

Univerza v Ljubljani
Filozofska fakulteta



OSNUTEK KONCEPTA NOVEGA VELIKEGA SLOVENSKO-MADŽARSKEGA SLOVARJA

Ljubljana, 1. marec 2018

Avtorji koncepta:

Iztok Kosem, Júlia Bálint Čeh, Vojko Gorjanc, Anna Kolláth, Attila Kovács, Simon Krek, Sonja Novak Lukanovič, Jutka Rudaš

Uredniški odbor:

mag. Júlia Bálint Čeh (Filozofska fakulteta UL), dr. Elizabeta Bernjak (Filozofska fakulteta UM – v pokoju), doc. dr. Polona Gantar (Filozofska fakulteta UL), prof. dr. Vojko Gorjanc (Filozofska fakulteta UL), red. prof. dr. Anna Kolláth (Filozofska fakulteta UM), dr. Iztok Kosem (Filozofska fakulteta UL in Institut »Jožef Stefan«), dr. Attila Kovács (Inštitut za narodnostna vprašanja), dr. Simon Krek (Filozofska fakulteta UL in Institut »Jožef Stefan«), dr. Cyprian Laskowski (Center za jezikovne vire in tehnologije UL), doc. dr. Nataša Logar (Fakulteta za družbene vede UL), dr. Katalin Márkus (Univerza Károli Gáspár, Madžarska), dr. Attila Martonfi (nekdanji Raziskovalni inštitut za jezikoslovje Madžarske akademije znanosti), Michal Boleslav Měchura (Univerza Masaryk, Češka republika), dr. Géza Németh (Fakulteta za elektrotehniko in informatiko Univerze za tehnologijo in ekonomijo v Budimpešti), red. prof. dr. Sonja Novak Lukanovič (Inštitut za narodnostna vprašanja), Miro Romih (Amebis d.o.o.), izr. prof. dr. Jutka Rudaš (Filozofska fakulteta UM), dr. Tamás Váradi (Raziskovalni inštitut za jezikoslovje Madžarske akademije znanosti)

Koncept je bil izdelan v okviru raziskovalnega projekta Ciljnega raziskovalnega programa z naslovom *Koncept madžarsko-slovenskega slovarja: od jezikovnega vira do uporabnika* (V6-1509), ki ga je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

Kazalo vsebine

1	Uvodna pojasnila in načela	5
1.1	Vrsta slovarja.....	5
1.2	Namen in vloga slovarja in slovarske baze.....	6
1.3	Mediji slovarja.....	7
1.4	Obseg slovarja in geslovnik.....	8
1.5	Slovarski uporabniki	8
1.6	Gradivo.....	10
1.7	Metodologija	12
1.7.1	Orodje za analizo korpusa	12
1.7.2	GDEX za luščenje kandidatov za slovarske zglede	13
1.7.3	Slovarsko orodje	15
2	Sestava slovarskega sestavka.....	17
2.1	Iztočnice	17
2.1.1	Lastna imena	18
2.1.2	Homonimija in homografija.....	19
2.1.3	Razvrstitev iztočnic	20
2.2	Podatki v geselski glavi	20
2.3	Geselski del.....	21
2.3.1	Pomen	21
2.3.2	Prevedki iztočnice	25
2.3.3	Kolokacije in skladenjske zveze.....	29
2.3.4	Večbesedne leksikalne enote	31
2.3.5	Zgledi	34
2.4	Slovarske oznake	36
2.4.1	Slovnične oznake.....	38
2.4.2	Stilne oznake	38
2.4.3	Področne oznake	39

2.4.4	Kontekstualne oznake	39
2.4.5	Pragmatične oznake	39
2.4.6	Časovne oznake	40
2.5	Zvočno gradivo	40
2.5.1	Slovensko zvočno gradivo	41
2.5.2	Madžarsko zvočno gradivo	41
3	Organizacija dela in opis leksikografskega procesa	42
3.1	Splošni opis	42
3.2	Izdelava slovarskih gesel	42
3.2.1	Priprava slovenskega dela gesel	43
3.2.2	Priprava madžarskega dela gesel	44
3.2.3	Priprava dodatnega gradiva	45
3.3	Objava slovarja	45
4	Literatura	46

