenščine kot tujega jezika na drugi ter zlasti to, ali je slovar namenjen dekodiranju, torej razumevanju, ali enkodiranju, torej tvorjenju besedila.

Za slovenščino različnih tipov oznak/kvalifikatorjev med seboj ne moremo primerjati, ker tako SSKJ kot SNB prinašata enak sistem kvalificiranja,¹⁴¹ lahko pa rečemo, da je v SSKJ leksika označena glede na knjižni jezik oz. normo, kot izhaja iz njegovih jezikovnosistemskih lastnosti.¹⁴² Ključen za prvo izdajo SSKJ (česar pa ne moremo več trditi za njegovo drugo izdajo; prim. Ahlin et al. 2014; Krek 2014) in dosledno izpeljan tako z geslovníkom kot sistemom kvalificiranja, je bil poudarek na kultiviranju slovenskega govorca, zato dejansko uporabljano besedišče takratnega časa v slovar ni bilo sprejeto zgolj na podlagi pogostosti, pač pa na podlagi presoje, ki je sledila kulturni in narodnopovezovalni vlogi knjižnega jezika. Zato je treba kljub predstavitvi različnih pojavnih oblik knjižnega jezika v SSKJ v zvezi s sistemom oznak, še toliko bolj pa s sprejetim besediščem, ki v SSKJ izhaja iz leposlovja 19. stoletja, opozoriti na drugačna izhodišča v LBS, kjer smo besede, zveze in pomene označevali glede na različne okoliščine sporočanja, torej predvsem

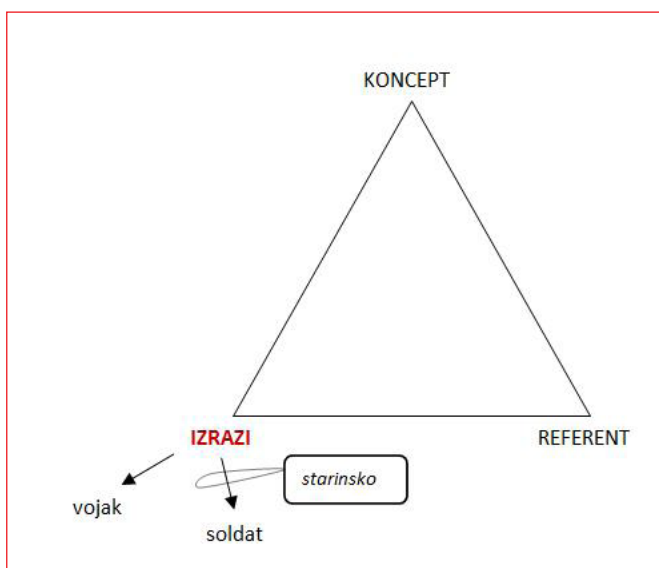
141 Za razliko od SNB ostaja izhodišče označevanja v 2. izdaji SSKJ, čeprav oba prinašata nekatere dopolnitve in spremembe (SNB: 34–38), glede na (jezikovnosistemsko) normo knjižnega jezika.

142 O razumevanju in uresničevanju normativno-informativne kodifikacije besedišča v SSKJ (in drugih slovarskih ter pravopisnih priročnikih za slovenščino) gl. Dobrovoljc (2014).

glede na doseganje stilnega učinka, različne lastnosti in posebnosti diskurza, zlasti ko je ta omejen na določeno področje, interesno skupino ipd. Poleg tega pa še glede na vrednotenje vsebine in udeležencev v sporočanju ter glede na sporočanski učinek besede, zveze ali pomena.

Pri uporabi oznak je pomembno poznati tudi razliko med označevanjem besede in označevanjem tistega, na kar se beseda nanaša.¹⁴³ Kot izpostavljata Atkins in Rundell (2008), si pri tem lahko pomagamo s pomenskim trikotnikom, ki ga ponazarja Slika 39, in v katerem so izpostavljeni naslednji pojmi:

- **referent** oz. **naslovnik**, ki zastopa resnično osebo ali stvar
- **koncept** ali **pojem**, ki zastopa to, na kar pomislimo, ko slišimo besedo,
- katerikoli **izraz**, ki se nanaša na ta koncept.



Slika 39: Povezanost pojmov koncept, referent in izraz pri označevanju besedišča

Izraza *vojak* in *soldat* se denimo nanašata na isti koncept (pripadnik vojske), na katerega pomislimo, ko slišimo ti besedi, in se lahko uporabita za poimenovanje različnih referentov (npr. sosedovega fanta, vojaka Ryana v filmu, neznanca v uniformi ipd.). Medtem ko je *vojak* nevtralen izraz, je *soldat* z vidika sodobnega jezika

¹⁴³ To razmerje določajo tudi pojmi *designat* (tj. spoznavni objekt), *denotat* (oz. denotativni pomen besede) in *leksem* (eno- ali večbesedni jezikovni izraz kot enota leksikalnega fonda jezika), ki jih uvaja strukturalno leksikalno pomenoslovje (prim. Vidovič Muha 2000: 29).

beseda, ki se je za pripadnika vojske uporabljala v preteklosti oz. v povezavi z določenimi vojaškimi formacijami ali vojnami (spodaj zgledi v (a)), zato lahko izraz *soldat* označimo npr. z oznako starinsko ali nekdanj, hkrati pa opozorimo, da se zaradi svoje časovne odmaknjenosti uporablja tudi za ustvarjanje stilnega učinka, kot spodaj v zgledih (b). Odločitev za oznako tipa starinsko, pa tudi za druge oznake, zlasti take, ki besedam dodajajo določeno stilno vrednost, mora biti povezana z analizo realne rabe, zlasti z opazovanjem diskurza, v katerem se tipično pojavlja, z opazovanjem učinka, ki ga dosega, in z opazovanjem neposrednega besedilnega okolja in besedilnega tipa, kot kažejo primeri iz korpusa Gigafida:

- (a)
- *Štiriletnemu sinu Jurčku so pod grožnjo, da bo sicer vržen v vodnjak, zabičali, da mora na vprašanje, kje je ata, odgovoriti: »V Stuttgartu je pri soldatih.« In res je tako odgovarjal vsakomur, tudi sosedom, ki so vedeli za resnico.*
 - *V pravo obsedenost pa se je sprevrglo iskanje vsega, kar je ostalo za soldati s soške fronte.*
- (b)
- *V tretjem delu nemški soldati Indiano in njegovega očeta lovijo z japonskimi motocikli namesto z nemškimi.*
 - *Varnost državljanov zagotavlja le splošna prepoved posesti orožja – razen jagrov, soldatov, policajev in čuvajev nihče ni upravičen do posedovanja orožja!*

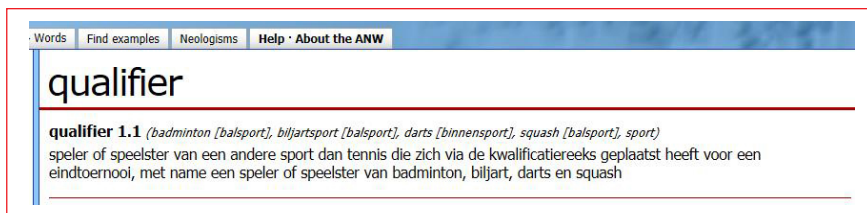
12.1 Oznake v slovarski bazi in slovarju

Kot smo izpostavili že v poglavju Leksikalna baza in slovar, je ena od poglavitnih značilnosti slovarske baze v odnosu do slovarja ta, da vsebuje več informacij, kot se jih dejansko prikazuje v slovarju. Sem sodi tudi beleženje jezikovne rabe v obliki oznak, kjer je med prednostmi, ki jih leksikografom ponuja slovarska baza, mogoče izpostaviti zlasti fleksibilnost in izčrpnost pri njihovem opisu in kategorizaciji ter možnost naknadnega grupiranja besedišča, opredeljenega z določeno oznako.

Fleksibilnost se kaže predvsem v možnosti razširitve prvotnega nabora oznak in relativne svobode pri njihovi ubeseditvi, saj se leksikografu ni treba togo držati seznama, če realna jezikovna raba kaže določene posebnosti, vezane le na posamezne pomene oz. njihove rabe. To sicer ne velja v enaki meri za vse tipe oznak, ampak predvsem za tiste, kjer jezikovna raba niha in kaže različne pomenske, stilne, pragmatične in druge omejitve, ki jih je težko ustrezno zajeti z vnaprej določenimi kvalifikatorji.

Ko govorimo o izčrpnosti, imamo v mislih težnjo, da se v fazi izdelave slovarske baze jezikovna raba opiše čim bolj podrobno, in sicer ne glede na to, kako in če sploh bodo ti opisi predstavljeni v slovarju. Na ta način leksikograf opozori na posebnosti v jezikovni rabi, odločitev o tem, ali in v kakšni obliki se bo podatek

prenesel v slovar, pa je stvar končne redakcije in načina njegove predstavitve (Gantar in Kosem 2013: 145). Dober primer takega postopka je beleženje podatkov o rabi v slovarski bazi, na kateri nastaja Splošni nizozemski slovar, kjer je beleženje področnih oznak zelo podrobno in poteka na več ravneh: geslo *qualifier* (sl. *kvalifikacije*), kot prikazuje Slika 40, ima oznako šport, ki je lahko še podrobneje členjena, npr. badminton, biljard, hkrati pa je ob vsaki ožji opredelitvi navedena tudi nadkategorija, npr. šport z žogo, dvoranski šport ipd.



Slika 40: Del gesla *qualifier* v slovarski bazi Splošnega nizozemskega slovarja glede na področne oznake

Podrobno beleženje oznak pa ne ponuja prednosti le leksikografom, temveč tudi slovarskim uporabnikom, saj lahko na podlagi podrobne kvalifikacije, ki ni nujno prikazana tudi v slovarju, omogočimo različna napredna iskanja v slovarski bazi.¹⁴⁴

Ena izmed lastnosti, ki odražajo prednost predhodno strukturirane elektronske slovarske baze, je označevanje področne rabe določene besede oz. njenega pomena na ravni pomenskega opisa, kjer smo pri snovanju LBS predvideli vključitev področne oznake neposredno v pomenski indikator:

```
<stalna_zveza>
<zveza>čelada za navidezno/virtualno resničnost</zveza>
<indikator><oznaka tip=<podrocje><računalništvo></oznaka></indikator>
<struktura></struktura><zgledi>
</zgled></zgledi>
</stalna_zveza>
```

stalna_zveza ☐

♦ čelada za navidezno/virtualno resničnost

indikator: računalništvo

struktura:

zgledi ☐ (3)

¹⁴⁴ Kot dober primer take prakse je mogoče navesti spletni enojezični angleški slovar Oxford, ki omogoča napredno iskanje po 15 različnih strokovnih področjih s številnimi podpodročji. Slovar je dostopen na spletni strani: <http://oxforddictionaries.com/>.

To pomeni, da končnemu uporabniku v slovarski bazi zabeležena oznaka ni nujno vidna, hkrati pa omogoča iskanje ali združevanje leksike z enakimi oznakami. Druga že omenjena prednost sodobne leksikografije je avtomatizacija in racionalizacija slovarskih postopkov, kjer je treba ugotoviti, katere leksikografske postopke je mogoče avtomatizirati, ter v kolikšni meri lahko računalnik pomaga leksikografu.

12.2 Avtomatsko pridobivanje oznak in opozoril o rabi iz korpusa

V LBS, ki temelji na korpusni analizi korpusnih podatkov, čim večji avtomatizaciji leksikografskih postopkov in je primarno zasnovana za izdelavo slovarja na spletu, se pridobivanje, beleženje ter prikazovanje podatkov o jezikovni rabi v marsičem razlikuje od tradicionalnih leksikografskih pristopov, vezanih na knjižni format. Velik del informacij tega tipa, zlasti slovničnih, ki so oblikovno prepoznavne, je namreč iz korpusa mogoče pridobiti avtomatsko. Rundell in Kilgarrieff (2011) pri razmišljanju o možnostih avtomatizacije različnih segmentov leksikografskega dela med drugim izpostavi, da je o avtomatskem beleženju oznak treba razmišljati že v času gradnje korpusa (prim. Kosem 2011). Eden od načinov avtomatskega luščenja oznak je namreč izdelava podkorpusov besedil, ki se uvrščajo pod obseg določene oznake, npr. za določeno strokovno področje, regijo ali časovno obdobje. Leksika, ki se pojavlja večinoma ali izključno v besedilih katerega od podkorpusov, pridobi nato v slovarski bazi oz. slovarju tudi z njim povezano oznako.

Nekoliko drugačen pristop zahteva avtomatsko luščenje slovničnih oznak, ki smo ga preizkusili v zaključni fazi izdelave LBS. Postopek je bil primarno namenjen luščenju slovničnih relacij, kolokacij in z njimi povezanih zgledov iz korpusa neposredno v programski vmesnik LBS, hkrati pa smo na ta način pridobili tudi podatke o obnašanju posameznih besed v njihovem besedilnem okolju in posledično o jezikovni rabi. Na eni strani gre za podatke, ki leksikografa opozarjajo na tipično sopojavljanje dane besede npr. s števili (*finalist*), količinskimi izrazi (*denar*) ali lastnimi imeni (*mesto*), tipično pojavljanje v povratnih se/si-stavčnih zgradbah (*izvoliti*), v nedoločniku (*izraziti*) ipd. Na drugi strani pa za informacije, ki jih je mogoče neposredno prepisati v slovarske oznake, npr. tipično v 3. os. (*boleti*), pogosto zanikano (*nakloniti*) ipd.

Preglednica 21: Gramatične relacije v orodju SkE za avtomatsko luščenje opozoril o jezikovni rabi iz korpusa

Tip odnosa v gramatični relaciji	Ime gramatične relacije v SkE	Iztočnica	Korpusni zgled
Constructions	O_z_lastnim_imenom	finalist	<i>Edini slovenski finalist Peter Prevc je osvojil 11. mesto.</i>
		mesto	<i>Maribor bo predstavil svojo kandidaturo skupaj s portugalskim mestom Guimarraes.</i>
	O_tretja_oseba	izvoliti	<i>"Ali se lahko razgledam naokoli?" "Seveda. Izvoli ."</i>
		boleti	<i>Mnogo ljudi je brez služb, kar jo boli.</i>
	O_povratni_se	izvoliti	<i>Slovenija je razdeljena na osem volilnih enot, v vsaki izmed njih se na volitvah izvoli enajst poslancev.</i>
	O_povratni_si	izvoliti	<i>H krofom si izvolite natočiti šilce borovničevca.</i>
Unary rels	O_s_števili	finalist	<i>Deset finalistov si bo naša komisija tudi ogledala in jih ocenila ter predstavila v Vestniku.</i>
	O_količina	denar	<i>Država ima namreč veliko gradov in malo denarja, zato sama ni najboljši skrbnik.</i>
	O_zanikanje	denar	<i>Igralka je namreč zatrdila, da ni denarja na tem svetu, ki bi jo premamil, da bi odvrгла vsa oblačila.</i>
		nakloniti	<i>Drug drugemu ne naklonijo nasmeha in prijazne besede.</i>
	O_nedoločnik_cs	izraziti	<i>Prizadeti starši smo se odločili temu upreti in izraziti svoj protest.</i>

Kot prikazuje Preglednica 21, smo tovrstne informacije pridobili iz korpusa prek gramatičnih relacij v slovnici besednih skic v orodju SkE, kjer posamezno relacijo v osnovi sestavljajo trije elementi: (procesna) direktiva, ki določa tip odnosa v relaciji (prvi stolpec), ime relacije (drugi stolpec) in poizvedba (tretji stolpec). Znotraj poizvedbe je mogoče skladenjske odnose definirati glede na en sam element, ki izpostavlja določen pojav v odnosu do vseh drugih elementov v korpusu. Za namene pridobivanja slovničnih oznak je glede na poizvedbo – tj. besedo v iztočnici – mogoče tako pridobiti podatek o tem, da je nek pojav, npr. sopojavljanje s količinskimi izrazi, zanikanje ipd., statistično izstopajoč (Krek 2010).

12.3 Moč oznak

Oznake obravnavamo kot pomenske informacije, zato se v LBS pripenjajo bodisi na pomenski indikator ali pomensko razlago določene leksikalne enote. To pomeni, da lahko oznake oz. več zaporednih oznak pričakujemo pri pomenih, podpomenih, stalnih zvezah in frazeoloških enotah. Izjema so slovnične oznake, ki bodisi kvalificirajo izrazno podobo iztočnice, tj. njeno samoedninsko obliko (oznaka *nešteвно*) ali samomnožinsko obliko (oznaka *množina*), in iztočnice, kjer oznaka (tako slovnična kot pomenska) velja za edini ali vse pomene besede v iztočnici.

Za pomenske oznake (tj. področne, diskurzne, stilne itd.), velja, da veljajo samo za konkretni pomen oz. podpomen. Če je s katero od pomenskih oznak označen krovni pomen, ki ima zgolj združevalno vlogo in ne vsebuje drugih elementov geselske zgradbe, velja oznaka za vse podpomene krovnega pomena. Sicer smo oznako pri podpomenih ponovili. Za slovnične oznake velja, da označujejo tako na ravni leme kot posameznega pomena samo negativne slovnične vrednosti, npr. *nešteвно*, v 3. osebi, oz. oznake, ki pomenijo odstopanja od osnovne oblike (npr. *množina*), ne pa obratno: oznake števno torej nismo uporabljali, ker razumemo števno kot inherentno lastnost samostalniških in pridevniških besed. Veljavnost in doseg posameznih oznak v geselski zgradbi LBS prikazuje Preglednica 22.

Preglednica 22: *Moč oznak*

Vrsta oznake	Oznaka	Mesto v LBS	Primer
področna		iztočnica	
		pomen/podpomen	avtonomija
		SZ- pomen	aktivno oglje
diskurzna		pomen/podpomen	matador, lajkati; kancler
		razlaga	naliti komu čistega vina; ikona
slovnična			
	nešteвно	pomen/podpomen	korespondenca, aplikacija
		iztočnica	divjačina, rukola
	množina	iztočnica	<i>spodnjice, tangice, hlačke</i>
	navadno v množini	pomen/podpomen	lisice
	navadno v velelniku	pomen/podpomen	iskati
	v 3. osebi	pomen/podpomen	boleti
	brezosebno	pomen/podpomen	klikniti
	ustaljena oblika	iztočnica	tešče, bore
		pomen/podpomen	otok
		v besedilni vlogi	jok
skladenjska	v povedni/prislovni/ pridevniški/predložni ipd. rabi	skladenjska struktura	jeans, dno; tipičen; tema, škoda; mimo

Vrsta oznake	Oznaka	Mesto v LBS	Primer
konotacijska	preneseno	pomen/podpomen	grajski, rujen; latino; barva, aktiven; naprodaj
		razlaga	potrošno blago
		pomen/podpomen	avtomatično, dedek
registrska	pragmatična	pomen/podpomen	izvažati, arhivirati
		pomen/podpomen	tip, topogleden
časovna		pomen/podpomen	soldat
		lema/iztočnica	

Moč oznake je smiselno omejiti na posamezni pomen oz. podpomen tudi zaradi možnosti prikaza posameznega (pod)pomena, ki ga v spletni postavitvi v meniju izbere uporabnik, izolirano od drugih (pod)pomenov. Taka možnost namreč predvideva, da se uporabnik osredotoči zgolj na informacije, ki so vezane na izbrani (pod)pomen, SZ ali FE.

12.4 Vrste oznak v LBS

V LBS smo jezikovno rabo beležili v obliki oznak, za katere smo predvidevali, da jih bo v spletni različici slovarja mogoče prikazovati bodisi v obliki krajših, čim bolj razumljivih opozoril, z različnimi možnostmi vizualizacije (prim. Krek et al. 2013: 35) ali kot del stavčnih razlag. Tako prakso poznajo zlasti slovarji, namenjeni učenju tujega jezika, ki večino pragmatičnih in skladenjskih elementov pomena vključujejo v razlago, zaradi česar uporabnik lažje uzavesti rabo besede ali njenega pomena (Barnbrook 2002: 178). Podobne ugotovitve veljajo za navajanje slovničnih podatkov, saj ti pogosto zaradi prezahtevnega metajezika ali krajšanj ne dosežejo uporabnikov (Cowie 1981: 206; Summers 1987: F8).

Oznake smo v LBS razdelili na več tipov: področne, diskurzne, slovnične oz. skladenjske, stilne, pragmatične in časovne. Nekateri tipi imajo dokaj omejen, drugi pa precej bolj odprt nabor oznak. V primeru stilnih, konotativnih in pragmatičnih lastnosti besed, zvez ali posameznih pomenov to pomeni, da ima leksikograf možnost svoja opažanja poljubno ubesediti glede na to, kaj razbere iz analize jezikovne rabe.

12.4.1 Področne oznake

Največ besedišča je v LBS označenega s t. i. področnimi oznakami, ki smo jih uporabili v zvezi z besedami, SZ ali njihovimi pomeni, pri katerih je bilo glede na

analizo gradiva mogoče ugotoviti omejitev rabe na določeno strokovno, interesno ali kako drugače specializirano področje:

blokirati *glagol*

1 omejiti gibanje

1.2 zaustaviti udarec

šport

blokirati [strel, napad]

blokirati z [roko]

cepiti *glagol*

4 izboljšati kakovost

4.1 o rastlinah

sadjarstvo

cepiti [drevje, trto]

cepiti s [cepiči]

aktivna varnost

avtomobilizem

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

Seznam področij in podpodročij ter oznake zanje smo predvideli na podlagi obstoječih taksonomij,¹⁴⁵ hkrati pa smo področja dodali, če se je ob pomenski analizi pokazala taka potreba.

V nekaterih primerih so določena beseda, zveza ali kateri od njenih pomenov rabljeni na več različnih specializiranih področjih. Če je taka raba vezana na iste pomenske vsebine, smo navedli več ločenih zaporednih oznak, kjer gre navadno za opredelitev širšega in ožjega strokovnega področja (*klavzula*), sicer smo besedo ali zvezo obravnavali kot samostojno leksikalno enoto, torej z različnimi pomenskimi opredelitvami, ustreznimi različnimi oznakami in drugimi deli geselskega članka, zlasti kolokacijami in zgledi (*topla greda*):

devizna klavzula

finance

bančništvo

topla greda | zaprta greda

vrtnarstvo

topla greda

ekologija

Pri opredeljevanju leksikalnih enot s področnimi oznakami, še posebej v primerih, ko smo področno omejenost pomena besede prenesli v indikator, npr.

¹⁴⁵ Za slovenščino je izdelan osnovni seznam terminoloških kvalifikatorjev v SSKJ in je slovarskemu uporabniku že znan, zato je smiselno osnovne kategorije ohraniti in jih nadgraditi s področji, ki za SSKJ-jevsko leksiko še niso bili aktualni.

panoga *samostalnik*

1 sorodne dejavnosti

1.1 v gospodarstvu

1.2 v športu

belina *samostalnik*

1 bela barva

2 pri fotografiranju,

smo bili pozorni, da označenosti pomena besede, zlasti ko ta vključuje tudi določene konotativne elemente, ki izhajajo iz prenosa vsebin na drugo področje, nismo prenesli na raven samostojnega (pod)pomena, čeprav tipično besedilno okolje kaže na omejitve rabe:

jokati *glagol*

3 oddajati neprijeten zvok

glasba

če rečemo, da GLASBENI INSTRUMENT joka, se zdi, kot da oddaja neprijeten, joku podoben zvok

mrtvi kot

avtomobilizem

nevidna točka ali položaj, ki nastane ob prekrivanju različnih zornih kotov, npr. pri gledanju v vzvratno in stransko ogledalo v avtomobilu

Pri označevanju leksike s področnimi oznakami smo bili pozorni tudi na to, da smo v primeru hkrati izkazanih neterminoloških in terminoloških rab – zadnje zlasti v specializiranih besedilih, ki so zastopana v korpusu, besedo označili s terminološkim kvalifikatorjem samo v strokovnih pomenih, ne pa tudi v splošno rabljenih, npr.

bazičen *pridevnik*

1 o kemični sestavi; alkalen

kemija

2 osnovni; temeljni

bazična ocena

gimnastika

bazični šport

šport

bazična industrija**bazična znanost****bazična raziskava****bazično cepljenje**

zdravstvo

bazična točka

bančništvo

Enopomenskih besed ali besed, ki jih je bilo glede na njihov osnovni pomen mogoče umestiti na konkretno (strokovno) področje, a se hkrati nezaznamovano rabijo v splošnem jeziku, npr. *nafta, fermentacija, glavkom, glavnica, ciroza, kurikulum*, načeloma nismo označevali s področnimi kvalifikatorji. Njihova vezanost na določeno strokovno področje je navadno razvidna iz pomenskega indikatorja in/ali razlage:

fermentacija *samostalnik*

biokemični proces

biokemični proces, pri katerem ob odsotnosti ali prisotnosti kisika mikroorganizmi razgrajujejo organske snovi, kar povzroči nastanek alkohola, bioplina, kisanje, oksidacijo ali zorenje

kurikulum *samostalnik*

izobraževalni program; učni načrt

izobraževalni program, ki vključuje predvidene učne cilje, vsebine in metode učenja v šolskem sistemu

Splošnejšo oznako strokovno smo uporabili za označevanje izraza, za katerega obstaja v splošni rabi poljudnejši izraz, ki smo ga zapisali v indikator, npr.

biceps *samostalnik*

strokovno

nadlaktna mišica

12.4.2 Diskurzne ali kontekstualne oznake

Z diskurzni ali kontekstualni oznaki smo za razliko od vezanosti na določeno strokovno področje v LBS označevali omejitve rabe besede oz. katerega od njenih pomenov na tipične sporočanje situacije (diskurze) ali besedilne tipe, pri čemer omejenosti rabe na diskurz pogosto ni bilo mogoče nedvoumno ločiti od strokovnega oz. interesnega področja. Za kontekstualne oznake smo se tako odločali predvsem v primeru, če so tako rabljene besede izkazovale določen delež konotativnih pomenskih elementov. Z vidika obstoječih funkcijskozvrstnih taksonomij (npr. v SSKJ in SNB) se ta tip oznake vsaj deloma prekriva z oznako žargonsko,¹⁴⁶ pri čemer v LBS na ta način označena leksika nima enotne oblike in lahko nastopa tudi kot sestavni del indikatorja (*kloniti*), npr.

¹⁴⁶ Z oznako žarg. so v SSKJ označene besede, pomeni ali zveze »iz govora posameznih stanov« (SSKJ: XXI, § 141). V SNB kvalifikator žargonsko nastopa skupaj z ustreznim terminološkim kvalifikatorjem.

matador <i>samostalnik</i>	kloniti <i>glagol</i>
2 najvidnejši predstavnik	v športu
zlasti v športu	[vratar, moštvo, košarkarji, obramba, ekipa] klone
[jadralski, teniški] matador	kloniti proti [prvaku, moštvu, zmagovalcu, tekmeču]
	kloniti v [dvoboju, končnici, derbiju, tekmi]
bobnati <i>glagol</i>	lajkati <i>glagol</i>
3 veliko govoriti	1 označiti z znakom za dvignjeni palec; všečkati
v političnem kontekstu	v družabnih omrežjih
bobnati v [parlamentu, vladi]	

Poseben status med kontekstualnimi smo namenili oznakam, ki opredeljujejo besedilni tip (*žlička, zbor*), ki je za rabo določenega pomena značilen, in oznakam, ki določajo kulturne, zgodovinske ali družbeno-politične okoliščine, znotraj katerih se uporablja določena beseda ali njen pomen (*ikona, pisanica, Veliki/veliki brat*). Taka pojasnila so sicer lahko vključena tudi v razlago (*kancler, ikona*) in postanejo tako njen sestavni del:

žlička <i>samostalnik</i>	zbor <i>samostalnik</i>	kancler <i>samostalnik</i>
1 del jedilnega pribora	1 skupina ljudi	predsednik vlade; premier
1.1 merska enota	1.5 skupina igralcev	v Nemčiji in Avstriji predsednik vlade
v kuharskih receptih	zlasti v grški drami	
ikona <i>samostalnik</i>		
2 religiozna slika		
v vzhodnoevropski in ruski pravoslavni cerkvi <i>religiozna podoba ali slika</i>		
Veliki/veliki brat Veliki brat te opazuje!		
po romanu G. Orwella »1984« <i>izraža popoln nadzor države ali ustanove nad posamezniki</i>		

H kontekstualnim oznakam štejemo tudi vezanost besede ali pomena na določeno predmetnost ali tematiko (*izsušen*). Take oznake so v LBS lahko tudi del indikatorja, saj omogočajo ločevanje posameznih pomenov že na ravni pomenškega menija, pri čemer smo v elementu <oznaka> ohranili atribut, ki določa tip oznake (*koalicija*):

izsušen *pridevnik*

1 ki ima premalo vode; dehidriran

o organizmu

izsušena [sluznica, koža, polt]

izsušene [ustnice]

izsušeno [grlo]

2 ki nima vode

o vodnih virih

izsušena [struga, puščava, zemlja, njiva]

izsušeno [močvirje, korito, jezero, blato]

koalicija *samostalnik*

1 zveza

1.1 o političnih strankah

1.2 o državah

<indikator><oznaka tip="kontekst">o
političnih strankah</oznaka></indikator>

<indikator><oznaka tip="kontekst">o
državah</oznaka></indikator>

12.4.3 Slovnice oznake

S slovnimi oznakami smo v LBS označevali omejitve besede ali pomena glede na (a) skladenjsko vlogo, npr. v samostalniški rabi (latino), v povedni rabi (balzam, tema); (b) določeno slovnično obliko, npr. navadno v množini (*brzica, ikra*), neštevno (*baseball*), in (c) glede na določeno vrsto zapisa, vezano na upovedovalno možnost, npr. navadno v velelniku (*klikniti, paziti*), ustaljena oblika (*tešče, bore*), za oba spola (*zborovodja, računovodja*) in (č) glede na pravopisne posebnosti.

12.4.3.1 Skladenjske oznake

S skladenjskimi oznakami smo označevali tipična odstopanja v pričakovani skladenjski vlogi določenega pomena besede, zato so te oznake tesno povezane z označevanjem besedne vrste na ravni iztočnice. Odločitve, ki smo jih pri označevanju konverzije besednih vrst sprejeli v LBS, smo podrobneje opisali že v istoimenskem poglavju, zato bomo na tem mestu posamezne odločitve le povzeli.

Skladenjske oznake, ki smo jih uporabljali v LBS, so (deloma z različnimi ubeseditvami) štiri: v samostalniški rabi, v pridevniški rabi, v povedni rabi in v prislovni rabi. V nadaljevanju jih opišemo glede na to, pri kateri besedni vrsti jih navajamo, kje v strukturi gesla jih zapisujemo in kaj z njimi označujemo.

(a) Pri samostalniki

Z oznako v pridevniški rabi smo označevali pomen samostalnikov, ki so prevzeli pridevniške pomenskoskladenjske vloge (*kaki*). V primeru, da sprememba skladenjske vloge ne vpliva na pomen, smo na ustrezno besedno vrsto opozorili v elementu <restrikcija> (*evro*):

kaki *samostalnik*

1 sadno drevo

1.1 sadež

2 barva ali tkanina

2.1 temno zelene barve

v pridevniški rabi

kaki [hlače]

[hlače, obleka] kaki barve

evro *samostalnik*

1 denarna enota

[sto, tisoč, milijon, milijarda] evrov

<struktura>SBZ1<r>**v pridevniški rabi**</r>sbz0</struktura>

evro [območje]

Na povedno rabo samostalnikov, kot v primerih *škoda*, *mraz*, *tema je*, *kaj je* (*pravi*) *balzam*, *blagor* ipd. opozarjajo v LBS skladenjske strukture z veznim glagolom biti in pripadajoče kolokacije, npr. Vez-gbz SBZ1: [biti, postati] avtoriteta, legenda, inspiracija, ikona. Pomensko osamosvojene rabe, kot npr. *ni vrag*; *dejstvo*, *absurd je*, *da ...*, so navadno obravnavane kot FE s prepoznavno pragmatično komponento in specifično besedilno funkcijo (npr. poudarno, v govoru). Redkeje smo oznako v povedni rabi uporabljali neposredno za indikatorjem konkretnega samostalniškega pomena, saj ustrezno skladenjsko vlogo vključuje tudi pomenska shema:

vonj *samostalnik*

1 zaznavna lastnost

1.1 dišava; parfum

1.2 občutek, lastnost

navadno v povedni rabi z glagoloma imeti in dajati

če ima PROSTOR ali PREDMET vonj po ZNAČILNOSTIH nekega ČASA ali DOGAJANJA, se to v njem na poseben način izraža ali občuti

Nekateri samostalniki v določenih pomenih prevzemajo funkcijo prislovov. Na take položaje smo opozarjali v restrikciji v skladenjskih strukturah in v pripadajočih kolokacijah, npr. gbz na SBZ4<r>v prislovni rabi</r>: [priplavati] na vrh, [spustiti se] na dno ipd.

(b) Pri pridevniki

Z oznako v samostalniški rabi smo označevali pomen, ki so nastali s prehodom pridevnika v samostalniško vlogo (*naglušen*, *racionalen*, *zapostavljen*) in posamezne pomenesklonljivih pridevnikov (*latino*, *afro*), kadar prevzamejo skladenjsko vlogo samostalnikov:

naglušen *pridevnik*

1 ki ima okvaro sluha

1.1 kdor ima okvaro sluha

v samostalniški rabi

gluhi in naglušni Slovenije

neprilagojen *pridevnik*

1 ki se ne obnaša v skladu s pravili

1.1 kdor se ne obnaša v skladu s pravili

v samostalniški rabi

Tak prehod navadno vpliva tudi na pomen obravnavane besede, vendar pa je pri posameznih pridevnikih različno pogosto zastopan. Če je torej taka raba vezana zgolj na konkretno skladiščno okolje, smo na to opozorili v restrikciji:

alkoholen *pridevnik*

1 o pijači

Kol-rbz PBZ2<r>

v samostalniški rabi</r>

[nekaj, nič] alkoholnega

revolucionaren *pridevnik*

1 pomemben, nov

Neg-GBZ PBZ2<r>

v samostalniški rabi</r>

ne prinašati [nič, kaj] revolucionarnega

Z oznako v povedni rabi smo pri posameznih pridevniških pomenih označevali tipično rabo pridevnika v povedkovem določilu, ki je razvidna tudi iz pomenske sheme:

pasji *pridevnik*

3 skrajno neprijeten

3.1 nesramen

v povedni rabi

če je ČLOVEK pasji, je neprijazen in v odnosu do drugih ljudi neprizanesljiv in zahteven

kriv *pridevnik*

1 odgovoren za kaj

v povedni rabi

če je ČLOVEK kriv, je odgovoren za nastanek NEGATIVNE SITUACIJE

Kadar predstavlja raba v povedkovem določilu le eno od skladišjskih možnosti, jo izkazuje skladišjska struktura s pripadajočimi kolokacijami in zgledi:

ambiciozen *pridevnik*

1 ki želi uspeti, napredovati

Vez-gbz PBZ1

[biti, postati, ostajati] ambiciozen

1.1 zahteven, velikopotezen

sbz1 Vez-gbz PBZ1

[načrt, cilj] je ambiciozen

(c) Pri prislovih

Tipično rabo prislovov v povedkovem določilu smo označevali neposredno za pomenskim indikatorjem z oznako v povedni rabi ali v restrikciji na ravni strukture, navadno pa je razvidna tudi iz stavčno ubesedene razlage v pomenski shemi:

zunaj *prislov*

1 na prostem

RBZ Vez-gbz sbz1<r>v povedni rabi</r>

- *Zunaj je lep dan, prečudovito moder ponedeljek.*
- *Zunaj je zima, v zakurjeni sobi pa sedemo za pogrnjeno mizo.*
- *Zunaj je še tema, jaz pa lačen kot volk.*

2 izven notranjosti

1.1 o rojstvu otroka

v povedni rabi

če je OTROK zunaj, se je ravnokar rodil

3 na površini

v povedni rabi

če je ZOB ali RASTLINA zunaj, je toliko zrasel, da je prišel na površje in ga je mogoče videti

Spremembo skladijske vloge prislovov označuje tudi opozorilo v predložni rabi, ki ga pripenjamo na ustrezni del skladijske strukture v elementu restrikcija:

mimo *prislov*

2 v smeri ob čem

RBZ<r>v predložni rabi</r>SBZ2

mimo [gola, koša]

mimo [stojnic, ruševin]

Posebna med skladijskimi oznakami je oznaka *v besedilni vlogi*, s katero smo označevali FE in pomene, ki so skladijsko gledano zaključena besedila, pomensko pa najpogostejše besedilni oz. znotrajstavčni modifikatorji, kar tako označenim, zlasti samostalniškim in prislovnim pomenom, daje členkovno vrednost:

jok *samostalnik*

1 izražanje žalosti z glasom in solzami

1.1 glas tistega, ki joka

1.2 zvok, podoben joku

2 za izražanje zanikanja

v besedilni vlogi

- *Glavni problem je v vas samih in nikar ne mislite, da je vaša žena kriva. Jok, ni!*
- *Kot invalid na vozičku si zaželite obiskati Ljudsko knjižnico. Jok, ne bo šlo.*

in pika | pa pika

izraža odločnost, dokončnost povedanega

v besedilni vlogi

- *Jaz grofa zagotovo nisem ustrelila, nisem in pika.*
- *Pa kaj potem, če ste strokovnjaki na svojem področju, najboljši policisti in psihologi, ne boste me ujeli, pa pika!*

12.4.3.2 Oblikoskladenjske oznake

Z oblikoskladenjskimi oznakami smo v LBS označevali posebnosti v slovničnih kategorijah, kot sta npr. števnost ali število, omejenost rabe besede ali njenega pomena na določeno slovnično obliko ter glede na vrsto zapisa, ki jo preferira konkretna upovedovalna možnost. Različne tipe slovničnih oznak ločimo v izhodišču glede na to, ali kvalificirajo sam izraz, tj. lemo, ali njen posamezni (pod)pomen.

(a) Slovnične oznake na ravni leme

Ker dojemamo oznake kot pomenska opozorila, s slovničnimi oznakami kvalificiramo načeloma le tiste omejitve ali odstopanja od paradigme, ki vplivajo tudi na spremembo pomena. Kadar se določene slovnične omejitve nanašajo na izraz, tj. lemo, je taka omejitev razvidna iz osnovne oblike besede v iztočnici ali pa je zapisana v elementu <ustaljena_oblika> neposredno za iztočnico. Iz tega razloga se tudi ne odločamo za opredeljevanje množinskih samostalnikov,¹⁴⁷ kadar je njihova osnovna oblika bodisi samoedninska (*divjačina, solata, očala, grozdje, srebrnina, rukola*) bodisi samomnožinska (*grablje, tangice*), ker pri tem sledimo odločitvam, ki so bile sprejete pri oblikoskladenjskem označevanju korpusa. Nas pa z vidika uporabnika v teh primerih zanima, v kateri od možnih oblik znotraj oblikoslovne paradigme se je osnovna oblika ustalila, saj uporabnik ne pozna nujno odstopanj

¹⁴⁷ Slovenska slovnica (Toporišič 2000: 271/272) kot množinske samostalnike, ki pa imajo navadno le edninsko obliko, redkeje samo množinsko, šteje snovna, skupna in pojmovna imena. Za slovarski opis je pomembno dejstvo, da niso nujno vsa skupna, pojmovna in snovna imena zastopana z množinskimi samostalniki, ampak tudi s troštevskimi, npr. *jajce* (prim. Hočevar Gregorič 1995: 442), in da zmožnost oz. nezmožnost štetja pomeni pogosto tudi pomensko spremembo (dobrota – dobre ipd.).

od sistema, npr. *tangica, *spodnjica ali *rukole. V takih primerih smo iztočnico v osnovni obliki označili z oznako *množina* že v t. i. zaglavju:

tangice *samostalnik*
množina
žensko spodnje perilo

spodnjice *samostalnik*
množina
spodnje perilo;
spodnje hlačke

grablje *samostalnik*
množina
vrtnarsko orodje

V primeru osnovne edninske oblike na ednino nismo posebej opozarjali, saj jo štejemo kot nevtrarno število osnovne oblike, smo pa pri enopomenskih množinskih samostalniki s samoedninsko obliko (*baseball*, *rukola*, *krama*) ali če do sprememb v zvezi s števnostjo ni prišlo pri nobenem od registriranih pomenov, uporabili oznako *neštevno* že na ravni iztočnice (*divjačina*, *ženskost*, *srebrnina*):

baseball *samostalnik*
neštevno
športna igra s kijem in žogo

rukola *samostalnik*
neštevno
rastlina

krama *samostalnik*
neštevno
neuporabni predmeti

divjačina *samostalnik*
neštevno
1 divje živali
1.1 divjačinsko meso

ženskost *samostalnik*
neštevno
1 lastnosti, bistvo žensk
1.1 žensko spolovilo

srebrnina *samostalnik*
neštevno
1 izdelki iz srebra
1.1 državno premoženje

(b) Slovnice oznake na ravni (pod)pomena

S slovarskega vidika je zanimiva možnost prehajanja množinskih samostalnikov iz svoje običajne (edninske ali množinske) oblike v needninsko oz. edninsko, še zlasti, ko je s tem povezana tudi sprememba pomena. V LBS smo za označevanje takih sprememb predvideli v elementu <oznaka> opozorila, ki sporočajo, da obstaja v določenem pomenu omejitev v oblikoslovni paradigmi. Ta opozorila so: neštevno, (navadno) v ednini, (navadno) v množini, samo v množini ipd. Z vidika mesta zapisa v zgradbi geselskega članka smo tovrstna opozorila vedno vezali na posamezne (pod)pomene:

kuža *samostalnik*
pes
navadno v ednini

brusnica *samostalnik*
1 rastlina
1.1 plod
navadno v množini

ligenj *samostalnik*
1 morska žival
1.1 jed
navadno v množini

Podoben način označevanja izraza na ravni iztočnice in pri posameznem (pod) pomenu smo uporabili pri glagolih, ki se uresničujejo v 3. osebi v vseh zabeleženih pomenih (*boleti*) – oznaka: v 3. osebi je zato vezana neposredno na iztočnico. Če gre za enopomenske besede (*daniti se*) ali pa je omejitev vezana samo na katerega od zabeleženih (pod)pomenov, smo oznako zapisali neposredno za indikatorjem (*klikniti*, *begati*). V spodnjih primerih pomene, ki se realizirajo v 3. osebi, ponazarjamo še s stavčnimi vzorci:

boleti *glagol*

v 3. osebi

1 o telesni bolečini

1.1 čutiti bolečino

kaj boli koga

boli koga

boli koga v čem

1.2 povzročati bolečino

kaj boli

boli koga

kaj boli koga

2 o duševni bolečini

boli koga

kaj boli koga

daniti se *glagol*

1 delati se dan; svitati se

v 3. osebi

dani se

kaj se dani

klikniti *glagol*

1 aktivirati z računalniško miško

2 ustvariti zvok

3 dojeti

v 3. osebi

klikne komu

klikne komu, da

4 postati si všeč

v 3. osebi

klikne med kom

kaj klikne med kom

klikne med kom in kom

begati *glagol*

1 vznemirjati, zanimati

v 3. osebi

kaj bega koga/kaj

Med oblikoskladenjska spada tudi opozorilo, da se beseda v iztočnici realizira zgolj v določeni zvezi (stežaj). Tako realizacijo smo zapisali v element <ustaljena_oblika> neposredno za indikator:

stežaj *samostalnik*

razdalja v obsegu raztega
na stežaj

<indikator> razdalja v obsegu raztega</indikator>

<ustaljena_oblika>na stežaj</ustaljena_oblika>

Med oblikoskladenjskimi oznakami obravnavano tudi opozorila, ki so vezana na določeno vrsto zapisa, ki jo preferira konkretna upovedovalna možnost pomena, npr. navadno v velelniku (*paziti*), navadno zanikano ali v nikalni rabi (*ljubiti*), za oba spola (*zborovodja*), in oznake, ki se nanašajo na pravopisne posebnosti, povezane s posameznim pomenom (*mati*):

paziti *glagol*

1 imeti pod nadzorom; skrbeti za kaj

1.3 o lastnini

navadno v velelniku

- *Pazite na osebne stvari, hišna vrata imejte zaprta.*

ljubiti *glagol*

2 biti pripravljen narediti

navadno zanikano

- *Tukaj sedim in ti kradem čas, ker se mi ne ljubi delati.*
- *Ni mu bilo jasno in ni se mu ljubilo razmišljati o tem.*

zborovodja *samostalnik*

vodja pevskega zbora

za oba spola

- *Pri tem je še kako pomembno sodelovanje s skladatelji, je povedala zborovodja, ki z zborom sodeluje od njegove ustanovitve.*
- *Posebna priznanja je vsem dosedanjim še živječim zborovodjem podelilo domače kulturno društvo.*
- *Leta 1983 je pevski zbor prevzel njihov sedanji zborovodja, ki je uspešen vodja že natanko 20 let.*

mati *samostalnik*

2 božja mati

navadno z veliko začetnico

- *Tolaži nas, Mati, v nevarnosti zdaj, ob smrti odpri nam preblaženi raj.*
- *Letos spomladi naj bi se bila Marija prikazala znova in aprila so v vasi zato postavili kip Matere božje.*
- *Naj se slovenski kristjani še naprej zatekamo k svoji duhovni Materi.*

Za vse tipe slovničnih oznak velja, da z njimi kvalificiramo besede oz. pomena, če se omejitve realizirajo v večini primerov.