1 Uvodna pojasnila in načela

Koncept novega Velikega slovensko-madžarskega slovarja (v nadaljevanju VSMS) opredeljuje vsebino in zgradbo načrtovanega slovarja ter z njim povezane slovarske podatkovne baze. Poleg tega koncept okvirno opredeljuje namen in vlogo slovarja ter glavne organizacijske vidike njegove priprave.

Koncept je nastal v okviru nacionalnega ciljnega raziskovalnega projekta *Koncept madžarsko-slovenskega slovarja: od jezikovnega vira do uporabnika* (V6-1509), ki je potekal od 1. 10. 2015 do 28. 2. 2018 in sta ga financirala Javna agencija za raziskovalno dejavnost Republike Slovenije in Ministrstvo za izobraževanje republike Republike Slovenije.¹ Projekt so izvajali Filozofska fakulteta Univerze v Ljubljani (vodilna institucija), Inštitut za narodnostna vprašanja in Filozofska fakulteta Univerze v Mariboru.

Izdelavo koncepta in samega slovarja je potrebno postaviti tudi v širši slovensko-madžarski kontekst, saj prva pobuda za novi slovensko-madžarski in madžarsko-slovenski slovar sega v leto 2009.² Predvideno je bilo sočasno delo na obeh delih slovarja, vendar pa so se iz različnih razlogov dela na Madžarskem začela prej kot v Sloveniji. Od leta 2015 v Budimpešti na univerzi ELTE tako poteka izdelava velikega madžarsko-slovenskega slovarja, za katerega je Madžarska zagotovila sredstva do leta 2019³ in katerega načrtovani obseg je 70.000 gesel. VSMS je zasnovan tako, da je slovar kompatibilen z madžarsko-slovenskim delom (pri čemer vsebuje še dodatne informacije) in tako omogoča kasnejšo združitev obeh delov in skupno objavo celotnega slovarja.

1.1 Vrsta slovarja

Novi Veliki slovensko-madžarski slovar je temeljni dvojezični slovensko-madžarski slovar velikega obsega, ki na eni strani beleži stanje sodobne slovenščine in madžarščine, na drugi pa prikazuje aktualna sporazumevalna razmerja med jezika. Priprava slovarja in slovarske baze sledi sodobnim smernicam in spoznanjem slovenske in tuje leksikografije, tako enojezične kot dvojezične.

¹ <https://www.cjvt.si/komass/>

² Skupna izjava vlad Republike Slovenije in Madžarske s srečanja v Keszthelyu, ki je potekalo 27. novembra 2009: [http://vrs-3.vlada.si/MANDAT12/VLADNAGRADIVA.NSF/aa3872cadf1c8356c1256efb00603606/7aa66fc2b526181bc1257abf00304256/\\$FILE/izjavaonamerislovar.pdf](http://vrs-3.vlada.si/MANDAT12/VLADNAGRADIVA.NSF/aa3872cadf1c8356c1256efb00603606/7aa66fc2b526181bc1257abf00304256/$FILE/izjavaonamerislovar.pdf)

³ http://www.uszs.gov.si/si/medijsko_sredisce/novica/archive/2015/2/browse/1/select/sporocilo_za_javnost/article/764/3092/

Ena glavnih temeljnih podstav VSMS je sistematična uporaba korpusnega pristopa, tako pri izbiri slovenskega geslovnika in pripravi slovenskega dela slovarskih gesel (od kolokacij do zgledov), kot pri izbiri in preverjanju madžarskih prevodov. Koncept slovarja predvideva tudi uporabo vzporednih slovensko-madžarskih korpusov, katerih uporabnost pa je trenutno zaradi njihovega majhnega obsega in pokrivanja zgolj določenih področij še omejena.