12.4.4 Stilne oznake

Stilne oznake so namenjene označevanju tistih pomenskih lastnosti besede, s katerimi dosegamo določen stilni učinek. Gre za lastnosti, ki so bodisi vsebovane v izrazu samem (npr. *kozlati*, *kuža*) ali pa njegovi pomenski vrednosti dodajajo vrednotenjske, asociativne ipd. vsebine, katerih razumevanje je za govorca pomembno, da jih v sporočanjski situaciji razume in uporabi na način, kot je v jeziku običajen. V LBS smo stilne oznake razdelili na tri relativno samostojne skupine: (a) konotacijske, (b) registrske in (c) pragmatične. V nadaljevanju jih opišemo glede na vsebinske lastnosti, ki jih določajo.

12.4.4.1 Konotacijske oznake

S konotacijskimi oznakami smo vrednotili zaznamovanost pomena besed in zvez, ki izhaja iz govorčevega vrednotenja ubesedene predmetnosti. Zapisa tovrstnih oznak nismo vnaprej določali, tako da so imeli uredniki pri ubeseditvi relativno proste roke. V grobem jih lahko ločimo na take s pozitivnim in na take z negativnim vrednotenjem, v strukturi gesela pa lahko nastopajo kot del razlage:

potrošno blago

z neodobravanjem ljudje, dejavnosti ali predmeti, ki se obravnavajo kot neakovostni in hitro pokvarljivi

avtomatično *prislov*

3 nezavedno; refleksno

3.1 rutinsko

slabšalno ali z neodobravanjem

če človek DELA ali NAREDI kaj avtomatično, o tem ne razmišlja in temu ne namenja pozornosti ali posebnega truda

dedek *samostalnik*

1 oče matere ali očeta

1.1 začetnik

1.2 starec

izraža naklonjenost

- *Neki stari dedek naju napoti v hotel.*
- *V hišici je živel star dedek z dolgo belo brado.*

V LBS smo uporabljali tudi oznako preneseno, ki napoveduje, da ob izhodiščnem pomenu obstaja še metaforični (tudi metonimični) pomen, ki je z izhodiščnim

tako ali drugače asociativno povezan. Vsebuje torej določene pomenske lastnosti, ki omogočajo prevrednotenje pomenskih vsebin na način, da dobijo novo pomensko vrednost. Na primer, glagol *mleti* v svojem najosnovnejšem pomenu pomeni 'drobiti kaj, zlasti zrnje ipd., na majhne delce', pri čemer izhodiščni pomenski koncept vključuje tudi predstavo o ponavljajočih se krožnih gibih, npr. v povezavi z mlinskim kamnom ali kolesom, ki vzbuja konotacijo dolgotrajnosti in ponavljanja: *sodni mlini počasi meljejo, birokracija melje počasi* ipd. Asociacija drobljenja večjih kosov v manjše pod težo mlinskega kamna pa omogoča tudi razumevanje primerov kot: *evropski prvaki meljejo tekmece* – pomen, ki se je uveljavil zlasti v športnem kontekstu, *ameriško kolesje smrti neutrudno melje naprej* ipd. Uporaba pomenskih lastnosti enega pomena pri nastanku drugega je do neke mere vključena v vsak pomen, ki je nastal po metaforični ali metonimični poti (prim. Snoj 2010: 67), zato bi bila oznaka upravičena pri veliko večjem številu pomenov večpomenskih besed, saj gre za enega najproduktivnejših pomenotvornih postopkov. V LBS smo z oznako preneseno označevali predvsem tiste pomene, kjer se pomenskim lastnostim izhodiščnega pomena dodajajo konotativne prvine, ki ključno določajo pomenski prenos, njegove pomenske vrednosti pa ni mogoče opredeliti s katero od stilnih oznak, kot so npr. ljubkovalno, slabšalno, z negativnim vrednotenjem ipd. Pogosto gre za pomenske rabe, ki prihajajo v splošno leksiko prek najbolj vplivnih besedil, zlasti tistih, ki se (ob tisku) širijo po tudi po elektronskih in spletnih medijih. Kot take so te oznake v slovarski bazi namenjene predvsem leksikografom in mu pomagajo pri pomenski členitvi besede in njenem opisu:

izvažati *glagol*

1 prodajati blago v tujino

1.1 prenašati značilnosti česa v druga okolja

preneseno

izvažati [demokracijo, stabilnost]

izvažati [terorizem]

arhivirati *glagol*

1 shraniti

2 zabeležiti

3 prenehati se ukvarjati

preneseno

- *Tekmo z Ukrajino je počasi že arhiviral in misli usmeril k Belgijcem, ki imajo po njegovem mnenju več igralcev z izkušnjami v evropski ligi in drugih tekmovanjih.*
- *Idejo o neodvisnem poročanju iz kriznega žarišča sem hitro, čeprav nerad arhiviral.*

migrirati *glagol*

2 prestaviti se iz ene dejavnosti ali oblike v drugo
preneseno

če rečemo, da LJUDJE, njihova DEJAVNOST ali VSEBINE migrirajo, hočemo povedati, da prehajajo iz enega OKOLJA, MEDIJA ali OBLIKE v DRUGEGA, navadno zato, da kaj izboljšajo ali lažje dosežejo

- *V prihodnjih dneh se obeta obsežna medijska kampanja, a naročniki ne migrirajo kar čez noč.*
- *Čeprav so kratki, večepizodni filmčki na mobilnih telefonih vse bolj priljubljeni, pa oglaševalski milijoni iz drugih zabavnih segmentov zelo počasi migrirajo na super mali zaslon.*

Med konotacijskimi oznakami je potrebno izpostaviti rabo nekaterih besed s pomensko vlogo poudarjanja, in sicer lastnosti ali količine, vsebovane v besedi, na katero se poudarek nanaša. V tej vlogi se znajdejo zlasti pridevniki, prislovi in nekateri samostalniki, ki vsebujejo pomen količine:

bajen *pridevnik*

1 zelo velik
izraža poudarek

bajna VSOTA DENARJA ali bajen ZASLUŽEK je tako velik, da si ga težko predstavljamo

- *Eno od področij, ki prinaša bajne zasluzke, je mobilna telefonija.*
- *Marsikdo bi mislil, da filmskim igralcem, ki zasluzijo bajne vsote in jih vsi poznajo, ni treba prav veliko delati.*

gladko *prislov*

1 brez težav; z lahkoto
za izražanje poudarka

če nekaj gladko DOSEŽEMO ali STORIMO, naredimo to brez težav ali obotavljanja

- *Trebanjci, ki so doma gladko ugnali Koprčane.*
- *»Če bi se hotela prijaviti na agenciji za manekenke in modele, bi me gladko zavrnili,« pravi Twiggy.*

gora *samostalnik*

2 velika količina; kup
izraža poudarek

gora je zelo velika količina zlasti neprijetnih ali neuporabnih STVARI

- *Povsod sama umazanija, gore smeti in množica ljudi.*
- *preiskovalci vse bolj izgubljali v gorah papirjev in nasprotujočih si podatkov.*
- *Njibov pralni stroj se tako rekoč nikoli ne ustavi, gore perila pa se kljub temu skoraj ne zmanjšajo.*

12.4.4.2 Registrske oznake

Registrske oznake smo navajali, kadar se beseda oz. njen pomen uporablja v sporočanjskih situacijah, ki jih določa družbeni status udeležencev in jezikovne

konvencije, npr. v neformalnih situacijah, tj. značilnih za sproščeno konverzacijo in vsakdanje govorne položaje, ki ni niso vezani na javno nastopanje, npr.

tip *samostalnik*

3 moški

v sproščeni neformalni komunikaciji

- *Njej se je zdel tip ful luškan.*
- *Tip je absolutno nor.*

Oznaka v formalnih situacijah nasprotno označuje rabo izraza v sporočanjških položajih, za katere je bodisi značilen določen protokol obnašanja, preciznost izražanja, odsotnost konotativnih jezikovnih sredstev, pa tudi težnja po izražanju jezikovne kultiviranosti, npr.

tolmačiti *glagol*

1 razlagati

v formalni komunikaciji in besedilih

- *Demokratski nemir in elan, ki budi samozavest, se pričenja napačno tolmačiti kot prebujanje slovenskega nacionalizma.*
- *Avtor meni, da v Sloveniji napak tolmačijo ustavno načelo o ločitvi cerkve od države.*
- *Ugotovila sem, kako različno in celo tendenciozno si bralci vašega časopisa tolmačijo vsebino te notice.*

topogleden *pridevnik*

tak; tovrsten

v formalni komunikaciji in besedilih

- *Zanimajo me tudi druga topogledna in meni še neznan tekmovalci.*
- *Morda bodo take knjige pomagale spremeniti topogledno topoglavno nacionalno identiteto.*

Z oznako **v govoru** smo opredeljevali rabo besede oz. pomena, ki ga dobi beseda zlasti v govorjeni komunikaciji (*cepec*), kamor štejemo tudi izraze v vlogi besedilnih povezovalcev (*mimogrede*), vzklike z besedilno modifikacijsko vrednostjo (*fant*) ipd.:

cepec *samostalnik*

2 neumen človek

2.1 kot zmerljivka

zlasti v govoru

v neformalnih situacijah

- *Pa saj ni mrtev, cepec!*
- *Ti cepec, ti idiot!*

Fant! | Fant moj!

zlasti v govoru

izraža poudarek ali presenečenje

- *Fant, kako je tečen!*
- *Fant, kako so si nas privoščili.*

Z oznakami, kot so v jeziku mladih, v jeziku otrok ali otroško, smo označevali rabo pomena med govorniki določene generacijske pripadnosti, npr.

keš *samostalnik*

denar, zlasti gotovina

neformalno; zlasti v komunikaciji mladih

[dost, ful] keša

- *On ma sigurno dost keša, sam ga skriva na računih v tujini zarad davkov*
- *»Radi bi imeli dost keša, hitre avtomobile in dobre ženske.«*
- *»Našla si bom novo službo za ful keša!«*
- *Ronaldo je edin izmed največjih, pa ne zato ker ma hud auto pa ful keša, ampak zato ker je pač dober.*

Pri oblikovanju LBS smo sicer težili k temu, da smo tovrstno označenost posameznih pomenov vključevali v pomensko razlago; oznake tega tipa smo ohranjali predvsem tam, kjer bodisi razlaga ob indikatorju ni bila potrebna ali pa je bilo smiselno označenost posebej izpostaviti, s čimer smo sugerirali tudi različne stopnje tovrstne zaznamovanosti v realni rabi.

12.4.4.3 Pragmatične oznake

Oznake: kot grožnja, kot pozdrav, kot opozorilo, dopolnjujejo pomen zlasti pri frazeoloških enotah, zato smo jih navajali neposredno za pomenskim indikatorjem, sicer pa so pragmatične vsebine in s tem tudi pragmatične oznake sestavni del pomenskega opisa:

fasati jih | Bi jih rad fasal?!

kot grožnja
zelo grobo

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

<indikator><oznaka tip="pragmatika">**kot grožnja**</oznaka></indikator>

Nabor oznak, kot smo jih prikazali, v času izdelave LBS ni bil dokončno izdelan niti sistemiziran, saj je bilo treba popolnoma na novo premisliti kvalificiranje jezikovne rabe z vidika sporočanje vloge in v skladu z gradivom, ki ne določa zgolj knjižnojezikovno sprejemljivega besedišča. Hkrati smo želeli upoštevati tudi

možnost, da nekatere oznake uporabniku, torej pri prenosu v slovar, bodisi ne bodo vidne, ker so primarno namenjene možnostim analiziranja določenega segmenta jezikovne rabe v slovarski bazi, bodisi bodo pri prenosu v slovar postale del pomenke razlage ali samostojnih zavihkov oz. opozoril (npr. slovnične in skladišne oznake), za katere obstajajo v spletnem mediju različne možnosti vizualizacije.

III. SKLADENJSKI NIVO

Skladenjska informacija je v LBS vezana na dejstvo, da je pomen neposredno povezan s skladnjo in da se lahko v besedilu uresničuje na različne skladijske načine: neskončno število pomenov je mogoče izraziti s končnim številom skladijskih kombinacij. Ti ugotovitvi izhajata iz dveh teoretično-metodoloških pristopov. Prvi je obravnavanje pomena kot (so)besedilnega pojava, ki postavlja v ospredje sintagmatsko-paradigmatske vzorce, povezane s pomenskim potencialom besede in temelji na Hanksovi (2004, 2006) teoriji jezikovnih konvencij ter možnosti njihove izrabe. Iz tega pristopa izhaja, da pomenske tendence besed vplivajo na svoje izbire v besedilnem okolju oz. da se vsak pomen oblikuje v konkretni sporočanski situaciji in je odvisen od vsakokratnega (so)besedilnega konteksta. Drugi pristop je probabilistični in obravnava pomen, kot pravi Fillmore (1976), v smislu najboljšega približka prototipu in ne v smislu iskanja niza zadostnih ali potrebnih pogojev. Oba pristopa, realizirana v projektih FrameNet (Fillmore et al. 2003) in Corpus Pattern Analysis (Hanks 2004), se odražata tudi v zgradbi LBS in opisu leksikal-nogramatičnih podatkov v njej.

Tako izhodišče zahteva zelo konkretne odločitve v zvezi s slovničnimi problemi leksike, zlasti tistimi, ki v strokovni literaturi nimajo enoznačnih rešitev, kot npr. pri odločanju o povedkovniku kot samostojni besedni vrsti, pri izrazito večfunkcijskih besedah oz. skladijskih rabah nekaterih besed.

V LBS smo na ravni leme obravnavali tiste slovnične oz. skladijske podatke, ki se nanašajo na izrazno podobo besede, zato so iztočnice opredeljene zgolj z besedno vrsto, kjer smo izhajali iz oblikoskladijskih oznak, uporabljenih pri označevanju korpusa Gigafida. Posebnosti pri označevanju besedne vrste, kot so homonimija, ustaljena oblika, konverznost ipd., smo obravnavali v makrostrukturi LBS glede na to, ali so vezane zgolj na izraz kot tak (raven iztočnice) ali pa imajo posledice tudi za pomen.

Opis skladijskih informacij v zvezi s posamezno leksikalno enoto je v LBS namenjen:

- (a) beleženju tipičnih skladijskih struktur in pri glagolu stavčnih vzorcev, s katerimi se uresničuje pomen ali podpomen besede, ki jo obravnavamo v iztočnici;
- (b) formalizaciji skladijskih struktur in stavčnih vzorcev za računalniško obdelavo naravnega jezika ter
- (c) beleženju slovničnih omejitev in preferenc, ki izhajajo iz skladijskega okolja besede, in jih beležimo ob posameznih delih struktur in vzorcev.

Schema DTD se glede vključevanja in načina zapisovanja skladijskih podatkov pri glagolskih iztočnicah razlikuje od sheme DTD pri neglagolskih iztočnicah. Pri

samostalnikih, pridevniki in prislovih smo beležili skladijske podatke v obliki skladijskih struktur, ki se nato potrjujejo s kolokacijami in zgledi, medtem ko smo pri glagolih beležili skladijske podatke v obliki struktur le v primeru glagolskih zvez, tj. takih, ki vključujejo pomožne glagole, nedoločnike ipd., sicer pa v stavčnih vzorcih, ki se prav tako potrjujejo s kolokacijami in zgledi.

Večina skladijskih podatkov, ki smo jih beležili v LBS, je namenjenih strojnemu procesiranju, temu pa je v celoti podrejen tudi njihov zapis, in skladijskim analizam slovenščine. V nadaljevanju se zato posvetimo načinom pridobivanja skladijskih struktur iz korpusa prek orodja SkE, njihovim povezavam s kolokacijami in vlogi, ki jo imajo pri avtomatskem pridobivanju kolokacij iz korpusa. Nato opišemo način zapisovanja skladijskih struktur na splošno ter glede na posebnosti znotraj posamezne besedne vrste. Na koncu poglavja izpostavimo povezanost skladijskih struktur z drugimi elementi slovarske baze in izpostavimo nadaljnje možnosti računalniške obdelave.

13 Skladijske strukture

Skladijska struktura je formaliziran zapis besedne zveze, ki predstavlja skladijsko podstavo določeni besedni kombinaciji, ki se v jeziku pojavlja relativno pogosto, je slovnično pravilna in pomensko smiselna,¹⁴⁸ npr. besedne kombinacije kot: armada brezposelnih, spekter znanja, gostoljubnost domačinov ipd., si delijo podstavno podredno besedno zvezo, ki jo tvori samostalnik v imenovalniku, ki predstavlja jedro besedne zveze in pri sklanjanju ohranja celotno sklonsko paradigmo, ter samostalnik v rodilniku, ki nastopa v vlogi desnega prilastka in ohranja rodilniško obliko v celotni sklonski paradigmi. Te podatke, z izjemo podatka o jedrni oz. prilastkovni vlogi posameznega dela besedne zveze, smo v LBS zapisali kot: SBZ0 sbz2 oz. sbz0 SBZ2, odvisno od tega, kateri del skladijske strukture zastopa besedo (zapisano z velikimi črkami) v iztočnici.

13.1 Pridobivanje skladijskih struktur iz korpusa

Vir skladijskih struktur so besedne skice v orodju SkE, natančneje gramatične relacije (ang. *grammatical relations*), ki so v *slovnici besednih skic* (ang. *sketch grammar*) določene za konkretni jezik, v našem primeru za slovenščino (Krek in Kilgariff 2006). Gramatične relacije določajo, katere podatke iz korpusa bo mehanizem upošteval pri pripravi besedne skice, zato je za njihovo izdelavo potrebno predhodno poznavanje tipičnih sobesedilnih pojavov v konkretnem jeziku pa tudi potreb, ki izhajajo iz namena uporabe besednih skic (npr. za leksikografske namene, za namene raziskav leksikalnogramatičnega obnašanja besed, za izdelavo kontrastivnih analiz ipd.).

Vsaka gramatična relacija je definirana z regularnimi izrazi in oblikoskladijskimi oznakami konkretnega korpusa. Če želimo npr. iz korpusa pridobiti kolokacije, ki so vezane na strukturo pridevnika in samostalnika, moramo definirati samostalnik kot jedro samostalniške zveze (tj. oznaka S v seznamu oznak za korpus Gigafida) in enega ali več predhodnih pridevnikov (oznaka P) v vlogi modifikatorja (Krek in Kilgariff 2006):

```
*DUAL
=S_kakšen?/S_kdo-kaj?
    2:pridevnik brez_GSD{0,5} 1:samostalnik
```

Prva vrstica, neposredno za =, določa ime gramatične relacije, in sicer prvi del za strukturo pridevnik + samostalnik, kjer je iskana beseda samostalnik, in drugi del,

148 Smiselnost konkretni skladijski zvezi določa šele njena leksikalna zapolnitev, ki se potrjuje s kolokacijami. S tega vidika je tudi slovnična pravilnost besedne zveze vedno določena z leksikalno zapolnitvijo elementov v skladijski podstavi.

tj. za poševnico, za primere, kjer je iskana beseda pridevnik. Številki 1: in 2: določata besedi, ki naj jih vsebuje besedna skica, pri čemer je pridevnik definiran kot:

```
define(`pridevnik', `[tag=«P.*«]')
```

samostalnik pa kot:

```
define(`samostalnik', `[tag=«S.*«]')
```

Konkretna gramatična relacija tudi določa, da se med pridevnikom in samostalnikom ne smejo pojaviti glagol, samostalnik ali predlog (oznake G, S, in D), in sicer v razponu od 0 do 5 mest:

```
define(`brez_GSD', `[tag!=«[GSD].*« & word!=«[:;]«]')
```

Vsaka struktura, ki smo jo v LBS beležili v elementu <struktura>, ima torej v slovnici besednih skic svojo gramatično relacijo oz. vsaka gramatična relacija, ki jo vsebuje slovnica besednih skic, generira eno ali več skladenjskih struktur, ki jih beležimo v elementu <struktura>. Zadnja različica slovnice besednih skic (različica 0.8), ki smo jo uporabljali za izdelavo LBS, vsebuje 103 gramatične relacije, ki zgenerirajo pribl. 370 skladenjskih struktur, in sicer 247 za samostalnike, 60 za pridevnike, 46 za glagole in 20 za prislove. Kot relevantno strukturo smo šteli tisto, ki se v korpusu pojavi več kot enkrat in ima potrditev v kolokaciji. Povezavo naključno izbranih gramatičnih relacij s skladenjskimi strukturami in primeri kolokacij v LBS prikazuje Preglednica 23:

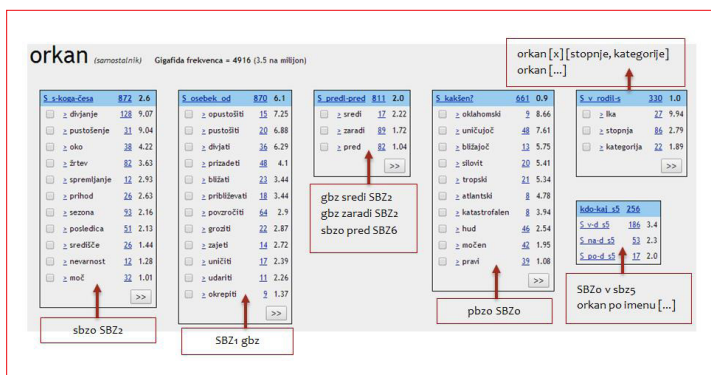
Preglednica 23: Povezava nekaterih gramatičnih relacij s skladenjskimi strukturami in kolokacijami

gramatična relacija	zapis v elementu <struktura> v LBS	besedna vrsta iztočnice	kolokacija
S_kako-kdaj_g?	<struktura>rbz GBZ</struktura>	glagol	[nemočno, tiho] besneti, [javno, grdo] blatiti
S_G-neg-inf	<struktura>neg-gbz Inf-GBZ</struktura>	glagol	[ne smeti se, ne nameravati se] vmešavati
S_kakšen-p?	<struktura>GBZ pbz1</struktura>	glagol	izkusiti [sam], roditi se [svoboden]
S_kako-kdaj_p?	<struktura>rbz PBZ0</struktura>	pridevnik	[srečno, nesrečno, sveže, pravkar, nikoli] poročen
S_%s_p2	<struktura>PBZ0 %s sbz2</struktura>	pridevnik	spačen od [besa, jeze, bolečine], usmiljen do [živali, ljudi]

gramatična relacija	zapis v elementu <struktura> v LBS	besedna vrsta iztočnice	kolokacija
S_nedoločnik	<struktura>GBZ Inf-gbz</struktura>	glagol	zahoteti se [plesati, jesti]
	<struktura>SBZ0 Inf-GBZ</struktura>	samostalnik	pravica [voliti, zahtevati]
	<struktura>PBZ0 Inf-gbz</struktura>	pridevnik	zasanjan [strmeti, zreti]
	<struktura>RBZ Inf-gbz</struktura>	prislov	korektno [obveščati, se obnašati]
S_%s_x_g2	<struktura>gbz %s SBZ2</struktura>	samostalnik	[izkopati se, izvleči se, rešiti se] iz težav, [pridrži, zapreti] do iztreznitve
S_kakšen?	<struktura>pbz0 SBZ0</struktura>	samostalnik	[skutni, jabolčni, sadni] narastek, [špinačne, ajdove, skutne] palačinke
S_komu-čemu	<struktura>gbz SBZ3</struktura>	samostalnik	[kljubovati, škoditi] konkurenci
S_osebek_od	<struktura>SBZ1 gbz</struktura>	samostalnik	oblaki [se zbirajo, se kopičijo], opozicija [očita, nasprotuje, kritizira, bojkotira]
S_kdo-kaj_r	<struktura>sbz1 Vez-gbz RBZ</struktura>	prislov	[poanta, bistvo] je drugje, [kri] je vsepovsod
S_količina_ob-s	<struktura>RBZ sbz2</struktura>	prislov	zgodaj [septembra]
S_z_vezajem	<struktura>pbz1{-}PBZ0</struktura>	pridevnik	[socialno-] emocionalen

Znak %s v gramatični relaciji označuje zastopanost konkretnega predloga ali veznika v skladenjski strukturi

Pri analizi gramatično-kolokacijskega obnašanja besed na podlagi izdelane besedne skice smo s pomočjo izkazanih gramatičnih relacij, ki so za obravnavano besedo tipične, določili ustrezne skladenjske strukture.



Slika 41: Del besedne skice za samostalnik orkan kot vir skladenjskih struktur

Kot prikazuje Slika 41, se posamezne gramatične relacije prepisujejo neposredno v skladenjske strukture: S_s-koga-česa --> sbz0 SBZ2 – [divjanje, pustošenje] orkana; S_osebek_od --> SBZ1 gbz – orkan [opustoši, pustoši, divja, prizadene]. Gramatične relacije, ki vključujejo predložne elemente, npr. S_predl-pred, lahko generirajo več skladenjskih struktur z različnimi predlogi, hkrati pa se lahko spreminja tudi besedna vrsta in sklon elementa strukture, ki se tipično pojavlja ob predlogu: gbz sredi SBZ2 – [biti, nahajati se] sredi orkana; sbz0 pred SBZ6 – [zatišje, strah] pred orkanom. Gramatične relacije, ki v kolokacijskih zapolnitvah predvidevajo količinske izraze ali lastna imena, oblikujejo besedne nize, ki smo jih beležili v elementu <zveza> znotraj skladenjskih zvez: orkan [prve, druge, tretje, četrte, pete] stopnje, orkan [4., pete] kategorije; orkan [Ika, Wilma, Mitch]; orkan po imenu [Frances, Igor].

13.2 Obravnava skladenjskih struktur po besednih vrstah

Pravila za določanje skladenjskih struktur v LBS temeljijo na pripadnosti besede, ki jo obravnavamo v iztočnici, konkretni besedni vrsti. V nadaljevanju so torej predstavljene glede na to, katere strukture navajamo v zgradbi geselskega članka pri samostalniških, glagolskih, pridevniških in prislovnih iztočnicah. V izhodišču smo ločili t. i. posebne strukture, kamor sodijo strukture s pomožnimi glagoli, nedoločniki, nikalnimi elementi in količinskimi izrazi, ter regularne strukture. Skladenjske strukture so načeloma dvodelne, trodelne strukture so lahko le predložne (pri glagolih tudi strukture z vezniki), strukture s pomožnim glagolom in strukture z nedoločniki. V elementu <struktura> smo zapisovali strukture glede na besedni red enotno, kar pomeni, da se npr. struktura Pbz0 SBZ5 lahko realizira kot: [dostopen] javnosti ali javnosti [neznan], razen če drugačen vrstni red ne sugerira samostojne strukture, npr. videti je imeniten – Vez-gbz Inf-gbz pbz1 in potrebno je izboljšati, pomembno je vedeti, treba¹⁴⁹ je improvizirati – pbz1 Vez-gbz Inf-GBZ.

Elemente skladenjskih struktur smo označevali z oznakami za besedne vrste, pri čemer smo sledili enakim merilom kot pri določanju besedne vrste pri iztočnici. Posebnosti v zvezi z zapisovanjem glagolskih elementov in nekaterih pomenskih skupin samostalnikov ter pridevnikov v skladenjskih strukturah so razvidne iz spodnjega seznama ter pojasnjene v nadaljevanju:

gbz/GBZ – glagol ali glagolska besedna zveza

Inf-gbz/Inf-GBZ – nedoločnik

Inf-za-GBZ/Inf-za-gbz – zloženi nedoločnik s predlogom za

Vez-gbz/Vez-GBZ – pomožni oz. vezni glagol

149 V LBO in korpusu je beseda treba sicer označena kot prislov.

Nam-gbz/Nam-GBZ – namenilnik
Neg-gbz¹⁵⁰/Neg-GBZ – zanikani glagol
Neg-Vez-gbz – zanikani vezni glagol
gbz-z¹⁵¹/GBZ-z – zloženi glagol – (npr. zagosti jo)
Sbz/SBZ – samostalnik ali samostalniška besedna zveza
Kol-sbz/Kol-SBZ – samostalnik ali samostalniška besedna zveza, ki izraža količino
pbz/PBZ – pridevnik ali pridevniška besedna zveza
Kol-pbz/Kol-PBZ – pridevnik ali pridevniška besedna zveza, ki izraža količino
num/NUM – glavni, vrstilni, ločilni ali množilni števniki
rbz/RBZ – prislov ali prislovna besedna zveza
Mod-rbz – modalni prislov: lahko
Kol-rbz/Kol-RBZ – prislov ali prislovna besedna zveza, ki izraža količino
Cit – citatni del strukture v obliki dobesednega navedka
RBZ-SM – beseda oz. prislov v vlogi stavčnega modifikatorja
'predlog' – konkretni predlog
'veznik' – konkretni veznik
Odv-'veznik' – odvisni stavek s konkretnim veznikom
0 – celotna sklonska paradigma besede
1–6 – določena sklonska oblika besede
{-} – vezaj (npr. slovensko-hrvaški: pbz1{-}pbz0)
 – – povezovalni del med elementi strukture, ki označuje zloženost elementov v strukturi, npr. Kol-rbz, gbz-z, RBZ-SM, Neg-Vez-gbz, Nam-gbz, Inf-za-GBZ ...). Uporabljali smo ga tudi v primeru zloženih besed v samostojnem pomenu: rbz gbz-Nam-GBZ: [včeraj] [priti, oditi, iti] iskat.

13.2.1 Posebne strukture

13.2.1.1 Strukture s pomožnim glagolom

Kot pomožni oz. vezni glagol (Vez-gbz) smo obravnavali samo glagol *biti* (kar ustreza tudi oznakam v korpusu Gigafida), in sicer kadar nastopa v slovnični vlogi. Druge glagole, ki nastopajo v tej vlogi oz. v pomensko šibkih pozicijah, npr. *postati*, *postajati*, *delati se*, *veljati za*, *iti za* ipd., smo obravnavali kot običajne glagole oz. smo jih v strukturi označevali z gbz oz. GBZ.

(a) Dvodelné

Vez-gbz sbz1	[je] tema; biti [avtoriteta, cepec]
Vez-gbz pbz1	biti [lep, pameten, ponosen, aktiven]; biti [nepozabno, čudovito]
Vez-gbz rbz	[biti] nazaj; [biti] še zgodaj; [je] mimo

150 Element Neg- se lahko kombinira z vsemi glagoli, pri katerih želimo opozoriti na prevladujoče zanikane realizacije.

151 Ob zloženem glagolu smo v elementu restrikcija zapisali, za kakšen tip zloženosti gre.

(b) Tri- oz. večdelne

sbz1 Vez-gbz sbz1	denar je [največji, glavni] problem; [skromnost, vztrajnost, hvaležnost] je [lepa] čednost
sbz1 Vez-gbz pbz1	[sodelovanje] je pomembno; [izboljšava, prenova] je potrebna
sbz1 Vez-gbz rbz	[šolanje, zdravstvo] je zastoj; [tekma] je nocoj
sbz2 Vez-gbz sbz1	agonije je [konec]
rbz Vez-gbz sbz1	[pozimi, zunaj] je mraz
rbz Vez-gbz pbz1	[pozimi] je hladno; uvodoma je [poudarjeno, zapisano, predstavljeno]
rbz Vez-gbz rbz	[najhuje] je mimo; [hitro, kmalu] bo mimo
pbz1 Vez-gbz Odv	pomembno je, da ...

Posebnost pri obravnavi predstavljajo strukture s pridevniškimi elementi, kadar ti nastopajo ob pomožniku. Za oznako pbz smo se namreč odločali v vseh primerih, kjer pridevnik nastopa v povedkovem določilu (gl. tudi poglavje 3.2 Konverzija besednih vrst – drugi težavnejši primeri), npr. *tako ravnanje je sporno; to je nemogoče/mogoče, odločilno, nujno ...* Pa tudi v primerih, kjer v pomožniški vlogi nastopa kateri drugi glagol, npr. *tako ravnanje se je pokazalo kot ustrezno; to se zdi odločilno, (ne)mogoče*.

Pri tri- oz. večdelnih strukturah, ki vsebujejo predloge, smo posebej obravnavali primere, kjer glagol *biti* ne nastopa kot pomožni glagol, ampak izraža pomen nahajanja v prostoru ali času, npr. *biti pri vojaki, v službi, na počitnicah, pred šolo* ipd. Predložne skladenjske strukture s pomožnikom navadno tvorijo skladenjsko podstavo FE ali fraznim glagolom, npr. *biti pri stvari, koritu; biti za luno, od hudiča; biti ob službo*. Med prave pomožnike lahko štejemo glagol *biti* ob predložnih zvezah predvsem takrat, kadar označuje lastnosti osebk, npr. Vez-gbz brez sbz2: *biti brez [apetita, konkurence, privilegijev]*; Vez-gbz ob sbz4: *biti ob [službo, zobe]*; Vez-gbz pri sbz5: *biti pri [zavesti]*; Vez-gbz pod sbz6: *biti pod [nadzorom, sankcijami]*. Kot kažejo tipične leksikalne zapolnitve tovrstnih skladenjskih struktur, gre za zveze, ki pogosto potrebujejo še dodatni element, zato se z vidika organizacije podatkov v LBS selijo na raven skladenjskih zvez, npr. *biti v igri za koga/kaj; biti v polnem cvetu; kaj je na vidiku* ipd. Če smo presodili, da take zveze potrebujejo pomensko razlago, smo jih obravnavali kot FE, npr. *biti/stati za katedrom* – 'poučevati'; *biti na konju* – 'uspeti'.

13.2.1.2 Strukture s pomožnim glagolom in z nedoločnikom

(a) Dvodelne

Strukture, ki vključujejo glagol *biti* in nedoločnik, so praviloma več kot dvodelne. V LBS smo samo s tema elementoma registrirali dve strukturi:

Inf-gbz Vez-gbz Vez-gbz Inf-za-gbz oz. Neg-Vez-gbz Inf-za-gbz	videti je, slišati je ni [za obupati, za računati]; se je za zamisliti
--	--

Prvo smo zapisovali le v primerih, ko nismo želeli posebej izpostaviti tretjega elementa, ki je navadno zastopan s pridevniki kot: *treba, potreben, možen, mogoč* (struktura: pbz1 Vez-gbz inf-gbz), s samostalnikom v ustreznem sklonu: *pričakovati je zmago* (Vez-gbz Inf-gbz sbz4), ali odvisnim stavkom: *sklepati je, da ...; slišati je, da ...; opaziti je, da ...* ipd.

Druga struktura, Vez-gbz Inf-za-GBZ, vključuje morfem *za*, ki se pripenja na nedoločnik, npr. *ni za računati, malo je za počakati* ipd. V vlogi pomožnika pogosto nastopajo tudi drugi glagoli, npr. *dati za poskusiti*, ki smo jih v skladenjskih strukturah označevali kot običajne glagole. Te strukture navadno zahtevajo še en vezljivostni element, kot je razvidno iz primerov: *(to) je za pozdraviti, (česa) ni za prezreti*. Če smo želeli izpostaviti ta del, smo navedli tudi ustrezno strukturo: Vez-gbz Inf-gbz-za sbz4, Neg-Vez-gbz Inf-gbz-za sbz2. Leksikalne zapolnitve pri teh strukturah kažejo na stilno zaznamovanost, značilno zlasti za izražanje v neformalnih položajih, v govoru, kar je pogosto razvidno tudi iz rabe kratkega nedoločnika: *za znoret, za popizdit* ipd.

(b) Tridelne

Vez-gbz Inf-gbz pbz1 pbz1 Vez-gbz Inf-gbz	videti je imeniten [potrebno] je izboljšati; [pomembno] je vedeti; [treba] je improvizirati
rbz Vez-gbz inf-gbz sbz1 Vez-gbz Inf-gbz gbz Inf-Vez-Gbz sbz1 gbz Inf-Vez-gbz pbz1	[najbolje, bolje] je saditi; [težko] je oporekati iluzija je [pričakovati]; sramota je [krasti] [hoteti] biti predsednik; [želeti biti] mama noče biti [zlobna, vsiljiva]; začeti biti [nevarno, tvegano]; [znati biti] pasji
Inf-gbz Vez-gbz sbz4	pričakovati je [zmago]; slišati je [govorice]

13.2.1.3 Strukture z nedoločnikom

(a) Dvodelne

Dvodelne strukture lahko poleg nedoločnika vsebujejo glagol, samostalnik ali pridevnik:

gbz Inf-gbz	[nehati, začeti] boleti; [znati] pasti, [dati se] rešiti; [nameravati] se preseliti; [morati] premagati, dati se [pregovoriti, prepričati]; uspeti [doseči, zmagati]
pbz0 Inf-gbz sbz0 Inf-gbz	[pripravljen, sposoben] pomagati [možnost, priložnost] zmagati; [pravica] pritožiti se; [namen, možnost] kandidirati

Glagolske strukture z nedoločnikom (gbz inf-gbz oz. Inf-gbz gbz) tipično vsebujejo fazne, npr. *začeti, začenjati, pričiti, nehati, prenehati* ipd., in naklonske glagole: *morati, moči (lahko), želei, hoteti, marati, dovoliti, smeti, drzniti si, znati, učiti, naučiti se, skušati, poskušati, prizadevati si, truditi se, uspeti, utegniti, pustiti se, pomagati, nameravati, hiteti, skleniti, dati se* ipd., če naštejemo le najpogostejše. Posebnost pri zapisu predstavlja glagol *moči*, ki smo ga v kolokacijah navajali v treh oblikah:

- kot modalni prislov *lahko*, za kar smo predvideli samostojni strukturni element: Mod-rbz gbz, npr. lahko [blokirati, aktivirati, fotografirati, garantirati];
- kot nedoločnik, kadar se rabi zanikano, tj. v strukturi Neg-gbz Inf-gbz, ki zastopa kolokacije kot: ne moči [pristati, dihati, funkcionirati, telefonirati];
- kot pridevnik ob pomožnem glagolu *biti*, tj. v strukturi pbz1 Vez-gbz Inf-gbz, npr. v kolokacijah: [mogoče, težko, lahko] je identificirati; [treba, mogoče, lahko] je igrati.

Strukture, kjer se ob nedoločniku pojavlja pridevnik (pbz0 Inf-gbz), se tipično realizirajo v povedkovem določilu. Zaradi možnosti nastopanja v vlogi desnega prilastka, glagola *biti* v strukturi in kolokacijah nismo navajali, npr. [pripravljen] intervenirati; [prisiljen] se odreči; [dolžen, sposoben] preživljati, razen, če se je taka realizacija kazala kot edino možna: Vez-gbz pbz1 Inf-gbz: [biti upravičen] terjati.

(b) Tridelne

Tridelne strukture z nedoločniki so vezane na realizacije naklonskega ali faznega glagola z nedoločnikom, ki predvideva še tretji element, ki je bodisi glagol, samostalni ali pridevnik, redkeje prislov, npr. gbz Inf-gbz Inf-gbz: utegniti začeti [snežiti, deževati]; [morati se naučiti] počakati; gbz Inf-gbz sbz1: hoteti postati [igralec, manekenka]; gbz Inf-gbz rbz: morati ostati [tukaj, doma]; sbz0 Odv-kako Inf-gbz: [recept, odgovor], kako obogateti.

13.2.1.4 Strukture s količinskimi izrazi

Z oznako Kol- smo označevali samostalnike, prislove in pridevnike (Kol-sbz, Kol-pbz, Kol-rbz), ki določajo količino sledečemu izrazu. Take oznake seznam oblikoskladenjskih oznak za korpus Gigafida ne predvideva, saj gre za dodatno pomenško opredelitev nekaterih samostalnikov, prislovov, pridevnikov in glagolov, ki je zanimiva predvsem s slovarskega vidika ter pri oblikovanju pravil za strojno učenje mehanizmov pri skladenjskem označevanju in razčlenjevanju. Z oznako Kol- so v LBS označeni naslednji izrazi: Kol-sbz0 sbz2: ura [hoje, rekreacije]; kos [potice,

torte]; žlička [čaja, sirupa]; [kozarček, kozarec, steklenica, požirek] alkohola; Kol-sbz1 pbz0: meter [dolga, velik]; uro [oddaljen, trajajoč]; Kol-sbz0 rbz: stoletja [dolga]; Kol-rbz pbz2:¹⁵² nič [alkoholnega]; Kol-rbz sbz2: [veliko, več, premalo] ambicij; Kol-rbz pbz0: [bolj, nekoliko, zelo, preveč] bazičen; [dvakrat, trikrat] poročen; Kol-rbz gbz: [precej, premalo] se angažirati; veliko [se smejeti, govoriti]; Kol-rbz rbz: [enkrat, dvakrat] dnevno; Kol-pbz0 sbz0: [različni, številni] biotopi; Kol-rbz na sbz4: trikrat na [leto, dan].

V LBS smo kot količinske izraze upoštevali tudi števnik (glavni, vrstilni, ločilni), ki smo jih v strukturah označevali z Num, in sicer kadar so se pojavljali v vlogi kolokatorjev v dvodelnih, npr. Num0 sbz0: [en] teden, mesec; [druga, tretja] armada; [prva] ljubezen, izkušnja; Num0 sbz2-6: [200, sto] podpisov; Num2 sbz2:¹⁵³ osmega [marca, februarja]; prvega [aprila, maja]; sbz0 Num2: petek [trinajstega]; pbz0 Num2: [pet] pogrešanih; rbz Num0: natančno [sto], in tridelnih strukturah: Num0 od SBZ2: [eden] od funkcionarjev; rbz ob Num2: popoldne ob [petih]; rbz od Num2: danes od [treh]; rbz pred Num6: jutraj pred [šesto] itd.

13.2.1.5 Strukture z nikalnimi elementi

(a) Pri glagolskih strukturah

Glagolski elementi skladenjskih struktur imajo lahko tudi nikalno obliko, saj obstaja zanikanje kot splošna upovedovalna možnost, zato neizstopajočih zanikanih realizacij nismo beležili v obliki posebnih struktur, ampak smo pod običajnimi strukturami navajali bodisi kolokacije z zanikanimi glagoli ali pa smo informacijo predstavili na raven zgledov, npr.

barvati *glagol*

1 spreminjati ali dodajati barvo

gbz Inf-GBZ

[začeti, pričeti] barvati

[želeti, ne moči] barvati

- *Poleti stez nismo mogli barvati, ker je bilo po mestu že tako ali tako zaprtih preveč cest.*

Kadar je bila nikalnost za obravnavani glagol bolj tipična kot nezanikana oblika, smo na to opozorili tudi v strukturi. Zanikani so lahko v tem primeru tudi pomožni glagoli:

¹⁵² Da gre za pridevnik v samostalniški rabi, smo opozorili v elementu restrikcija.

¹⁵³ Strukture s to oznako nastopajo v prislovni ali prilastkovni vlogi kot celota in se razlikujejo od struktur tipa: Num0 sbz0: *osmi marec, prvi april* ipd.

blatiti *glagol*

1 grdo govoriti o kom

Neg-gbz Inf-GBZ

[ne želeti, ne hoteti, ne smeti] blatiti

izdati *glagol*

4 razkriti

4.1 skrivne podatke

Neg-gbz GBZ

[ne upati, ne upati si] izdati

Neg-Vez-gbz pbz1 Inf-GBZ

[ne biti pripravljen] izdati

(b) Pri stavčnih vzorcih

Kadar so se obravnavani glagoli tipično pojavljali zanikano v primerih, kjer nismo beležili glagolskih struktur, ampak stavčne vzorce, smo uporabili naslednji možnosti:

(b1) Če se je nikalna oblika glagola kazala samo kot upovedovalna možnost in glede na druge ni izstopala, hkrati pa je ostal tudi sklon samostalniškega elementa nespremenjen, smo navedli ob običajnih tudi kolokacije z zanikanimi glagoli in ob pozitivnih vzorcih tudi zanikane (*dvomiti*) oz. smo med zgledi navedli tudi primere z zanikanim glagolom (*bremeniti*):

dvomiti *glagol*

1 ne verjeti

kdo dvomi o kom/čem

kdo dvomi v koga/kaj

kdo ne dvomi o kom/čem

kdo ne dvomi v koga, kaj

[mnogi] dvomijo; [strokovnjaki, poznavalci] dvomijo

dvomiti v [uspeh, sposobnost]

ne dvomiti v [uspeh]

ne dvomiti o [sposobnosti, uspehu, pravilnosti, upravičenosti, poštenosti]

bremeniti *glagol*

2 obremenjevati; težiti

kaj bremeni koga

- *Moja preteklost je čista, nikomur nisem storil nič slabega, ne bremeni me nikakršna krivda.*

(b2) V primeru, da se je ob zanikanem glagolu spremenil tudi sklon sledečega elementa, smo ob pozitivnih vzorcih in pripadajočih kolokacijah navedli tudi strukturo z zanikanim glagolom ter pripadajoče kolokacije in/ali vzorce:

pričakovati <i>glagol</i>	doživeti <i>glagol</i>
1 predvidevati, obetati si	1 biti deležen česa
kdo pričakuje kaj	1.1 o izkušnji
kdo pričakuje kaj od koga/česa	kdo doživi kaj
kdo pričakuje, da ...	kdo doživi, kako ...
pričakuje se kaj	doživeti [poraz, polom, razočaranje]
[organizatorji, prireditelji] pričakujejo	doživeti [kritiko, pohvalo]
pričakovati [porast, rast, povečanje]	Neg-GBZ sbz2
pričakovati [naval, gnečo]	kdo ne doživi česa
Neg-GBZ sbz2	ne doživeti [poraza]
ne pričakovati [čudežev, sprememb]	ne doživeti [orgazma]
ne pričakovati [zapletov, težav]	
ne pričakovati [bistvenih, večjih] sprememb]	

V primeru, da se je zanikani zvezi pridružil še tretji element, npr. ne iskati [krivcev, vzrokov] za kaj; ne kaže dvomiti v koga/kaj; ne dvomiti o [uspehu, sposobnostih] koga/česa ipd., smo zvezo uvrstili med skladenjske zveze.

Strukture z zanikanimi glagoli smo navajali tudi pri samostalniških in pridevniških iztočnicah, če je negacija glagola zahtevala spremembo sklona pri samostalniškem ali pridevniškem delu strukture, ki je naša iztočnica:

čas <i>samostalnik</i>
1 obdobje trajanja
1.1 glede na porabo
gbz SBZ4
[vzeti si, rezervirati si, imeti] časa
[krajšati si, preživljati] časa
[prihraniti] časa
ne imeti [dovolj, veliko] časa
ne najti [dovolj, veliko] časa
Neg-gbz SBZ2
[ne izgubljati, ne zapravljati, ne imeti] časa
ne najti [dovolj, veliko] časa

T. i. razširjene kolokacije s količinskimi izrazi, ki prav tako spreminjajo sklon iztočnice, npr. [imeti] časa, imeti [veliko] časa, smo navajali pod pozitivno strukturo (zgoraj pod GBZ sbz4) kljub spremembi sklona.

Pri nikalnih strukturah smo želeli – vsaj na ravni slovarske baze – opozoriti tudi na zanikanje s tožilnikom (z opozorilom v restrikciji), čeprav ni v skladu s trenutno normo, še zlasti v primerih, kjer taka raba izstopa:

Neg-gbz SBZ4

[ne izgublјati, ne zapravlјati, ne imeti] čas

- *Vedel je, da nima smisla zapravlјati čas.*
- *Če nimaš premožnih staršev z vezami, ne zapravlјaj čas na fakultetah.*
- *Če delam ves dan, še pomisliti nimam čas na kaj takega.*

13.2.2 Regularne strukture

Kot regularne strukture smo navajali pri posameznih besednih vrstah skladenjske kombinacije, ki vsebujejo samostalniške, glagolske, pridevniške in prislovne elemente, lahko tudi s predlogi in tipičnimi vezniki.

13.2.2.1 Samostalniške strukture

Dvodielne

(a) S samostalniškimi elementi

sbz0 sbz0	[brat] dvojček; mesec [maj, april]
sbz0 sbz1	[hiša, vila] dvojček; [Ljubljana, Maribor] mesto; solata [endivija, berivka, kristalka]

Če so zgornjo strukturo potrjevale kolokacije, kjer je bil drugi element fiksni v imenovalniku, in hkrati kolokacije, kjer sta se sklanjala po celotni paradigmi oba elementa, npr. solate endivija in solate endivije ali v hiši dvojček in v hiši dvojčku, smo ob tem elementu uporabili element <r> kot razširitev, npr. sbz0 SBZ1<r> tudi sbz0<r> ali sbz1 r> tudi sbz0<r> SBZ0.

sbz0 sbz2	iskanje [zaposlitve, zaklada]; konec [jeseni, poletja] [segrevanje, ohlajanje] ozračja
sbz1 sbzo	[golf, tenis] igrišče; tenis [lopar]; fitnes [center] [roza, bež] barva; celofan [papir]

Besede, kot so *fitnes*, *kaki*, *mini*, *roza*, *žiro*, *latino* ipd., ki lahko nastopajo v samostalniški ali pridevniški vlogi (npr. latino ritmi ali ritmi latina), smo v strukturah zapisovali s tisto besednovrstno oznako, s katero je lema označena na ravni iztočnice, torej: sbz1 sbz0: celofan papir; jeans barva; pbz1 sbz0: latino glasba; [roza, bež] barva.

sbz2 sbz2	meseca [maja, aprila]
sbz4 sbz2	konec [jeseni, ¹⁵⁴ poletja]; [konec] poletja
sbz6 sbz2	koncem [jeseni, poletja, leta]; vrh [gore, poslopja]; [vrh, dno] poslopja

154 V kolokaciji konec jeseni je beseda jeseni določena kot samostalnik, in sicer ločimo: sbz0 sbz2: [konec, začetek] jeseni in rbz rbz: [lani, letos, pozno] jeseni.

Zgornje strukture so skladenjska podlaga kolokacijam, ki navadno nastopajo v prislovni ali prilastkovni vlogi kot celota.

(b) Z glagolskimi elementi

Samostalniške strukture z glagolskimi elementi smo navajali zgolj pri samostalniških iztočnicah, pri glagolskih iztočnicah smo tovrstne kombinacije zapisovali s stavčnimi vzorci.

SBZ0 Odv-xy	vprašanje, ali ...; čas, ko ...; slutnja, da ...; polica, na kateri ...
SBZ1 gbz	rana [se zaceli, zaraste]; ljubosumje [razjeda, muči]
SBZ1 gbz-z	dež [jo zagode]
gbz SBZ1	[vrniti se, odhajati] junak; [postati] rdečelaska
gbz SBZ2	[naučiti se] plesa; [udeležiti] se avdicije; [razveseliti se] deklice
gbz SBZ3	[predajati se, prepuščati se] spominom; [uiti, izogniti se] atentatu
gbz SBZ4	[zapravljati, izgubljati] čas; [krotiti, brzdati] apetite
	[preklicati, umakniti] embargo

(c) S pridevniškimi elementi

SBZ0 pbz2	izobraževanje [brezposelnih]; sleng [mladih]
SBZ0 pbz3	spomenik [padlim]; svoboda [narodu, ljudstvu]

V zgornjih strukturah je pridevnik v rodilniku ali dajalniku navadno pridevnik v samostalniški rabi (tj. posamostaljeni pridevnik), na kar smo opozarjali v restrikciji.

pbz0 SBZ0	[boleč, lep, tih] spomin; [srečno, brezskrbno] otroštvo
pbz0 SBZ2	[deležen, vreden] pozornosti
pbz0 SBZ3	[prijazen, razumljiv] uporabniku
pbz0 SBZ4	[dolžen] denar
pbz1{-} SBZ0	[e-]arhiv; [cd-]gramofon; [U-, I-, T-, L-] profil
pbz2 SBZ2	[šibkega, trdnega] zdravja
pbz4 SBZ4	[letošnje, lansko, prihodnjo] jesen

(č) S prislovnimi elementi

SBZ0 rbz	potovanje [domov, nazaj]; ponedeljek [popoldne, dopoldne]; porod [doma]
----------	---

Tridelne

(a) S samostalniškimi elementi

(a1) roditeljske: SBZ0 *predlog* sbz2 s predlogi: *blizu, brez, do, glede, iz, mimo, od, okoli, okrog, prek/preko, s/z, sredi, zaradi, znotraj, zunaj* v kolokacijah:

vir blizu [ministrstva]; otrok brez [staršev]; čas do [volitev]

(a2) dajaljske: SBZ0 *predlog* sbz3 s predlogi: *k/h, proti, nasproti, kljub* v kolokacijah: poziv k [dialogu]; soglasje h [gradnji]; boj proti [korupciji, diskriminaciji]

(a3) tožiljske: SBZ0 *predlog* sbz4 s predlogi: *čez, na, skozi, v, za, zoper* v kolokacijah: most čez [reko]; spomin na [otročnost]; potovanje skozi [čas]

(a4) mestniške: SBZ0 *predlog* sbz5 s predlogi: *na, o, ob, po, pri, v* v kolokacijah: senca na [vrtu, terasi]; film o [ljubezni]; incident ob [meji]

(a5) orodniške: SBZ0 *predlog* sbz6 s predlogi: *med, nad, pod, pred, z/s, za* v kolokacijah:

razlika med [spoloma]; skrbništvo nad [otrokom]; rezultat pod [pričakovanji]

(b) Z glagolskimi elementi

(b1) imenovalniške: gbz kot SBZ1 v kolokacijah:

[počutiti se, živeti, obnašati se] kot kralj; [zaposeliti se] kot učiteljica; [zveneti] kot šala

(b2) roditeljske: gbz *predlog* SBZ2 s predlogi: *blizu, brez, do, glede, iz, izven, mimo, namesto, od, okoli/okrog, poleg, prek/preko, s/z, sredi, zaradi, znotraj, zunaj* v kolokacijah:

[stati] blizu odra; [shajati] brez denarja; [stopiti] do sosedu

(b3) dajaljske: gbz *predlog* SBZ3 s predlogi: *k/h, proti* v kolokacijah:

[privezati] k postelji; [glasovati] proti zakonu

(b4) tožiljske: gbz *predlog* SBZ4 s predlogi: *čez, kot, med, na, nad, ob, pod, pred, skozi, v, za* v kolokacijah:

[stopiti] čez lužo; [zaposeliti] kot hišnika; [pomešati se] med gledalce

(b5) mestniške: gbz *predlog* SBZ5 s predlogi: *na, o, ob, po, pri, v* v kolokacijah:

[gójiti] na gredah; [spregovoriti] o nasilju; [stati] ob ograji

(b6) orodniške: gbz *predlog* SB62 s predlogi: *med, nad, pod, pred, z/s, za* v kolokacijah:

[znajti se] med osumljenci; [leteti] nad mestom; [umreti, končati] pod giljotino

(c) S pridevniškimi elementi

(c1) imenovalniške: pbz0 kot SBZ1 v kolokacijah:

[priden, marljiv] kot čebela; [ponosen] kot pav

(c2) roditeljske: pbz0 *predlog* SBZ2 s predlogi: *glede, do, iz, od, okrog/okoli, prek/preko, spričo, zaradi* v kolokacijah:

[negotov] glede prihodnosti; [kritičen] do medijev; [odstranjen] iz prostora

(c3) dajalniške: pbz0 *predlog* SBZ3 s predlogi: *k/h, proti* v kolokacijah:

[nagnjen] k alergijam; [namenjen] proti jugu

(c4) tožilniške: pbz0 *predlog* SBZ4 s predlogi: *na, med, pod, pred, v, za* v kolokacijah:

[uperjen] na vasico; [uvrščen] med finaliste; [stisnjen] v šop

(c5) mestniške: pbz0 *predlog* SBZ5 s predlogi: *na, o, ob, po, pri, v* v kolokacijah:

[najboljši, največji] na svetu; [obveščen] o dogajanju; [zgrajen] ob železnici

(c6) orodniške: pbz0 *predlog* SBZ6 s predlogi: *med, nad, pod, pred, z/s* v kolokacijah:

[srečen, blažen] med ženami; navdušen nad [izvedbo, lepoto]; [poznan] pod nazivom

(č) S prislovnimi elementi

(č1) roditeljske: rbz *predlog* SBZ2 s predlogi: *do, od* v kolokacijah:

[najpozneje] do februarja; [slabše, bolje, manj] od pričakovanj

(č2) tožilniške: rbz *predlog* SBZ4 s predlogi: *čez, za, zaradi* v kolokacijah:

[očitno] čez noč; [posebej] za goste; [prav, ravno] zaradi zdravja

(č3) mestniške: rbz *predlog* SBZ5 s predlogi: *na, ob, po, pri, v* v kolokacijah:

[najverjetneje] na morju; [lani, vedno] ob srečanju; [povsod, nikjer] po svetu

(č4) orodniške: rbz *predlog* SBZ6 s predlogi: *pred, z/s* v kolokacijah:

[zjutraj, neposredno] pred sestankom; [hkrati, skladno] s predlogom

13.2.2.2 Glagolske strukture

Regularne glagolske strukture so v LBS omejene na t. i. glagolske zveze z glagolom (Toporišič 2000: 587) oz. na zveze glagola s količinskimi izrazi, ki smo jih že obravnavali, sicer smo namesto skladenjskih struktur navajali stavčne vzorce, ki jih obravnavamo v samostojnem poglavju.

Poleg glagolskih struktur z glagoli smo pri glagolskih iztočnicah zapisovali tudi glagolske strukture, ki vsebujejo (a) pridevniške, prislovne in samostalniške elemente v vlogi povedkovega prilastka ali določila, (b) glagole v tipični kombinaciji s prislovi, (c) namenilnike in (č) citatne elemente ter odvisne stavke:

(a) Glagolske strukture s pridevniškim, prislovnim ali samostalniškim elementom v vlogi povedkovega prilastka ali določila

gbz sbz1	[vrniti se] junak; [postati] darovalec; [postajati, postati] metropola
gbz pbz1	fotografirati se [gol]; [vesel, presenečen] razlagati; [postati] lahkomišeln; [obležati, zgruditi se] nezavesten
GBZ pbz2	delati se [pametnega, neumnega]
GBZ kot sbz1	vrniti se kot [junak, tujec]; uveljaviti se kot [znanstvenik, raziskovalec]
GBZ kot pbz2	odpustiti kot [nezanesljivega, nesposobnega] ¹⁵⁵
GBZ za pbz2	veljati za [krivega, nedolžnega]; imeti se za [najboljšega, pomembnega]
GBZ za sbz4	veljati za [prevaranta, legendo]

155 Po Toporišiču (2000: 269) gre za roditeljski povedkovski prilastek: Bala se ga je preveselega, kar ustreza primerom kot: *igralca so označili kot/za smešnega*, vendar: *igralko so označili kot/za smešno*, torej tožilnik. Pri različnih rabah tožilnika Toporišič ne navaja primera s pridevnikom v funkciji povedkovskega prilastka, pač pa zgolj samostalnik: *Prijatelj sem ob vrnitvi iz tujine našel pijanca* (prav tam: 270).

(b) Glagolske strukture s prislovi

rbz GBZ	[nesrečno, hudo] pasti; [rad] slikati; [pridno, nenehno, vztrajno] pitati; [silovito, močno] pljuskni; [počasi, hitro, rahlo, malce] počepniti
Mod-rbz GBZ	[lahko] oteči
GBZ kot rbz	pokazati se kot [ustrezno, pomembno, odločilno]

(c) Glagolske strukture z namenilniki

gbz Nam-gbz	priti [spodbujati, navijati]; [priti, oditi, hoditi] iskati; [dati] skenirati
gbz Nam-gbz rbz	vrtniti se študirati domov

(č) Glagolske strukture s citatnimi elementi in odvisniki

GBZ Odv- xy	žareti, kadar ...; žvižgati, kolikor ...; žalovati, [dokler, kajti, saj]; zažugati, [da, naj]
Cit GBZ	»dober dan« pozdraviti

13.2.2.3 Pridevniške strukture

Dvodelne

(a) S samostalniškimi elementi

PBZ0 sbz0	svetla [soba, prihodnost]; rdeča [vrtnica, zvezda]
PBZ0 sbz2	potreben [ljubezni, topline]; poln [denarja]
PBZ0 sbz3	nevaren [državi]; zvest [gospodarju]
PBZ0 sbz4	dolžen [denar]; vreden [milijardo, milijon]
PBZ0 Odv-xy	vesel, ker ...
sbz0 PBZ2	[izobraževanje] odraslih
sbz0 PBZ3	[spomenik] padlim
PBZ1 sbz0	mini [krilo, hlačke]
PBZ4 sbz4	prihodnjo [jesen, zimo]

(b) Z glagolskimi elementi

gbz PBZ1	[plezati] navezan; [fotografirati se] gol
gbz PBZ4	[narediti] pričakovano

(c) S pridevniškimi elementi

pbz0 pbz0	[lesen, okrogel] rezljan; [najboljši, edini] slovenski
pbz1{-} pbz0	[poljsko-, slovensko-, nemško-] ruski; [pomladno-, spomladansko-] poleten
pbz1 pbz1	[ultra, ekstra, super] mini

(č) S prislovnimi elementi

PBZ0 rbz	aktiven [ponoči]; končan [jeseni, poleti]
rbz0 PBZ0	[kristalno, povsem] čist; [neznansko] vesel

Tridelne

(a) S samostalniškimi elementi

(a1) imenovalniške: PBZ0 kot sbz1

v kolokacijah:

bled kot [zid, stena]

(a2) rodilniške: PBZ0 *predlog* sbz2 s predlogi: *do, glede, iz, od, okrog, okoli, prek/ preko, spričo, zaradi* v kolokacijah:

kritičen do [nastopajočih, udeležencev]; negotov glede [pogajanj, izida]; narejen iz [lesa, gline]

(a3) dajalniške: PBZ0 *predlog* sbz3 s predlogi: *k/h, proti* v kolokacijah:

nagnjen k [samomorom]; namenjen proti [severu, domu]

(a4) tožilniške: PBZ0 *predlog* sbz4 s predlogi: *na, v, za* v kolokacijah:

mahnjen na [denar]; stisnjen v [kalup, model]; usoden za [obiskovalce, turiste]

(a5) mestniške: PBZ0 *predlog* sbz5 s predlogi: *na, o, ob, po, pri, v* v kolokacijah:

najboljši na [tekmovanju, prvenstvu]; obveščen o [dogajanju, dogodkih]; pritisnjen ob [zid, steno]

(a6) orodniške: PBZ0 *predlog* sbz6 s predlogi: *med, nad, pod, pred, z/s* v kolokacijah:

edini med [kandidati, udeleženci]; ogorčen nad [potekom, dogajanjem]; znan pod [imenom, nazivom]

(b) S pridevniškimi in glagolskimi elementi

PBZ0 kot pbz0	stigmatiziran kot [protinacionalni]
gbz kot PBZ2	odpustiti kot [nezanesljivega]

(c) S prislovnimi elementi

PBZ0 rbz	aktiven [ponoči]; ukrivljen [navzven, navzgor]
----------	--

13.2.2.4 Prislovne strukture

Dvodelne

(a) S samostalniškimi elementi

RBZ sbz0	narobe [svet]
RBZ sbz1	pasje [mrz]; točno [polnoč, poldne]

(b) Z glagolskimi elementi

RBZ gbz	nesrečno [pasti, izbrati]; močno [deževati, snežiti]; rad [slikati]
---------	---

(c) S pridevniškimi elementi

RBZ pbz0	neznansko [vesel, razočaran]; kristalno [čist, jasen]
----------	---

(č) S prislovnimi elementi

RBZ rbz	naprodaj [spomladi]; [nadvse, izjemno] preprosto
RBZ Mod-rbz	praviloma [lahko]
RBZ Odv-xy	preprosto, da ...

Tridelne

(a) rodilniške: RBZ *predlog* sbz2 s predlogi: brez, do, iz, izpod, od, mimo, sredi, zaradi v kolokacijah:

praviloma brez [zapletov]; najpozneje do [februarja, jeseni]; nekam iz [mesta]

(b) dajalniške: RBZ *predlog* sbz3 s predlogi: *k/h*, *proti*, *kljub* v kolokacijah:

tesno k [prsim, telesu]; odločno proti [potvarjanju, kloniranju]; žal kljub [naporom]

(c) tožilniške: RBZ *predlog* sbz4 s predlogi: *čez, med, na, pod, v, za* v kolokacijah:

očitno čez [noč]; enakomerno med [uporabnike]; drugam na [dopust, morje]

(č) mestniške: RBZ *predlog* sbz5 s predlogi: *na, ob, po, v* v kolokacijah:

nikjer na [svetu]; poleti ob [večerih]; takoj po [sestanku, srečanju]

(d) orodniške: RBZ *predlog* sbz6 s predlogi: *med, pred, pod, z/s* v kolokacijah:

četrtič med [deseterico]; zunaj pred [vrati]; deloma pod [vplivom, nadzorom]

14 Skladijske zveze

Med večbesedne enote, ki imajo status slovarske, ne pa tudi leksikalne enote, smo v LBS uvrščali tudi statistično prepoznavne večbesedne nize – t. i. skladijske zveze (SkZ), ki izkazujejo predvsem skladijsko ustaljenost, hkrati pa – vsaj z vidika maternih govorcev slovenščine – ne izkazujejo samostojne pomenske vrednosti. Gre za strukturno različne zveze, ki glede na nivojsko zgradbo LBS in odsotnost pomena sodijo na raven skladijskega opisa, saj jih, kot rečeno, ne obravnavamo kot samostojnih leksikalnih enot. Med drugim to pomeni, da ne predvidevajo pomenske razlage (indikatorja in/ali razlage), zato smo jih uvrščali pod posamezni (pod)pomen besede, ki jo obravnavamo v iztočnici. V zvezi z njimi smo predvidevali naslednje informacije:

- identifikacija SkZ (predvsem v odnosu do kolokacij, razširjenih kolokacij in SZ) v okviru posameznega (pod)pomena
- oblike rabe (variantnost, fakultativnost sestavin in tipično besedilno okolje)
- zgledi rabe

Za SkZ je značilno, da so razmeroma pogost jezikovni pojav, hkrati pa so zaradi svoje strukturnoskladijske raznolikosti in relativne pomenske transparentnosti znotraj teoretičnih opisov različno obravnavane, posledično pa zanje tudi ne obstaja enotno poimenovanje, npr. ang. *grammatical collocation* (Benson et al. 1986), *fixed phrase* (Hartmann 2003), *chunk*, (*language*) *chunk* (Krishnamurthy 2002: 289), *lexical bundle* (Biber et al. 2003), *formulaic phrase/expression* (Columas 1994; Wray 2002), *recurrent phrase/string* (Stubbs 2007). Hkrati pa se jim ni mogoče izogniti v slovarskih opisih, ki temeljijo na prepoznavanju tipičnih jezikovnih vzorcev v korpusu, kot npr. v leksikalni bazi DANTE in v splošnem nizozemskem slovarju (ANW).

V leksikalni bazi DANTE, katere namen je zagotoviti opis angleščine za izdelavo novega angleško–irskega slovarja na podlagi 1,7 milijardnega korpusa sodobne angleščine (Rundell in Atkins 2011: 233), sodijo med različne tipe večbesednih enot (ang. *phrase*, *collocations*, *compounds*, *support prepositions* itd.) tudi zveze t. i. *chunks*,¹⁵⁶ ki so opredeljene kot dovolj pogosti neidiomatični besedni nizi, ki jih je mogoče prepoznati s korpusno analizo, in prinašajo v zvezi z lemo, na katero se nanašajo, pomembne leksikalne informacije (Rundell in Atkins 2011: 245). V svoji zgradbi naj bi imeli tovrstni besedni nizi nekatere elemente stalnih zvez, tj. določene leksikalizirane sestavine, katerih izbor je omejen, vendar pa je njihov pomen mogoče

¹⁵⁶ Za namene nadaljnjega razpravljanja bomo v zvezi z angleškim izrazom *chunk* govorili kot o ustaljenih besednih nizih.

razbrati iz sestavnih delov, zaradi česar ne potrebujejo pomenske razlage in ne ustrezajo samostojnim leksikalnim enotam. V strukturi slovarske baze so zato uvrščeni pod posamezni pomen iztočnice. Za angleško besedo *next*¹⁵⁷ so v DANTE navedeni naslednji besedni nizi:

next

7 (adj.) marking the following stage in a series

next up

11 (adj.) the following person or thing

next to do

be next to do

be next in line

be next in line to sth

be next in line to do

next up

from one ... to the next

3 (prep.)

next to

Podobno obravnava take besedne nize tudi Splošni nizozemski slovar, ki ločuje med t. i. besednimi kombinacijami in trdnimi oz. stalnimi besednimi zvezami. Besedne kombinacije obsegajo slovnico ustaljene zveze s predlogi in zveze, za katere je značilno, da so pomensko prozorne in hkrati skladenjsko ustaljene, npr. *z užitkom, do nadaljnjega* ipd.

14.1 Prepoznavanje skladenjskih zvez na podlagi korpusa

Glede na to da so SkZ leksikalni pojav, ki pride do izraza pri statističnem vzorčenju besedila in da semantično niso problematične na enak način kot SZ ali FE, je za način njihovega identificiranja v korpusu smiselno uporabiti orodja, ki prepoznajo leksikálnogramatične vzorce, ki presegajo kombinacijo zgolj dveh elementov. Z analizo korpusnih konkordanc v prvi fazi izdelave LBS smo tako zaznali vrsto besednih nizov, ki so presegali zgradbo dvodelnih in trodelnih kolokacij. Na podlagi ugotovitev, ki so izhajale iz strukturnoskladenjskih lastnosti takih kombinacij, je bila nato izpopolnjena slovnica besednih skic, ki je vključevala gramatične relacije (gl. Preglednico 24), s pomočjo katerih smo nato v 2. fazi izdelave LBS avtomatsko prepoznavali take zveze v korpusu.

¹⁵⁷ Beseda *next* ima v slovenščini lahko več prevodov, npr. *naslednji, sosednji, nato, potem, naslednjič, spet*. Zveze, v katerih nastopa v angleščini, pa imajo v slovenščini samostojni pomen, npr. *next to – poleg, zraven, tik ob; skoraj; v primerjavi z.* (Povzeto po Veliki angleško-slovenski slovar Oxford DZS, 2006).

Preglednica 24: *Seznam gramatičnih relacij za element <skladenjska_zveza> v shemi DTD*

gramatična relacija v SkE	besedna vrsta iztočnice	prepis v element DTD pri avtomatskem luščenju SkZ	primer v LBS
Z_biti_videti	samostalnik	<zveza>biti videti %s</zveza>	kaj je videti kot igra; kaj je videti kot [nedolžna, neslana] šala
Z_predl-pred	samostalnik	<zveza>%(1.lemma)_(2.lemma)</zveza>	avtoriteta na področju [fotografije, mednarodžmenta]; kviz o poznavanju [česa]
Z_primer	samostalnik	<zveza>SBZ0 kot sbz0</zveza>	
	glagol	<zveza>GBZ kot sbz0</zveza>	cveteti kot rožica
	pridevnik	<zveza>PBZ0 kot sbz0</zveza>	biti znan kot [velik, nepopoljšljiv] deloholik
	prislov	<zveza>RBZ kot sbz0</zveza>	[več, manj] kot letos
Z_priredje	samostalnik	<zveza>SBZ0 in/ali SBZ0</zveza>	[venčki, šopki, venci, darila] in aranžmaji
	glagol	<zveza>GBZ in/ali GBZ</zveza>	begati in/ali vznemirjati
	pridevnik	<zveza>PBZ0 in/ali PBZ0</zveza>	[rdečelas, lisast] in pegast
	prislov	<zveza>RBZ in/ali RBZ0</zveza>	cenovno in [kakovostno, količinsko, tehnološko, časovno]
Z_simile	samostalnik	<zveza>sbz0 kot SBZ0</zveza>	
		<zveza>gbz kot SBZ0</zveza>	skakati kot kenguru
		<zveza>pbz0 kot SBZ0</zveza>	oster kot britev
		<zveza>rbz kot SBZ0</zveza>	[več, manj, tako, podobno] kot lani

14.2 Merila za določanje skladenjskih zvez

SkZ, kot smo jih evidentirali v LBS, so navadno več kot dvobesedni leksikalni nizi z vsaj nekaj ustaljenimi leksikalnimi elementi in skladenjsko zgradbo, ki lahko predvideva prosta skladenjska mesta, zamejena z določenimi slovničnimi in slovničnopomenskimi kategorijami, kot so npr. sklon, število, živost oz. neživost, njihov pomen pa je bolj ali manj vsota pomenov posameznih sestavin. Povedano je mogoče strniti v naslednja načela:

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

Čeprav obravnavamo SkZ kot pomensko prozorne, pa zlasti predložne zveze tega tipa nimajo vedno povsem identičnih tujejezičnih ustreznikov, zato je v primerih, kot jih navajamo v Preglednici 25, in z vidika nematernih govorcev smiselno nekatere takih zvez uvrstiti (tudi) med SZ, ki predvidevajo pomensko razlago:

Preglednica 25: *Predložne skladenjske zveze v različnih jezikih*

slovenščina	češčina	nemščina	angleščina
čez čas/ sčasoma	po čase/ začas	im Laufe der Zeit/ über die Zeit	over time
v hipu/kot bi trenil/ takoj	na prvni pohled	im Überblick	in (less than) no time
skrajni čas	nejvyšší čas/ načase	höchste Zeit	high time/ about time

14.3 Vrste skladenjskih zvez

Na lestvici od pomensko popolnoma transparentnih kolokacij do pomensko ne-transparentnih FE zavzemajo SkZ mesto med razširjenimi kolokacijami in SZ, vendar pa zanje, tako kot smo v zvezi z različnimi tipi večbesednih enot že ugotavljali, veljajo različne stopnje pomenske trdnosti. Pri SkZ je stopnja pomenske

prozornosti, če odmislimo ločevanje med rojenimi govorci in učenci slovenščine kot tujega jezika, v veliki meri povezana s strukturnoskladenjskimi lastnostmi zveze. V nadaljevanju si bomo posamezne tipe SkZ ogledali podrobneje, pri čemer bomo prehajali od bolj k manj pomensko transparentnim.

Zveze s številskimi elementi

Med SkZ predstavljajo opazen delež zveze, ki vključujejo številske elemente, ki so navadno spremenljivi glede na skladenjsko prilagajanje v stavku (npr. ustrezna sklonska oblika, vrstni, glavni števniki ipd.) in pomen besede, ki predstavlja eno od leksikaliziranih sestavin SkZ. Ta je lahko, vendar ne nujno, beseda, ki jo obravnavamo v iztočnici:

pomlad <i>samostalnik</i> 1.4 leto šteti [x] pomladi	žepnina <i>samostalnik</i> [x] [dolarjev, tolarjev, evrov] žepnine	žlica <i>samostalnik</i> zmešati kaj z/s [x] žlicami česa
datiran <i>pridevnik</i> datiran v čas od [x.] stoletja do [x.] stoletja	diplomirati <i>glagol</i> diplomirati leta [x]	

Zveze z lastnoimenskimi elementi

Ob lastnih imenih, ki smo jih znotraj kolokacijskega niza vedno označili s tremi pikami [...], se pogosto pojavlja predvidljivo skladenjsko in pomensko besedilno okolje, npr.

(železniška) proga med postajama [...] in [...]
film z naslovom [...]
biti znan po filmu [...]
izdati [zgoščenko, kaseto, brošuro, zloženko, knjigo] z naslovom [...]
članek izpod peresa [...]

Predložne zveze

Najbolj tipične so predložne SkZ v različnih prislovnih vlogah, npr. kraj: *na [vseh, raznih, različnih] koncih*; čas: *od časa do časa, iz meseca v mesec, do pozne jeseni, na jesen, ob vikendih*; način: *brez težav, po želji, v skladu s kom/čim, pod vplivom koga/česa* ipd. Glede na značilnosti zgradbe ali pomenskih lastnosti smo v LBS prepoznali naslednje skupine:

(a) SkZ s pomensko-skladenjsko predvidljivim vezljivostnim mestom

angažirati se v prid koga/česa
angažirati se na področju česa
[imeti, opravljati] aktivno vlogo pri čem
biti akreditiran v skladu s čim

(b) Pomensko odvisne ali neodvisne od iztočnice

Z vidika ugotavljanja skladenjskega jedra oz. z vidika ločevanja SkZ od njenega kolokabilnega okolja, smo ločevali zveze, kjer si v kolokatorskem nizu sledijo besede, ki se tipično pojavljajo kot besedilno okolje trdnjšega jedra skladenjske zveze. Iztočnica je v teh primerih lahko zastopana kot sestavina zveze (*cigareta*, *spomin*) ali pa kot njen kolokabilni del (*aktivirati*, *cveteti*):

cigareta

[kaditi, prižigati] cigareto za cigareto

aktivirati

aktivirati po potrebi

spomin

[slikati se, ohraniti, imeti] za spomin

cveteti

vse bolj cveteti
cveteti kot še nikoli

Priredne zveze (in/ali)

Obsežno pomensko skupino predstavljajo priredne SkZ, ki vključujejo veznik *in* ali *ali*. V LBS smo jih ločevali glede na pomenske lastnosti kolokatorjev v kolokatorskem nizu. Združevanje kolokatorjev po različnih pomenskih merilih zagotavlja tudi podatke o pomensko sorodnih leksikalnih elementih, ki jih je mogoče uporabiti kot učno množico pri strojnem procesiranju jezika in pomenskem gručenju leksikalnih elementov:

(a) S pomensko sorodnimi ali sopomenskimi kolokatorji

morala in [etika, vera, religija, vrednota, politika, kultura]
alkohol in [tobak, cigarete, droge, nikotin]
akreditiran in [verificiran, pooblaščen]
karizma in [avtoriteta, šarm, lepota, energija]
privilegij in [imuniteta, boniteta, ugodnost, pravice, prednost]
aluzije in [asociacije]
avtonomen in [neodvisen, samostojen]

(b) S protipomenskimi kolokatorji

argumenti in protiargumenti
bazičen in kisel
absoluten in relativen
kisel in grenek
vzleteti in pristati
združevati in razdvajati

Primerjalne zveze (kot/kakor)

Med SkZ smo beležili samo tiste primerjalne zveze, ki z vidika pomenske transparentnosti ne potrebujejo pomenske razlage, npr.

(a) S pomenom imenovanja, funkcije

častiti koga kot [zavetnika, boginjo] česa
biti znan kot [velik, nepopoljšljiv] deloholik
[delati, sodelovati] kot dramaturg
roditi se kot [katolik, kralj, Slovenec]

(b) S pomenom lastnosti

biti kot brat in sestra
biti (komu) kot brat
biti (si) kot bratje
dihati kot [celota, eno/eden]

(c) S pomenom posnemanja; podobnosti

cviliti kot [prašiček, prašič]
[obletavati, brenčati] kot čebela
skakati kot kenguru
preluknjan kot ementalec

(č) S pomenom poudarjanja

oster kot [britev, britvica]
[priden, marljiv] kot čebela

(d) S prenesenim pomenom

brusiti koga kot diamant
kaj cveti kot še nikoli

(e) S pomenom časovne umestitve

kot dojenček
kot [mlad] fant

(f) S pomenom naštevanja

tako dečki kot deklice

Primerjalne SkZ se pogosto približujejo FE, saj njihov pomen ni vedno razviden iz posameznih sestavin. Posebne v tem smislu so primere, ki se uvrščajo pod katerega od metaforičnih pomenov sestavin (zgoraj: *brusiti*, *cveteti*) ali pa potrebujejo kot celota pragmatična ali žanrska pojasnila. V takih primerih smo zvezo navedli v frazeološkem delu.

14.4 Opis skladenjskih zvez v LBS

V zgradbi LBS smo SkZ znotraj posameznega pomena ali podpomena beležili v elementu <skladenjska_zveza> znotraj nadelementa <skladenjske_zveze>, in sicer lahko več zaporednih zvez, ki pripadajo določenemu (pod)pomenu besede v iztočnici, v elementih <zveza>:

zmagati *glagol*

1 biti, najboljši, najuspešnejši med vsemi

1.1 v športu, na tekmovanju

zmagati za las

zmagati za vsako ceno

Kadar obstaja več oblikovno različnih, a hkrati pomensko sorodnih SkZ, tj. takih, ki se nanašajo na sorodne pomenske vsebine, smo jih beležili v zaporednih elementih <zveza> znotraj enega nadelementa. Tako rešitev smo uporabili tudi pri navajanju oblik rabe in pretvorb pri FE. Zgledi si v takih primerih sledijo na koncu za vsemi navedenimi SkZ:

zmagati *glagol*

1 biti najboljši, najuspešnejši med vsemi

1.1 v športu, na tekmovanju

zmagati na [x] metrov

zmagati v teku na [x] metrov

zmagati (v teku) na [x] metrov s časom [x]

zmagati na [x] metrov [prosto, hrbtno, prsno]

zmagati z [izidom, rezultatom] [x : y]

zmagati v kategoriji do [x] [kg, let]

zmagati v kategoriji do [x.] leta

Za razliko od kolokacij in razširjenih kolokacij predvidevajo SkZ več zaporednih kolokabilnih nizov v elementu, ki predstavljajo kolokabilne sestavine SkZ. Beseda, ki je obravnavana v iztočnici, nikoli ni del takega kolokabilnega elementa (tj. nikoli ni zapisana v elementu <k></k>), pri čemer je s strukturnega in pomenskega vidika mogoče predvideti dve tipični situaciji:

(a) Beseda v iztočnici je le kolokabilni del določene SkZ, kar pomeni, da lahko v besedilih na njenem mestu nastopa niz, navadno pomensko sorodnih besed, v analizi pa je konkretna iztočniška beseda prišla na površje kot element besedilnega okolja:

internet *samostalnik*
v dobi interneta

zmagati *glagol*
zmagati za las
zmagati za vsako ceno

doba *samostalnik*
v dobi [odraščanja, globalizacije, interneta,
računalništva, informatike, vodnarja ...]

las *samostalnik*
za las [zgrešiti, uiti, izogniti, izmuzniti, prehiteti,
ujeti, zmagati]

(b) Beseda, ki jo obravnavamo v iztočnici, je leksikalizirani del SkZ, ob katerem lahko pride do izraza tudi njen kolokabilni zunanji del, npr.

čas *samostalnik*
[človek, glasba, ljudje, pot,
potovanje, sprehod ...] skozi čas

absurdnost *samostalnik*
do absurdnosti [kičast, karikiran]

14.5 Skladenjske zveze v razmerju do drugih večbesednih enot v LBS

Pomen ali skladenjska zveza

Glede na to da SkZ pomensko nismo razlagali je bila tudi dilema, ali gre za samostojni pomen besede ali za SkZ, ki to besedo vsebuje, na tej stopnji leksikalne analize že razrešena. Gre bolj za vprašanje, ali lahko govorimo o samostojnem pomenu besede ali o frazeološkem pomenu zveze, ki pa smo ga morali rešiti že pri pomenski členitvi. Če je bilo namreč mogoče ugotoviti, da je pomen omejen na relativno trdno zvezo, zunaj nje pa ne živi in posledično potrebuje razlago, smo se odločali za FE. Če pa je bil realen samostojni, navadno preneseni pomen besede, smo imeli možnost v njegovem razdelku navesti tudi SkZ brez razlage:

žezlo *samostalnik*

1.4 mesto voditelja, zmagovalca, zlasti v organizaciji, podjetju
vzeti/prevzeti žezlo v svoje roke
imeti žezlo v rokah

žezlo v ognju | imeti več žezel v ognju

Kolokacija – razširjena kolokacija – skladenjska zveza

Jasne in nedvoumne meje med SkZ in razširjenimi kolokacijami ni vedno mogoče določiti. Kot razločevalno lastnost smo upoštevali stopnjo ustaljenosti posameznih elementov zveze. Če torej enega ali več elementov besedne zveze ni bilo mogoče

zamenjati oz. smo ga lahko zamenjali z nekaj točno določenimi leksikalnimi elementi ali če je notranja zgradba besednega niza predvidevala oblikoskladenjsko omejena prosta mesta, smo obravnavali zvezo kot skladenjsko in ne kot razširjeno kolokacijo. V razširjenih kolokacijah, na drugi strani, posamezni elementi tipično kolocirajo s svojimi lastnimi kolokatorji, npr. pridevniki in prislovi, ki lahko določajo tudi obliko sledečega elementa, npr. deževati [več, nekaj] dni – deževati [cel, ves] dan. SkZ lahko za razliko od razširjenih kolokacij vsebujejo več kolokabilnih nizov, zabeleženih v več zaporednih elementih <k></k>, ter prosta skladenjska mesta, ki jih napovedujejo notranji elementi. Preglednica 26 prikazuje nekaj tipičnih primerov:

Preglednica 26: Razmerje med kolokacijami, razširjenimi kolokacijami in skladenjskimi zvezami

kolokacija	razširjena kolokacija	skladenjska zveza
aktivirati s [pritiskom, potegom]		aktivirati kaj s pritiskom na kaj
v fazi [pogajanj, dogovarjanja]	v fazi [pristopnih, širitvenih] pogajanj	v fazi [pristopnih, širitvenih] pogajanj s/z [kandidatkami, Evropsko unijo]
[delovno, prosto] mesto	[prosto, novo, plačano, vodilno] delovno mesto	[objava, prijava, zasedba] prostega delovnega mesta razpis za [prosto, vodilno] delovno mesto koga [objava, ponudba] prostega delovnega mesta v [medijih, javni upravi]
cerkveni [zbor, organist]	cerkveni [mešani, pevski] zbor	[peti, prepevati] v cerkvenem zboru

SkZ in razširjene kolokacije imajo tudi pomembno skupno lastnost: so pomensko prozorne in kot take ne potrebujejo razlage, čeprav v različnih jezikih, kot smo že izpostavili, nimajo nujno skladenjsko in pomensko enakih ustreznikov. Na osnovi pomenske prepoznavnosti, ki izhaja iz sestavnih elementov zveze, se tako SkZ kot razširjene in navadne dvodelne kolokacije razlikujejo od SZ in FE.

15 Stavčni vzorci

Pojem stavčnega vzorca v slovenistični literaturi ni nov. S stavčnimi vzorci (oz. stavčnimi tipi) se srečamo npr. pri B. Pogorelec, ko preučuje slogovna sredstva in izražanje govorečevega oz. piščevega sporočevalnega namena v besedilih 16. stoletja in moderne (Pogorelec 1968; 1969). Avtorica se v teh delih osredotoča bodisi na zgradbo prostega stavka, npr. na dativne konstrukcije (Pogorelec 1968), ali na jedra stavčne zgradbe, kadar je stavek zložen. Preučevanje stilističnih lastnosti je sicer, kot izpostavlja, stvar organizacije besedila, kjer je v izhodišču proučevanja bodisi več zaporednih stavkov (npr. z vidika ponavljanja ali uvajanja določenih stavčnih ali znotrajstavčnih konstrukcij, besednega reda ipd.) bodisi razmerja znotraj zloženega stavka (npr. raba veznikov, tipi odvisnikov ipd.).

Z vidika vezljivostne teorije in njenega prenosa v slovar obravnava stavčne vzorce A. Žele. Pri obravnavi vezljivosti kot zmožnosti določene besede, da veže nase določeno število vezljivostnih položajev, s čimer je določena obveznost skladijskih mest nasproti družljivosti, ki označuje prosta skladijska mesta, so stavčni vzorci obravnavani kot skladijsko okolje, ki določa strukturoskladijsko obveznost določil (Žele 2001: 13). Na podlagi skladijske in oblikoskladijske obveznosti določil in oblikoskladijske neobveznosti dopolnil nadalje ločuje med slovničnim in pomenskim stavčnim vzorcem (Žele 2001: 50, 58/59, 111). Slovnični stavčni vzorec označuje obvezne zapolnitve napovedljivih praznih skladijskih mest oz. obvezno izrazitev udeleženskih vlog ali t. i. stavčni minimum (Žele 2001: 239), pomenski stavčni vzorec pa izkazuje besedno usmerjenost na pomenski ravni povedi oz. vsebinski/pomenski minimum.

V slovenistični literaturi je pojmovanje stavčnega vzorca najbližje pojmovanju v LBS takrat, ko se ga razume kot formalizacijo skladijskih razmerij za izražanje pomensko-skladijskih lastnosti glagola znotraj prostega stavka. Toporišič (2000: 488) v zvezi s stavčnimi (skladijskimi) vzorci pravi takole: »Skladijski vzorec najmanjše možne oblike povedi ima obliko stavka določene vrste (simbol S) (torej obstajajo stavčni skladijski vzorci). /.../ V tem glavnem stavčnem vzorcu so besede zbrane okoli osebne glagolske oblike, tj. glagol odpira določena skladijska mesta ali pa dovoljuje njihova neobvezna druženja.« Vzorci so torej obrazci, v katerih so skladijska razmerja, ki se ustvarjajo med glagolom in njegovimi določili ter dopolnili, formalizirana. Skladijska razmerja se znotraj stavka ustvarjajo tudi znotraj neglagolskih besednih vrst z določenimi stavčnočlen-skimi vlogami, s tem pa tudi znotraj zloženega povedka – po Toporišiču (2000: 557) gre v teh primerih za besedne zveze ali stavčne fraze. V LBS je stavčna skladnja (in s tem stavčna vezljivost) od besednozvezne (in s tem tudi povedkove

vezljivosti) ločena tako, da je stavčna vezljivost prikazana s stavčnimi vzorci, povedkova vezljivost in besednozvezna skladnja pa s skladenjskimi strukturami.

Stavčni vzorci, kot jih razumemo v kontekstu izdelave LBS, so formalni prikaz vezljivosti posameznega glagolskega pomena znotraj stavčne zgradbe, ali natančneje, prikaz različnih skladenjskih in oblikoskladenjskih načinov uresničevanja določenega glagolskega pomena v zgradbi prostega stavka. Na ta način vzpostavljajo na eni strani povezavo z (a) pomensko shemo, ki vsebuje največji (v smislu pomensko in skladenjsko nujnih, čeprav v besedilu ne vedno izraženih določil in dopolnil) stavčni vzorec, izražen v svoji tipični skladenjski in oblikoslovni realizaciji, in na drugi (b) s kolokacijami kot leksikalnimi zapolnitvami posameznih vezljivostnih mest. Stavčni vzorci na ta način potrjujejo udeležensko zgradbo, zapisano v pomenski shemi, hkrati pa razkrivajo možnosti opuščanja elementov kot tudi tipične izbire znotraj več skladenjskih možnosti, kar je zlasti očitno pri predložnih zvezah v skladenjsko različnih vlogah. Vezljivostna mesta, kot jih predvideva posamezni glagolski pomen, so v stavčnih vzorcih izražena s poljubnostnim zaimkom *kdo/kaj*, ki poleg sklonov izraža tudi kategoriji neživosti in človeškosti (Toporišič 2000: 311). V LBS smo upoštevali samo pomensko dimenzijo te kategorije, kar pomeni, da smo z zaimkom *kdo* označevali vse, kar je s slovničnega vidika živo, torej tudi nečloveško, in vse, kar je človeško. Metonimijo, kot je npr. ustanova – ljudje v ustanovi, smo prav tako obravnavali s pomenskega in ne s slovničnega vidika, torej kot *kdo*, vso neživo predmetnost pa smo v vzorcih zapisovali s *kaj*.

Stavčnih vzorcev nismo zapisovali pri vezljivih pridevniki in samostalnikih, smo pa v stavčnih vzorcih glagolskih pomenov posebej opredeljevali tiste dele, ki izražajo okoliščine, kot so kraj, smer in čas, ter lastnosti, kot so način pa tudi namen in vzrok. Načini skladenjskega in oblikoskladenjskega izražanja teh pomenskih vsebin so v LBS prikazani z vzorci, združenimi pod pomenske skupine, ki jih označujejo izrazi: NAČIN, LOKACIJA, SMER, ČAS in VZROK.

Prikazano pojmovanje stavčnih vzorcev v LBS je glede na teoretična izhodišča najbližje prepoznavanju stavčnih vzorcev v povezavi s konkretnimi pomeni obravnavane besede, kot izhajajo iz projektov FrameNet in CPA, ki smo ju opisali v samostojnem poglavju in v poglavju, kjer opisujemo teoretična izhodišča za oblikovanje pomenskih shem. Naj na tem mestu ponovno izpostavimo le tista izhodišča, ki se nanašajo na pojmovanje stavčnih vzorcev v LBS.

Pri FrameNetu so za razumevanje stavčnih vzorcev (in za oblikovanje pomenske sheme) pomembni predvsem t. i. shemski elementi (ang. *frame elements*). Kot smo že dejali, so ti obravnavani glede na obvezno (ang. *core elements*) in

neobvezno (ang. *non-core elements*) prisotnost v določeni shemi. Prepoznavanje elementov za določeno shemo oz. za posamezne glagolske pomene, ki tej shemi pripadajo (t. i. leksikalne enote), je v izhodišču pomensko (in ne skladenjsko): da bi bilo mogoče ugotovili, kateri so shemski elementi za določen glagolski pomen, je treba pregledati razmeroma obsežno število naključno izbranih primerov iz korpusa. Če izhajamo iz za slovenščino prirejenega primera za glagol *pritrđiti* (Krek 2008: 39), lahko ugotovimo, da je v scenariju pritrdjevanja vedno prisoten vršilec, ki pa ni nujno eksplicitno izražen, nadalje predmet, ki ga vršilec pritrdi na nek drugi predmet, in t. i. vez, ki je navadno predložna zveza s predlogom *s/z*, ki je prav tako lahko izražena ali pa ne. Iz tega izhaja, kot ugotavlja Krek, da ima pomenska shema za glagol *pritrđiti* 4 obvezne elemente, in sicer VRŠILCA, ki kaj pritrdi, PREDMET, ki ga pritrdjuje, DRUGI PREDMET, na katerega pritrdjuje prvi predmet, in SREDSTVO, ki oba predmeta povezuje v trdno zvezo. Vsi drugi elementi, ki se pojavljajo v okolici pomenske sheme, so t. i. obrobni elementi, navadno so to stavčna določila, ki izražajo čas, prostor, način, vzrok ipd. (Krek 2008: 40).

Postopek določanja pomenske sheme, ki velja za FrameNet, tj. prepoznavanje shemskih elementov, formalno definiranje sheme in njenih elementov, ima vzporednice tudi v postopku prepoznavanja in oblikovanja stavčnih vzorcev v LBS, kjer smo na podlagi korpusnih primerov za posamezni glagol ugotovili pomen in stavčne vzorce, čeprav je logično, da se posamezni vzorci znotraj različnih pomenov tudi prekrivajo. Pri tem smo ugotavljali, kateri elementi stavčnega vzorca so obvezni in jih želimo vključiti tudi v pomensko shemo, kateri elementi so obrobni oz. kateri elementi se tipično opuščajo in kateri ne ter kateri elementi so medsebojno zamenljivi, npr. pritrditi kaj/na kaj/ob kaj/v kaj/pod kaj/okoli česa itd. Poleg tega pa še semantične tipe posameznih delov stavčnih vzorcev, ki so za določen glagolski pomen tipični in jih abstrahiramo iz kolokacij. Za razliko od postopka v FrameNetu v LBS nismo ugotavljali udeleženskih vlog. Če ostanemo pri glagolu *pritrđiti*, bi se pomenska shema z obveznimi elementi, izraženimi s semantičnimi tipi, glasila: če ČLOVEK pritrdi nek PREDMET na DRUG PREDMET, ga z njim poveže v trdno celoto, navadno s PRIPOMOČKOM ali uporabo SNOVI. Na podlagi tipičnih realizacij stavčnega vzorca v pomenski shemi: kdo pritrdi koga/kaj na koga/kaj s čim, je mogoče ugotoviti naslednje realizacije:¹⁵⁸

158 Med 200 pregledanimi konkordancami so bile izbrane samo tiste, ki so ustrezale pomenu 'namestiti kaj na določen predmet, mesto'.

kdo pritrdi kaj	obloge bodo močnejše pritrdili
kdo pritrdi kaj na kaj	na Columbus sta pritrdila opremo
kdo pritrdi koga na kaj	na drugi konec je pritrdil sebe, skočil čez plot in rodil se je bungee jumping
kdo pritrdi kaj na koga	zdravniki pritrdijo tablice na paciente
kdo pritrdi kaj nad kaj	nad mizo pritrdimo še polico za odlaganje
kdo pritrdi kaj ob kaj	platinasti žički pritrdimo ob steno petrijevke
kdo pritrdi kaj v kaj	okrasje, ki ti ga je v lase pritrdil frizer, lahko odstrani le on
kdo pritrdi kaj pod kaj	vzemi rezervni ključ avtomobila in ga nekako pritrdi pod pokrov platišča
kdo pritrdi kaj pod kaj (s čim)	ženske so pritrdile ročno pletene nogavice pod kolenom s podvezo
kdo pritrdi kaj med kaj	med tramove pritrdimo toplotno izolacijo
kdo pritrdi kaj čez kaj	čez stopalo pritrdimo elastičen trak
kdo pritrdi kaj okoli česa	Tsaya je torbica, ki jo ženska pritrdi okoli stegna
kdo pritrdi kaj za kaj	domiselna naprava, ki jo pritrdite za pas
kdo pritrdi kaj s čim	napravo postavimo na trdno mizo in jo pritrdimo s prižemami
kdo pritrdi kaj na kaj (s čim)	nato na boleče mesto s povojem pritrdimo led
kaj se pritrdi na kaj	virus se pritrdi na steno celice
kaj se pritrdi s čim	korenine ne ozelenijo. Z njimi se rastlina pritrdi in sprejema iz okolja vodo
kaj se pritrdi v kaj	nove rastline se pritrdijo v tla
kdo si pritrdi kaj v kaj	pramen si spleti v kitko in si jo pritrdi v lase
LOKACIJA	
kdo pritrdi kaj v čem	tečaje pritrdimo v višini povezovalnih letev
kdo pritrdi kaj na čem	na vrhu razgledne ploščadi so pritrdili antene za mobilno telefonijo

Za identifikacijo semantičnih tipov smo nato v orodju SkE uporabili besedno skico in določili kolokacije, s katerimi se potrjujejo posamezni deli stavčnih vzorcev:

[ličinka, virus] se pritrdi; [samica] pritrdi; [potapljač] si pritrdi
 pritrditi [letev, letvico, kljukico, elektrodo, zaponko, zanko, vrvico]
 pritrditi z [bucikami, lasnico, zobotrebcem, žebličkom, vijakom, elastiko, gumico, žico, lepilom, silikonom]
 pritrditi na [deščico, ovratnico, trup, steno, stojalo, obod, ogrodje, podstavek, desko, vrvico, strop]
 pritrditi v [akvarij, trup, tla, avto, notranjost, dno]
 pritrditi za [pas]
 pritrditi pod [strop]
 pritrditi ob [rob]

Pri CPA gre v izhodišču za to, kot pravi Hanks (2010), da je vsaka beseda, ki jo uporabi govorec, da bi z njo nekaj povedal, uporabljena v skladu z neko normo (dogovorom), ki jo določa njen stavčni vzorec. Na podlagi tega stavčne vzorce deli na osnovne, ki predstavljajo veliko večino pomenskih realizacij določenega glagola (ang. *norms*), na vzorce, ki predstavljajo različne skladijske realizacije izhodiščnega glagolskega pomena (alternacije), in na ustvarjalno rabo ali t. i. eksploatacije (ang. *exploations*). Osnovna poudarka CPA sta skladijska vezljivost in formalizirano povzemanje semantičnih tipov v obliki t. i. implikatur, ki pomensko opredeljujejo vezljivostna mesta, kar v LBS pokrivajo semantični tipi. Znotraj nabora osnovnih vzorcev tudi CPA razlikuje oba nivoja: v vzorcu/patternu nakaže skladijsko opcijskost elementa, v opisu/implikaturi pa navede vse elemente kot (semantično) obvezne, npr. za glagol *vaccinate* (cepiti):

PATTERN: [[Human 1 = Health Professional]] vaccinate [[Human 2 = Patient]]
{(against [[Disease]])}

IMPLICATURE: [[Human 1 = Health Professional]] injects a vaccine into [[Human 2 = Patient]] to produce immunity against [[Disease]]

V LBS opis za isti glagolski pomen izgleda takole:

cepiti *glagol*

1 prejeti ali vnesti cepivo

če se ČLOVEK cepi, prejme CEPIVO proti določeni BOLEZNI |ali| če ZDRAVNIK cepi ČLOVEKA, mu v telo vnese CEPIVO proti določeni BOLEZNI

kdo cepi

kdo se cepi

kdo se cepi proti čemu

kdo cepi koga/kaj

kdo cepi koga proti čemu

kdo cepi koga s čim

Iz prikazanega je razvidno, da so izhodišča FrameNeta z vidika prenosa v LBS pomembna predvsem pri oblikovanju pomenske sheme, kjer izhajamo iz t. i. pomenske vezljivosti, ki določa obveznost (izraženo, neizraženo in implicirano) udeleženskih mest in okoliščin, medtem ko se pri oblikovanju stavčnih vzorcev približujemo CPA, ki izhaja neposredno iz konkretnih jezikovnih situacij, kjer se pomensko določena udeleženska zgradba realizira na različne načine, ko se spreminja število in razporeditev skladijskih udeležencev.

15.1 Vrste stavčnih vzorcev v LBS

Stavčne vzorce smo v LBS glede na to, ali vključujejo zgolj udeležence ali tudi okoliščine in lastnosti, potrebne/tipične za izražanje obravnavanega glagolskega pomena, obravnavali v dveh skupinah. V prvi skupini so za vsak glagolski pomen, ki ima svoj najširši skladiškopomenski stavčni vzorec vključen v pomensko shemo (tj. tako udeležence kot okoliščine), navedeni stavčni vzorci, ki ne vključujejo predložitnih zvez, s katerimi se izražajo časovne in krajevne okoliščine, način ali vzrok glagolskega dejanja, pač pa samo udeleženci. To konkretno pomeni, da bo stavčni vzorec *kaj boli koga*, realiziran v kolokaciji [križ, sklepi, trebuh] boli, obravnavan v prvi skupini, stavčni vzorec *bolí koga v čem*: boli v [križu, sklepih, trebuhu] pa v drugi skupini stavčnih vzorcev pod skupnim imenovalcem LOKACIJA, o čemer več v nadaljevanju.

Na tej prvi ravni izražajo stavčni vzorci različne realizacije prototipičnega vzorca v pomenski shemi v smislu povratnosvojilnih zgradb, brezoseb(kov)nih in tretjeosebnih realizacij in z vidika izražanja oz. opuščanja določil. Vzorce potrjujemo v spodnjih primerih s kolokacijami, če se posamezni deli realizirajo s tipičnimi leksikalnimi zapolnitvami, sicer pa s prilagojenimi korpusnimi zgledi:

boleti *glagol*

1 o telesni bolečini

1.1 čutiti bolečino

če ČLOVEKA boli DEL TELESA, čuti v njem bolečino

boli koga	<i>tako jo je bolelo, da ni mogla spati</i>
kaj boli koga	[glava, trebuh, hrbet] boli; če otroka boli trebuh, mu pripravimo kamilični čaj

LOKACIJA:

boli koga v čem ¹⁵⁹	boli v [prsih, križu, trebuhu, sklepih]
--------------------------------	---

2 o duševni stiski

če ČLOVEKA boli neko DEJSTVO ali DOGODEK, se čuti prizadetega in nemočnega

kaj boli koga	[krivica, resnica] boli
boli koga, da	<i>najbolj ga boli, da še danes ne pozna razloga</i>
boli koga, ker	<i>boli jih, ker niso mogli pomagati</i>
boli koga, ko	<i>vsakega normalnega človeka boli, ko je izdan</i>

159 Posamezni deli stavčnega vzorca so med seboj zamenljivi oz. izključujoči, npr. kaj boli koga in v čem boli koga nikoli ne moreta tvoriti vzorca: *kaj boli koga v čem, zato je za pripravo vzorcev za avtomatsko procesiranje potrebno posamezne zamenljive oz. izključujoče dele posebej označiti. Leksikalne zapolnitve na teh mestih niso arbitrarne in vedno vzajemno zamenljive: ušesa/v ušesih boli, križ/v križu boli : boli v prsih, glava boli (neizkazano: prsa bolijo, v glavi boli).

absorbirati *glagol*

1 vsrkati; vpiti

če SNOV, ORGANIZEM ali PREDMET absorbira drugo SNOV, ENERGIJO ali TEKOČINO, jo vsrka vase |ali| če se SNOV, ENERGIJA ali TEKOČINA absorbira v drugo SNOV, ORGANIZEM ali PREDMET, se vanjo integrira

kaj absorbira kaj	[telo, organizem] absorbira; absorbirati [svetlobo, energijo, toploto]
-------------------	--

kaj se absorbira	[svetloba, vonj, kalcij, snov] se absorbira
------------------	---

Za izražanje brezpredložnega primičnega skladenjskega razmerja, ki označuje pretežni del družljivosti in kjer pomenskoskladenjska povezanost z jedrnim glagolom nima sklonске vplivnosti in posledično ni končniško izražena (Žele 2003: 11), smo uporabljali glagolske strukture s prislovnimi elementi:

rbz GBZ:

[hitro, dobro, bolje, najbolje, slabo, slabše, počasi, učinkovito, popolnoma] absorbirati
 [ponovno, znova, hitro, maksimalno] aktivirati; [politično] se aktivirati
 [počasi, pomalem] se svitati
 [sprva, dolgo, resnično] begati

Mod-rbz GBZ:

[lahko] absorbirati; [lahko] se absorbirati
 [lahko] aktivirati

V drugi skupini smo v vzorcih navajali predložne realizacije krajevnih in časovnih okoliščin ter načine izražanja vzroka, namena in načina pri obravnavanem glagolskem pomenu, kjer smo vzorce združevali pod skupni pomenski imenovalec. Sem sodijo tudi primeri, kjer predlogi – Žele (2003: 11) jih imenuje vezavnodružljivi predložni glagolski morfemi – pomensko sodijo k predložnim prislovnim določilom, od glagola pa so vplivani le sklonsko, npr. pasti pod kolesje, in primeri, kjer gre za predložna prislovna določila, ki jih je v nekaterih primerih mogoče pretvoriti v zveze s prislovi: delati z veseljem – veselo delati. Tovrstne primere smo v LBS obravnavali le s pomenskega vidika, načinov skladenjskih razmerij, tj. ali gre za vezavo ali primik pa nismo ugotavljali. Nekaj primerov:

aktivirati *glagol*

1 povzročiti, da postane kaj dejavno

1.1 o človeku, dejavnosti

če ČLOVEK aktivira drugega ČLOVEKA, njegove MISLI, OBČUTJA ali SPOSOBNOSTI, povzroči, da postane dejaven |ali| če ČLOVEK aktivira določeno DEJAVNOST, povzroči, da se začne odvijati

kdo/kaj se aktivira
kdo/kaj aktivira koga/kaj

[potencial] se aktivira
slovenska vlada bo aktivirala zakon o zaščiti tujcev

NAMEN

kdo se aktivira pri čem
kdo aktivira koga pri čem
kdo aktivira kaj ob čem

aktivirati se pri reševanju krize
aktivirati alkoholika pri reševanju problemov
aktivirati znanje ob konkretnem problemu

NAČIN

kdo aktivira koga/kaj s čim

z ustreznimi metodami aktivirati miselne potencialne

Kadar se tovrstne vsebine izražajo z ustaljenimi zvezami ali potrebujejo še dodatni element, smo jih navajali v segmentu skladijskih zvez, ki veljajo za obravnavani pomen glagola, npr.

aktivirati kaj na [nek, tak, svojstven] način

begati sem in/ter tja

demonstrirati v [prijem, spomin] komu/čemu

demonstrirati v znamenje [podpore, protesta] komu/čemu

V zvezi z navajanjem stavčnih vzorcev glede na opuščanje predložitnih določil in dopolnil je treba povedati, da smo navajali le tiste, za katere smo našli potrditve v dejanskih besedilnih realizacijah. To seveda ne pomeni, da se nekatere sistemske možnosti, ki izhajajo iz navedenih stavčnih vzorcev, pa jih v slovarski bazi nismo navedli, v besedilu nikoli ne uresničujejo. Take vzorce je, če imamo v pomenski shemi naveden pomensko-skladijski maksimum, v posameznih vzorcih pa različne predložne realizacije, mogoče avtomatsko zgenerirati, kar smo prikazali pri enem od pomenov glagola *zmagati* v samostojnem poglavju.

15.2 Prepoznavanje stavčnih vzorcev na podlagi korpusa

15.2.1 Analiza konkordančnega niza

Relevantne stavčne vzorce smo pri oblikovanju LBS evidentirali na dva načina. V prvi fazi je analiza potekala ročno hkrati s pomensko členitvijo ob analizi

konkordančnega niza 150 do 300 konkordanc, saj so stavčni vzorci, ob tem da so lahko identični znotraj posameznih pomenov in med pomeni različnih glagolov, običajno znanilci glagolske večpomenskosti. V drugi fazi smo skušali vzorce iz korpusa pridobiti avtomatsko s pomočjo prilagojene slovnice besednih skic v SkE.

15.2.2 Avtomatsko pridobivanje stavčnih vzorcev iz korpusa

Zadnja različica gramatičnih relacij v slovnici besednih skic v orodju SkE – verzija ALLP 0.8 (Krek 2012a), ki je bila izdelana za avtomatsko luščenje podatkov iz korpusa, vsebuje med drugim tudi t. i. direktivi CONSTRUCTION in SEPARATEPAGE, ki omogočata podrobnejšo razčlenitev določenih skladenjskih razmerij znotraj stavka. Tako npr. direktiva CONSTRUCTION omogoča prepoznavanje skladenjskih struktur brez kolokacij, kar je primerno predvsem za luščenje vezljivostnih vzorcev pri glagolih (Preglednica 27).

Preglednica 27: Gramatične relacije za avtomatsko luščenje stavčnih vzorcev iz korpusa

Gramatična relacija v SkE	Prepis v element DTD pri avtomatskem luščenju vzorcev	Korpusni primer
V_S_V_O3	<vzorec>kdo/kaj G komu</vzorec>	kdo čestita komu; kdo/kaj dominira komu/čemu; kdo dvori komu; kdo gospodinji komu
V_S_V_O3_O2	<vzorec>kdo/kaj G komu koga/česa</vzorec>	kdo ne dovoli komu česa
V_S_V_O3_O4	<vzorec>kdo/kaj G komu kaj</vzorec>	kdo predlaga komu kaj; kdo nakloni komu kaj, kdo prekriža komu kaj
V_S_V_O4_predl	<vzorec>kdo/kaj G koga/kaj + predlog</vzorec>	kdo obvaruje koga pred (čim); kdo odvrča koga od (česa)

Direktiva SEPARATEPAGE pa je namenjena odpiranju relacij s tremi elementi (*TRINARY) na novi spletni strani znotraj SkE, kar pomeni, da je za konkretno slovnično relacijo, ki vključuje npr. razmerje besede, obravnavane v iztočnici (npr. glagola), z drugim elementom (npr. samostalnikom) v določenem sklonu (npr. mestniku: kom-čem_g5), mogoče v ločenih konkordančnih nizih analizirati vse predloge, ki se s tem sklonom vežejo. Na novi strani odprta besedna skica poleg tipičnih predlogov izkazuje tudi kolokatorje za drugi element relacije, kot prikazuje Slika 42.

gojiti (glagol)		
kom-čem g5	4165	
<u>S_v-d_g5</u>	<u>2579</u>	5.1
<u>S_na-d_g5</u>	<u>1192</u>	5.9
<u>S_po-d_g5</u>	<u>220</u>	1.8
<u>S_ob-d_g5</u>	<u>88</u>	2.3
<u>S_pri-d_g5</u>	<u>59</u>	1.1
<u>S_o-d_g5</u>	<u>27</u>	0.6

S_v-d_g5
gojiti v [rastlinjaku, lončku nasadu, vrtnariji]
gojiti v [ribogojnici, akvariju, epruveti]
gojiti v [Sredozemlju]

S_na-d_g5
gojiti na [plantažah, gredicah, vrtu, balkonu]
gojiti na [farmi, posestvu]
gojiti na [Kitajskem, Primorskem]

S_po-d_g5
gojiti po [vrtovih]

S_ob-d_g5
gojiti ob [opori]

S_pri-d_g5
gojiti pri temperaturi [x] stopinj

S_o-d_g5
kdo goji kaj o čem
gojiti stereotipe o harmoniki
gojiti predsodke o duševnem zdravlju

Slika 42: Direktiva SEPARATEPAGE v slovnici besednih skic

Tovrstne relacije prinašajo zelo podrobno razčlenitev predložnih zvez tipa samostalnik-predlog-samostalnik, glagol-predlog-samostalnik, pridevnik-predlog-samostalnik itd., ki je z vidika ročne leksikografske obdelave preprosta in prenametna. Pri avtomatskem prenosu teh relacij ter pripadajočih kolokacij in zgledov iz korpusa neposredno v slovarski vmesnik pa mora leksikograf oceniti le, ali je relacija ter z njo povezana kolokacijska informacija za pomen dane besede, obravnavane v iztočnici, relevantna ali ne. Take odločitve navadno ne zahtevajo poglobljenih leksikografskih analiz, zato je za odstranjevanje nerelevantnih podatkov smiselno uporabiti tudi druge oblike čiščenja slovarske baze, kot je denimo množičenje ali postopno izdelovanje slovarskih gesel, kjer tovrstne preglede lahko opravijo leksikografi začetniki.

Postopek je mogoče ponazoriti s spodnjim primerom, ki prikazuje gramatične relacije, ki predvidevata različna stavčna vzorca za glagol absorbirati, ko se ta veže s predlogom v in samostalnikom v mestniku ali tožilniku: kaj se absorbira v čem in kaj se absorbira v kaj. Iz konkordanc, ki jih na podlagi slovnične relacije zgenerira orodje SkE v besedni skici, je nato mogoče ugotoviti, ali posamezni primeri dejansko ustrezajo predvidenemu stavčnemu vzorcu, kar označujemo s kljukico, in so torej za obravnavano besedo relevantni, ali ne (označeno z x):

kom-čem_g5

- S_v-d_g5

- kaj se absorbira v čem

- ✓ Če ob pijaci tudi ješ, se alkohol absorbira v hrani, zato traja dlje časa, da dospe v kri.
- ✗ Olja v čistilni kremi absorbirajo maščobe in umazanijo z obraza.
- ✓ Del svetlobe sploh ne prodre v akvarij, ker se odbije že na vodni gladini, precejšen del se absorbira v vodi.
- ✓ Najbolje naj bi se v prebavilih absorbirali sladkorji v približno šestodstotni raztopini.
- ✓ Hrana se ob uporabi teh preparatov hitreje prebavi, tako se del zaužitih maščob ne absorbira v črevesju.
- ✓ Vлага se v montažni biši absorbira v mavčno-vlaknenih oblogah.
- ✗ Kozarec vode se v petnajstih minutah absorbira v kri.

koga-kaj_g4

S_v-d_g4

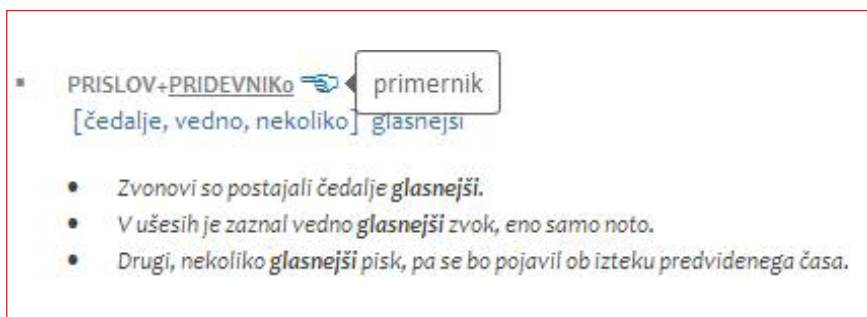
- kaj se absorbira v kaj

- ✓ Raztopina iz modnika se potem absorbira v spužvasto tkivo penisa in znatno poveča njegov obseg.
- ✓ Je tudi vam prišlo na uho, da se kalcij v homogeniziranem mleku težje absorbira v telo?
- ✗ Ob nastanku je ta energija v obliki žarkov gama, potem pa se na svoji poti proti zunanjim plastem Sonca večkrat absorbira v snovi in spet sprosti pri vedno nižjih temperaturah.
- ✓ Alkohol, ki se absorbira v materino kri, preide na rastočega otroka.

Avtomatsko luščenje stavčnih vzorcev iz korpusa omogočajo tudi druga orodja, zlasti skladenjski razčlenjevalniki, ki prepoznavajo skladenjska razmerja med stavčnimi gradniki. V tem smislu je razčlenitev posameznega glagolskega pomena na stavčne vzorce in navedba prototipičnega stavčnega vzorca v pomenski shemi uporabna v navezavi z razdelanimi skladenjskimi razmerji, kot jih predvideva skladenjski razčlenjevalnik. Možnost uporabe stavčnih vzorcev v LBS v povezavi s skladenjskim označevanjem in pripisom udeleženskih vlog je predmet samostojne raziskave, ki je v času nastanka tega besedila še v teku.

17 Restrikcije

Restrikcije v LBS razumemo kot poseben tip slovničnih oz. skladenjskih oznak, ki jih pripenjamo na posamezne dele skladenjskih struktur ali stavčnih vzorcev in so v prvi vrsti namenjene označevanju posebnosti znotraj predvidljive slovnične paradigme, skladenjske vloge ipd., zaradi česar so uporabne za prepoznavanje izjem pri strojnem procesiranju podatkov, ter za slovnične in skladenjske analize, zato so zanimive za zahtevnejše uporabnike (jezikoslovce, učitelje), ki se osredotočajo na slovnične in skladenjske posebnosti v jeziku (npr. pri pripravi učnih gradiv ipd.). Kot take so restrikcije tisti element slovarske baze, ki v slovarju navadno ni prikazan na prvi ravni, ampak bodisi z zahtevnejšimi iskalnimi možnostmi ali na drugem nivoju – pri spletni postavitvi LBS npr. z možnostjo klika ali prečjenja z miško prek besedila na zaslonu, kot prikazuje Slika 43:



Slika 43: Prikaz slovničnih omejitev v demo spletni postavitvi LBS

Označevanje slovničnih in skladenjskih posebnosti pa ni namenjeno le strojnemu procesiranju, pač pa tudi ustrezni rabi besede ali zveze v določenem besedilnem okolju in v določenem pomenu, zato so tovrstna opozorila namenjena tudi šolskemu uporabniku in pri učenju slovenščine kot tujega jezika. V LBS smo v fazi strukturiranja in določanja relevantnih leksikalnogramatičnih podatkov uporabljali pri označevanju omejitev v slovenski leksikografski tradiciji (prim. SSKJ in SNB) bolj ali manj uveljavljene ubeseditve kot npr. v povedni, prislovni, predložni, pridevniški ipd. rabi, vendar pa bi bilo z vidika razumljivosti slovarskih navodil potrebno vsaj v zvezi z nekaterimi razmisliti o drugačnih ubeseditvah, ki ne predvidevajo poznavanja slovničnega oz. slovarskega metajezika, npr. v položaju za glagolom biti, v vlogi predloga, prislova, pridevnika ipd. Ustrezne odgovore v tej smeri bi prav gotovo ponudila raziskava uporabniškega mnenja in razumevanja različnih ubeseditvenih možnosti.

17.1 Vrste restrikcij v LBS

(a) **Slovnične omejitve** se nanašajo na posamezni del skladenjske strukture, ki je v konkretni kolokaciji vezan zgolj na določeno slovnično obliko.

(a1) V večini primerov gre za omejitev na slovnično število, zlasti na množino, redkeje na dvojino. Omejitev na množino je zlasti produktivna pri strukturah, ki vsebujejo količinski element, pri strukturah, ki vključujejo samostalnik v rodilniku in pri strukturah, kjer predlog (lahko) predvideva množinsko udeležbo (*med, iz-med, po* itd.)

Kol-sbz0 SBZ2<r>**v množini**</r>

[čreda] antilop; [kopica, kup, vrsta] nasvetov; [škatla, komplet] barvic

Kol-rbz SBZ2<r>**v množini**</r>

[več, veliko, precej] avtomobilov

pbz0 SBZ2<r>**v množini**</r>

[poln] absurdov; poln [lepih] čednosti

sbz0 izmed SBZ2<r>**v množini**</r>

[eden] izmed novinarjev; [ena] izmed zvezdnic

SBZ0 med sbz6<r>**v množini**</r>

[potka, poti] med gredami; [nasilje, razlike] med otroki; [žrtve] med civilisti

gbz po SBZ5<r>**v množini**</r>

[posedati] po bifejih

V redkih primerih smo opozarjali tudi na edninsko omejitev, zlasti takrat, ko se je različna omejitev števila pri konkretnem elementu znotraj identične strukture pokazala kot pomensko relevantna:

identiteta *samostalnik*

1 prepoznavne lastnosti

1.1 pripadnost na podlagi nacionalnih in verskih vrednot

Kol-rbz SBZ2<r>**v množini**</r>

[veliko, več] identitet

- *Ljudje na tem ozemlju imajo več identitet: vsaj arabsko, plemensko in versko.*

1.1 prepoznavnost na podlagi osebnih, specifičnih lastnosti

Kol-rbz SBZ2<r>**v ednini**</r>

[veliko, več] identitete

- *Tako je croma zgledno oblikovan avto s karavansko obliko, a ne izraža veliko Fiatove identitete.*

(a2) Omejitve na določeno obliko pri stopnjevanju smo pripenjali na pridevniške in prislovne elemente skladenjskih struktur:

RBZ< r> **navadno primernik**< /r> gbz
 aktivneje [poseči, posegati, vključiti]
 PBZ0< r> **primernik ali presežnik**< /r> sbz0
 ožja [okolica, domovina]
 pbz0< r> **primernik**< /r> SBZ0
 [višja, nižja] najemnina

(a3) Omejitve na določeno obliko glagola

gbz< r> **brezosebno**< /r> SBZ2
 argumentov [zmanjkuje, zmanjka]; časa [zmanjkuje]; [naredi] klik
 gbz< r> **navadno v velelniku množine**< /r> SBZ4
 • *Kar kliknite! Na zaslonu se vam bo pokazal popisan oblaček.*
 iskati< r> **v velelniku**< /r> po čem
 iskati po [ključnih] besedah; iskati po [avtorju, izrazih]
 • *Išči po: naslovu igralcev režiserju zvrsti letu državi.*

(b) Skladienske oz. pomensko-skladienske omejitve

Kol-rbz PBZ2< r> **samostalniški rabi**< /r>
 [nič] alkoholnega
 sbz1< r> **v pridevniški rabi**< /r> SBZ0
 [DOS] aplikacija; čarter [plovilo, jadrnica]; čili [paprika, omaka]; džez [scena, festival]
 sbz1 Vez-gbz PBZ1< r> **v povedni rabi**
 • *Stvari so tako povsem evidentne.*
 RBZ Vez-gbz sbz1< r> **v povedni rabi**< /r>
 • *Zunaj je zima, v zakurjeni sobi pa sedemo za pogrnjeno mizo.*
 PBZ0< r> **v prislovni rabi**< /r> rbz
 kostanjevo [rjav]
 Kol-sbz< r> množina< /r> < r> **v prislovni rabi**< /r> rbz
 stoletja [dolgo]

(c) Druge omejitve

pbz0< r> **poudarno**< /r> SBZ0
 [pravi, čisti] bolnik
 SBZ0 na sbz5< r> **navadno z lastnim imenom**< /r>
 dopust na [morju, Hrvaškem, Pagu, Jamajki]
 SBZ0 v SBZ5< r> **navadno z lastnim imenom**< /r> v sbz5
 intervencija v [Iraku, Vietnamu, ZRJ, Somaliji, Afganistanu, Bosni]

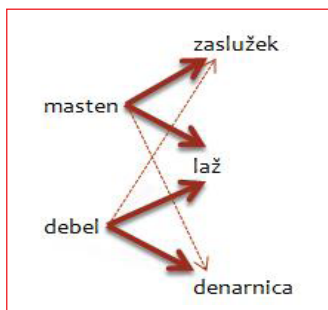
IV. KOLOKACIJSKI NIVO

Kolokacije so, od kar je mogoče opazovati leksikalno obnašanje besed v njihovem neposrednem besedilnem okolju v realnih besedilih in s pomočjo statističnih orodij, nepogrešljiv del pomenskega opisa besed.¹⁶⁰ Kot smo lahko spremljali skozi celoten nivojski opis leksikalnogramatičnih podatkov v LBS, imajo kolokacije pomembno vlogo tako pri pomenski členitvi oz. pomenskem razdvoumljanju besed, pri njihovem pomenskem opisu kot tudi pri določanju in medsebojnem ločevanju večbesednih leksikalnih enot, pri prepoznavanju sinonimije ipd. Kolokacije so zato ključni del sodobno zasnovanih slovarjev in imajo posebno vlogo zlasti pri učenju tujega jezika ter na sploh pri empiričnih leksikalnih analizah.

160 Kot je mogoče razbrati iz gradiva, je bilo merilo pomenske delitve na podlagi tipičnih kolokacij, vsaj pri nekaterih besedah, uporabljeno tudi v SSKJ.

18 Kaj so kolokacije

Kolokacije so izrazi, sestavljeni iz dveh ali več besed, s katerimi na konvencionalen, v jeziku običajen način izražamo stvari (Manning in Schütze 1999: 141). Kot take so kolokacije lahko zelo različnih vrst in vključujejo različne skladenjske in oblikoskladenjske možnosti. Posebno pozornost zbujejo jezikovnospecifični in ne vedno enoznačno razložljivi primeri jezikovne rabe, ki so posebej dobro znani na eni strani maternim govorcem, ki z njimi nimajo težav, in na drugi strani učencem tujega jezika, ki jih prepoznavajo kot jezikovnospecifične leksikalne celote, ki se jih je potrebno naučiti na pamet. Gre namreč za to, da so nekatere besedne povezave za rojene govorce sprejemljive, druge pa ne, čeprav so oblikovane pomensko in skladenjsko analogno, kot prikazuje Slika 44.¹⁶¹



Slika 44: Primer jezikovnospecifične kolokacijske izbire

Bistvo omenjenih sintagmatskih pogledov na jezik je spoznanje, da je razmerje med pomenom in obliko zelo subtilno, saj različni pomeni leksikalne enote pogosto ustrezajo samo izbranim besednim kombinacijam. Prav tako je očitno, da lahko dokazi o besedni povezovalnosti spodbijajo introspekcijo rojenih govorcev in da se večina naravnega jezika pojavlja v vsaj deloma vnaprej oblikovanih frazah, ki jih določajo omejene izbire (ang. *single choices* in *idiom principle*; Sinclair 1987; 1991), čeprav se sicer na prvi pogled zdi, da jih je mogoče razstaviti na posamezne elemente. Pojem kolokacije, ki se je oblikoval na tej podlagi, je tudi v nadaljnjih obravnavah pri posameznih avtorjih izpostavljal bodisi statistično prepoznavanje karakterističnih vzorcev sopojavitve besed (t. i. formalnoverjetnostni, statistični pristop) oziroma se je skušalo določiti karakteristične vzorce tudi z vidika pomen-ske odvisnosti znotraj zveze, torej kot pojav, ki je poleg svoje strukturne povezovalne trdnosti tudi pomensko prepoznaven.

¹⁶¹ Primer za slovenščino je prirejen po Hallidayjevemu primeru za angleščino (1966: 105), kjer izpostavlja konvencionalnost kolokacije *strong tea*, ne pa tudi *powerful tea*, čeprav je mogoče pričakovati, da je zveza tudi v taki obliki razumljiva večini angleških maternih govorcev.

18.1 Kolokacija kot statistični pojav

S tem ko se je pozornost s preučevanja besednega pomena z vidika (ne)razstavljalivosti, komponentne analize in pomenskih razmerij med leksikalnimi enotami preusmerila v preučevanje distribucije besed v besedilu in predvidljivosti pogostejšega sopojavljanja nekaterih besed z enimi kot z drugimi, so bila dana izhodišča za obravnavanje besednega povezovanja kot statistično predvidljivega jezikovnega pojava.

Pri preučevanju značilnih vzorcev besednih sopojavitev znotraj korpusnega okolja, ki temelji na upoštevanju absolutne frekvence besedne povezovalnosti, so na površje prišle besedne sopojavitve, t. i. bi- in trigrami, ki za leksikalne analize niso preveč zanimive in navadno tudi niso del osrednje slovarske problematike. Slovar angleških kolokacij (Kjellmer 1994), ki je nastal izključno na taki statistični podlagi,¹⁶² je denimo doživel precej nasprotujočih si ocen, zlasti v smislu svoje klasične slovarske uporabnosti. Za boljše rezultate v smislu luščenja kolokacijskih kandidatov iz korpusa glede na njihovo leksikalno zanimivost so bile postopoma razvite hevrstike, ki so omogočile filtriranje pogostih besednih sopojavitev, kjer se je pokazalo, da je za leksikalno zanimive kolokacije potrebna identifikacija leksikalno zanimivih skladenjskih vzorcev ter t. i. kolokacijskega okvira, znotraj katerega je mogoče predvideti, da se med leksikalna elementa, ki se tipično sopojavljata, lahko vrine določeno število drugih elementov, ter da je za ugotavljanje besedne sopovezovalnosti smiselno uporabiti statistične parametre, ki omogočajo merljivost medsebojne besedne soodvisnosti in ne zgolj absolutnih statističnih vrednosti.

18.2 Kolokacija kot pomensko-skladenjski pojav

Firth (1957) pojmuje kolokacije kot zveze, ki se oblikujejo v besedilu na podlagi besedne družljivosti, pomen v povezavi s kolokacijo pa je treba obravnavati kot abstrakcijo na sintagmatski ravni, ki sama na sebi ni neposredno povezana s konceptom besednega pomena: tako se denimo eden od pomenov besede *noč* kaže v njeni kolokabilnosti z besedo *temen*, eden od pomenov besede *temen* pa v njeni kolokabilnosti z besedo *noč*. Tak vidik je tudi v soglasju z že večkrat omenjenim Hanksovim razumevanjem pomenskega potenciala besed, ki se aktivira šele v besedilu, s tem ko pride beseda v stik z drugimi besedami.

Pomensko-skladenjski vidik (prim. Čermák 1982) pri kolokacijah izpostavlja pomensko in skladenjsko povezanost najmanj dveh besed, pri čemer celostnega

162 V slovar so denimo sprejete številne zveze, ki jih v angleščini določata npr. člena *the* in *a*: samo zveze s *the* predstavljajo skupaj 15.000 kolokacij (prim. Johansson 1998).

pomena kolokacije naj ne bi bilo mogoče izpeljati iz pomena besed, ki jo sestavljajo. Na ta način se kolokacije približujejo leksikalnim (tj. pomenskim) enotam, hkrati pa tak vidik sproža tudi potrebo po opredelitvi kolokacij glede na druge tipe večbesednih leksikalnih enot, npr. SZ in FE. Kolokacije se na podlagi tega vidika različno uvrščajo na lestvico med obe skrajnostni točki, ki jo na eni strani tvorijo (proste) besedne zveze tipa *kupiti knjigo*, na drugi pa pomensko popolnoma ne-transparentne FE in idiomi kot npr. *slamnata vdova*, *fant od fare* ipd.

Glede na to da je pomenska (ne)transparentnost neulovljiv in z vidika govorcev subjektiven pojav, povezan z različnim dojetjem stopnje pomenske asociativnosti posameznih besed v zvezi, z različnimi stopnjami govorčevega poznavanja pomena besed, njihovega izvora ipd., tudi v strokovni literaturi ni konsenza pri določanju definicijskih lastnosti kolokacij nasproti drugim večbesednim leksikalnim enotam. O merilih določanja kolokacij pri izdelavi LBS smo govorili že v razdelku o večbesednih leksikalnih enotah, nekaj o tem pa sledi še v nadaljevanju.

19 Pridobivanje kolokacij iz korpusa

Za pridobivanje kolokacij iz korpusa smo uporabili funkcijo besedne skice v orodju SkE. Postopek smo podrobneje predstavili v samostojnem poglavju, zato se na tem mestu omejujemo le na posamezne faze v procesu pridobivanja relevantnih kolokacij iz korpusa.

The image shows two identical screenshots of the SkE software interface, side-by-side. Each window displays a list of collocates for the word 'svetolola'. The interface includes a sidebar with a tree view of the corpus structure, a main area with a list of collocates, and a 'Slovenski slovar' (Slovene dictionary) button. The collocates are listed with their frequency and a 'Slovenski slovar' button. The windows are titled 'svetolola' and 'svetolola'.

Slika 45: Primerjava besedne skice v posameznih različicah slovnice besednih skic

Najpomembnejše za pridobivanje kolokacij iz korpusa je, kot rečeno, ustrezna predhodna priprava korpusa (lematiziranost in oblikoskladenjska označenost) in izdelava slovnice besednih skic, ki je ključna za optimalno pokritje vseh v jeziku relevantnih skladenjskih struktur, ki bi utegnile biti skladenjska podlaga za slovenščino tipičnim besednim sopojavitvam. Do takega nabora gramatičnih relacij smo pri izdelavi LBS prišli postopoma. Prva različica slovnice besednih skic, ki je bila izdelana ob vključitvi korpusa FidaPLUS v orodje SkE (Krek in Kilgarriř 2006), je bila prvič nadgrajena po končani prvi fazi izdelave LBS, ko je bilo mogoče na podlagi pomensko-skladenjskih analiz pri poskusno izdelanih geslih ugotoviti, katere skladenjske strukture so za leksikalnogramatični opis slovenščine relevantne in jih je potrebno vključiti v novo različico. Večino leksikografskega dela je potekalo na tej različici besednih skic. Druga nadgradnja je potekala v povezavi z izvedbo postopka avtomatizacije, ki je poleg prilagojenih in dodanih gramatičnih relacij upoštevala tudi večino možnosti, ki so bile na novo razvite v samem orodju, hkrati pa je bila ta različica aplicirana tudi na nov korpus Gigafida. Ob tej priložnosti smo SkE prenesli tudi na lokalni strežnik in ga lokalizirali na slovenščino. Primerjavo med obema korpusoma z vidika količine in vrste kolokacijskih podatkov, ki jih ponudi besedna skica, si je mogoče ogledati na Sliki 45.

20 Kolokacije v LBS

Razumevanje kolokacij v LBS se približuje pojmovanju, kot ga je v izhodišču zastavil Firth (1957), kasneje pa so v svojih leksikalnih in komunikacijskih študijah iz tega izhajali tudi drugi, zlasti npr. Halliday (1966) in Sinclair (1987a). Tako izhodišče med drugim določa, da kolokacij ne obravnavamo kot samostojnih leksikalnih enot, imajo pa pomembno vlogo pri pomenski členitvi besed in njihovem pomenskem opisu. Kot takih kolokacij v slovarski bazi nismo obravnavali na ravni iztočnice, pač pa se kot tipične besedne sopolovitve približujejo korpusnim zgledom. Kolokacije imajo pomembno vlogo tudi na skladenjskem nivoju: kot prikaz tipičnih besednih kombinacij so hkrati tudi odraz tipičnih skladenjskih razmerij v jeziku.

Tesna povezanost zgradbene in skladenjske ustaljenosti s semantičnimi omejitvami je navadno indikator stopnje leksikaliziranosti določene besedne kombinacije, ki nam pomaga ločevati kolokacije od večbesednih leksikalnih enot. Znotraj slovarja oz. slovarske baze je namreč identifikacija določene besedne kombinacije kot leksikalne enote ključna, zato je v zvezi s kolokacijami pomembna predvsem njihova opredelitev do vseh drugih večbesednih leksikalnih enot. Pri določanju parametrov za prepoznavanje kolokacij se pri oblikovanju LBS nismo izognili pomenskemu merilu, saj so kolokacije, kot smo dejali, za pomensko prepoznavanje in pomenski opis besed ključne, hkrati pa ne obstajajo nobena formalna merila (razen statističnih), npr. skladenjska, ki bi omogočila ločevanje kolokacij od prostih zvez. Te so namreč v veliki meri strukturirane enako kot proste ali stalne zveze ali celo frazeološke enote, kjer lahko sicer pričakujemo več strukturnih in pomenskih anomalij (prim. Čermák 1985: 176, 172).

Glede na to da so večbesedne leksikalne enote eno najtežje jasno ulovljivih elementov jezika, smo se v LBS odločili za organizacijo, kjer lahko posamezne enote zamenjujejo položaj na lestvici, ki zajema prehajanje besed in zvez med registriranimi (pod)pomeni, zlasti metaforičnimi, kolokacijami in FE. Tak način organizacije večbesednih enot morda zbujajo dvome o konsistentnosti uporabljenih meril pri določanju posameznih tipov, vendar pa je bil naš primarni namen podatke v bazi organizirati tako, da bo uporabnik lahko do njih prišel z različnih vidikov in z različnimi iskalnimi možnostmi, pri čemer menimo, da uporabnika načeloma ne zanimajo posamezne kategorije večbesednih enot, pač pa njihove pomenske lastnosti in ustrezna raba v besedilu. Iskalno izhodišče, prek katerega uporabnik dostopa do informacije, ki ga zanima, pa glede na položaj v slovarski bazi ponudi tudi različne informacije oz. stopnjo podrobnosti informacije, pri čemer si informacije med seboj ne nasprotujejo, ampak se dopolnjujejo. Na primer samostalnik *vreča* se

v slovenščini zelo pogosto uporablja kot metafora za količino denarja, zato bo uporabnik z vidika pomenskega opisa pri samostalniku *vreča* med drugim naletel na pomen 'vir finančnih sredstev', z vidika kolokacij pa na primere kot: [proračunska, državna, finančna] vreča, [luknja] v vreči, vreča z [denarjem], [polniti, napolniti] vrečo, [odvezati] vrečo, pri čemer bo lahko ugotovil, da sodijo te kolokacije le h konkretnemu pomenu samostalnika *vreča* in se po tem ločujejo od drugih kolokacij, ki vsebujejo isti samostalnik, npr. vreča [moke, žita, riža], [plastična, polivinilasta, papirnata, platnena] vreča itd. Z vidika FE – in lahko tudi neodvisno od predhodno navedenih podatkov – bo v zvezi s samostalnikom *vreča* naletel na FE: *vreča brez dna* in njeno variantno obliko *kaj ni vreča brez dna*, ki se uporablja v pomenu: projekt ali podjetje, ki zahteva nenehne finančne vloške, a ne daje zelenih rezultatov: [metanje] denarja v vrečo brez dna, [polniti, napolniti] vrečo brez dna.

20.1 Vrste kolokacij v LBS

V LBS smo registrirali dve vrsti kolokacij: regularne ali binarne kolokacije ter razširjene kolokacije. Izraze regularna ter razširjena kolokacija smo uvedli za namene izdelave LBS in ne izhajajo iz ustaljenega izrazoslovja, čeprav sta oba tipa zastopana v različnih kolokacijskih tipologijah, ki načeloma ločujejo med t. i. leksikalnimi in gramatičnimi kolokacijami (prim. Benson et al. 1986). Oba tipa nekoliko podrobneje predstavimo v nadaljevanju.

20.1.1 Regularne kolokacije

Osnovni tip kolokacije nam je predstavljala zveza dveh polnopomenskih besed,¹⁶³ ki lahko kot tretjo besedo vključuje tudi predlog. Kolokacije so določene s podstavno skladenjsko strukturo, ki omogoča na semantični ravni smiselno besedno zvezo, kar pomeni, da so kolokacije omejene tudi pomensko in da je medsebojno druženje besed odvisno tudi od tega, katere besedne vrste je beseda, ki jo obravnavamo v iztočnici. Skladenjske podlage kolokacij smo podrobno obravnavali pri skladenjskih strukturah, zato bomo na tem mestu navedli le nekaj najpogostejših strukturnih tipov regularnih kolokacij pri posameznih besednih vrstah.

(a) Samostalnik:

pbz0 SBZ0: [orožarska, korupcijska, podkupovalna] afera; [habsburška, avstro-ogrska, bizantinska] cesarica

SBZ0 sbz2: [šotor, bivališče] nomadov; [odkritje, otvoritev, odprtje] obeležja

SBZ1 gbz: zastava [plapola, vihra, se dviga], zbor [obravnavo, razpravlja, zaseda]

¹⁶³ Izjema so kolokacije s pomožnimi glagoli.

Med predložnimi po pogostosti izstopajo zlasti:

SBZ0 za sbz4: [opravičilo, krivec] za grozote; [raj, poslastica] za gurmane
SBZ0 v sbz5: etika v [znanosti, zdravstvu, športu]; [vadba, telovadba] v fitnesu
SBZ0 z/s sbz6: [obračunavanje, žuganje] s pestmi; ogrlica z [diamantom]

(b) Pri glagolih so med regularnimi kolokacijami, ki se nanašajo na strukturo glagolske zveze, najpogostejše dvo- in trodelne:

gbz in GBZ: [bliskati se, treskati, deževati] in grmeti; cviliti in [lajati, tuliti, praskati]
gbz Inf-GBZ: [začeti, pričeti, hoteti] dvoriti; [začeti se, pričeti se] odmikati**pbz1**
Vez-gbz Inf-GBZ: [treba, potrebno] je spoštovati; [dovoljeno, mogoče] je prisluškovati

Sicer pa so s kolokacijami najbolj tipično pokriti stavčni vzorci oz. njihovi posamezni deli. Spodaj navajamo 10 najpogostejših in pripadajoče kolokacije:

kdo – glagol – kaj

kdo sadi kaj	[kmetje, vinogradniki, ljubitelji] sadijo, saditi [rožice, drevo, cvetlice]
kdo sanja kaj	[večina, ženska, moški] sanja, sanjati [sanje, prizor]
kdo lika kaj	likati [srajce, perilo, uniformo, oblačilo, hlače, majice, obleko]

kdo – glagol

kdo demonstrira	[študenti, privrženci, protestniki, sindikati, delavci] demonstrirajo
kdo shodi	[otrok, malček] shodi

kaj – glagol

kaj blokira	[kolo, zavora] blokira
kaj besni	[orkan, nevihta, vihar, neurje, požar] besni
kaj dežuje	[čestitke, pohvale] dežujejo

kaj – glagol-se

kaj se sproži	[plaz, kamenje, podor] se sproži
kaj se mečka	[oblačila] se mečkajo
kaj se stabilizira	[stanje, vreme] se stabilizira
kaj se cepi	[lasje, nohti] se cepijo

kdo – glagol-se

kdo se češe	[ženska] se češe
kdo se cepi	[otroci, ljudje, bolniki] se cepijo
kdo se brije	[moški, mož] se brije

kdo – glagol – kaj – s čim

kdo natre kaj s čim	natreti [piščanca, meso, ribo], natreti s/z [soljo, česnom]
kdo nadgradi kaj s čim	[uporabnik, kupec] nadgradi, nadgraditi [programsko, strojno] opremo
kdo enači kaj s čim	[mnogi, nekateri, ljudje] enačijo, enačiti z/s [razizmom, terorizmom]

kdo – glagol – koga

kdo angažira koga	[podjetje, režiser] angažira, angažirati; [igralca, trenerja]
kdo časti koga	[ljudstvo] časti; častiti [junaka, prednike]

kdo/kaj – glagol – kaj

kdo/kaj blokira kaj	[policija, vojska] blokira, [tovornjak, avto] blokira; blokirati [cesto, promet, ulico]
kdo/kaj blaži kaj	[analgetik, obkladek] blaži; blažiti [krče, bolečino, vnetje]

kaj – glagol – kaj

kaj blokira kaj	[protitelesce, zdravilo] blokira; blokirati [receptorje, encime]
kaj odplakne kaj	[val, morje, voda, reka] odplakne; odplakniti [pesek]
kaj para kaj	[bliski, strele] parajo; parati [nebo, tišino, noč, zrak]

kdo – glagol – s čim

kdo bobna s čim	bobnati s [prsti, nogami]
kdo barva s čim	barvati z/s [barvicami, čopičem]
kdo upravlja s čim	[komunala, upravnik, zavod] upravlja; upravljati z/s [premoženjem, smučiščem]

(c) Pri **pridevniki** se kolokacije najpogostejše pojavljajo v naslednjih strukturah:

PBZ0 sbz0	svilena [ruta, kravata, obleka, srajca, posteljnina]
rbz PBZ	[zelo, izjemno, pretirano, bolešno] ambiciozen
sbz1 Vez-Gbz PBZ1	[lasje] so razmršeni; [strasti] so razvnete

(d) Pri **prislovi**h so najpogostejše kolokacije s strukturami:

RBZ gbz	voljno [leči, prižeti se, prepustiti se]; uvodoma [povedati, poudariti, dejati]
RBZ pbz0	brezglavo [zaljubljen]; strašno [prijazen, dolgočasen]
RBZ rbz in rbz RBZ	fizično [nemogoče]; [čimbolj, dokaj] enakomerno

20.1.2 Razširjene kolokacije

V procesu oblikovanja LBS se je pri analizi besednih skic pokazalo, da se nekatere besedne kombinacije zelo pogosto pojavljajo še s tretjim elementom, pri čemer je ta element bodisi tipično dopolnilo katere od sestavin kolokacije, lahko pa je nujno potreben za ustrezno razumevanje kolokacije. Na tej podlagi smo uvedli podkategorijo kolokacij, t. i. razširjene kolokacije, ki smo jih navajali v elementu <r_kolokacija> skupaj z navadnimi kolokacijami pod ustreznim pomenom in skladiščno strukturo. Zadnje je pomembno predvsem pri tistih razširjenih kolokacijah, kot bomo pokazali v nadaljevanju, kjer dodatni element spreminja skladiščno strukturo izhodiščni dvodelni kolokaciji, npr. gbz sbz4: opeči kruh – opeči [rezino] kruha. Dodajanje tretjega elementa v zgradbo dvodelne kolokacije pa sproži tudi vprašanje njihovega ločevanja od skladišjskih zvez, zlasti v primerih, kjer na novo nastala struktura izven razširjene kolokacije ni produktivna, npr. jeseni [prihodnje] leto, nadalje od SZ s tipičnim besedilnim okoljem, npr. kurilno olje – [lahko] kurilno olje – [ekstra] lahko kurilno olje, in od kolokacij, ki vsebujejo SZ v vlogi kolokatorja, npr. fotokopirati na [fotokopirnem stroju], o čemer bomo spregovorili v nadaljevanju.

20.1.2.1 Razširjene kolokacije z vidika zgradbe

Strukturno gledano, predstavljajo razširjene kolokacije razširitev obstoječe dvodelne ali predložne kolokacije z dodatnim elementom v vseh tistih položajih, kjer obstaja možnost strukturne razširitve z obstoječimi skladišjskimi strukturami (spodaj so ti deli navedeni v oklepaju):

gbz SBZ4 -> gbz SBZ4(= Prid+Sam)

[ohraniti, ohranjati, obdržati, pretrgati, nadaljevati] tradicijo -> ohranjati ipd.

[družinsko] tradicijo

[dodati, primešati, priliti, zmešati, vmešati, umešati] smetano -> dodati ipd.

[kislo, sladko] smetano

pbz0 SBZ0 -> pbz0(=Prid0+Prid0) sbz0

[laserski, brizgalni] tiskalnik -> [barvni, zmogljiv] laserski tiskalnik

pbz0 SBZ0 -> pbz0(=Prisl+Prid0) sbz0:

[narezan, sesekljan] ingver -> [drobno] sesekljan ingver

sbz0 sbz2 -> sbz2(Sam0+Sam2)

[tekma, tekmovanje] smučarjev -> tekmovanje smučarjev [tekačev, skakačev]

sbz0 sbz2 -> sbz0(=Prid0+Sam0) sbz2

[kos, del] poti -> [dobršen, lep, precejšen] kos poti

sbz0 sbz2 -> sbz2(=Prid0+Sam0)

mreža postaj -> mreža [seizmoloških, merilnih] postaj

sbz0 za sbz4 -> sbzo za sbz4(=Prid0+Sam0)

avdicija za [napovedovalca, voditelja -> avdicija za [radijskega, televizijskega]
napovedovalca

sbzo v sbz5 -> sbz0(=Prid0+Sam0) v sbz5

kvalifikacije v [badmintonu, nogometu] -> [dodatne, reprezentančne] kvalifikacije
v nogometu

gbz v sbz5 -> gbz v sbz5(=Prid0+Sam0)

živeti v [anonimnosti, osami] -> živeti v [popolni, precejšnji] anonimnosti

sbz1 gbz -> gbz(=Prisl+Glag)

cvet [diši, dehti] -> cvet [opojno, lepo] diši

20.1.2.2 Razširjene kolokacije s pomenskega vidika

Razširjene kolokacije smo razdelili v dve skupini glede na to, ali (a) razširitveni element določa katerega od elementov osnovne kolokacije, kar pomeni da ima taka kolokacija lahko tudi nerazširjeno različico in je kot taka tudi brez razširitvenega elementa smiselna, in (b) na kolokacije, kjer je razširitveni element za dosego obvestilnosti obvezen in tako skupaj z drugimi deli kolokacije tvori celoto.

(a) Razširjene regularne kolokacije

[najti, potrebovati] spremljevalca -> najti [stalnega] spremljevalca
omleta iz [beljakov, jajc] -> omleta iz [stepenih] jajc
[preseliti se] v palačo -> preseliti se v [predsedniško, kraljevo] palačo
plavati pod [vodo, gladino] -> plavati pod [vodno, morsko] gladino
[znova, spet] razočarati -> [vedno] znova razočarati

V posebno podskupino lahko štejemo razširjene kolokacije, pri katerih sicer prihaja do razširitve notranjega elementa in se na prvi pogled zdi, da lahko obstajajo tudi brez njega, vendar so v realnih besedilih skoraj vedno trodelne, kar kaže tudi na večjo trdnost kolokatorja z razširitvenim elementom, ki se v nekaterih primerih približuje SZ, prim. kaliti [mir] – kaliti [nočni] mir, kaliti [javni red in mir].

obvladati [jezik] -> obvladati [slovenski, nemški, angleški] jezik
zastopati [barve] -> zastopati [slovenske, olimpijske] barve, zastopati barve [Slovenije, kluba]
osvojiti [kolajno, medaljo] -> osvojiti [srebrno, zlato] kolajno
pita iz [krompirja, mesa, bučk] -> vendar samo: pita iz [krhkega, listnatega, vlečenega] testa

(b) Kolokacije z obveznim razširitvenim elementom

gostiti [svetovno] smetano
prevesti v [slovenski, angleški, tuj] jezik
barve v [različnih, vseh] odtenkih
veljati za [donosno, gospodarsko, neperspektivno] panogo
vroč poletni [dan, mesec, večer]
obetajo se [boljši] časi
obiskovati v [zadnjem, prostem] času
opustelo [mestno] središče, opustelo [mestno] jedro
briti se [vsak] dan, briti se [vsako] jutro
počivati [nekaj, mesec, leto, dva] dni

To merilo je bilo mogoče uporabiti tudi pri ločevanju razširjenih kolokacij od skladenjskih zvez.

V skupini razširjenih kolokacij, kjer vpliva razširitveni element na spremembo skladenjske strukture, prepoznamo predvsem kolokacije z dodanim količinskim elementom

opeči [kruh] -> opeči [rezino] kruha
[imeti] pogum - > imeti [dovolj, toliko] poguma

in kolokacije, kjer nastopa razširitveni element v vlogi osebka ali povedka in s tem razširi kolokacijo na dve skladenjski razmerji znotraj stavka, npr.

gbz pod sbz4 -> [temperatura] pade pod ledišče
pbz0 sbz0 -> [imeti] neizpodbitne dokaze
gbz sbz4 -> [redno] brati časopis

V takih primerih smo se raje odločali za dve kolokaciji, odvisno od besede, obravnavane v iztočnici, torej: [predložiti, imeti] dokaze in [trden, neizpodbiten] dokaz; brati [časopis ipd.] in [redno ipd.] brati.

20.1.2.3 Razširjene kolokacije v razmerju do stalnih zvez

V LBS kot razširjenih kolokacij nismo obravnavali primerov, kjer predstavlja razširitveni element del SZ, npr. zemljepisnega imena. Take primere smo obravnavali kot (a) regularne kolokacije, kjer nastopa SZ v vlogi kolokatorja, (b) kot SZ ali (c) kot kolokabilni del SZ:

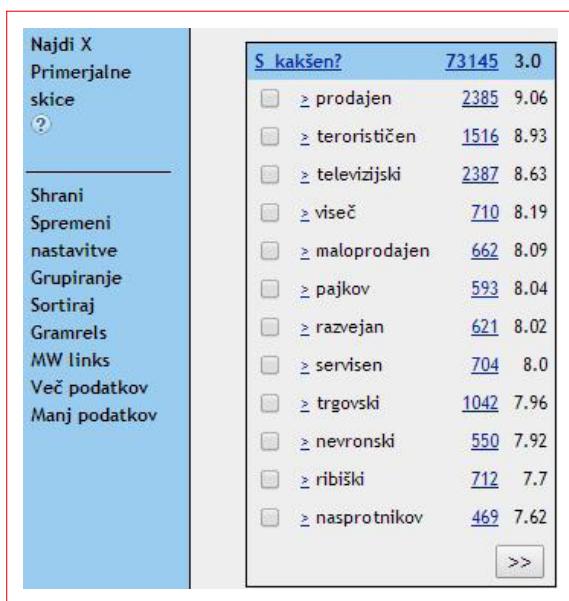
- (a) pripluti v [Perzijski zaliv, Mehiški zaliv, Piranski zaliv]
doseči [Severni tečaj, Južni tečaj]
pokopališče iz [prve svetovne vojne, druge svetovne vojne]
brzdati [spolno slo]
zaveden v [poslovni knjigi]
- (b) duševno prizadet otrok
gensko spremenjena živila
ekstra deviško olivno olje
- (c) [avtomatski, samodejni] telefonski odzivnik
[belo, rdeče] sortno vino
[bernski, švicarski] planšarski pes

20.1.2.4 Razširjene kolokacije v razmerju do skladenjskih zvez

Kontekstualni pristop v obravnavi leksike logično izpostavlja tudi daljše tipične besedne nize, ki jih je težko enoznačno zamejiti. Razširjene kolokacije tako trčijo vsaj še ob skladenjske zveze, tj. večbesedne nize, ki izkazujejo skladenjsko in relativno pomensko ustaljenost, hkrati pa navadno nimajo samostojne pomenske vrednosti. Vpis navedenih primerov v ustrezen segment slovarskega gesla v bazi je zato prepuščen presoji leksikografa o tem, kolikšen delež besedilnega okolja je zanimiv z vidika leksikalne informacije in na kakšen način enoto umestiti v zgradbo gesla, da bo informacija optimalna in objektivna. Navadno se je bilo mogoče zanašati na odločilnost razširitvenega elementa pri celostni pomensko-skladenjski vlogi zveze kot celote, kot smo prikazali zgoraj: notranja strukturna trdnost z omejeno predvidljivostjo sestavin nakazuje torej skladenjsko zvezo, razširitev, ki ni pomensko odločilna oz. tvori samostojne leksikalne enote, pa kot razširjeno kolokacijo oz. SZ.

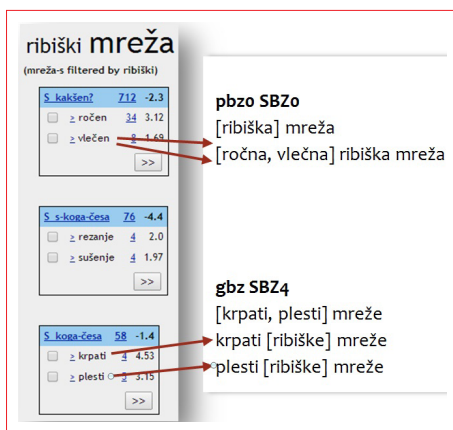
20.1.3 Pridobivanje razširjenih kolokacij iz korpusa

Razširjene kolokacije smo iz korpusa pridobivali s pomočjo orodja SkE, in sicer prek besednih skic s funkcijo povezave na večbesedne enote (ang. *MW links*). Postopek prikazuje Slika 46:



Slika 46: Povezava na večbesedne enote v orodju SkE

Povezava na večbesedne enote pri posameznem kolokatorju odpre novo besedno skico, ki izkazuje tipične kolokatorje za osnovno kolokacijo, npr. za [ribiška] mreža, kot prikazuje Slika 47. V slovarsko bazo smo vključili samo tiste povezave, ki so se pokazale za statistično relevantne.



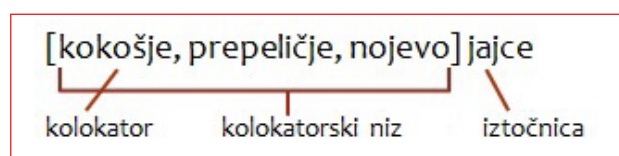
Slika 47: Besedna skica s povezavo na večbesedno enoto

Na ta način je bilo mogoče iz korpusa pridobiti razširjene kolokacije za vse registrirane slovnicihne relacije, pri čemer seveda vse niso leksikalno zanimive. Omenjena funkcija v orodju SkE omogoča tudi nadaljnje širjenje besedilnega okolja, tako da je mogoče beležiti tudi širše leksikalno relevantne zveze in prepoznavati kolokabilnost na ravni SZ ali FE:

[gospodarska, svetovna] recesija → [globalna, splošna, svetovna] gospodarska recesija
 kislo jabolko → [ugrizniti, zagristi] v kislo jabolko → zagristi v kislo jabolko [spora]

20.2 Zapisovanje kolokacij v LBS

Kolokacije smo zapisovali v nadelementu <kolokacije>, ki združuje v elementu <kolokacija> vse navadne in razširjene kolokacije. Znotraj elementa <kolokacija> smo zapisovali kolokatorje v kolokatorskem nizu <k></k>. Znotraj geselske zgradbe, kot jo določa shema DTD, smo kolokacije zapisovali vedno pod ustrezno strukturo in registrirani (pod)pomen besede v iztočnici. Zgradbo kolokacije prikazuje Slika 48.



Slika 48: Zgradba kolokacije

Zapis kolokacij smo v zgradbi LBS predvidevali povsod tam, kjer so se deli predhodno registriranih vzorcev, struktur in skladenjskih zvez kazali kot kolokabilni, in tam, kjer so večbesedne leksikalne enote, tj. SZ in FE, izkazovale lastno kolokabilnost. Kolokatorski niz sestavljajo tiste besede (kolokatorji), ki so glede na obravnavano besedo (iztočnico) spremenljivi, kar je zlasti pomembno pri zapisu predlogov, npr. marmelada iz [pomaranč, jagod, sliv], čas pred [volitvami, prazniki], odporen proti [vodi, vlagi], itd.

Pri zapisovanju kolokacij in razširjenih kolokacij, ki zapolnjujejo stavčne vzorce pri glagolih, smo potrjevali prisotnost osebkovega mesta z navajanjem iztočnice v nedoločniku, npr. pomešati z/s [pecilnim praškom, vodo, sladkorjem]; oklevati pri [izbiri, nakupu]. Glagol smo navajali v nedoločniku tudi pri glagolskih strukturah, npr. [začeti, znati, naučiti se] leteti. Pomožni glagol, kadar smo ga obravnavali kot del povedka, smo navajali v tretji osebi: marmelada je [okusna, aromatična]; [nevihta, nevarnost] je mimo.

Kadar se je kot kolokabilen kazal osebkov ali predmetni del stavčnega vzorca (zadnje pri neosebni glagoli oz. glagolskih pomenih), smo navedli glagolsko iztočnico v 3. osebi ednine oz. množine, če so bili kolokatorji tipično množinski: [glava, noga, trebuh, hrbet, želodec, zob] boli, [domotožje, dvom, občutek, strah, mora, ljubosumje] muči; svita se na [obzorju, vzhodu]; [ptice, štoklje] letijo. Pri brezosebni rabi, kot jo izkazujejo primeri: *gre za, priti do* ipd. smo se odločali za zapis skladenjskih zvez: *priti do intervencije*.

Pri zapisovanju kolokatorjev znotraj kolokatorskega niza smo sledili berljivosti celotne kolokacije. Kolokatorske nize smo zato ločevali oz. združevali glede na pripadnost kolokatorjev sorodnemu pomenskemu polju, sicer že znotraj določenega pomena, npr.

bife *samostalnik*

1 manjši gostinski lokal

[zakajen, zanikrn, zakoten] bife

[bližnji, skupščinski, interni, pravniški] bife

2 izbor pripravljene hrane

[zajtrkovalni, samopostrežni] bife

[solatni] bife

[dobro] založen bife

[hladni, topli] bife

Glagolske kolokatorje smo pri samostalniških in pridevniških iztočnicah združevali v kolokatorske nize tudi glede na povratnost glagolov, npr. [kljubovati, pripadati] času, [prilagoditi se, prilagajati se] času; rana [se celi, se zaraste].

Poleg pomenskega merila smo pri razvrstiti kolokatorjev upoštevali tudi slovnično ujemanje, zato smo kolokatorje pri pridevniških iztočnicah razvrščali tudi glede na ujemanje v spolu in številu, pri samostalnikih, kjer kolokacija vključuje predlog, pa tudi v sklonu, npr. hudomušna [domislica, opazka, pripomba], hudomušen [nasmeh, nasmeh]; veljati za [ikono], igra pod [košem, obročem].

Pridevniške kolokatorje smo v nizu razporedili v ustrezni obliki tudi glede na to, ali izražajo vrsto ali lastnost: globalni [trg, problem, proces] : [postajati, postati] globalen; gorati [sever, jug] : [pretežno, najbolj, bolj, zelo] gorati.

V kolokatorskem nizu smo zaporedoma navajali tudi leksikalne variante, npr. stisniti za [ramo, ramena], kamor sodi tudi zapis dovršnih in nedovršnih glagolov: [obujati, obuditi] spomin, [ohranjati, ohraniti] spomin, [obujati, obuditi, zbujati, prebuditi] spomine ipd.

Niso redki primeri, ko tudi kolokator sam nastopa kot SZ, npr. [berlinski zid, železna zavesa] pade; roditi se z/s [Downovim sindromom].

Izstopanje kolokatorja po pogostosti v različnih skladijskih relacijah, ki ga izkazujejo besedne skice, je navadno znak, da nastopa isti kolokator pri različnih pomenih obravnavane besede. Take večpomenske kolokatorje smo navajali v kolokatorskem nizu pri vsakem pomenu:

sesti *glagol*

1 prenesti težo telesa na zadnjico

sesti za [mizo]

sesti na [kolo, invalidski voziček]

2 začeti izvajati neko dejavnost

sesti za [pisalno, delovno, igralno] mizo

sesti v [šolske] klopi

2.1 odpeljati se

sesti na [kolo, vlak, avtobus]

2.2 opravljati funkcijo

sesti v [poslanske, parlamentarne] klopi

V. PONAŽARJALNI NIVO

Zgledi rabe so pomemben, če ne kar ključni del slovarske informacije, zato imajo tudi v zasnovi slovarske baze pomembno mesto. Njihova temeljna vloga je potrditi in ponazoriti vsa jezikovna dejstva, predhodno navedena v slovarski bazi, zato je treba leksikografom zagotoviti zadostno gradivo, na podlagi katerega bodo lahko pri izdelavi konkretnega slovarja izbrali najustreznejši slovarski zgled.

21 Zgledi rabe v slovarski bazi in slovarju

Izbira zgledov rabe se z vidika oblikovanja slovarske baze in izdelave slovarja razlikuje. V slovarski bazi težimo k izbiri večje količine korpusnih zgledov, ki so tipično celi stavki, vzeti iz korpusa, in predstavljajo kandidate za dobre slovarske zgleds. Ker pa spletna oblika slovarja tako kot tudi slovarska baza prostorsko ni omejena, je tudi količina ponazarjalnega gradiva lahko obsežna, hkrati pa jo lahko dopolnjuje povezava obravnavane besede, npr. v konkretni kolokaciji, na korpusne konkordance. Pri spletni obliki slovarja je zato smiselno razmisliti predvsem o možnosti leksikografskega preoblikovanja določenega števila dobrih korpusnih zgledov v slovarske, uporabniku pa omogočiti tudi dostop do leksikografsko nepreoblikovanih korpusnih zgledov. V spletni vizualizaciji LBS smo združili tri možnosti, najprej smo navedli zgleds, ki jih je leksikograf pridobil iz korpusa s pomočjo ustrezne aplikacije in jih preoblikoval glede na optimalnost informacije, ki jo je želel z zgledom pojasniti. Z izbiro možnosti *Več zgledov* je nato mogoče dostopati do zgledov, ki so na podlagi kolokacij avtomatsko izluščeni iz korpusa, leksikograf pa jih ni prilagajal slovarskim potrebam. Tretjo stopnjo predstavljajo korpusne konkordance, do katerih vodijo povezave pri kolokacijah, kot prikazuje Slika 49.

The screenshot displays the online Slovenian dictionary interface. At the top, it reads 'SPLETNI SLOVAR SLOVENSKEGA JEZIKA' and 'TESTNI SPLETNI PRIKAZ GESELEKSIKALNE BAZE ZA SLOVENŠČINO'. Below this is a navigation bar with letters A through Ž, and a search bar. The main content area shows the entry for 'izumrl' (pridevnik). Under the heading '1 ki ne obstaja več', there are two sub-entries: '1.1. kjer nihče več ne živi ali se nič ne dogaja' and '1.2. ki se več ne uporablja ali izvaja; zastarel'. A red arrow points from the '1.2.' sub-entry to a list of related terms on the right. This list includes 'izumrla ŽIVAL, RASTLINA ali SKUPINA LJUDI ne obstaja več', 'PRIDEVNIKO-SAMOSTALNIKO', and a list of related terms: 'izumrl [glavonožec, plazilec, mamut, dinosaver, sesalec, tiger, ptič]', 'izumrl [iglavc]', 'izumrla [civilizacija, rasa]', 'izumrla [pleme]', 'izumrl [predniki]', and 'izumrla [žival, vrsta, ptica]'. Below this list, there are several bullet points providing additional context, such as 'Ljudje so se najbolj potegovali za amonite, izumrle glavonožce, in druge prebivalce morja kakor tudi za lepo ohranjene fosile drevesnega listja.' and 'AMONIT - ripsi podoben izumrl glavonožec'. At the bottom of the list, there is a link 'Več zgledov...' circled in red.

Slika 49: Zgledi rabe v demo spletni postavitvi LBS s povezavo na korpus Gigafida

22 Vloga slovarskih zgledov

Kot pravita Atkinsova in Rundell (2008: 453), bi moral biti jezik opisan na podlagi objektivnih dokazov o tem, kako ga dejansko uporabljajo njegovi govorci. V zgodovinskih slovarjih imajo zgledi predvsem to vlogo, da dokazujejo obstoj neke besede v določenem času: zgled, ki izhaja iz avtentičnega besedila, namreč objektivno dokazuje, da je (bilo) besedo dejansko mogoče najti v jeziku (Simpson 2003: 268). V sodobnih enojezičnih slovarjih pa so zgledi primarno namenjeni olajšanju razumevanja pomenov in so zato koristen dodatek k razlagi (Frankenberg-Garcia 2014).¹⁶⁴ Dobro izbran zgled pogosto pomaga razjasniti pomenske distinkcije pri večpomenskih besedah, bolj problematično pa je, če posameznih pomenov oz. njihovih razlag ni mogoče razumeti, ne da bi prebrali tudi zglede. Idealno je, če sta razlaga in zgled samozadostna; razlaga, ki je brez zgleда ni mogoče razumeti, je daleč od optimalne rešitve. Vendar pa je slovarska razlaga po svoji naravi abstrakten konstrukt in številni so primeri, kjer postane celostni pomen zapletenega koncepta razumljiv šele, ko ga dopolnjuje dober zgled.

Poleg dopolnjevanja pomenskih lastnosti, sicer zajetih v slovarsko razlago, je temeljna vloga zgledov, da odlikavajo sobesedilne lastnosti posameznih pomenov, kot so tipična skladijska struktura ali stavčni vzorec, kolokacije in tipičen tematski, stilistični in pragmatični kontekst. Dobro organizirana slovarska baza mora zato registrirati celotni spekter sobesedilnih elementov, tako leksikalnih kot skladijskih, znotraj katerih se beseda ali njen pomen tipično pojavlja. To je posebej pomembno v slovarjih, namenjenih učenju jezika, kjer imajo lahko še tako preproste besede tipično besedilno okolje, ki določa naravno rabo jezika. V slovenščini se npr. glagol *gledati* v povezavi s tožilnikom pogosto pojavlja v velelniku v vlogi opozorila, npr. ob besedah kot: *glej preglednico, opombo, priložo*, glagol *videti* pa predvsem ob besedah kot: *videti priložnost, rešitev, možnost* in pogosto zanikano: *ne videti razloga, smisla, konca, potrebe, zadržka*. V večini slovarjev za tujce in tudi v nekaterih slovarjih za materno govorce (Atkins in Rundell 2008: 454) je običajno, da zgledi podpirajo vsako trditev o skladijskem obnašanju besede: če zgled ilustrira tipično rabo besede, potem navadno prikazuje besedo v kombinaciji s katero od njenih pogostih sopojavitev. Tako je npr. pričakovano, da bodo zgledi za samostalnik *nasvet* prikazovali, da ljudje *gredo oz. se obrnemo, zatečemo po nasvet*, da *delamo oz. se ravnamo po nasvetu koga*, ki je navadno strokovnjak za določeno področje (*odvetnik, zdravnik, farmacevt, veterinar*) ali nam je blizu (*prijatelj, prijateljica*).

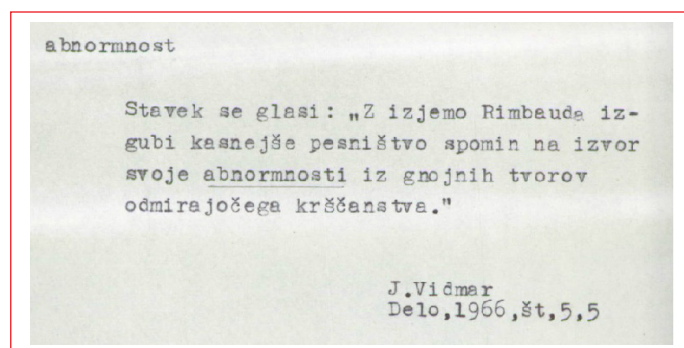
¹⁶⁴ Avtorica ugotavlja, da predstavljajo zgledi pomemben vir informacij pri tvorjenju besedila. S pomočjo zgledov pisci izbirajo ustrezne kolokacije, uporabijo ustrezno skladijsko zgradbo in se izognejo prepoznamim napakam.

Kadar je za določen pomen značilen poseben stil, register ali regionalna distribucija, je naloga dobrega zgleda, da to pokaže v naravnem okolju, npr. za glagola *rotiti* in *kalkulirati*:

- *Rotila sem jih, naj ga pustijo z menoj, ker je bolan, vendar so bile vse prošnje zaman.*
- *»Prosim,« me je rotil »daj mi priložnost, da spoznam svojega sina ...«*
- *Bili smo v težki skupini, a nismo želeli kalkulirati. To ne bi bilo športno.*
- *Tako kot se politične stranke zavedajo, da se jim bo najbrž obrestovalo, če bo na njihovi listi nastopil kandidat, ki je všeč volivcem, enako tudi potencialni kandidati kalkulirajo, na listi katere stranke se jim »še najbolj splača« kandidirati.*

22.1 Avtentični ali ustvarjeni slovarski zgledi

Zgledi, kot jih pozna npr. SSKJ, so zajeti z listkovnega gradiva¹⁶⁵ (Slika 50) in navadno opremljeni s podatkom o viru:



Slika 50: Ekscerpcijski listek listkovne kartoteke ISJFR ZRC SAZU

V sodobnih na korpusu temelječih slovarjih navajanje vira ob zgledih ni običajno,¹⁶⁶ saj to informacijo omogoča povezava s korpusom. Zgledi se lahko navajajo v različnih oblikah: od kratkih fragmentov (npr. iztržkov v SSKJ, gl. Slika 51) do celih stavkov in iz različnih virov: od avtentičnih besedil, leksikografovih izdelkov do kombinacije obojega. V angleškem okolju je za pedagoške slovarje dolgo časa veljala izključna praksa navajanja ustvarjenih (ang. *invented*) zgledov, ker so leksikografom omogočali ponazoritev različnih relevantnih leksikalnih in slovničnih lastnosti v eni sami, pazljivo in natančno oblikovani ne preobsežni zvezi ali stavku (prim. Rundell 1998: 316–317).

¹⁶⁵ Vir: digitalizirana listkovna kartoteta Inštituta za slovenski jezik Frana Ramovša Slovensko besedišče. Dostopno na: <http://www.arzenal.si/sobe/arhiv/slovensko-besedisce>.

¹⁶⁶ Izjema je denimo Wielki słownik języka polskiego, ki ob citatih (zavihek Cytaty), avtomatsko pridobljenih iz korpusov, vedno navaja tudi vir.

fantazija -e ž (i)

1. *prosto kombiniranje misli in predstav, domišljija*: fantazija ga je zanesla predaleč; v fantaziji si slikati lepoto tujih dežel; neusahljiva, ustvarjalna, živahna fantazija // *sposobnost za tako kombiniranje*: otrok ima fantazijo; dramiti, krotiti fantazijo; dela brez fantazije; bujna fantazija; vse to je le plod njegove fantazije
2. *predstava, ki ni osnovana na resničnosti*: kar je rekel, ni resnica, ampak fantazija; meglene teorije in fantazije; ekspr. živi v praznih fantazijah / bolezenske, vročične fantazije
3. *muz. instrumentalna skladba iz glasbenih motivov drugih avtorjev*: igrati fantazijo; fantazija na narodno pesem

Slika 51: Ponazarjalno gradivo za samostalnik fantazija v SSKJ

V angleškem okolju, z vplivom na celotno slovarsko produkcijo korpusnega tipa, je bila ta praksa presežena s prvim v celoti na korpusu temelječim Cobuildovim angleškim slovarjem, ki je sprožil tudi debato o tem, ali morajo slovarske zglede izdelati leksikografi ali naj bodo vzeti neposredno iz avtentičnih besedil. Argumenti za avtentične zglede temeljijo na stališču, da je nesmiselno preučevati realni jezik z namenom, da bi ugotovili dejstva o njegovem delovanju, nato pa dokaze zavreči in uporabiti umetne zglede. Vendar pa, če navajanje izključno avtentičnih zgledov velja predvsem za zgodovinske slovarje, imajo slovarji za tujce, šolski slovarji kot tudi slovarji za matere govorce drugačne naloge. Zanje je, kot trdita Atkins in Rundell (2008: 456), jasnost in razumljivost slovarskega zgleada vsaj tako pomembna kot prikazovanje jezika v naravnem okolju. Vendar pa je tudi avtentičen zgled izvzet iz svojega naravnega okolja, zato prav tako potrebuje skrbno izbiro in upoštevanje več načel, ki optimizirajo izbor. Debata o ustvarjenih nasproti avtentičnim zgledom je zato, kot trdita Atkins in Rundell (2008: 457), že v izhodišču napačna, saj predvideva izključno izbiro med dvema skrajnima možnostima, od katerih prva predvideva izmišljanje zgledov neodvisno od realnega življenja, druga pa kopiranje zgledov iz besedil, ne da bi bila spremenjena ena sama beseda ali podrobnost.

Realnost leksikografskega dela je v resnici drugačna in pogosto zelo pragmatična. Tudi če zagovarjamo avtentičnost zgledov, je presenetljivo redko – in to kljub obstoju zelo obsežnih korpusov – mogoče najti korpusni zgled, ki izpolnjuje vse pogoje dobrega slovarskega zgleada. Navadno leksikograf med avtentičnimi korpusnimi primeri znotraj konkordančnega niza izbira med takimi, ki vsebujejo obravnavano besedo in približno štiri do šest besed na vsaki strani, ki predstavljajo njeno sobesedilo, ki ga je s pomočjo korpusnih orodij mogoče določiti glede na tipičnost. Ko želimo k izbranemu delu, ki se zdi z vidika meril za dober slovarski

zgled ustrezen, poiskati še ustrezno širšo okolico stavka, se pogosto srečamo bodisi z nizom odvisnih in/ali prirednih stavkov, ki vsebujejo lastna imena, anafore, nenavadne besedne oblike, fraze ipd., ki v celoti pokvarijo obvestilnost ali sploh primernost za opravljanje vloge slovarskega zgleda.

Za leksikografa je zato bistveno zavedanje, da predstavljajo korpusi v prvi vrsti vir dokazov in ne vir zgledov (Atkins in Rundell 2008: 458). Tako neizpodbitno velja, da mora vse, kar v slovarju (in slovarski bazi) povemo o tem, kaj besede pomenijo in kako se v jeziku obnašajo, objektivno izhajati iz tega, kar nam jezik kaže – tudi kadar so naša intuitivna prepričanja ali celo nazor drugačni. Ob tem pa je optimalna kompatibilnost med avtentičnim in ustvarjenim zgledom dosežena takrat, ko je na podlagi korpusa izdelan pomenski opis dopolnjen z zgledi, ki odražajo tipične vzorce besedne rabe, a v obliki, ki je uporabniku dostopna in lahko razumljiva.

23 Merila za določanje dobrih slovarskih zgledov

Lastnosti zgledov se spreminjajo glede na tip slovarja ter glede na potrebe in pričakovanja uporabnikov. Načela, ki jih navajamo v nadaljevanju, pa, kot pravita Atkins in Rundell (2008: 458), veljajo za večino slovarjev. Po njunih besedah naj bi bili zgledi v slovarjih predvsem naravni in tipični, informativni in razumljivi.

Tipičnost je razmeroma lahko prepoznavna, saj obsežni korpusi in orodja omogočajo prepoznavnost sobesedila, stavčnih vzorcev, kolokacij in SZ, v katerih se obravnavana beseda tipično pojavlja.

Naravnost je bolj intuitiven in težje formalno in objektivno določljiv pojem, vendar pa v kombinaciji s slovničnimi in skladenjskimi vidiki, kot je prevladujoča izbira časa, števila, glagolskega načina ali položaja v stavku (npr. atributivna ali povedkovodoločilna raba), pripomore k prepoznavanju tega, kaj je za določeno besedo ali njen pomen naravno. Čeprav uporabniki radi uporabljajo jezik na idiosinkratičen način, je prikazovanje ustvarjalne rabe na ravni zgleda manj primerno zlasti v slovarjih za tujce in sploh šolskih slovarjih.

Naravnost je pogosto povezana s količino konteksta, ki je za dano besedo ali pomen značilen, saj se pogosto zgodi, da je določen kontekst za pojasnitev besede zelo primeren, vendar pa je celotni stavek, v katerem se beseda nahaja, preobsežen in zamegli bistvo tega, kar želimo z njim pokazati. Pri prenosu korpusnih zgledov iz slovarske baze v slovar je zato smiselno predolge zglede, ki vsebujejo odvisnike, krajšati, če na ta način ohranimo naravnost konteksta v obvladljivi količini besedila. Vendar pa tak postopek skriva drugo težavo, namreč obstoj anaforičnih elementov, bodisi znotraj bodisi zunaj stavčne zgradbe. Krajšanje lahko tako povzroči nekonsistentnost zgleda ali opustitev kake informacije, ki je za ustrezno pojasnitev obravnavanega pomena ključna.

Če želi biti zgled informativen, mora dopolnjevati pomensko razlago in jo pomagati bolje razumeti. Če razlage ni mogoče razumeti, ne da bi prebrali tudi zgled, je potrebno razmisliti o ustreznosti pomenske razlage. Prav tako zgled (gl. spodaj) ni informativen, če ne potrjuje tega, kar je izpostavljeno v razlagi, ali če temu celo nasprotuje. Razlaga pridevnika *dvoumen* spodaj sugerira pomen 'nesporazum', zato bi pričakovali, da se nanaša predvsem na izjave, zapise ipd., česar pa izbrani zgledi ne potrjujejo:

dvoumen pridevnik

pomensko nejasen

dvoumen je tak, katerega pomen lahko razumemo na najmanj dva različna načina

- *Poteza je sicer precej **dvoumna** in veliko je zraven hipokrizije, je pa hkrati najbolj učinkovit način, da kak problem postane planetarna obsesija.*
- *Evropska unija, za katero bi pričakovali, da bo zaveznik Slovenije, pa je tako **dvoumna** in nejasna, da je včasih prav mučno.*
- *Nihče ne ve, zakaj, vedno pravijo, da je njen smehljaj tako **dvoumen**, ali se res smehlja, ali je žalostna, ali ve več, kot si mislimo, da ve?*

Prvi zgled, ki se nanaša na potezo, izraža prej dvom v njeno ustreznost kot pa pomensko nejasnost, poleg tega ni razvidno, na kaj se nanaša drugi del zgleда. Podobno velja za drugi zgled, kjer je dvoumnost pripisana *Evropski uniji*, kar za rabo konkretnega pridevnika ni tipično. Prav tako je pridevnik *dvoumen* ob samostalniku *smehljaj* uporabljen netipično, saj v konkretnem primeru ne gre za zbujanje pomenske nejasnosti, pač pa za ustvarjanje občutka skrivnostnosti in nedorečenosti, kar potrjuje tudi drugi del zgleда: *ali ve več, kot si mislimo, da ve?*

V prikazanem primeru je torej bodisi napačna razlaga ali pa izbira zgledov, ki naj bi razlago potrjevali, zato predlagamo rešitev s tremi različnimi pomenskimi indikatorji, ki jih po našem mnenju dokumentirajo tri različne skupine sicer avtentičnih zgledov, ki pa smo jih leksikografsko priredili:

dvoumen pridevnik

1 nejasen; nerazumljiv

- *Pravilen odgovor ne sme biti **dvoumen** in strokovno oporečen.*
- *Težave jim povzroča tudi zapletena in **dvoumna** zakonodaja, ki je podjetniki ne morejo sami obvladati v celoti.*

2 pomenljiv; zagoneten

ki omogoča več interpretacij

- *Švejska je naredil nesmrtnega njegov **dvoumni** humor, ki je hkrati izražal tragične in komične občutke.*
- *Kadar sem prišla v njegovo bližino, mi je vedno namenjal **dvoumne** pripombe.*
- *Javnost je ugibala, ali je mogoče ukradene uranove palice uporabiti tudi v vojaške namene, odgovori učejakov pa so bili **dvoumni***
- *Nihče ne ve, zakaj, vedno pravijo, da je njen smehljaj tako **dvoumen**, ali se res smehlja, ali je žalostna, ali ve več, kot si mislimo, da ve?*

3 sporen; vprašljiv

ki vzbuja dvom v primernost, ustreznost

- *Potencialne koristi reprodukcijskega kloniranja so moralno **dvoumne***

Kot smo že poudarili, je zgled lahko povsem avtentičen, vendar ne more opraviiti svoje naloge, če je za uporabnika prezapleten in nerazumljiv. Razumljiv zgled je zato razmeroma kratek in ne vključuje zapletenih besed in nerazumljivih ali celo slovnično napačnih struktur, kot npr. prvi zgled v prikazanem primeru za glagol *garantirati*:

garantirati *glagol*

2 zagotavljati kakovost

če PROIZVAJALEC s POTRDILOM garantira določene LASTNOSTI izdelka ali STORITVE, jamči, da bo v primeru okvare, če bi do nje prišlo v določenem ČASOVNEM OBDOBJU, zagotovil brezplačno popravilo ali izdelek zamenjal

- ✱ *Mislim, da so uradne ure za stranke zaradi urejanja zadev, ki jih ljudje morajo opraviti po nalogu upravnega organa oziroma zaradi predpisov, ki **garantirajo** to tudi urediti.*
- ✓ *Prodajalec se je s pisnim potrdilom obvezal, da za novo kritno **garantira** deset let.*

K zgoraj izpostavljenim načelom, ki jih omenjata Atkins in Rundell, je treba dodati še politično in družbeno korektnost zgledov, ki je za slovar, ki predstavlja določeno družbeno avtoriteto (Gorjanc 2013/14: 12), zelo pomembna. Če izhajamo s stališča, da je vsak jezikovni opis tako ali drugače ideološki (Gorjanc 2012: 38), so pomenske razlage (Moon 2014) in zgledi rabe (Béjoint 2000) gotovo tisti elementi slovarja, kjer še posebej prihaja do izraza ideologija, čeprav, kot izpostavlja Gorjanc (prav tam: 43), na izrazito drugačna načina: ker so pomenske razlage bolj strogo podvržene pravilom razlagalnega jezika, je v njih prostora za ideologijo manj, medtem ko je izbor zgledov svobodnejši in, vsaj kakor kaže analiza v SSKJ, velikokrat odraz družbene resničnosti, kot jo vidijo leksikografi (Gorjanc 2012: 43).

Glede na to, da kulturološko, družbeno ter politično specifičnih vidikov leksike (ang. *non-denotative aspects of word meaning*, Moon 2014: 86), kamor sodi zlasti besedišče, ki odraža politične koncepte, odnos do etnične pripadnosti, starosti in staranja, seksualnosti ipd., ni mogoče formalno opredeliti za namene avtomatske ekstrakcije, se poraja vrsta vprašanj – tako z metodološkega vidika kot tudi z vidika korpusnih dokazov – glede ustrezne pomenske opredelitve in izbire najustrežnejših zgledov rabe že na ravni slovarske baze, še toliko bolj pa pri prenosu v slovar. Znotraj tega na primer, ali naj se zgledi in razlage osredotočajo na odnos, kot ga do leksikalnih enot, kot so npr. *priseljenec*, *materializem*, *kapitalizem*, *komunizem*, *kolonializem*, *feminizem* ipd., izkazujejo predvsem govorci določenega jezika ali naj prikazujejo bolj univerzalne koncepte, če vemo, da se znotraj posameznih družb lahko vzpostavljajo zelo različni odnosi do teh pojmov (prim. Moon 2014: 88). V povezavi s korpusnimi dokazi je v zvezi s to problematiko pomembno zlasti vprašanje navajanja kolokacij, tipičnega besedilnega okolja in tematskega konteksta v smislu prikazovanja bodisi čim bolj realnega stanja, zavestnega nevtraliziranja ali skrajnega radikaliziranja pri prikazovanju konotativnih in pragmatičnih elementov leksike.

Če vzamemo za primer samostalnik *cigan*, za katerega lahko predvidevamo etnično in kulturološko pogojeno konotacijo, in ga opišemo na podlagi tipičnega besedilnega okolja, kot ga določa besedna skica v orodju SkE, ter posamezne kolokacije

potrdimo z zgledi rabe, lahko dobimo v nekaterih segmentih precej radikalno predstavo o tem, kaj slovenski jezikovni skupnosti ta beseda predstavlja:

cigan *samostalni*

cigani [kradejo]

- *V vrtcu so me naučili, da cigani kradejo.*
- *In tam me starši niso pustili na cesto, ker so cigani kradli otroke, jih pobabili in jih poslali beračit.*

Slika bi bila glede na besedila v korpusu Gigafida lahko tudi drugačna, npr. z izbiro spodnjih, prav tako avtentičnih korpusnih zgledov:

- *Poznavalci literature se bodo spomnili, da Cigani pogosto nastopajo v delih slovenske književnosti kot neprijetni potepuhi, razbojniki, tatovi konj in celo otrok, ki jih prodajajo v turško sužnost.*
- *Spomnimo se izreka, da se cigan smeje, kadar dežuje, kajti za dežjem vedno pride sonce.*
- *Beseda Cigan verjetno izvira iz grške besede athinganos, ki pomeni nedotakljivo.*

Vendar pa bi v zgornjih zgledih težko govorili o tipičnosti in potrjevanju predhodno navedenih kolokacij, ki so vedno statistično dejstvo. Prav tako je navajanje zgledov, ki vključujejo tipične kolokatorje, eden od pogojev za njihov izbor v LBS in hkrati izhodišče za avtomatsko luščenje zgledov iz korpusa.

Kako torej kljub statističnim dejstvom, ki pogosto kažejo radikalno sliko pomenske vrednosti družbeno občutljive leksike, najti za slovarsko bazo in slovar ustrezno rešitev, se kar v največji meri izogniti ideološkim preferencam ali celo sovražnemu govoru?

Eden izmed pomembnih korakov v to smer je diferenciran izbor besedil, ki jih vključuje (referenčni) korpus. Izbira zgledov (in še prej pomenskih opisov) na podlagi realnih besedil je namreč tesno povezana z izbiro korpusnih besedil, kjer se je med drugim treba zavedati tudi vprašanja izrazite ideološke opredeljenosti avtorja ter njegovega vpliva na besedilno recepcijo. Na pomembnost vloge enega samega avtorja oz. prevajalca v manjši jezikovni skupnosti opozarja npr. Gorjanc (2012: 40), ko izpostavlja vlogo Antona Sovreta, ki je med obema vojnama veljal za največjega poznavalca antike pri nas, posledično pa so bila kot referenčna besedila v SSKJ sprejeta tudi njegova dela, v katerih eksplicitno izraža odklonilen odnos do homoerotike, ki se je nato tako prek pomenskih razlag kot zgledov rabe prenesel tudi v SSKJ.

Drugo možnost ponazarja spodnji primer analize besedilnega okolja za samostalni *feministka* v korpusu Gigafida. Analiza besedne skice, ki je povzetek tipičnega

besedilnega okolja, prikazuje z navedbo petih najpogostejših kolokatorjev¹⁶⁷ v izbranih skladenjskih strukturah štiri možne klasifikacije¹⁶⁸ neposrednega besedilnega okolja: pozitivno, negativno, opazno in nevtrarno (Preglednica 28). Posebej je potrebno opozoriti na kategorijo opazno, saj vključuje besede, ki imajo potencialno velik družbeno-politični naboj, njihovo pozitivno ali negativno dožemanje pa je odvisno od tega, katero nazorsko izhodišče je značilno za določeno jezikovno skupnost (npr. ameriški vs. vzhodnoevropski odnos do komunizma, socializma, imperializma ipd.). V primeru pridevnikov je stopnja intenzivnosti, ki jo izražajo, vezana na pozitivno ali negativno konotacijo glede na to, kako se ta odraža v določenem sociolingvističnem okolju.

Preglednica 28: Besedilno okolje samostalnika feministka z vidika pozitivne/negativne konotacije

feministka – Gigafida 2.450 pojavitev

Struktura	Nevtrarno	Opazno	Pozitivno	Negativno
pridevnik0 + samostalnik0		socialistična, marksistična, radikalna, goreča, zagrizena		militantna, zadrtá, jezikava, frustrirana, lezbična
samostalnik1 + glagol	trditi, strinjati se			kritizirati, očitati, sovražiti
samostalnik0 in samostalnik0	pisateljica, filozofinja	socialistka, levičarka		
glagol + z + samostalnik6			sodelovati	zmerjati

Na podlagi stanja, kot je razviden iz preglednice, ima leksikograf – vsaj na ravni zgledov, če že pri kolokacijah ostaja zvest tipičnosti – možnost izbirati med vsemi kategorijami enakovredno oz. se odločati prednostno za nevtralne besedilne kontekste, kot so spodnji:

- *Hčerka je zagrizena feministka in se udeležuje protestnih protiglobalističnih zborovanj.*
- *Nekatere sodobne feministke trdijo, da je celotna tradicija razlikovanja na »moške« in »ženske« morale kulturni mit, ki nima empirične podlage.*
- *Pobudo za praznovanje 8. marca je dala nemška feministka in socialistka Clara Zetkin.*
- *Ljudje na splošno pravijo, da feministke sovražijo moške.*

167 V tabeli smo navajali le absolutno pogoste kolokatorje, ob upoštevanju statistične kolokabilne trdnosti bi bila slika drugačna, izpostavila pa bi tudi specifične kolokatorje, vezane na določene filmske naslove ipd., ki v danem primeru ne kažejo objektivne slike.

168 Omenjene štiri kategorije se navadno pojavljajo tudi v raziskavah, ki se ukvarjajo z analizo razpoloženja (ang. *sentiment analysis*), prim. Šniderlová et al. 2014.

24 Avtomatski postopki luščenja korpusnih zgledov

Če vzamemo za izhodišče merila dobrih zgledov, ki sta jih postavila Atkins in Rundell (2008: 458), je iskanje dobrih zgledov zamudna naloga, saj iskanje najboljšega primera v realni situaciji, kot smo jo opisali zgoraj, zahteva pregledovanje velikega števila konkordanc in upoštevanje različnih parametrov, kot so: polna stavčna zgradba, dolžina stavka, odsotnost redkih, manj znanih ali kako drugače težko razumljivih besed oz. besednih oblik in anaforičnih elementov. Pri tem seveda ne moremo odmislliti osnovne leksikografove naloge, ki zahteva pri izbiri zgledov njegovo osredotočenost na to, da bodo zgledi ustrezno pojasnili pomen, tipično besedilno okolje, žanr, upovedovalno izbiro itd. Ker korpusi postajajo vedno večji, morajo tudi leksikografi pregledati vedno več gradiva. Eno od možnih rešitev, ki tako delo olajšajo, ponuja funkcija GDEX (Kilgarrieff et al. 2008) v orodju SkE, ki smo jo že opisali, zato se bomo na tem mestu osredotočili na izvedbo postopka in določitev meril.

Dobre korpusne zglede je glede na merila, kot jih izpostavljata Atkins in Rundell (2008: 458) težko avtomatsko prepoznati in meriti, zato orodje GDEX za izbiro kandidatov za dobre korpusne zglede uporablja formalno prepoznavne lastnosti, kot so dolžina besed v zgledu, dolžina stavka, prisotnost oz. odsotnost besed s t. i. črne liste, nejezikovnih elementov (številke, URL-ji, spletni in elektronski naslovi ipd.).

Pri oblikovanju LBS smo težili k vključitvi velikega števila zgledov, in sicer najmanj dva za vsak element geselske zgradbe, tj. za potrditev kolokacije v skladijski strukturi ter za potrditev stavčnega vzorca, skladijske zveze, SZ in FE. Poleg tega so leksikografi glede na razlago, stilne in druge lastnosti, povezane s tipičnostjo, naravnostjo in informativnostjo, izbirali zglede, ki so te lastnosti prikazovali v realnem besedilu. To je pomenilo, da je bilo npr. pri pomenu, ki je vseboval najmanj štiri skladijske strukture, znotraj tega pa najmanj tri kolokatorske nize, pri čemer sta bila v kolokatorskem nizu vsaj dva kolokatorja, izbrati skupaj kar 48 zgledov. Pri večpomenskih geslih, ki navadno vsebujejo tudi veliko število skladijskih struktur in daljše kolokatorske nize, glagolske iztočnice pa še stavčne vzorce, poleg tega pa še skladijske zveze, SZ in FE, bi število zgledov, ki naj bi jih leksikografi vključili v slovarsko bazo, krepko preseglo številko 300, kar pa je izredno zamudno, neekonomično in v praksi domala neizvedljivo.

V prvi fazi izdelave LBS smo zato z zgledi potrjevali sicer vse skladijske strukture, skladijske zveze in FE, znotraj kolokatorskih nizov in stavčnih vzorcev pa samo tiste kolokatorje (oz. vzorce), ki so se kazali kot najbolj tipični. Pri

semiavtomatskem izboru zgledov, tj. s pomočjo prvotne za slovenščino prilagojene funkcije GDEX v orodju SkE in z uporabo kliksikografije, je tako nastal korpus ročno izbranih zgledov, ki je bil nato uporabljen kot učna množica, na kateri so bili izdelani parametri za luščenje dobrih korpusnih zgledov v postopku avtomatizacije. Pri določanju hevristik za slovensko različico funkcije GDEX je bil cilj upoštevati tudi jezikovne specifičnosti slovenščine. Na podlagi evalvacije izhodiščne angleške različice GDEX-a in lastnosti ročno izbranih zgledov so bili za slovensko različico določeni naslednji parametri, s pomočjo katerih so se točkovali¹⁶⁹ zgledi na lestvici od najbolj do najmanj primerne (Kosem et al. 2011: 154):

- točkovanje zgleda glede na dolžino stavka od 8 do 30 besed;
- točkovanje zgleda glede na količino besed z manj kot 140 pojavitvami v korpusu;
- točkovanje pozicije ključne besede v stavku: na začetku stavka oz. med prvimi 20 pojavnicami;
- odbitne točke za zglede, ki vključujejo posebne znake (*, /, & ipd.)
- odbitne točke za zglede, ki vključujejo url-je in elektronske poštne naslove;
- odbitne točke za zglede, ki vključujejo besede s samimi velikimi črkami;
- odbitne točke za zglede, ki vključujejo lastna imena;
- odbitne točke za zglede, ki vključujejo zaimke;
- zgled mora imeti celostavno obliko;
- zgled ne sme vsebovati besed, ki imajo v korpusu manj kot 3 pojavitve.

Glede na leksikografske izkušnje pri ročni izbiri zgledov je bilo poleg omenjenih hevristik treba upoštevati še težnjo po izbiri med seboj čim bolj različnih zgledov: velika verjetnost namreč je, da bo leksikograf med 10 zgledi, ki so si med seboj zelo podobni, za pojasnitev konkretnega podatka v slovarski bazi izbral samo enega. Različnost zgledov je bila zato v konfiguraciji zagotovljena z implementacijo t. i. Levenhsteinove razdalje, ki zagotavlja 30-% razlikovalnost med izluščenimi zgledi. Evalvacija različnih konfiguracij je pokazala, da so posamezni parametri za določitev dobrega zgleda bolj pomembni kot drugi. Med najpomembnejšimi se je pokazala dolžina stavka, kjer je bilo mogoče ugotoviti, da zgledi, ki predvidevajo najmanj 8 in največ 30 besed, pogosto ne zagotavljajo zadostnega sobesedila. S tem, ko smo dolžino stavka povečali na najmanj 15 in največ 35 besed, se je povečalo tudi število dobro ocenjenih zgledov med 10 avtomatsko izločenimi. Kot zelo pomemben parameter se je pokazal tudi položaj ključne (tj. obravnavane) besede v stavku. Med zgledi, ki so bili ocenjeni kot dobri, se je ključna beseda skoraj vedno

169 Odbitne točke pomenijo rangiranje zgleda na lestvici najustreznejših, ne pa njegove izločitve.

pojavljala nekje v sredini oz. drugi polovici izluščenega zgleda, vendar pa to v večini primerov ne drži za glagole, kjer postavitev na konec stavka navadno pomeni odsotnost ustreznega sobesedila. Ob omenjenih parametrih se je tudi pokazalo, da zgledi, ki vsebujejo več kot eno pojavitev ključne besede, navadno niso dobri, zato je bil dodan še parameter, po katerem so zgledi z več kot enkratno ponovitvijo ključne besede dobili dodatne odbitne točke. Ocena konkretnih primerov avtomatsko pridobljenih zgledov je omogočila tudi izdelavo t. i. črnih list besed, kot npr. prislovov: *nato, tako, torej, potem, poleg tega, zaradi tega* itd., ki, zlasti, ko se pojavljajo na začetku stavka, povezujejo vsebino s predhodnim stavkom, kar zmanjša razumljivost in informativnost izluščenega zgleda.

Pri izboru zgledov smo imeli seveda ves čas v mislih dejstvo, da so namenjeni polnjenju slovarske baze in ne že končnemu slovarju. Pri avtomatskem izboru korpusnih zgledov so kandidate za dober slovarski zgled zato predstavljali tudi tisti primeri, ki so izkazovali možnost prilagoditve v smislu krajšanja, opuščanja zapletenih podrednih razmerij, osebnih lastnih imen ipd.

Sklep

Prikazani opis leksikografske obdelave sodobnega slovenskega besedišča, kot se je izoblikoval v procesu izdelave LBS, vključuje vse temeljne prvine slovarske zgradbe, kot so iztočnica, pomenski opis, inherentni slovnični podatki, podatki o besedilnem okolju, podatki o pomenskih elementih sobesedila, področna, žanrska in stilna kvalifikacija in prepoznavna večbesedne leksike. Ker lahko informacija o normi iz takega opisa šele izhaja, je v LBS nismo predvideli kot izhodiščne slovarske informacije. Čeprav so informacije o jezikovnem standardu v slovarju dobrodošle, še posebno, če jih je uporabnik vajen, je jasno, da ni mogoče sodobne slovenščine standardizirati na podlagi meril, ki so veljala za jezikovni opis v 60. in 70. letih, ampak je teba vzpostaviti nova izhodišča standardizacije, ki bi na takem (empiričnem) opisu šele temeljila.

V primerjavi z obstoječo slovarsko prakso (tj. slovensko slovarsko tradicijo), kjer je izhodišče slovarske obravnave leksem kot neločljiva celota izraza in vsebine, se je opis v LBS premaknil k njegovemu vsebinskemu delu, tj. k pomenu, ki tako postane osrednja slovarska enota. Vse slovarske enote, ki potrebujejo pomenski opis, tj. posamezni (pod)pomeni, SZ in FE, pa postanejo pomenske, tj. leksikalne enote. Ker je mogoče pripenjati več pomenov na en sam izraz, je njegova vloga v LBS omejena na združevanje pomenov pod skupni imenovalac, tj. pod lemo z enotno oblikoskladenjsko in naglasno paradigmo. Tak pristop pa pomeni vrsto sprememb, ki se tičejo prikazovanja homonimije, obravnave večbesednih leksikalnih enot, besednih vrst in besednih oblik glede na obstoječo slovarsko prakso. Zlasti možnosti, ki jih ponuja spletni medij pri prikazovanju pomenskih podatkov in njihovih medsebojnih povezav, so pokazale, da organizacija gesla in njegove naslovne enote v razmerju do različnih pomenskih vsebin, ki temelji na intuitivnem prepoznavanju pomenske povezanosti oz. nepovezanosti, ni dobro izhodišče za prikazovanje pomenskih lastnosti, saj se leksikografovo (teoretično) vedenje, ki temelji na prepoznavanju diahronih procesov, besedotvornih postopkov ipd. ne ujema nujno z uporabnikovim vedenjem in predstavami. Posledično lahko predvidevamo, da bo zaradi tega dostop uporabnika do slovarske informacije omejen ali celo onemogočen. Potreba po razumevanju leksikografskih odločitev s strani uporabnikov, ki pa so ključne za to, da uporabnik do informacije sploh pride, izhaja po vsej verjetnosti iz predstave, ki jo ima tiskana oblika slovarske iztočnice, kjer se predvideva, da ima uporabnik vedno pred seboj celotno podobo gesla, ne pa denimo samo njegov posamezni (pomenski) del. Očitek, da je na ta način uporabnik prikrajšan za določeno vedenje o leksikalni naravi pomena pa ne vzdrži, saj je v spletnem mediju

te podatke mogoče navesti samostojno – tj. v obliki pojasnil o pomenskem izvoru, stopnji prevzetosti, vezanosti na različne jezikovne skupnosti, kulture itd.

Posebno pozornost smo pri oblikovanju LBS namenili pomenski členitvi in pomenskemu opisu besed, zlasti prenosu stavčnih razlag v slovenščino, in vključevanju vezljivostnih, pragmatičnih, stilističnih ter konotativnih sestavin pomena v slovarsko razlago. Kot ključni element povezovanja pomena s skladnjo na način, ki omogoča tudi računalniško procesljivost, smo uvedli zapis pomenske sheme in nakazali možnosti nadgradnje, kot je denimo pripis udeleženskih vlog udeležencem (ang. *Semantic Role Labeling*) in avtomatsko generiranje stavčnih vzorcev na ravni pomena. Opisani postopki avtomatskega pridobivanja leksikalnogramatičnih podatkov iz korpusa so ključni pri uvajanju jezikovnotehnoloških postopkov v proces izdelave slovarja in predstavljajo vrsto možnosti za nadgradnjo, ki bi še izboljšala kakovost izluščenih podatkov in pospešila postopek izdelave slovarja.

Obravnava leksikalnogramatičnih podatkov je bila pri izdelavi LBS v celoti podrejena predstavi slovarja, ki deluje v digitalnem okolju in izkorišča prednosti spletnega medija. V leksikografsko teorijo in prakso so namreč vstopile nove tehnologije in s tem povezana znanja, ki omogočajo obsežnejše jezikovne analize in ponujajo nove poglede na predstavitev in kompleksnost slovarske informacije. Pri tem je logično, da koncept slovarja, kot je bil še do nedavnega zasidran v jezikovni zavesti leksikografov in uporabnikov, v takih okoliščinah ne more ostati nespremenjen.

Povzetek

V monografiji je opisana zasnova, zgradba in vsebina Leksikalne baze za slovenščino (LBS), slovarske oz. leksikalne podatkovne zbirke (Arhar Holdt 2011: 19–20, Gorjanc et al. 2005: 139), ki je bila izdelana pri projektu Sporazumevanje v slovenskem jeziku in v kateri so leksikalnogramatični podatki zbrani ter organizirani glede na možnost izrabe pri slovarskih opisih sodobne slovenščine, ki bi sledili potrebam slovenske jezikovne skupnosti in bi omogočali nadaljnje računalniške obdelave ter izdelavo jezikovnih aplikacij za slovenščino.

Teoretična izhodišča

LBS smo zasnovali na dobrih praksah leksikalnih baz in sorodnih projektov evropskih jezikov zadnjih 20, 25 let, obenem pa smo upoštevali specifično stanje slovenščine tako glede jezikovnih lastnosti kot glede obstoječih slovarskih priročnikov in trenutno razpoložljivih računalniško procesljivih leksikalnih virov. Zgledovali smo se predvsem po dveh projektih, ki sta z združitvijo skladenjskih in pomenskih informacij najbolj ustrezala predvidenemu na korpusu temelječemu opisu sodobne slovenske leksike. Gre za projekt FrameNet, ki ga je zasnoval Charles J. Fillmore (Fillmore et al. 2003), in projekt Corpus Patterns Analysis (CPA) Patricia Hanksa (Hanks 2004, 2008; Hanks in Pustejovsky 2005). Z vidika teoretičnih izhodišč, ki jih vključujeta, je pomembno izpostaviti logiko pomenskega opisa, ki temelji na ugotavljanju pomensko odločujočih elementov sobesedila, kamor sodi prepoznavanje skladenjskih vzorcev, kolokacij in pragmatičnih lastnosti pomena v sporočanjških situacijah.

Temeljne značilnosti

Pri oblikovanju LBS smo posebno pozornost namenili prilagoditvi jezikovnotehnoloških orodij specifičnim leksikalnogramatičnim lastnostim slovenščine in si prizadevali preizkusiti čim več možnosti avtomatiziranja postopkov pri pridobivanju relevantnih leksikografskih podatkov iz korpusa in pri njihovem vključevanju v slovarsko bazo. Na podlagi analiz, ki so bile opravljene v prvi fazi z uporabo besednih skic in drugih funkcij v orodju Sketch Engine (Kilgarriff et al. 2004), smo postopoma oblikovali parametre, ki so omogočili izločitev relevantnih skladenjskih struktur s pripadajočimi kolokacijami in zgledi iz korpusa ter njihov neposredni prenos v slovarski vmesnik. Na ta način smo leksikografovo presojo osredotočili na analizo obvladljive količine relevantnih podatkov ter na ključna leksikografska opravila, kot je pomenska členitev, pomenski opis, prepoznavanje večbesednih leksikalnih enot ter oblikoskladenjskih in skladenjskih značilnosti pomena.

Pri pomenskem opisu smo izhajali iz ugotovitve, da pomenskih lastnosti ni mogoče ločiti od skladenjskega obnašanja besede in da je smiselno – tudi z vidika možnosti, ki jih ponuja spletni medij in narekujejo potrebe sodobnega uporabnika – ločiti izrazno podobo besede od njene vsebine (pomena). Slovarsko iztočnico smo tako obravnavali kot formalno prepoznavno besedno obliko – tj. lemo, kot je definirana v korpusu, saj nam ta predstavlja vir informacij, na podlagi katerih sprejemamo leksikografske zaključke. Vsebina izraza, tj. njegova pomenska vrednost, pa je zastopana s posameznimi pomeni besed in večbesednih enot, kot so stalne zveze in frazeološke enote. Te slovarske enote, ki so bile v tiskanih slovarjih nujno prikazane enodimenzionalno v odvisnosti od naslovne enote geselskega članka, postajajo v novih okoliščinah samostojni elementi slovarja, ki pa so lahko organizirani in povezani različno, pač glede na to, katero skupno točko postavimo za izhodišče. Tak pristop je narekoval tudi obravnavo besednovrstne konverzije, kjer smo na spremenjeno skladenjsko vlogo opozarjali pri posameznih pomenih, in obravnavo homonimije, kjer smo kot samostojne iztočnice prikazovali samo leme z neprekrivnimi oblikoskladenjskimi kategorijami, npr. *plesen* – sam. : *plesen* – prid., *prst* – sam. ž. spola : *prst* – sam. m. spola, ter leme z neprekrivno spregatveno in naglasno paradigmo, npr. *pasti* – *pasem* : *pasti* – *padem*. Različne pomene takih lem pa smo obravnavali kot pomene, tj. znotraj večpomenskosti, in sicer ne glede na etimološko (ne)povezanost, npr.

vila *samostalnik*

1. lepa, razkošna hiša
2. pravljичno ali mitološko bitje

Zgradba LBS in tipi leksikalnogramatičnih podatkov

LBS, kot smo jo zasnovali pri projektu SSJ, je v prvi vrsti namenjena slovarskemu tipu jezikovnih podatkov. Predvsem nas je torej zanimalo, kaj določena beseda oz. leksikalna enota pomeni, posledično pa tudi vse, kar izhaja iz njenih pomenskih lastnosti, tj. obnašanje v sobesedilu, stilistični potencial, omejitve v rabi ipd. Pri tem smo upoštevali dejstvo, da je opis pomena v slovarju nujno leksikografski konstrukt, ki ni enak razbiranju pomenov v sporočanjškem procesu, in da pomenske tendence besed vplivajo na izbire v besedilnem okolju, kar se odraža v stavčnih vzorcih, skladenjskih strukturah in kolokacijah. V okviru dvostopenjske pomenske hierarhije, ki jo določa glavni pomen oz. osnovni pomeni in podpomeni, smo težili k podrobnemu pomenskemu členjenju (Atkins in Rundell 2008: 268). Pri tem smo se zanašali na analizo vzorca korpusnih konkordanc, ter na prepoznavanje pomenskih in slovničnih elementov pomena,

kot jih je mogoče izbrati iz besednih skic v orodju Sketch Engine. Pri tem se je za zagotovitev optimalnih in leksikografsko relevantnih podatkov kot izredno pomembna pokazala prilagoditev orodja specifikam slovenščine, kar je bilo mogoče narediti s postopno analizo besedišča in implementiranjem ugotovitev v korpusno orodje.

Pomenski nivo

Pomensko informacijo smo v LBS razdelili na več segmentov: pomenski meni, pomenski indikator, pomenska shema, slovarska razlaga ter oznaka.

Pomenski meni, kot ga poznamo zlasti iz slovarjev, namenjenih učenju angleščine kot tujega jezika (Ptasznik in Lew 2014), prikazuje pomensko zgradbo večpomenskih gesel in logična razmerja med pomeni in podpomeni, z razvrstitvijo osnovnih pomenov pa ponuja tudi grobo predstavo o zastopanosti posameznega pomena v jezikovni rabi. Zato je pomembno, da je razporeditev in opredelitev pomenov znotraj pomenskega menija logična in na prvi pogled razumljiva. To smo skušali doseči s t. i. pomenskimi indikatorji, ki so najpogostejše enobesedni ali kratki pomenski opisi leksikalnih enot. Za vse vrste indikatorjev, ki smo jih uporabljali v LBS, tj. za sinonimne, opisne, nanašalne in ontološke, velja, da morajo biti kratki, razumljivi in asociativni s pomensko vrednostjo, ki jo zastopajo:

jagoda *samostalnik*

1 rastlina

1.1 sadež

2 majhen okrogel plod

2.1 o grozdju

3 predmet

Za beleženje pomenskoskladenjskih lastnosti pomena smo uvedli zapisovanje t. i. pomenskih shem, ki temeljijo na predpostavki, da pomena besede ni mogoče ločiti od njegove skladenjske realizacije. Ker zapis pomenske sheme nosi v sebi ključne informacije o realizaciji pomena v okviru skladenjsko zaključene zgradbe, kot je stavek ali besedna zveza, predstavlja izhodišče za (a) oblikovanje stavčnih razlag, ki vključujejo udeležence in okoliščine ter konotativne in pragmatične pomenske elemente, ter (b) za strojno procesiranje in avtomatsko prepoznavanje pomenov in semantično označevanje korpusov. Pomenska shema vključuje naslednje elemente: udeležence in okoliščine, zapisane v obliki semantičnih tipov (z velikimi črkami), ter pomensko relevantne prvine sporočanje situacije, kar smo poimenovali pomenski scenarij:

sesti *glagol*
ustrezati

če rečemo, da neka DEJAVNOST, STANJE ali LASTNOST česa ČLOVEKU sede, s tem mislimo, da mu ugaja ali ustreza

Stavčni vzorec, ki ga vsebuje pomenska shema, praviloma vključuje vse pomensko relevantne udeležence in okoliščine oz. prikazuje polno udeležensko zgradbo danega pomena, ki jo je mogoče ugotoviti iz večjega števila realnih besedilnih rab, pri čemer se vsi pomensko relevantni udeleženci bodisi ne realizirajo v vseh rabah bodisi se ne uresničuje nujno le znotraj enostavčne povedi, pač pa lahko tudi v okviru širšega sobesedila. Udeleženska mesta so v pomenski shemi zapisana s semantičnimi tipi, ki jih razumemo kot abstraktne zastopnike realnih leksikalnih zapolnitev, ki se nato na kolokacijskem nivoju realizirajo v obliki kolokacij:

če ŠPORTNIK ali ŠPORTNA EKIPA^[1] zmaga, doseže na ŠPORTNEM TEKMOVANJU^[2] v določenem ŠPORTU^[3] ali glede na določena MERILA^[4] v odnosu do drugih ŠPORTNIKOV ali ŠPORTNE EKIPE^[5] NAJBOLJŠI REZULTAT^[6]

[1], [5] ŠPORTNIK, ŠPORTNA EKIPA: [skakalec, smučar, domačin, gostitelj, barca, favorit, ekipa] zmaga

[2] ŠPORTNO TEKMOVANJE: zmagati na [tekmi, turnirju, dirki, tekmovanju, mitingu]; zmagati v [tekmi]

[3] ŠPORT: zmagati na [slalomu, smuku, rallyju]; zmagati v [veleslalomu, superveleslalomu, smuku]

[4] MERILA: zmagati v [kategoriji, konkurenci nizu, etapi, disciplini]

[6] REZULTAT: zmagati z/s [točko, golom, zadetkom]

Pomenski scenarij predstavlja tretji sestavni del pomenske sheme in je primarno namenjen opisu okoliščin sporočanja, vrednotenju vsebine sporočila in razmerij med udeleženci sporočanja, kot jih zapisujemo v pomenskih razlagah. V LBS smo ločili dva tipa pomenskih razlag: t. i. pomenskosestavinske razlage ali klasične (aristotelovske) razlage, kot jih poznamo iz SSKJ in so tudi sicer v slovarjih pogosto zastopane, ter stavčne razlage, značilne zlasti za angleške slovarje, namenjene učenju angleščine kot tujega jezika.

Pomenskosestavinske razlage smo uporabili pri razlaganju velike večine konkretnih in abstraktnih samostalnikov z ustaljeno pomensko vrednostjo oz. pri vseh tistih pomenih, ki se ne izražajo prek sporočajske funkcije, npr.

skodelica *samostalnik*

manjša, navadno okrogla keramična posoda, namenjena shranjevanju in postrežbi tople hrane ali pijače

Stavčne razlage so se na drugi strani pokazale kot učinkovite pri tistih besedah oz. njihovih pomenih, kjer je za ustrezno pomensko interpretacijo pomembna integracija skladijske informacije v sobesedilo, zlasti torej pri vezljivostnih pomenih (a) samostalnikov in (b) pridevnikov ter povsod tam, kjer konverzijski obrazec »če ...« stavka omogoča vključitev pragmatičnih in konotativnih pomenskih elementov, tipično torej pri razlaganju (c) frazeoloških enot, pa tudi vsebin, ki se tipično realizirajo (d) v govorjenem diskurzu:

(a) komentar 1 izražanje mnenja; komentiranje
če da ČLOVEK komentar na DOGODEK, SITUACIJO ali TEMO, izrazi o tem svoje mnenje
(b) dovzeten če je ČLOVEK dovzeten za nek VPLIV, IDEJO ali SPREMEMBO, jo je pripravljen sprejeti in uresničiti, navadno zato ker ustreza njegovim nazorom ali osebnostnim lastnostim
(c) naliti/natočiti komu čistega vina <i>če rečemo, da smo komu nalili čistega vina, želimo poudariti, da smo mu povedali resnico ali dejstva o kaki neprijetni stvari ali dogodkih</i>
(d) cepec <i>če rečemo komu cepec, ga imamo za neumnega in nepremišljenega</i>

Naslednja informacija, ki je v LBS vezana na opis pomena, se v zgradbi gesla nahaja v elementu oznaka, kjer smo v obliki kratkih pojasnil beležili, kako je posamezni pomen besede, stalne zveze, frazeološke enote ali njene variante treba razumeti oz. uporabiti v besedilu glede na to, katere slovnične omejitve predvideva (slovnične in skladijske oznake), na katerem strokovnem področju se rabi (področne oznake), kakšen odnos do vsebine ali udeležencev izraža (konotativne, stilne in pragmatične oznake), na katero časovno obdobje (časovne oznake), besedilni tip, ter regionalne ali kulturološke posebnosti je vezan (kontekstualne oznake). Oznake smo posameznim pomenom dodajali z upoštevanjem tega, da jih bo v spletnem slovarju mogoče prikazovati bodisi v obliki krajših, čim bolj razumljivih opozoril z različnimi možnostmi vizualizacije (prim. Krek et al. 2013: 35), ali kot del pomenskih razlag. Bistvena novost pri označevanju besedišča v LBS glede na obstoječ kvalifikatorski sistem v SSKJ je, da v LBS besedišča ne opisujemo z vidika knjižnojezikovne norme in z namenom uporabnika kultivirati, ampak izključno z vidika lastnosti besedišča, kot izhajajo iz empirično preverljivih sporočanjških situacij in z namenom uporabnika predvsem informirati.

Na pomenskem nivoju LBS obravnavamo tudi večbesedne leksikalne enote, ki na podlagi leksikografske presoje potrebujejo pomenski opis. V LBS jih formalno ločimo od drugih večbesednih enot, kot so denimo kolokacije in strukturno ustaljeni skladenjski nizi, ki pomenskega opisa ne potrebujejo.

Pri prepoznavanju stalnih zvez (SZ) kot samostojnih leksikalnih enot smo upoštevali dejstvo, da njihovega celostnega pomena ali sporočanje funkcije ni mogoče izpeljati iz pomenov ali funkcij posameznih sestavin. To lastnost si SZ delijo tudi s FE, hkrati pa za razliko od FE njihov pomen kot celota ni metaforičen oz. v besedilu nima ekspresivne ali vrednotenjske sporočanje vloge. V prvo skupino sodijo zveze kot npr. spalna srajca, glasbeni stolp, osnovna šola, bančna kartica ipd., v drugo pa zveze, katerih pomenski opis vključuje tudi pojasnitev konotativnih in pragmatičnih pomenskih elementov, npr. princ na belem konju, naliti komu čistega vina, biti krvav pod kožo ipd. Ker gre pri ločevanju SZ od FE na eni strani in SZ od prostih zvez in kolokacij na drugi za polje, ki v jeziku nima jasno določljivih mej (prim. Rundell in Atkins 2011), saj je njihovo prepoznavanje v smislu pomenske (ne)prozornosti povezano s konkretnim besedilnim kontekstom in predispozicijami jezikovnega uporabnika, npr. ali gre za šolarja, učenca slovenščine kot tujega jezika, materne govorce ipd., smo v strukturi gesla predvideli možnost uvrstitve SZ pod posamezni registrirani pomen katere od sestavin SZ bodisi od registriranih pomenov besede v iztočnici neodvisno:

mreža *samostalnik*

1 preplet niti, žic

1.2 predmet

visočna mreža

naprava za ležanje iz trpežne tkanine in vrvi, ki se napne med dve drevesi ali stojali

2 o ribolovu

vlečna mreža

stojčna mreža

5 vzorec

koordinatna mreža

geografija

geometrija

nevronska mreža

računalništvo

kristalna mreža

kemija

matna mreža

šah

Taka obravnava SZ v strukturi gesla omogoča različne predstavitve v spletnem slovarju. Uporabniku je tako v primeru pomenske podrejenosti SZ kateremu od pomenov na voljo tudi informacija o pomenu katere od sestavin zveze, če taka povezava seveda obstaja. Zaradi možnosti povezav med enotami slovarske baze smo pri opisu SZ predvideli tudi možnost, da se SZ kot samostojna leksikalna enota pojavlja hkrati tudi v kolokacijskem nizu pri katerem od pomenov besede, ki je njen sestavni del, npr.

šola *samostalnik*

1 o izobraževanju

1.4 proces

[osnovna, srednja, višja, visoka, poklicna] šola

osnovna šola

vzgojno-izobraževalni proces, ki traja 8 oz. 9 let in je obvezen

V LBS smo kot frazeološke enote (FE) obravnavali tiste zveze besed, ki poleg več-besednosti izkazujejo glede na sestavne elemente neodvisni – tj. frazeološki pomen, ki ga je v besedilu mogoče prepoznati na podlagi težnje po opaznem, t. i. »slikovitem« izražanju. Zgradbeno so FE bolj zapletene kot SZ, saj so lahko strukturirane kot nestavčna podredna in priredna besedna zveza ter kot stavčna ali nestavčna poved. V prvi skupini gre za FE, ki so v svoji osnovni obliki prekrivne s samostalniškimi, pridevniškimi, glagolskimi in prislovnimi besednimi zvezami, npr. *star maček*, *vonj po denarju*; *blažen med ženami*; *iti kot po maslu*, *jasno in glasno*, v drugi skupini pa za FE, ki so pomensko odvisne od vsebine širšega sobesedila in imajo navadno zgradbo nestavčne povedi, npr. *korajža velja!*, *odrasti že (enkrat)!*, *daj/dajte (no) mir!*, njihova sobesedilnost pa je pogosto vezana na govorjeni diskurz. Tipična značilnost FE je, da kljub pomenski in strukturni trdnosti izkazujejo veliko stopnjo variantnosti, in sicer tako pri posameznih sestavinah kot pri besednih oblikah in predvidenih prostih skladenjskih mestih. V LBS smo take primere beležili znotraj istega nadelementa, če ni prihajalo do bistvenih pomenskih sprememb in seveda le, če so se take variante kazale kot tipične in za posamezno FE relevantne v sodobni slovenščini:

rešilna bilka | **videti v čem rešilno bilko** | **ponuditi komu rešilno bilko** | **oprijeti se česa kot rešilne bilke** | **oprijeti se/okleniti se edine/zadnje/vsake rešilne bilke**

Kolokacijski nivo

Kolokacije so, od kar je mogoče opazovati leksikalno obnašanje besed v njihovem besedilnem okolju v realnih besedilih in s pomočjo statističnih orodij, nepogrešljiv del pomenskega opisa besed. Razumevanje kolokacij v LBS se približuje pojmovanju, kot ga je v izhodišču zastavil Firth (1957), kasneje pa so v svojih leksikalnih in komunikacijskih študijah iz tega izhajali tudi drugi, zlasti npr. Halliday (1966) in Sinclair (1987). Tako izhodišče med drugim opredeljuje kolokacije kot izraze, sestavljene iz dveh ali več besed, s katerimi na konvencionalen, v jeziku običajen način izražamo stvari, hkrati pa gre za kombinacije, ki so statistično prepoznavne in pomensko smiselne. Zato v LBS kolokacij ne obravnavamo kot samostojnih leksikalnih enot, imajo pa pomembno vlogo pri pomenski členitvi besed in njihovem pomenskem opisu. V LBS smo registrirali dve vrsti kolokacij: regularne ali binarne kolokacije ter razširjene kolokacije. Osnovni tip kolokacije nam je predstavljala zveza dveh polnopomenskih besed, ki lahko kot tretjo besedo vključuje tudi predlog, npr.

[orožarska, korupcijska, podkupovalna] afera
[šotor, bivališče] nomadov; [otvoritev, odprtje] razstave
zastava [plapola, vihra, se dviga], zbor [obravnavo, razpravlja, zaseda]
[opravičilo, krivec] za grozote; [raj, poslastica] za gurmane
etika v [znanosti, zdravstvu, športu]; [vadba, telovadba, razgibavanje] v fitnesu
[obračunavanje, žuganje] s pestmi; ogrlica z [diamantom, biseri]

V procesu oblikovanja LBS se je pri analizi besednih skic pokazalo, da se nekatere besedne kombinacije zelo pogosto pojavljajo še s tretjim elementom, pri čemer je ta element bodisi (a) tipično dopolnilo katere od sestavin kolokacije, lahko pa je (b) nujno potreben za njeno ustrezno razumevanje:

- (a) [ohraniti, ohranjati] tradicijo -> ohranjati [družinsko] tradicijo
[tekma, tekmovanje] smučarjev -> tekmovanje smučarjev [tekačev, skakačev]
- (b) gostiti [svetovno] smetano
obiskovati v [zadnjem, prostem] času

Skladenjski nivo

Skladenjska informacija je v LBS vezana na dejstvo, da je pomen neposredno povezan s skladnjo in da se lahko v besedilu uresničuje na različne skladenjske načine. Opis skladenjskih informacij v zvezi s posamezno leksikalno enoto tako vključuje formalizacijo skladenjskih struktur, v katerih se obravnavana beseda pojavlja, opis slovničnih omejitev in preferenc, pri glagolih pa še zapis stavčnega vzorca.

Skladenjske lastnosti pomena, tj. tiste, ki izhajajo iz sintagmatskih razmerij, smo v LBS beležili s pomočjo t. i. skladenjskih struktur pri vseh leksikalnih besednih vrstah, tj. samostalniki, glagolih, pridevniki in prislovih. Skladenjske strukture so neposredno povezane z gramatičnimi relacijami, ki so prilagojene avtomatskemu pridobivanju kolokacijskih in drugih podatkov iz korpusa s pomočjo besednih skic, zato je pomembno, da temelji njihova identifikacija na predhodnih skladenjsko-pomenskih analizah. Vsaka struktura, ki smo jo beležili v LBS, ima torej v slovnici besednih skic svojo gramatično relacijo oz. vsaka gramatična relacija, ki jo vsebuje slovnica besednih skic, zgenerira eno ali več skladenjskih struktur v slovarski bazi. Različica slovnice besednih skic (Krek 2012a), ki smo jo uporabljali za izdelavo LBS, vsebuje 103 gramatične relacije, ki zgenerirajo pribl. 370 skladenjskih struktur. V izhodišču smo ločili t. i. posebne strukture, kamor sodijo strukture s pomožnimi glagoli, nedoločniki, nikalnimi elementi in količinskimi izrazi, ter regularne strukture, ki so lahko predložne ali brezpredložne.

Podobno vlogo, kot imajo strukture pri samostalniki, pridevniki in prislovih, imajo pri glagolih stavčni vzorci, ki jih v kontekstu izdelave LBS razumemo kot formalni prikaz vezljivosti posameznega glagolskega pomena znotraj stavčne zgradbe. Na ta način stavčni vzorci potrjujejo udeležensko zgradbo, zapisano v pomenski shemi, hkrati pa razkrivajo možnosti opuščanja elementov kot tudi tipične izbire med več skladenjskimi možnostmi, kar je zlasti očitno pri predložnih zvezah v skladenjsko različnih vlogah:

pritrđiti *glagol*

trđno namestiti na določen predmet, mesto

če ČLOVEK pritrđi nek PREDMET na DRUG PREDMET, ga z njim poveže v trđno celoto, navadno s PRIPOMOČKOM ali uporabo SNOVI

kdo pritrđi kaj
 kdo pritrđi kaj na kaj
 kdo pritrđi koga na kaj
 kdo pritrđi kaj na koga
 kdo pritrđi kaj nad kaj
 kdo pritrđi kaj ob kaj
 kdo pritrđi kaj v kaj
 kdo pritrđi kaj pod kaj
 kdo pritrđi kaj med kaj
 kdo pritrđi kaj čez kaj
 kdo pritrđi kaj okoli česa
 kdo pritrđi kaj za kaj
 kdo pritrđi kaj s čim

Slovnične omejitve in preference smo v LBS uvedli kot poseben tip slovničnih in skladenjskih oznak, ki jih pripenjamo na posamezne dele skladenjskih struktur ali stavčnih vzorcev in so namenjene označevanju posebnosti znotraj predvidljive slovnične paradigme, skladenjske vloge ipd., zaradi česar so uporabne za ustrezno prepoznavanje izjem pri strojnem procesiranju podatkov ter za slovnične in skladenjske analize:

Kol-sbz0 SBZ2<r>v množini</r>
 [čreda] antilop; [kopica, kup, vrsta] nasvetov; [škaf, komplet] barvic
 pbz0<r>primernik</r> SBZ0
 [višja, nižja] najemnina
 sbz1<r>v pridevniški rabi</r>SBZ0
 [DOS] aplikacija; čarter [plovilo, jadrnica]; čili [paprika, omaka]; džez [scena, festival]
 SBZ0 v SBZ5<r> navadno z lastnim imenom</r> v sbz5
 intervencija v [Iraku, Vietnamu, ZRJ, Somaliji, Afganistanu, Bosni]

Na skladenjskem nivoju smo poleg skladenjskih struktur beležili tudi t. i. skladenjske *zveze*, tj. statistično prepoznavne večbesedne nize, ki izkazujejo predvsem strukturno ustaljenost, hkrati pa – vsaj z vidika maternih govorcev slovenščine – ne izkazujejo samostojne pomenske vrednosti. Med drugim to pomeni, da ne predvidevajo pomenske razlage, zato smo jih uvrščali pod posamezni (pod)pomen besede, ki jo obravnavamo v iztočnici, npr.

žlica *samostalnik*
 zmešati kaj z/s [x] a
 žlicami čes

angažirati *glagol*
 angažirati se na
 področju česa

znan *pridevnik*
 znan kot [velik, nepopoljšljiv]
 [deloholik, ženskar]

Ponazarjalni nivo

Zgledi rabe so pomembni, če ne kar ključni del slovarske informacije, zato imajo tudi v zasnovi LBS pomembno mesto. Njihova temeljna vloga je potrditi in ponazoriti vsa jezikovna dejstva, predhodno navedena v slovarski bazi. Izbira zgledov rabe se z vidika oblikovanja slovarske baze in izdelave slovarja razlikuje. V slovarski bazi težimo k izbiri večje količine korpusnih zgledov, ki so tipično celi stavki, vzeti iz korpusa, in predstavljajo kandidate za dobre slovarske zgledne. Kot pravita Atkins in Rundell (2008: 458), veljajo za dobre slovarske zgledne taki, ki so predvsem naravni in tipični, informativni in razumljivi. K zgoraj izpostavljenim načelom je treba dodati še politično in družbeno korektnost, ki je za slovar, ki predstavlja določeno družbeno avtoriteto (Gorjanc 2013/14: 12), zelo pomembna. Eden izmed pomembnih korakov v to smer je diferenciran izbor besedil, ki jih vključuje (referenčni) korpus, drugo pa možnost izbire nevtralnih a hkrati tipičnih sobesedilnih elementov na ravni kolokacij.

Ker je iskanje dobrih slovarskih zgledov med najbolj zamudnimi leksikografskimi nalogami, smo pri vključevanju zgledov težili k čim večji avtomatizaciji postopkov tako pri njihovi izbiri kot pri vključevanju v slovarsko bazo. Orodje SkE predstavlja s funkcijama GDEX in Tickbox Lexicography dobro izhodišče za nadgradnjo z upoštevanjem specifik slovenščine (Kosem et al. 2011). Z določitvijo parametrov, kot so npr. dolžina besed v zgledu, dolžina stavka, prisotnost besed s t. i. črne liste, manj pogostih besed, nejezikovnih elementov ipd., je mogoče razmeroma hitro priti do zadostne količine dobrih korpusnih zgledov. Kot jezikovno specifično pomembno se je pokazalo zlasti merilo pozicije ključne besede v stavku, ki v slovenščini varira glede na besedno vrsto, in upoštevanje t. i. tretjega kolokatorja - tipičnega elementa širšega besedilnega konteksta besede v pomenu, ki ga ponazarja korpusni zgled. Postopek avtomatskega pridobivanja korpusnih zgledov prek skladenjskih struktur in kolokacij, ki je bil preizkušen pri izdelavi LBS, je bil deležen velikega zanimanja tudi znotraj širše slovarske skupnosti, povezane v evropsko pobudo Enel (European Network of e-Lexicography), saj ga je s prilagajanjem parametrov in možnostmi nadaljnega procesiranja izluščenih podatkov mogoče aplicirati tudi na druge jezike, ki slovarski opis gradijo na prepoznavanju pomenskih lastnosti sobesedila z uporabo korpusnih orodij.

Predstavljena struktura in vrsta leksikalnih podatkov v LBS je v slovenski leksikografiji nova v tem, da v celoti temelji na korpusnih izhodiščih, kar pomeni, da so vsi jezikovni opisi empirično utemeljeni na sodobnem jeziku in da omogočajo vpogled v to, kaj je v jeziku naravno, hkrati pa opozarjajo tudi na posebnosti, ki od regularnosti odstopajo in predstavljajo potencialni uporabniški problem. V slovenski leksikografiji predstavlja vzorčni opis 2.500 gesel v slovarski bazi tudi nov pogled na organizacijo in hierarhizacijo podatkov, ki že v zasnovi predvideva možnost nadaljnega strojnega procesiranja in je že v izhodišču oblikovana z mislijo na možnosti spletne vizualizacije slovarskih vsebin, ki jih je mogoče prilagajati prepoznanim uporabniškim potrebam.

Summary

This book describes the concept, structure and contents of the Slovene Lexical Database (SLD) (cf. Arhar Holdt 2011: 19–20, Gorjanc et al. 2005: 139), a description of modern Slovene developed within the Communication in Slovene project and containing lexico-grammatical data intended for the purposes of lexicographic description of Slovene, as well as for the development of natural language processing (NLP) applications and tools for Slovene.

Theoretical background

The concept of SLD is based on best practices in the creation of lexical databases and similar projects for European languages in the last 20–25 years. On the other hand, specific features of Slovene and the state of available machine-readable lexical resources were also considered. Two projects are most prominent as the inspiration of the SLD concept, as they represented a good point of departure by combining syntactic and semantic information based on empirical data from text corpora. The first one is FrameNet (FN) developed by Charles J. Fillmore (Fillmore et al. 2003), and the other one is the Corpus Pattern Analysis project (CPA) by Patrick Hanks (Hanks 2004, 2008; Hanks and Pustejovsky 2005). From the theoretical point of view it is important to emphasize the prevailing role of semantic description based on detection of semantically determinative contextual elements which include syntactic patterns, collocations and pragmatic properties of particular senses in communicative situations.

Basic features

In the process of SLD conceptualization, special attention was given to the adaptation of NLP tools to specific lexico-grammatical characteristics of Slovene, and to the techniques of automatic extraction of lexicographically relevant data from annotated corpora and their integration in the SLD. Based on the analyses of word sketches and other modules available in the Sketch Engine tool in the first stage of the project, parameters of the extraction of collocations and corresponding corpus examples in relevant syntactic structures were formulated, and their automatic inclusion in the database was devised. This enabled lexicographers to concentrate on the analysis of manageable amounts of relevant data and on key lexicographic tasks such as sense division and description, identification of multi-word expressions, and lexico-grammatical and syntactic characteristics of particular senses.

Semantic descriptions in SLD are based on the premise that semantic properties of a headword cannot be separated from its syntactic behavior, and that it is sensible – also in the light of the new possibilities offered by the digital medium and new user needs – to distinguish between purely formal appearance of the headword as a string of characters, and its substance – meaning. A headword is treated as a recognizable word form i.e. a lemma, as defined in the resources providing lexicographically relevant information for SLD: the linguistically annotated corpus, which is in turn dependent on the information from a morphological lexicon. The substance – the meaning of formalized headword items – is presented by different senses of single-word and multi-word expressions, such as compounds and phraseology. These dictionary elements were traditionally treated as subordinate to the headword in a flat structure of printed dictionaries, but in the digital database they become independent elements that can be re-organized and interlinked in different ways depending on various possible starting points. This approach is characterized by the treatment of conversion or zero derivation, as the change in syntactic roles is signaled within particular senses, not as different headwords, and of homonymy, as only lemmas with distinct morphosyntactic categories are treated as headwords, e.g. plesen (blight) – noun : plesen (of dance) – adjective, prst (finger) noun masculine : prst (soil) noun feminine. The same is true for lemmas with dissimilar inflectional or accentual paradigms, e.g. pasti (to fall) – padem (1st pers sing) : pasti (to graze) – pasem (1st pers sing). Therefore, senses of lemmas with common morphosyntactic features and inflectional or accentual paradigms are always treated in one lexical entry, regardless of etymology.

vila *samostalnik (noun)*

1 lepa, razkošna hiša (villa, luxury house)

2 pravljичno ali mitološko bitje (fairy, imaginary being)

SLD structure and types of lexico-grammatical information

SLD as part of the Communication in Slovene project was devised primarily for lexicographic purposes, to organize dictionary type of language data. Our first interest was description of meaning of a word or a lexical unit, and consequently all its semantic features, such as contextual behavior, stylistic potential, restrictions in use, etc. One of the starting points was the fact that by necessity sense descriptions in dictionaries are lexicographic constructs, different from decoding meaning in a communicative process, and that meaning potentials have impact on the choice of words in textual environment, which is reflected in verbal patterns, syntactic

structures and collocations. Within the limits of a two-level semantic hierarchy with main senses and sub-senses, the bias was on the sense splitting side (Atkins in Rundell 2008: 268). Identification of senses was based on the analysis of a sample of corpus concordances, and of word sketches provided by the Sketch Engine tool. In the process, it was found that it is important to adapt the tool to specific characteristics of Slovene, which was made possible through a gradual analysis of the vocabulary and the implementation of findings in the tool.

Semantic level

Semantic information is segmented into several elements: sense menu, sense indicator, sense frame, dictionary definition and labels.

Sense menu, known primarily from EFL dictionaries (Ptasznik in Lew 2014) shows the structure of senses in polysemous entries and logical relations between senses and sub-senses. The ordering of senses crudely signifies the importance of a particular sense in real language use. It is important that the ordering and description of senses within the menu is logical and comprehensible at first glance. This is achieved by using sense indicators, single-word or condensed descriptions of senses of lexical units. As a rule, all types of indicators used in SLD – synonymous, descriptive, referential and ontological – have to be short, easy to comprehend and providing proper associations for the relevant sense:

jagoda (berry; strawberry; bead)

- 1 rastlina (plant)
 - 1.1 sadež (fruit)
- 2 majhen okrogel plod (small round seed)
 - 2.1 o grozdu (of grapes)
- 3 predmet (object)

Sense frames were introduced to record semanto-syntactic characteristics of senses. The concept is based on the premise that word meaning cannot be separated from its syntactic realization. Sense frames incorporate key information about a particular sense within a syntactically complete formation, such as a sentence or a phrase, and represent a starting point for (a) the formation of whole-sentence definitions which integrate arguments and adjuncts, as well as connotative and pragmatic elements of senses, and (b) natural language processing, particularly word-sense disambiguation and semantic role labelling. In the example below, the sense frame includes: arguments and adjuncts recorded as semantic types (in capital

letters), and semantically relevant elements of communicative situation, termed as a semantic scenario (in *italics*):

sesti glagol (to sit)

ustrezati (to suit, to agree with)

če rečemo, da neka DEJAVNOST, STANJE ali LASTNOST česa ČLOVEKU sede, s tem mislimo, da mu ugaaja ali ustreza

if we say that an ACTION, STATE or PROPERTY of something sits *to a PERSON, we mean that it suits them or agrees with them

The pattern included in the semantic frame usually includes all semantically relevant arguments and adjuncts or it shows the valency pattern of a particular sense which can be inferred from a larger number of instances of real use. It is not necessary that all semantically relevant arguments are realized in all particular instances or within one sentence, they can also be found in a larger context. Arguments are described as semantic types which are perceived as abstract representatives of real lexical items recorded on the next hierarchical level as collocations:

če ŠPORTNIK ali ŠPORTNA EKIPA_[1] zmaga, doseže na ŠPORTNEM TEKMOVANJU_[2] v določenem ŠPORTU_[3] ali glede na določena MERILA_[4] v odnosu do drugih ŠPORTNIKOV ali ŠPORTNE EKIPE_[5] NAJBOLJŠI REZULTAT_[6]

if a SPORTSMAN or a SPORTS TEAM_[1] wins, they achieve the BEST RESULT_[6] in a SPORTS COMPETITION_[2] in a particular SPORT_[3] or according to given RULES_[4] in relation to other SPORTSMEN or SPORTS TEAMS_[5]

[1], [5] ŠPORTNIK, ŠPORTNA EKIPA: [skakalec, smučar, domačin, gostitelj, favorit, ekipa] zmaga

[1], [5] SPORTSMAN, SPORTS TEAM: [jumper, skier, home team, host team, favourite, team] wins

[2] ŠPORTNO TEKMOVANJE: zmagati na [tekmi, turnirju, dirki, tekmovanju, mitingu]; zmagati v [tekmi]

[2] SPORTS COMPETITION: to win *at the [game, tournament, race, competition, meeting]; to win *in the [game]

[3] SPORT: zmagati na [slalomu, smuku, rallyju]; zmagati v [veleslalomu, superveleslalomu, smuku]

[3] SPORT: to win *at [slalom, downhill, rally]; to win *in [giant slalom, super-G, downhill]

[4] MERILA: zmagati v [kategoriji, konkurenci nizu, etapi, disciplini]

[4] RULES: to win in the [category, lap, leg, discipline]

[6] REZULTAT: zmagati z/s [točko, golom, zadetkom]

[6] RESULT: to win *with a [point, goal, hit]

A sense scenario represents the third part of the sense frame. Its primary role is the description of the circumstances of communication, evaluation of the content of the message and the relations between the participants in communication, as parts of meaning explanations. In SLD two types of explanations are used: genus-differentia or Aristotelian definitions, widely used in lexicography, also in SSKJ, and whole-sentence definitions, typical of EFL dictionaries.

Genus-differentia definitions are used to explain the meaning of most of concrete and abstract nouns with stable semantic values, and all senses that do not express some communication function, e.g.

skodelica *samostalnik* (a cup, a mug)

manjša, navadno okrogla keramična posoda, namenjena shranjevanju in postrežbi tople hrane ali pijače

small, usually round ceramic container used for storing or serving warm food or drinks

On the other hand, whole-sentence definitions are efficient in explaining words or senses that require the integration of syntactic co-text in the definition for its meaning to be interpreted. In particular, this is the case with (a) nouns and (b) adjectives forming valency patterns, and in cases where the conversational if-clause form enables the integration of pragmatic and connotative semantic elements, typically in defining (c) phraseology, and also elements usually realized in (d) spoken discourse.

(a) komentar (a comment)

1 izražanje mnenja; komentiranje (expressing opinion; commenting)

če da ČLOVEK komentar na DOGODEK, SITUACIJO ali TEMO, izrazi o tem svoje mnenje

if a PERSON gives a comment on an EVENT, SITUATION or THEME, they express an opinion on it

(b) dovzeten (susceptible, receptive, prone)

če je ČLOVEK dovzeten za nek VPLIV, IDEJO ali SPREMEMBO, jo je pripravljen sprejeti in uresničiti, navadno zato ker ustreza njegovim mnenjem ali osebnostnim lastnostim

if a PERSON is susceptible to INFLUENCE, an IDEA or CHANGE, they are ready to accept it or realize it, usually because it conforms to their views or character

(c) naliti/natočiti komu čistega vina (*to pour pure wine to somebody)

če rečemo, da smo komu nalili čistega vina, želimo poudariti, da smo mu povedali resnico ali dejstva o kaki neprijetni stvari ali dogodkih

if we say that we poured pure wine to somebody, we want to emphasize that we told them the truth or facts about an unpleasant thing or event

(d) cepec (idiot, nitwit, schmuck)

če rečemo komu cepec, ga imamo za neumnega in nepremišljenega

if we call somebody an idiot we consider them as stupid and simple-minded

The next type of information related to the description of meaning in SLD is the label. In the form of a condensed explanation it clarifies how a particular sense of a word, compound, phraseological unit or its variant is to be understood or used in a text, with regard to its grammatical or syntactic restrictions (grammatical labels), its use within a particular professional domain (field labels), what relation to the semantic value or participants in communication process it expresses (connotative, pragmatic labels), its limitations in relation to the time period (temporal labels), text type, regional or cultural characteristics (contextual labels). Labels were designed with the option in mind that they can be visualized in various digital media either in the form of condensed warnings or as parts of definitions (cf. Krek et al. 2013: 35). The main difference between the use of labels in traditional dictionaries, such as SSKJ, and in SLD is in their different role: in SLD vocabulary is not described with the aim to cultivate the users from the point of view of (literary) language norm. Labels are used to describe semantic properties of vocabulary as identified in empirically verifiable communicative acts found in corpora, with the aim to inform the user about these properties.

Multi-word expressions that need semantic description are included in the semantic level, based on lexicographers' assessment. In SLD they are formally separated from multi-word expressions that do not need a special semantic description, such as collocations and standard syntactic structures or patterns.

Compounds as independent lexical units are identified on the basis of the compositionality or non-compositionality of the meaning of their components. This is a shared feature with phraseology, but contrary to PUs the meaning of compounds as a whole is not metaphorical and does not have expressive or evaluative function. The first group includes compounds such as spalna srajca (chemise), glasbeni stolp (hi-fi system), osnovna šola (primary school), bančna kartica (bank card), and in the second group there are combinations whose semantic description includes

explanation of connotative and pragmatic elements, such as *princ na belem konju* (prince charming), *naliti komu čistega vina* ('to pour pure wine to somebody'), *biti krvav pod kožo* ('to have blood under the skin') etc. Since separating compounds from phraseology on one hand, and compounds from free combinations and collocations on the other, is a difficult task, with no clear boundaries (cf. Rundell in Atkins 2011), as the identification of (non-)compositionality and categorization is connected with the interpretation of context and predispositions of end users (whether they are school children, foreign language speakers, adult native speakers etc.), the structure allows to list compounds under a particular sense of one of its components or independently of all registered senses of the headword:

mreža *samostalnik* (**net, network**)

1 preplet niti, žic (mash of strand, wire)

1.2 predmet (object)

viseča mreža (hammock)

naprava za ležanje iz trpežne tkanine in vrvi, ki se napne med dve drevesi ali stojali

2 o ribolovu (fishing)

vlečna mreža (trawl)

stoječa mreža (trammel)

5 vzorec (pattern)

koordinatna mreža (grid map)

geografija (geography)

geometrija (geometry)

nevronska mreža (**neural network**)

računalništvo (computer science)

kristalna mreža (**crystal lattice**)

kemija (chemistry)

matna mreža (**mating net**)

šah (chess)

This treatment of compounds in SLD allows for different visualizations on the web or in other media. If one of the components is semantically dependent on one of its single-(head)word senses, users can be offered information about this relation. To enable linking between different elements of SLD, it is possible to include compounds also as collocations under particular senses, in addition to their inclusion as separate lexical entries:

šola *samostalnik* (school)

1 o izobraževanju (of education)

1.4 proces (process)

[osnovna, srednja, višja, visoka, poklicna] šola

[primary, secondary, college, high, vocational] school

osnovna šola (primary school)

vzgojno-izobraževalni proces, ki traja 8 oz. 9 let in je obvezen

obligatory education process lasting 8 or 9 years

In SLD, phraseological units (PU) are combinations of words that demonstrate non-compositional meaning, independent of the meaning of its parts. In texts, phraseology can also be identified due to its tendency to be expressive and conspicuous. In terms of their composition, PUs are more complex than compounds, as their structure can vary from different types of phrases to subordinate or coordinate clauses. The first group includes PUs structured as nominal, adjectival, verbal and adverbial phrases, e.g. *star maček* ('old cat' – experienced person), *vonj po denarju* ('smell of money' – tendency to prefer material goods, prestige) ; *blažen med ženami* ('blessed among women' – of a man, be regularly found in the company of women); *iti kot po maslu* ('to go like on the butter' – to proceed smoothly), *jasno in glasno* ('loudly and clearly'). The second group includes PUs that are semantically dependent on the context and usually form frozen expressions such as *korajža velja!* ('courage!'), *odraستي že (enkrat)!* ('grow up!'), *daj/dajte (no) mir!* ('come on!', 'give it a rest'), which often can be found in spoken discourse. PUs typically show a high degree of variability, both in terms of their phrasal components and their morphological forms, and in their syntactic arguments. In SLD these variants were recorded under the same element if no semantic changes were noticeable and if variants were frequent:

rešilna bilka | videti v čem rešilno bilko | ponuditi komu rešilno bilko | oprijeti se česa kot rešilne bilke | oprijeti se/okleniti se edine/zadnje/vsake rešilne bilke

(to clutch at straws, to be one's last resort)

Collocation level

Since the time when it was first possible to study lexical behavior of words in their context on a large scale in text corpora with the help of statistical tools, collocations are an indispensable part of lexicographic description of words. The view on collocations in SLD comes close to the one set up by Firth (1957), and later recapitulated by others in their lexical and communicational studies, such as Halliday

(1966) and Sinclair (1987). In these, collocations are defined as combinations of two or more words that are used in to convey meaning in real language in a conventional, regular manner. This means that combinations have to be statistically identifiable and semantically meaningful. Which is also the reason that in SLD collocations are not considered as independent lexical units, but they do have an important role in the identifications of senses and their description. There are two types of collocations registered in SLD: regular or binary collocations and extended collocations. The basic type is represented by a combination of two content words which can incorporate prepositions as the third element:

[orožarska, korupcijska, podkupovalna] afera
[monitions, corruption, bribery] affair

[šotor, bivališče] nomadov; [otvoritev, odprtje] razstave
[tent, residence] *of nomads; [opening, inauguration] *of an exhibition

zastava [plapola, vihra, se dviga], zbor [obravnavo, razpravlja, zaseda]
a flag [waves, billows, is hoisted], a council [handles, discusses, is held]

[opravičilo, krivec] za grozote; [raj, poslastica] za gurmane
[excuse, cause] for/of atrocities; [heaven, feast] for gourmets

etika v [znanosti, zdravstvu, športu]; [vadba, telovadba, razgibavanje] v fitnessu
ethics in [science, health, sports]; [drill, training, exercise] in a fitness centre

[obračunavanje, žuganje] s pestmi; ogrlica z [diamantom, biseri] [fighting, shaking] with fists; neckless with [diamonds, pearls]

In the process of creating SLD word sketch analyses revealed that some combinations often need also the third element. This element typically functions as (a) a modifier of one of the collocation components, or (b) it can be an indispensable element of a fixed combination:

(a) [ohraniti, ohranjati] tradicijo -> ohranjati [družinsko] tradicijo
[keep] tradition - > keep [family] tradition

[tekma, tekmovanje] smučarjev -> tekmovanje smučarjev [tekačev, skakačev]
[race, competition] *of skiers - > race, competition *of skiers [runners, jumpers]

(b) gostiti [svetovno] smetano
to host the [upper] crust

obiskovati v [prostem] času
to visit in [free] time

Syntactic level

In SLD, information on syntax is based on the premise that meaning is directly connected with syntax and that it can be realized in the text in different syntactic structures. In relation to a particular lexical unit, description of syntactic information includes the formal notification of syntactic structures, description of grammatical restrictions and preferences, and in the case of verbal headwords also their valency patterns.

Syntactic components of meaning – those originating from syntagmatic relations – are recorded in SLD as syntactic structures in four included parts-of-speech entries: nouns, verbs, adjectives and adverbs. Syntactic structures are directly connected with gramrels (grammatical relations) as defined in the formal grammar of the word sketches module in the Sketch Engine tool. The tool enables collocation and other data to be automatically extracted from the corpus, therefore it was crucial that the identification of syntactic structures was based on previous analyses. Each structure recorded in SLD has its own gramrel in the sketch grammar and each gramrel in the sketch grammar produces data for one or several syntactic structures in the database. The version of sketch grammar (Krek 2012a) that was used for creating SLD includes 103 gramrels generating data for about 370 different syntactic structures. These are divided into s.c. special structures which include structures with auxiliary verbs, infinitives, negation elements and quantifiers, and regular structures, with and without prepositions.

In lexical entries with verbs as headwords, sentence patterns have a similar role as syntactic structures in the case of nouns, adjectives and adverbs. In the context of SLD, they are understood as formal representations of valency patterns in each particular sense of the verb. Sentence patterns serve to confirm argument structures recorded in the sense frame. At the same time, they show possible use of ellipsis, and express typical choices among various syntactic possibilities, which is most prominently shown in the case of prepositional phrases in various valency roles.

pritrđiti *glagol* (to attach)

trdno namestiti na določen predmet, mesto (to attach firmly to an object or place)

če ČLOVEK pritrđi nek PREDMET na DRUG PREDMET, ga z njim poveže v trdno celoto, navadno s PRIPOMOČKOM ali uporabo SNOVI

if a PERSON attaches an OBJECT on ANOTHER OBJECT, they make a firm connection between them, usually using a DEVICE or some MATERIAL

kdo pritrđi kaj
somebody attaches something
kdo pritrđi kaj na kaj
somebody attaches something on something
kdo pritrđi koga na kaj

somebody attaches somebody on something
 kdo pritrdi kaj na koga
 somebody attaches something on somebody
 kdo pritrdi kaj nad kaj
 somebody attaches something above something
 kdo pritrdi kaj ob kaj
 somebody attaches something next to something
 kdo pritrdi kaj v kaj
 somebody attaches something in something
 kdo pritrdi kaj pod kaj
 somebody attaches something under something
 kdo pritrdi kaj med kaj
 somebody attaches something between something
 kdo pritrdi kaj čez kaj
 somebody attaches something over something
 kdo pritrdi kaj okoli česa
 somebody attaches something around something
 kdo pritrdi kaj za kaj
 somebody attaches something behind something
 kdo pritrdi kaj s čim
 somebody attaches something with something

Grammatical restrictions and preferences were introduced in SLD as a special type of grammatical or syntactic labels attached to syntactic structures or sentence patterns. They are intended to mark special cases within predictable morphological paradigms, syntactic roles, etc. This is useful both for automatic detection of exceptions in NLP tasks, and for the purposes of grammatical or syntactic analyses:

Kol-sbz0 SBZ2<r>v množini</r> (in plural)
 [čreda] antilop; [kopica, kup, vrsta] nasvetov; [škaf, komplet] barvic
 pbz0<r>primernik</r> SBZ0 (comparative form)
 [višja, nižja] najemnina
 sbz1<r>v pridevniški rabi</r> SBZ0 (in attributive use)
 [DOS] aplikacija; čarter [plovilo, jadrnica]; čili [paprika, omaka]; džez [scena, festival]
 SBZ0 v SBZ5<r> navadno z lastnim imenom</r> v sbz5 (usually with proper nouns)
 intervencija v [Iraku, Vietnamu, ZRJ, Somaliji, Afganistanu, Bosni]

In addition to syntactic structures, elements in the syntactic level also include syntactic combinations, statistically salient multi-word expressions that show structural fixedness, and on the other hand – from the point of view of native speakers of Slovene – do not show independent semantic value. Following from that they do not require semantic description and are placed under particular (sub-)senses of relevant headwords:

žlica *samostalnik*

zmešati kaj z/s [x] žlicami česa
to mix sth with [x] spoons of sth

angažirati *glagol*

angažirati se na področju česa
to be active on a field of sth

Exemplification level

Examples of use are an important if not crucial part of dictionary information, therefore they also have an important role in SLD. Their basic function is to confirm or exemplify all language facts mentioned in other levels in SLD. However, choice of examples is different in lexical databases and in dictionaries. Conceptually, a lexical database requires a larger number of corpus examples, typically these are whole sentences taken directly from the corpus. They represent candidates for good dictionary examples. As Atkins and Rundell mention (2008: 458), good examples are natural, typical, informative and intelligible. Political or social correctness can be added to these principles, as this is an important feature of dictionaries representing authority in the society (Gorjanc 2013/14: 12).

Search for and selection of good examples is one of the most time-consuming lexicographic tasks. Therefore a lot of effort was dedicated to the automation of this task during the compilation of SLD. With its GDEX and Tickbox Lexicography functions, the Sketch Engine tool represents a good starting point for upgrading the initial parameters to fit to the characteristics of Slovene (Kosem et al. 2011). With general parameters such as the preferred length of sentences, length of words, blacklisting rare words, non-textual elements etc. it is possible to acquire a sufficient amounts of good corpus examples. However, as a language-specific parameter the position of the headword in the sentence turned out to be a good measure for Slovene, as well as the inclusion of the third collocate in the parameters – a typical element in the wider context of the headword. The automation process was also recognized in the European lexicographic community, in particular in the European Network of e-Lexicography, as it was possible to adapt the parameters for different languages with good results.

In Slovene lexicography, the structure and the typology of lexical data included in SLD is a novelty in several aspects. The concept is corpus-driven, which means that the entire lexical description is empirically grounded on data from modern Slovene texts. The new description enables an insight into what is natural in language, but it also exposes idiosyncrasies which potentially represent language problems for users. The sample of 2,500 entries in the SLD also offers a new perspective on the organization and hierarchy of data, conceptualized with a view of its use in NLP from the beginning, and in the way that different visualizations of data in various digital media are possible, primarily driven by user needs and user-friendliness.

Literatura

- AHLIN, Martin, LAZAR, Branka, PRAZNIK, Zvonka, SNOJ, Jerica, 2014: Slovar slovenskega knjižnega jezika. Druga, dopolnjena in deloma prenovljena izdaja. Izdali Slovenska akademija znanosti in umetnosti, Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti, Inštitut za slovenski jezik Frana Ramovša. Ljubljana: Cankarjeva založba, 2014. 1. knjiga 1152 str., 2. knjiga 1150 str. *Jezik in slovstvo* 59/4. 121–127.
- APRESJAN, Juri, D., 2002: Principles of Systematic Lexicography. Corr  ard (ur.): *Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins*. UK: Euralex. 91–104.
- APRESJAN, Juri, D., 1992: Systemic Lexicography. Tommola et al. (ur.): *EURALEX'92. Proceedings*. Tampere: University of Tampere. 3–16.
- APRESJAN, Juri, D., 1973: Regular Polysemy. *Linguistics* 142. 5–39.
- ARHAR Špela, 2009: U ni korpus SSJ in leksikon besednih oblik za sloven  ino. *Jezik in slovstvo* 54/3–4. 43–56.
- ARHAR, Špela, HOLOZAN, Peter, 2009: ASES – leksikalna podatkovna zbirka za razvoj slovenskih jezikovnih tehnologij. Mikoli , Vesna (ur.): *Jezikovni korpusi v medkulturni komunikaciji*. Koper: Znanstveno-raziskovalno sredi  e, Zalo ba Annales. 30–51.
- ARHAR, Špela, GORJANC, Vojko, 2007: Korpus FidaPLUS: Nova generacija slovenskega referen nega korpusa. *Jezik in slovstvo* 52/2. 95–110.
- ARHAR HOLDT, Špela, 2012: Pedago ka korpusna slovnica med Scilo in Karibdo – jezikovna norma in korpusni podatki. *Predavanje na posvetu: Pravopis ve  kot le predpis. Fakulteta za dru bene vede, 15. februar 2012*. http://videlectures.net/pravopis2012_arhar_holdt_slovnica/.
- ARHAR HOLDT, Špela, KOSEM, Iztok, KRAP  VODOPIVEC, Irena, LEDINEK, Nina, MO E, Sara, STRITAR KU UK, Mojca, SVEN SEK, Tanja, ZWITTER VITEZ, Ana, 2012: *Pedago ka slovnica pri projektu Sporazumevanje v slovenskem jeziku: kazalnik 16 – Standard za korpusno analizo slovnini nih pojavov*. Ljubljana: Ministrstvo za  olstvo in  port: Amebis. Dostopno na: http://www.slovenscina.eu/Media/Kazalniki/Kazalnik16/Kazalnik_16_Pedagoska_slovnica_SSJ.pdf
- ARHAR HOLDT, Špela, 2011: *Lu  enje besednih zvez iz besedilnega korpusa z uporabo dvodelnih in tridelnih oblikoskladenjskih vzorcev*. Zbirka Trojinski konj. Ljubljana: Trojina, zavod za uporabno slovenistiko.

- ATKINS, Sue, B. T., KILGARRIFF, Adam, RUNDELL, Michael, 2010: Database of ANalysed Texts of English (DANTE): the NEID database project. A. Dykstra in T. Schoonheim (ur.): *Proceedings of the XIV. Euralex International Congress (Leeuwarden, 6–10 July 2010)*. Ljouwert: Fryske Akademy. 549–556.
- ATKINS, Sue, B. T. in RUNDELL, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- ATKINS, Sue, B. T., 1992/93: Theoretical Lexicography and its relation to Dictionary-making. Dictionaries. *Journal of the Dictionary Society of North America* 14. 4–43.
- BAJEC, Anton, 1955/56: Vezanje več osebkov s povedkom. *Jezik in slovstvo* 1/1. 12–14.
- BARNBROOK, Geoff, 2002: *Defining language: A local grammar of definition sentences*. Studies in Corpus Linguistics. Amsterdam: John Benjamins Publishing Company.
- BARONI, Marco, KILGARRIFF, Adam, 2006: Large linguistically-processed Web corpora for multiple languages. *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics*. 87–90.
- BÉJOINT Henry, 2000: *Modern Lexicography. An Introduction*. Oxford: Oxford University Press.
- BÉJOINT Henry, 1988: Monosemy and the Dictionary. T. Magay in J. Zigány (ur.): *BudaLEX '88 Proceedings: Papers from the EURALEX Third International Congress, Budapest, 4–9 September 1988*. Budapest: Akadémiai Kiadó. 13–26.
- BENSON, Morton, BENSON, Evelyn, ILSO, Robert, 1986: *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- BIBER, Douglas, CONRAD, Susan, CORTES, Viviana, 2003: Lexical Bundles in Speech and Writing: An Initial Taxonomy. Wilson, A., Rayson, P., McEnery, T. (ur.): *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*. Frankfurt/Main: Peter Lang. 71–92.
- BIZJAK KONČAR, Aleksandra, DOBROVOLJC, Helena, DOBROVOLJC, Kaja, LOGAR, Nataša, KOCJANČIČ, Polonca, KREK, Simon, ROZMAN, Tadeja, 2011: *Slogovni priročnik: sporazumevanje v slovenskem jeziku: kazalnik 17 – Standard za korpusno analizo težav pri tvorbi besedil*. Ljubljana: Ministrstvo za šolstvo in šport: Amebis. Dostopno na: http://www.slovenscina.eu/Media/Kazalniki/Kazalnik17/Kazalnik_17_Slogovni_prirocnik_SSJ.pdf

- BOLINGER, Dwight, 1965: The Atomization of Meaning. *Language* 41. 555–573.
- BORIN Lars, FORSBERG, Markus, ROXENDAL, Johan, 2012: Korp – the corpus infrastructure of Språkbanken. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. 474–478.
- BORIN, Lars, FORSBERG, Markus, OLSSON, Leif-Jöran, UPPSTRÖM, Jonatan, 2012a: *The open lexical infrastructure of Språkbanken. Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. 3598–3602.
- CCED 2001³: *Collins Cobuild English Dictionary for advanced Learners*. Glasgow: Harper Collins Publishers.
- CLEAR, Jeremy, 1987: Overview of the Role of Computing in Cobuild. Sinclair, J. M. (ur.): *Looking up: An Account of the COBUILD Project in Lexical Computing*. London: Collins.
- COULMAS, Florian, 1994: Formulaic language. Asher, R. E. (ur.): *Encyclopedia of Language and Linguistics*. Oxford: Pergamon. 1292–1293.
- COWIE Anthony, P. 1998: *Phraseology: Theory, Analysis, and Applications. Introduction*. Cowie, A. P. (ur.). Oxford: Oxford University Press. 1–20.
- COWIE, Anthony, P., 1981: The treatment of collocations and idioms in learners' dictionaries. Cowie, A. P. (ur.): *Lexicography and its Pedagogical Applications. Applied Linguistics* 2/3 (thematic issue). 223–235.
- CRYSTAL, David, 1985: *A Dictionary of Linguistics and Phonetics*. Oxford: Basil Blackwell.
- CSD 2009⁴: *Collins School Dictionary*. London: Collins.
- ČERMÁK, František, 2010: *Lexikon a sémantika*. Praha: Nakladatelství Lidové Noviny.
- ČERMÁK, František, 2009: Leksikografovi zapiski o korpusnem slovarju. *Jezik in slovstvo* 54/3–4. 25–42.
- ČERMÁK, František, 2001: Substance of idioms: perennial problems, lack of data or theory? *International Journal of Lexicography* 14/1. Oxford: Oxford University Press. 1–20.
- ČERMÁK, František, 1985: Frazologie a idiomatika. František Čermák, Josef Filipec: *Česká lexikologie*. Praha: Academia. 166–248.
- ČERMÁK, František, HOLUB, Jan, 1987: *Syntagmatika a paradigmatica českého slova. I. Valence a kolokabilita*. Praha.

- ČERMÁK, František, 1982: *Idiomatika a frazeologie češtiny*. Praha, UK.
- ČERMÁKOVÁ, Anna, 2009: *Valence českých substantiv*. Praga: Nakladatelství Lidové novin.
- ČERNELIČ-KOZLEVČAR, Ivanka, 1975: O funkciji glagolov z oslavljenim pomenom tipa *biti*. *Jezik in slovstvo* 21/3. 76–81.
- DANEŠ, František, HLAVSA, Zdeněk, 1987: *Větné vzorce v češtině*. Praha: Academia.
- DOBROVOLJC, Helena, 2014: Normativna informacija v slovarju. Bergoč, S. Grahek, I. (ur.): *Novi slovar za 21. stoletje. E-zbornik prispevkov s Posveta o novem slovarju slovenskega jezika na Ministrstvu za kulturo 12. 2. 2014, Ljubljana, april 2014*. Ljubljana: Ministrstvo za kulturo. Dostopno na: http://www.mk.gov.si/si/delovna_podrocja/sluzba_za_slovenski_jezik/predstavitev_podrocja/dogodki/posvet_o_novem_slovarju_slovenskega_jezika/.
- DOBROVOLJC, Helena, JAKOP, Nataša (ur.), 2011: *Sodobni pravopisni priročnik med normo in predpisom*. Ljubljana: Založba ZRC, ZRC SAZU.
- DOBROVOLJC, Helena, KREK, Simon, 2011: Normativne zadrege – empirični pristop. Kranjc, Simona (ur.): *Meddisciplinarnost v slovenistiki (Obdobja 30)*. Ljubljana: Znanstvena založba Filozofske fakultete. 89–97.
- DZIEMIANKO, Anna, LEW, Robert, 2013: When-definitions revisited. *International Journal of Lexicography* 26/2. 154–175.
- ERJAVEC, Tomaž, 2014: Odprt dostop do podatkovne baze slovarja. Bergoč, S. Grahek, I. (ur.): *Novi slovar za 21. stoletje E-zbornik prispevkov s Posveta o novem slovarju slovenskega jezika na Ministrstvu za kulturo 12. 2. 2014, Ljubljana, april 2014*. Ljubljana: Ministrstvo za kulturo. Dostopno na: http://www.mk.gov.si/fileadmin/mk.gov.si/pageuploads/Ministrstvo/slovenski_jezik/E_zbornik/20-_Tomasz_Erjavec-SlovarPosvet.pdf.
- ERJAVEC, Tomaž, HOLOZAN, Peter, KREK, Simon, PIVEC, Matej, RIGAČ, Simon, ROZMAN, Simon, VELUŠČEK Aleš, 2008: *Specifikacije za leksikon besednih oblik (kazalnik 3). Projekt Sporazumevanje v slovenskem jeziku*. Ljubljana: Ministrstvo za šolstvo in šport: Amebis. Dostopno na: <http://projekt.slovenscina.eu/Vsebine/SI/Kazalniki/K3.aspx>.
- ERLANDSEN, Jens, 2004: iLex – new DWS. *Third International Workshop on Dictionary Writing systems (DWS 2004)*. Brno, 6. – 7. September 2004.
- FELLBAUM, Christiane (ur.), 1998: *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

- FILIPEC, Josef, ČERMÁK, František, 1985: *Česká Lexikologie*. Praha: Academia.
- FILLMORE, Charles, J., JOHNSON, Christopher, R., PETRUCK, Miriam, R. L., 2003: Background to Framenet. *International Journal of Lexicography* 16/3. 235–250.
- FILLMORE, Charles, J., 1976: Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech* 280. 20–32.
- FIRTH, John, R., 1957: Modes of Meaning. *Papers in Linguistics*. London: Oxford University Press. 1934–51.
- FIŠER, Darja, GANTAR, Polona, KREK, Simon, 2012: Using explicitly and implicitly encoded semantic relations to map Slovene wordnet and Slovene lexical database. *8th International Conference on Language Resources and Evaluation, 21–27 May 2012, Istanbul, Turkey. LREC 2012: proceedings (Workshops: Semantic relations II)*. Istanbul: ELRA, 2012. 77–84.
- FIŠER, Darja, NOVAK, Jernej, 2011: Visualizing sloWNet. Kosem, I., Kosem, K. (ur.): *Proceedings of eLex 2011, Bled, 10–12 November 2011: Electronic Lexicography in the 21st Century – New Applications for New Users*. 76–82.
- FIŠER, Darja, ERJAVEC, Tomaž, 2010: sloWNet: construction and corpus annotation. *Proceedings of the 5th Global WordNet Conference, GWC 2010, January 31 – February 4, 2010, Mumbai, India*. 177–183.
- FIŠER, Darja, VINTAR, Špela, 2010: Uporaba wordneta za boljše razdvoumljanje pri strojnem prevajanju. Erjavec, T., Žganec Gros, J. (ur.): *Zbornik 13. mednarodne multikonference INFORMACIJSKA DRUŽBA – IS 2010 Zvezek C*. 53–57
- FIŠER, Darja, 2009: sloWNET – slovenski semantični leksikon. Stabej, M. (ur.): *Infrastruktura slovenščine in slovenistike, (Obdobja, 28)*. Ljubljana: Znanstvena založba Filozofske fakultete, 2009. 145–149.
- FIŠER, Darja, 2009a: Leveraging parallel corpora and existing wordnets for automatic construction of the Slovene wordnet. *Human language technology: challenges of the information society, (LNCS 5603)*. Berlin, Heidelberg: Springer. 359–368.
- FIŠER, Darja, 2005: Pristopi k izdelavi leksikalnih podatkovnih zbirk. *Jezik in slovstvo* 50/6. 17–32.
- FN 2003: Framenet and Frame Semantics. *International Journal of Lexicography* 16/3. Special Issue.

- FRANKENBERG-GARCIA, Anna, 2014: The Use of Corpus Examples for Language Comprehension and Production. *Special Issue on Researching Uses of Corpora for Language Teaching and Learning*. ReCALL, 26. 128–146.
- FRATH, Pierre, 2000: Polysemy, Homonymy and Reference. A. Hamm (ur.): *Proceedings of the JASGIL Seminar, Strasbourg 5–6 May 2000, 34/2001*. Strasbourg: RANAM (Recherches Anglaises et Nord-Américaines). 43–56.
- GANGEMI, Aldo, GUARINO, Nicola, OLTRAMARI, Alessandro, 2001: Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-Level. *Proceedings of the International Conference on Formal Ontology in Information Systems FOIS 2001, Ogunquit, Maine, October 17–19, 2001*. 285–296.
- GANTAR, Polona, 2014: Moč in nemoč korpusnega pristopa k analizi pomenov. Bergoč, S. Grahek, I. (ur.): *Novi slovar za 21. stoletje E-zbornik prispevkov s Posveta o novem slovarju slovenskega jezika na Ministrstvu za kulturo 12. 2. 2014, Ljubljana, april 2014*. Ljubljana: Ministrstvo za kulturo. Dostopno na: http://www.mk.gov.si/si/delovna_podrocja/sluzba_za_slovenski_jezik/predstavitev_podrocja/dogodki/posvet_o_novem_slovarju_slovenskega_jezika/.
- GANTAR, Polona, KOSEM, Iztok, 2013: Beleženje in prikazovanje podatkov o jezikovni rabi: od leksikalne baze do spletnega slovarja. Žele, A. (ur.): *Družbena funkcijskost jezika: vidiki, merila, opredelitve. (Obdobja 32)*. Ljubljana: Znanstvena založba Filozofske fakultete. 133–139.
- GANTAR, Polona, 2011: Leksikalna baza za slovenščino: komu, zakaj in kako (naprej)? *Jezikoslovni zapiski* 17/2. 77–92.
- GANTAR, Polona, GRABNAR, Katja, KOCJANČIČ, Polona, KREK, Simon, POBIRK, Olga, REJC, Rok, ŠORLI, Mojca, ŠUSTER, Simon, ZARANŠEK, Petra, 2009: *Specifikacije za izdelavo leksikalne baze za slovenščino: standard za izdelavo posamezne leksikalne enote v leksikalni bazi. (Kazalnik 6). Projekt Sporazumevanje v slovenskem jeziku*. Ljubljana: Ministrstvo za šolstvo in šport: Amebis. Dostopno na: http://projekt.slovenscina.eu/Media/Kazalniki/Kazalnik6/SSJ_Kazalnik_6_Specifikacije-leksikalna-baza_v1.pdf.
- GANTAR, Polona, GRABNAR, Katja, KOCJANČIČ, Polona, KREK, Simon, POBIRK, Olga, REJC, Rok, ŠORLI, Mojca, ŠUSTER, Simon, ZARANŠEK, Petra, 2009a: Specifikacije za izdelavo leksikalne baze za slovenščino: opis analize referenčnega korpusa. *(Kazalnik 5). Projekt Sporazumevanje v slovenskem jeziku*. Ljubljana: Ministrstvo za šolstvo

- in šport: Amebis. Dostopno na: http://projekt.slovenscina.eu/Media/Kazalniki/Kazalnik5/SSJ_Kazalnik_5_Specifikacije-opis-analize-korpusa_v1.pdf.
- GANTAR, Polona, 2008: (Slovenska) leksika med leksikonom in slovnico. *Jezik in slovnstvo* 53/5. 19–35.
- GANTAR, Polona, 2007: *Stalne besedne zveze v slovenščini: korpusni pristop*. Lingua Slovenica. Ljubljana: Založba ZRC, ZRC SAZU.
- GORJANC, Vojko, 2013/2014: O heteronormativnosti slovarskega opisa slovenskega jezika: homoseksualnost, ekshibicionizem in druge perversnosti. *Narobe* 7, 27/28. 12–15.
- GORJANC, Vojko, 2012: Ideologija heteronormativnosti, prevodna in jezikovna norma. Bjelčevič, A. (ur.): *Ideologije v slovenskem jeziku, literaturi in kulturi: zbornik predavanj*. Ljubljana: Znanstvena založba Filozofske fakultete. 38–44.
- GORJANC, Vojko, KREK, Simon, GANTAR, Polona, 2005: Slovenska leksikalna podatkovna zbirka. *Jezik in slovnstvo* 50/2. 3–19.
- GRABNAR, Katja, ŠORLI, Mojca, 2003: Novi veliki angleško-slovenski slovar Oxford-DZS. *Jezik in slovnstvo* 48/3–4. 126–133.
- GRØNVIK, Oddrun, ORE, Christian-Emil Smith, 2013: What should the electronic dictionary do for you – and how?. Iztok Kosem, I., Kallas, J., Gantar, P., Krek, S., Langements, M., Tuulik, M. (ur.): *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut. 243–260.
- HALLIDAY, Michael, A. K., 1966: Lexis as a linguistic level. M. A. K. Halliday, C. E. Bazell, J. C. Catford, R. H. Robins (ur.): *In Memory of J. R. Firth*. London: Longman.
- HANKS, Patrick, 2013: *Lexical Analysis: Norms and Exploitations*. Cambridge: MIT Press.
- HANKS, Patrick, 2010: Elliptical Arguments: a Problem in relating Meaning to Use. S. Granger, M. Paquot (ur.): *eLexicography in the 21st century: New challenges, new applications. Proceedings of ELEX2009. Cahiers du CENTAL*. Louvain-la-Neuve: Presses universitaires de Louvain. 109–125.
- HANKS, Patrick, 2009: Sestavljanje enojezičnega slovarja za domače govorce. *Jezik in slovnstvo* 54/ 3–4. 7–24.
- HANKS, Patrick, 2008: *Mapping meaning onto use: a Pattern Dictionary of English Verbs*. AACL 2008, Utah.

- HANKS, Patrick, 2006: Metaphorocity is gradable. Stefanowitsch, A., Gries, S. Th. (eds.): *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin, New York: Mouton de Gruyter. 17–34.
- HANKS, Patrick, 2005: Definition. K. Brown (ur.): *Encyclopedia of Language and Linguistics*. (Second edition.) Oxford: Elsevier. 399–402.
- HANKS, Patrick, PUSTEJOVSKY, James, 2005: A Pattern Dictionary for Natural Language Processing. *Revue Française de Linguistique Appliquée* 10/2. 63–82.
- HANKS, Patrick, 2004: Corpus Pattern Analysis. Williams, G., Vessier, S. (ur.): *EURALEX 2004 Proceedings*. Lorient: Université de Bretagne-Sud.
- HANKS, Patrick, PUSTEJOVSKY, James, 2004: Common Sense About Word Meaning. *Sense in Context*. TSD 004: Brno.
- HANKS, Patrick, 1994: Linguistic norms and pragmatic exploitations, or why lexicographers need prototype theory and vice versa. F. Keifer, G. Kiss, J. Pajzs (ur.): *Papers in Computational Lexicography. Complex '94*. Budapest: Hungarian Academy of Sciences.
- HANKS, Patrick, 1987: Definitions and explanations. J. M. Sinclair (ur.), *Looking up. An account of the COBUILD project*. London, Glasgow: Collins ELT. 116–136.
- HARTMANN, Reinhard, R. K. (ur.), 2003: *Lexicography Critical Concepts. Vol. III: Lexicography, Metalexicography and Reference Science*. London, New York: Routledge.
- HOČEVAR GREGORIČ, Mateja, 1995: Slovenski neštevni samostalniki in njihova obravnava v SSKJ. *Slavistična revija* 43/4. 414–458.
- JOHANSSON, Stig, 1998: Goran Kjellmer. A Dictionary of English Collocations, Based on the Brown Corpus. 3 vols. Oxford: Clarendon Press, 1994. *International journal of Corpus Linguistics* 3/2. 338–348.
- KATZ, Jerrold, J., FODOR, Jerry, A., 1963: The structure of a semantic theory. *Language* 39. 170–210.
- KILGARRIFF, Adam, HUSÁK, Miloš, McADAM, Katy, RUNDELL, Michael, RYCHLÝ, Pavel, 2008: GDEX: Automatically finding good dictionary examples in a corpus. E. Bernal, J. DeCesaris (ur.): *Proceedings of the 13th EURALEX International Congress*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. 425–432.
- KILGARRIFF, Adam, RYCHLÝ, Pavel, SMRZ, Pavel, TUGWELL, David, 2004: The Sketch Engine. G. Williams, S. Vessier (ur.): *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6–10, 2004*. Lorient: Université de Bretagne - sud. 105–116.

- KILGARRIFF Adam, RUNDELL, Michael, 2002: Lexical Profiling Software and its lexicographic applications: a case study. *Proceedings EURALEX 2002, Copenhagen*. 807–818.
- KJELLMER, Goran. 1994. *A Dictionary of English Collocations*. Oxford: Clarendon Press.
- KOPŘIVOVÁ, Marie, 2006: *Valence českých adjektiv*. Praga: Nakladatelství Lidové noviny.
- KOROŠEC, Tomo, 1985: *Slovenski jezik 4*. Maribor: Založba Obzorja.
- KOSEM, Iztok, 2011: Prihodnost leksikografije: dinamični slovar. Jesenšek, M. (ur.): *Izzivi sodobnega slovenskega slovaropisja*. Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti, Filozofska fakulteta. 38–48.
- KOSEM, Iztok, 2006: Definijski jezik v Slovarju slovenskega knjižnega jezika s stališča sodobnih leksikografskih načel. *Jezik in slovnost* 51/5. 25–45.
- KOSEM, Iztok, GANTAR, Polona, KREK, Simon, 2013: Automation of lexicographic work: an opportunity for both lexicographers and crowdsourcing. Kosem, I., Kallas, J., Gantar, P., Krek, S., Langemets, M., Tuulik, M. (ur.): *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut. 32–48.
- KOSEM, Iztok, GANTAR, Polona, KREK, Simon, 2012: Avtomatsko luščenje leksikalnih podatkov iz korpusa. Erjavec, T., Žganec Gros, J. (ur.): *Zbornik 15. mednarodne multikonference Informacijska družba – IS 2012, zvezek C*. Ljubljana: Institut Jožef Stefan. 117–122.
- KOSEM, Iztok, HUSÁK, Miloš, MCCARTHY, Diana, 2011: GDEX for Slovene. Kosem, I., Kosem K. (ur.): *Electronic Lexicography in the 21st Century: New applications for new users. Proceedings of eLex 2011, Bled, 10–12 November 2011*. Ljubljana: Trojina, zavod za uporabno slovenistiko. 151–159.
- KREK, Simon, 2014: Prva in druga izdaja SSKJ. *Slovenščina 2.0* 2/2. 114–160.
- KREK, Simon, KOSEM, Iztok, GANTAR, Polona, 2013: Predlog za izdelavo Slovarja sodobnega slovenskega jezika. Dostopno na: http://www.sssj.si/datoteke/Predlog_SSSJ_v1.1.pdf.
- KREK, Simon, 2012: *Slovenski jezik v digitalni dobi* (Zbirka Bela knjiga). Rehm, G., Uszkoreit, H. (ur.). Berlin, Heilderberg: Springer Verlag.
- KREK, Simon, 2012a: 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 2011: Koliko je v slovenskih korpusih besednih vrst, ampak čisto zares?. *Predavanje na posvetu: Slovnica, več kot le sistem. Fakulteta za družbene vede, 21. februar 2011*. Videlectures: http://videlectures.net/slovnica2011_krek_ksk/.
- KREK, Simon, 2010: Pridobivanje jezikovnih podatkov iz besedilnih korpusov za namen izdelave enojezičnih slovarjev in slovníc. *Doktorska disertacija*. Ljubljana: Filozofska fakulteta UL, Oddelek za slovenistiko.
- KREK, Simon, 2010a: O stvareh, ki jih ljudje ne izrečemo. *Predavanje na posvetu: Korpusi, več kot le statistika. Fakulteta za družbene vede, 5. februar 2010*. Videlectures: http://videlectures.net/korpusi2010_krek_sli/.
- KREK, Simon, 2009: Od SSKJ do spletnega portala standardne slovenščine. *Jezik in slovstvo* 54/ 3–4. 95–113.
- KREK, Simon, 2008: FrameNet in slovenščina. *Jezik in slovstvo* 53/5. 37–54.
- KREK, Simon, ŠORLI, Mojca, KOCJANČIČ, Polonca, 2008: The Funny Mirror of Language: the Process of Reversing the English-Slovenian Dictionary to Build the Framework for Compiling the New Slovenian-English Dictionary. E. Bernal, J. DeCesaris (eds.): *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)*. Barcelona: Documenta Universitaria: Institut universitari de lingüística aplicada, Universitat Pompeu Fabra.
- KREK, Simon in KILGARRIFF, Adam, 2006: Slovene Word Sketches. Erjavec, T., Žganec Gros, J. (ur.): *Proceedings of the 5th Slovenian and 1st International Languages Technology Conference. Ljubljana, Slovenia*. 62–67.
- KREK, Simon, 2004: Slovarji serije COBUILD in formalizacija definicijskega jezika. *Jezik in slovstvo* 49/2. 3–16.
- KRISHNAMURTHY, Ramesh, 2002: Language as chunks, not words. Swanson, M., Hill, K. (eds.): *JALT 2002: conference proceedings: Waves of the future*. JALT: Tokyo (JP). 288–294.
- KRŽIŠNIK, Erika, 1996: Norma v frazeologiji in odstopi od nje v besedilih. *Slavistična revija* 44/2. 133–154.
- KRŽIŠNIK, Erika, 1994: Slovenski glagolski frazemi (ob primeru glagolov govorjenja): *Doktorska disertacija*. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- KRŽIŠNIK, Erika, 1990: Tipologija frazeoloških prenovitev v Cankarjevih proznih besedilih. *Slavistična revija* 38/4. 399–420.

- KRŽIŠNIK, Erika, 1987: Prenovitve stalnih besednih zvez v Kosmačevi prozi 30. let. Zadavec, F. et al. (ur.): *Socialni realizem v slovenskem jeziku, književnosti in kulturi: mednarodni simpozij v Ljubljani od 26. do 28. junija 1985 (Obdobja 7)*. Ljubljana: Filozofska fakulteta. 517–532.
- LDOCE 2003⁵: *Longman Dictionary of Contemporary English*. Pearson Longman.
- LEW, Robert, De SCHRYVER, Gilles-Maurice, 2014: Dictionary users in the digital revolution. *International Journal of Lexicography*, 27/4. 1–19.
- LEW, Robert, DZIEMIANKO, Anna, 2006: A New Type of Folk-Inspired Definition in English Monolingual Learners' Dictionaries and its Usefulness for Conveying Syntactic Information. *International Journal of Lexicography*, 19/3. 225–242.
- LOGAR, Nataša, GRČAR, Miha, BRAKUS, Marko, ERJAVEC, Tomaž, ARHAR HOLDT, Špela, KREK, Simon, 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Zbirka Sporazumevanje. Ljubljana: Trojina, zavod za uporabno slovenistiko, Fakulteta za družbene vede.
- LOGAR, Nataša, ŠUSTER, Simon, 2009: Gradnja novega korpusa slovenščine. *Jezik in slovstvo* 54/ 3–4. 57–68.
- LÖNNEKER-RODMAN, Birte, BAKER, Collin, 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*. 147–154.
- LYONS, John, 1977: *Semantics*. Cambridge: Cambridge University Press.
- MANNING, Christopher, SCHÜTZE, Hinrich, 1999: *Foundations of Statistical Natural Language Processing*. MIT Press. Cambridge, MA.
- MARUŠIČ, Franc, NEVINS, Andrew, SAKSIDA, Amanda, 2007: Last-Conjunct Agreement in Slovenian. Compton, R., Goledzinowska, M., Savchenko, U. (ur.): *Annual Workshop on Formal Approaches to Slavic Linguistics: the Toronto meeting 2006*. Ann Arbor: Michigan Slavic Publications. 210–227.
- MED 2002, 2007²: *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Publishers Ltd.
- MOERDIJK, Fons, TIBERIUS, Carole, NIESTADT, Jan, 2008: Accessing the ANW Dictionary. *Proceedings of the Workshop on Cognitive Aspects of the Lexicon. COGALEX '08*. Manchester: Association for Computational Linguistics. 18–24.

- MOON, Rosamund, 2014: Meanings, Ideologies, and Learners' Dictionaries. A. Abel, Ch. Vettori, N. Ralli (ur.): *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15–19 July 2014, Bolzano/Bozen*. Institute for Specialised Communication and Multilingualism: Bolzano/Bozen. 85–105.
- MOON, Rosamund, 2004: On Specifying Metaphor: An Idea and its Implementation. *International Journal of Lexicography* 17/2. Oxford: Oxford University Press. 195–222.
- MOON, Rosamund, 1987: The Analysis of Meaning. J. M. Sinclair (ur.): *Looking up: An account of the COBUILD Project in Lexical Computing*. London, Glasgow: Collins ELT. 86–103.
- MOON Rosamund, 1987a: Monosemous words and the dictionary. A. P. Cowie (ur.): *The dictionary and the language learner. Papers from the Euralex Seminar at the University of Leeds. Lexicographica Series Major* 17. Niemeyer, Tübingen. 173–182.
- MOŽE, Sara, 2013: FrameNet in večjezičnost: kontrastivna analiza glagolov premikanja v slovenščini in angleščini. *Doktorska disertacija*. Univerza v Ljubljani, Filozofska fakulteta, Oddelek za anglistiko in amerikanistiko.
- MOŽE, Sara, 2009: Semantično označevanje korpusa slovenščine po modelu FrameNet. M. Stabej (ur.): *Obdobja 28: Infrastruktura slovenščine in slovenistike*. Znanstvena založba FF UL in Center za slovenščino kot drugi/tuji jezik. 265–269.
- MÜLLER, Jakob, 2009: Kritične misli in zamisli o SSKJ. Perdih, A. (ur.): *Strokovni posvet o novem slovarju slovenskega jezika, 23. in 24. oktober 2008*. Ljubljana: Založba ZRC, ZRC SAZU. 17–25.
- MÜLLER, Jakob, 1980: Pomenske skupine in pomenska zgradba besed (samostalnikov). *XVI. seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta. 35–47.
- 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-SPITZER, Carolin, KOPLENIG, Alexander, TÖPEL, Antje, 2011: What Makes a Good Online Dictionary? – Empirical Insights from an Interdisciplinary Research Project. Kosem, I., Kosem, K. (ur.): *Proceedings of eLex 2011, Bled, 10–12 November 2011: Electronic Lexicography in the 21st Century – New Applications for New Users*. 203–208.

- OOI, B. Y., Vincent 1998: *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.
- OREŠNIK, Janez, 1972: Formalizacija semantičnih definicij najmanjših jezikovnih enot s pomenom. *Problemi semantike, sintakse in obravnave tekstov*. Ljubljana: Inštitut Jožef Štefan. 27–33.
- PALA, Karel, ŠEVEČEK, Pavel, 1997: Valence českých sloves. *Sborník prací FFUB*. Brno. 41–54.
- PANEVOVÁ, Jarmila: 1994. Valency Frames and the Meaning of the Sentence. P. A. Luelsdorff (ur.): *The Prague School of Structural and Functional Linguistics*. John Benjamins Publishing Company. 223–243.
- PANEVOVÁ, Jarmila, 1974: On Verbal Frames in Functional Generative Description. *The Prague Bulletin of Mathematical Linguistics* 22. 3–40.
- POGORELEC, Breda, 1969: Zgradba stavka v prozi Ivana Cankarja. *V seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta Univerze v Ljubljani, Oddelek za slovanske jezike in književnosti.
- POGORELEC, Breda, 1968: Razvoj prostega stavka v slovenskem knjižnem jeziku. *Jezik in slovstvo* 8. 145–150.
- POLONČIČ RUPARČIČ, Andraž, 2012: Sprejemljivost ujemalnih in anaforičnih konstrukcij v pomenskih shemah leksikalne baze za slovenščino. *Diplomsko delo*. Filozofska fakulteta Univerze v Ljubljani: Oddelek za Slovenistiko.
- PRINSLOO, Daniel, J., HEID, Ulrich, BOTHMA, Theo, FAAß, Gertrud, 2011: Interactive, dynamic electronic dictionaries for text production. Kosem, I., Kosem, K. (ur.): *Proceedings of eLex 2011, Bled, 10–12 November 2011: Electronic Lexicography in the 21st Century – New Applications for New Users*. 215–220.
- PTASZNIK, Bartosz, LEW, Robert, 2014: Do menus provide added value to signposts in monolingual dictionary entries? *XVI EURALEX International Congress: The User in Focus, 15–19 July 2014, Bolzano/Bozen*.
- PUSTEJOVSKY, James, 1995: *The Generative Lexicon*. Massachusetts Institute of Technology: Cambridge, Massachusetts.
- ROMIH, Miro, HOLOZAN, Peter, 2002: Infrastruktura za razvoj jezikovnih tehnologij – korpus FIDA in sistem ASSES. Erjavec, T., Žganec Gros, J. (ur.): *Jezikovne tehnologije/Language Technologies. Zbornik konference/Proceedings of the Conference. 14. do 15. oktober 2002*. Ljubljana: Institut Jožef Štefan. 166.

- ROZMAN, Tadeja, 2010: Vloga enojezičnega razlagalnega slovarja slovenščine pri razvoju jezikovne zmožnosti. *Doktorska disertacija*. Ljubljana: Filozofska fakulteta UL.
- RUNDELL, Michael, ATKINS, Sue, 2011: The DANTE database: a User Guide. Kosem, I., Kosem, K. (ur.): *Proceedings of eLex 2011, Bled, 10–12 November 2011: Electronic Lexicography in the 21st Century – New Applications for New Users*. 233–246.
- RUNDELL, Michael, KILGARRIFF, Adam, 2011: Automating the creation of dictionaries: where will it all end? Meunier F., De Cock S., Gilquin G. in Paquot M. (ur.): *A Taste for Corpora. A tribute to Professor Sylviane Granger*. Benjamins. 257–281.
- RUNDELL, Michael, 2009: The Road to Automated Lexicography: First Banish the Drudgery then the Drudges? S. Granger, M. Paquot (ur.): *eLexicography in the 21 st century: New challenges, new applications*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université catholique de Louvain. 9–10.
- RUNDELL, Michael, 2006: More than one way to skin a cat: why full-sentence definitions have not been universally adopted. Corino, E., Marelllo C., Onesti, C. (ur.): *EURALEX 2006 Proceedings*. Torino: EURALEX. 323–337.
- RUNDELL, Michael, 1998: Recent trends in English pedagogical lexicography. *International Journal of Lexicography* 11/4. 315–342.
- RUNDELL, Michael, 1988: Changing the rules: why the monolingual learner's dictionary should move away from the native-speaker tradition. M. Snell-Hornby (ur.): *ZüriLEX '86 Proceedings*. Tübingen: Franke Verlag. 127–137.
- SANDFORD PEDERSEN, Bolette, 2012: Lexicography in Language Technology (LT). R. Vatvedt Fjeld, J. M. Torjusen (ur.): *Proceedings of the 15th EURALEX International Congress 7–11 August, 2012, Oslo*. Department of Linguistics and Scandinavian Studies, University of Oslo. 31–46.
- SIEPMAN, Dirk, 2006: Collocation, colligation and encoding dictionaries. Part II: Lexicographical Aspects. *International Journal of Lexicography* 19/1. 1–39.
- SIEPMAN, Dirk, 2005: Collocation, colligation and encoding dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography* 18/4. 409–444.
- SIMPSON, John, 2003: The production and Use of Occurrence Examples. *Van Sterkenburg 2003*. 260–272.
- SINCLAIR, John, 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- SINCLAIR, John, Mc., 1987: *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London and Glasgow: Collins ELT.
- SINCLAIR, John, Mc., 1987a: Collocation: a Progres Report. S. Threadgold (ur.): *Language Topics: Essays in Honour of Michael Halliday II*. Amsterdam, Philadelphia: John Benjamins. 319–331.
- SINCLAIR, Mc., J., COULTHARD, Malcolm, 1975: *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford: Oxford University Press.
- SNB 2012: *Slovar novejšega besedja slovenskega jezika*. Bizjak Končar, Aleksandra, Snoj, Marko (ur.). Ljubljana: Založba ZRC, ZRC SAZU.
- SNOJ, Jerica, 2010: *Metafora v leksikalnem sistemu*. Ljubljana: Založba ZRC, ZRC SAZU.
- SP 2001: *Slovenski pravopis*. Ljubljana: SAZU, ZRC SAZU, Založba ZRC.
- SSKJ: *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.
- SSKJ2: *Slovar slovenskega knjižnega jezika: druga izdaja* (2014/spletna različica: 2014). Ljubljana: ZRC SAZU. <http://www.sskj2.si>.
- STUBBS, Michael, 2007: Quantative data on multi-word sequences in English: the case of of the word *world*. Hoey, M., Mahlberg, M., Stubbs, M., Teubert, W. (ur.): *Text, Discourse and Corpora, theory and Analysis*. London, New York: Continuum. 163–190.
- SUMMERS, Della, 1987: *Introduction to LDOCE2*. Harlow: Longman.
- SVOZILOVÁ, Nada, PROUZOVÁ, Hana, JIRSOVÁ, Anna, 2005: *Slovník slovesných, substantivních a adjektivních vazeb a spojení*. Praha.
- ŠINDLEROVÁ, Jana, VESELOVSKÁ, Kateřina, HAJIČ, Jan, 2014: Tracing Sentiments: Syntactic and Semantic Features in a Subjectivity Lexicon. A. Abel, Ch. Vettori, N. Ralli (eds.): *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15–19 July 2014, Bolzano/Bozen. Institute for Specialised Communication and Multilingualism*. Bolzano/Bozen. 405–413.
- ŠORLI, Mojca, 2014: Sodobni sporazumevalni slovar slovenskega jezika: izhodišča, viri, izvedba. S. Bergoč, I. Grahek (ur.): *Novi slovar za 21. stoletje, E-zbornik prispevkov s Posveta o novem slovarju slovenskega jezika na Ministrstvu za kulturo 12. 2. 2014, Ljubljana, april 2014*. Ljubljana: Ministrstvo za kulturo.

- ŠORLI, Mojca, 2012: Semantična prozodija v teoriji in praksi – korpusni pristop k proučevanju pragmatičnega pomena: primer slovenščine in angleščine. M. Šorli (ur.): *Dvojezična korpusna leksikografija. Slovenščina v kontrastu: novi izzivi, novi obeti*. Ljubljana: Trojina, zavod za uporabno slovenistiko. 90–116.
- ŠORLI, Mojca, 2011: Pragmatic Components in the Slovene Lexical Database Descriptions. Kosem, I., Kosem K. (ur.): *Electronic lexicography in the 21st century: new applications for new users. Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ljubljana: Trojina, Institute for Applied Slovene Studies. 251–259.
- ŠORLI, Mojca, GRABNAR, Katarina, KREK, Simon, KOŠIR, Tomato, 2006: The Oxford-DZS Comprehensive English-Slovenian Dictionary. Corino, E., Marelllo, C., Onesti, C. (eds.): *Atti del XII congresso internazionale di lessicografia: proceedings XII EURALEX international congress, Torino, Italia September 6th-9th, 2006. Torino: Edizioni dell'Orso. Università di Torino: Accademia della Crusca*. 631–637.
- TAVČAR, Aleš, FIŠER, Darja in ERJAVEC, Tomaž, 2012: sloWCrowd: orodje za popraviljanje wordneta z izkoriščanjem moči množic. Erjavec, T., Žganec Gros, J. (ur.): *Zbornik 15. mednarodne multikonference Informacijska družba – IS 2012, zvezek C*. Ljubljana: Institut Jožef Stefan, 2012. 185–190
- TOPORIŠIČ, Jože, 2000: *Slovenska slovnica*. Maribor: Obzorja
- TOPORIŠIČ, Jože, 1981: Strukturalno pomenoslovje besed. *XVII. seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta. 95–112.
- TOPORIŠIČ, Jože, 1980: O strukturalnem določanju besednih pomenov (ob glagolu *biti*). *Linguistica*. 151–167.
- VATVEDT FJELD, Ruth, 2000: An outline of Norwegian Lexical Database (LDB) and its classification of adjectives. S. Evert, U. Heid, E. Lehmann, Ch. Rohrer (ur.): *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000 Stuttgart, Germany, August 8th – 12th, 2000*. Institut für Maschinelle Sprachverarbeitung. Universität Stuttgart. 671–678.
- VIDOVIČ MUHA, Ada, 2014: Nastanek Seminarja slovenskega jezika, literature in kulture – čas spremembe jezikoslovnega koncepta. H. Tivadar (ur.): *50. seminar slovenskega jezika literature in kulture: Prihodnost v slovenskem jeziku, literaturi in kulturi*. Ljubljana: FF: Oddelek za slovenistiko. 127–132.

- VIDOVIČ MUHA, Ada, 2009: Poskus določitve meril slovarskega pomena. Perdih, A. (ur.): *Strokovni posvet o novem slovarju slovenskega jezika*, 23. in 24. oktober 2008. Ljubljana: Založba ZRC SAZU, ZRC SAZU. 27–35.
- VIDOVIČ MUHA, Ada, 2000: *Slovensko leksikalno pomenoslovje – Govorica slovarja*. Ljubljana: Znanstveni inštitut Filozofske fakultete.
- VIDOVIČ MUHA, Ada, 1999: Čas in prostor, ujeta v slovenski slovar 20. stoletja. E. Kržišnik (ur.): *35. seminar slovenskega jezika, literature in kulture*. Ljubljana: Oddelek za slovanske jezike in književnosti. 7–26.
- VIDOVIČ MUHA, Ada 1997: Razmerja med leksemi in homonimija. Júlija Bálint, *Slovar slovenskih homonimov: Na podlagi gesel Slovarja slovenskega knjižnega jezika*. Ljubljana: Znanstveni inštitut Filozofske fakultete. 7–16.
- VIDOVIČ MUHA, Ada, 1996: Določnost kot besedilna prvina v slovničnem opisu slovenskega jezika: (ob Kopitarjevi slovnic). Toporišič J. (ur.): *Kopitarjev zbornik: mednarodni simpozij v Ljubljani, 29. junij do 1. julij 1994, Jernej Kopitar in njegova doba: simpozij ob stopetdesetletnici njegove smrti*. Ljubljana: Univerza v Ljubljani. 115–130.
- VIDOVIČ MUHA, Ada, 1978: Pomenske skupine nekakovostnih izpeljanih pridevnikov. *Slavistična revija* 3/26. Str. 253–277.
- WRAY, Alison, 2002: *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- YEROŠINA, Olga, ZARANŠEK, Petra, ŠUSTER, Simon, 2009: Besedne skice za slovenščino: Kritični pogled. M. Stabej (ur.): *Infrastruktura slovenščine in slovenistike (Obdobja 28)*. Ljubljana: Znanstvena založba Filozofske fakultete. 413–422.
- ŽELE, Andreja, 2012: *Pomensko-skladenjske lastnosti slovenskega glagola*. Zbirka *Linguistica et philologica*. Ljubljana: Založba ZRC, ZRC SAZU
- ŽELE, Andreja, 2011: Skladenjske posebnosti slovenščine v razmerjih glagol – vezljivost – stavčni vzorec. *Slavia Centralis* IV/1. 48–68.
- ŽELE, Andreja, 2010: Elipsa med glagolsko intenco in besedilno koherenco (Izpust med glagolsko usmerjenostjo in besedilno soveznostjo). *Slavistična revija* 58/1. 117–131.
- ŽELE, Andreja, 2003: *Glagolska vezljivost: iz teorije v slovar*. Ljubljana: Založba ZRC, ZRC SAZU.
- ŽELE, Andreja, 2001: *Vezljivost v slovenskem jeziku (s poudarkom na glagolu)*. Ljubljana: Založba ZRC, ZRC SAZU.

Imensko kazalo

A

Ahlin, M. 281
Apresjan, J. D. 130, 150
Arhar, Š. 12, 18, 38, 39, 57, 65, 104
Arhar Holdt, Š. 12, 13, 15, 18, 19, 57, 389, 400
Atkins, S. 22, 42, 91, 125, 128, 129, 241, 242, 244, 252, 263, 282, 330, 375, 377–379, 384, 390, 394, 398, 400, 406, 411

B

Bajec, A. 233
Barnbrook, G. 244, 247, 250, 288
Baroni, M. 80
Béjoint, H. 58, 128, 381
Benson, M. 330, 361
Biber, D. 330
Bizjak, K. 13
Bizjak Končar, A. 57
Blažević, Ž. 13
Bolinger, D. 125, 161
Borin, L. 28

C

Clear, J. 40, 41
Columas, F. 330
Coulthard, M. 242, 244
Cowie, A. 253, 278, 288
Crystal, D. 128

Č

Čermák, F. 80, 138, 177, 181, 263, 266, 278, 357, 360
Čermáková, A. 138, 139, 181

Černelič-Kozlevčar, I. 207
Červ, G. 13

D

Daneš, F. 138
Dobrovoljc, H. 281
Dobrovoljc, K. 67
Doroszewski, W. 49
Drinovec Sever, N. 13
Drstvenšek, N. 13
Dušej, T. 13
Dziemianko, A. 241

E

Erjavec, T. 15, 36, 57, 104
Erlandsen, J. 89

F

Fellbaum, C. 36
Filipec, J. 138
Fillmore, Ch. 21, 308, 389, 400
Firth, J. 128, 357, 360, 396, 407
Fišer, D. 16, 35, 36, 214
Fodor, J. 142
Frat, P. 147
Frankenberg-Garcia, A. 375

G

Gangemi, A. 213
Gantar, P. 12, 13, 24, 62, 63, 82, 84, 85, 102, 136, 253, 262, 268, 269, 284
Gorjanc, V. 12, 13, 55, 65, 79, 381, 382, 389, 398, 400, 411
Grabnar, K. 13, 55
Grønvik, O. 80
Guna, P. 13

H

Halliday, M. 356, 360, 396, 407
Hanks, P. 21, 23, 24, 58, 125, 127,
146, 147, 157, 178, 185, 241, 243,
244, 250, 308, 344, 389, 400
Hartmann, R. 330
Hlavsa, Z. 138
Hočevar Gregorič, M. 297
Hollman, J. 51
Holozan, P. 12, 18, 38, 39
Holub, J. 138
Holz, N. 13

J

Jemec Tomazin, M. 13
Johansson, S. 357

K

Kamenšek, U. 13
Kann, V. 51
Katz, J. 142
Kilgarrieff, A. 51, 79, 80, 84, 87, 285,
310, 359, 389
Kjellmer, G. 357
Kocjančič, P. 13
Kopřivova, M. 138, 139, 140, 181
Korošec, T. 232
Kosem, I. 13, 87, 89, 91, 241, 284,
285, 385, 399, 411
Kosem, K. 13
Košir, A. 13
Krek, S. 13, 15, 20, 53, 54, 55, 67, 84,
90, 106, 107, 120, 128, 243, 256,
281, 286, 288, 310, 342, 348, 359,
393, 397, 405, 409
Krishnamurthy, R. 330
Kržišnik, E. 245, 268, 271

L

Laharnar, A. 13
Lew, R. 162, 241, 391, 400
Logar, N. 13, 24, 65, 79
Lönneker-Rodman, B. 22
Lyons, J. 147

M

Manning, Ch. 356
Marušič, F. 233
Miller, G. A. 36
Modrijan, B. 13
Modrijan, N. 13
Moerdijk, F. 43
Moon, R. 128, 136, 148, 381
Može, S. 13, 16, 22, 214
Müller, J. 63, 245
Müller-Spitzer, C. 15

N

Novak, T. 13
Novak, J. 36

O

Och, F. J. 80
Ooi, V. 18
Ore, Ch. 80
Orešnik, J. 142

P

Pala, K. 137
Panevová, J. 213
Pogorelec, B. 340
Polončič Ruparčič, A. 13, 234
Ponikvar, P. 13
Pori, E. 13
Prinsloo, D. 15

Ptasznik, B. 162, 391, 400
Pustejovsky, J. 21, 23, 30, 389, 400

R

Radovič, T. 13
Romih, M. 38
Rozman, T. 67
Rundell, M. 15, 22, 42, 84, 91,
125, 128, 194, 241, 242, 244,
245, 252, 263, 282, 285, 330,
375–379, 384, 390, 394, 398,
400, 406, 411

S

Sandford Pedersen, B. 26
Shütze, H. 356
Siepmann, D. 136
Simpson, J. 375
Sinclair, J. 242, 243, 244, 247, 356,
360, 396, 408
Snoj, J. 147, 157, 302
Strajnar, A. 13
Stritar, M. 13
Stubbs, M. 330
Summers, D. 288
Svozilová, N. 177

Š

Ševeček, P. 137
Šniderlova, J. 383
Šorli, M. 13, 55, 239, 247
Šter, K. 13
Šuster, S. 13, 79

T

Toporišič, J. 106, 107, 109, 110, 111,
117, 120, 142, 233, 237, 297, 325,
340, 341

Trebar, B. 13
Tavčar, A. 36

V

Vatved Fjeld, R. 27
Vidovič Muha, A. 54, 55, 112, 130,
133, 135, 138, 142, 145, 152, 241,
242, 244, 245, 246, 266, 282
Vranjek, U. 13
Vintar, Š. 36

W

Wray 330
Yerošina, O. 13, 84

Z

Zaranšek, P. 13

Ž

Žalodec, P. 13
Žele, A. 138, 177, 182, 184, 185, 200,
340, 346