Organiziranost podatkov ter njihov opis sledita sodobnim leksikografskim smernicam, torej težimo k čim večji uporabnosti, razumljivosti in preglednosti slovarskih gesel, hkrati pa v bazi beležimo številne podatke, ki so koristni tako za jezikovne kot jezikovnotehnološke namene.

1.2 Namen in vloga slovarja in slovarske baze

VSMS bo izdelan povsem na novo, pri izdelavi gesel bo uporabljeno sodobno jezikovno gradivo. To je glede na stanje na področju slovensko-madžarskih slovarjev (gl. Bálint Čeh in Kosem 2017; Rudaš in Kolláth 2017) nujno potrebno, saj so obstoječi slovensko-madžarski in madžarsko-slovenski slovarji stari več kot 20 let, poleg tega pa so tudi slovarji, izdelani v 90. letih 20. stoletja, močno temeljili na svojih še starejših predhodnikih. Obenem so bili vsi slovarji dokaj majhnega obsega, saj nobeden ni presegel 40.000 gesel. Vendar pa to ne pomeni, da v njih ne najdemo koristnih rešitev in podatkov, ki pa jih je treba v vsakem primeru ponovno preveriti v sodobnih korpusih. Pri pripravi koncepta smo tako opravili obsežno analizo obstoječih slovensko-madžarskih slovarjev (Bálint Čeh in Kosem 2017), izpostavili njihove prednosti in slabosti, pregledali pa smo tudi primere dobrih tujih praks, zlasti britanskih založb Collins in Oxford.

Resolucija o nacionalnem programu za jezikovno politiko 2014–2018 (v nadaljevanju Resolucija) izpostavlja močno pomanjkanje sodobnih večjezičnih virov nasploh in poudarja:

»Pri načrtovanju večjezičnih virov za prihodnje obdobje ne moremo več govoriti samo o klasičnih dvojezičnih ali terminoloških slovarjih, temveč tudi o slovarskih bazah ali večjezičnih bazah znanja, ki se izkoriščajo bodisi neposredno za preverjanje informacij v ustrezni spletni ali drugi aplikaciji (ali izvedeni tiskani obliki) bodisi za uporabo v orodjih, ki na druge načine pomagajo pri učenju tujih jezikov, prevajanju ali tolmačenju, kot so sistemi za strojno prevajanje, sistemi za računalniško podprto prevajanje, sistemi za podporo tolmačenju itd. Opisane vire je mogoče izkoristiti, če je vzpostavljena ustrezna (spletna) infrastruktura.« (str. 37)

Iz tega vidika ima VSMS oz. njegova podatkovna baza več temeljnih funkcij: kot prvo, izdelava VSMS bo pomembna za ohranjanje kulturne identitete madžarske manjšine,

prav tako pa tudi vitalnosti jezikovne skupnosti, ki ima s pomočjo kvalitetnega jezikovnega vira možnost celovitega ustvarjalnega sobivanja v dvojezičnem okolju v svojem prvem jeziku, s čimer je zagotovljeno enakopravno sobivanje obeh jezikov in kultur. Drugič, slovarska baza VSMS mora biti zasnovana tako, da bo omogočala enostavno povezovanje z enojezičnimi viri, od slovarjev do korpusov, ter vsebovati informacije in biti na voljo v formatu, ki podpira povezovanje z drugimi, tujimi, večjezičnimi ali enojezičnimi viri. In tretjič, VSMS je prvi slovar kombinacije slovenščina-tuji jezik, ki bo vseboval podatke o sodobnem slovenskem jeziku in bo izdelan s pomočjo sodobnih jezikovnotehnoloških metod, kar pomeni, da mora biti baza zasnovana tako, da bo omogočala enostavno kasnejšo izdelavo drugih dvojezičnih slovarjev para slovenščina-tuji jezik, tudi za jezike, ki do zdaj v kombinaciji s slovenščino še niso bili obdelani.

1.3 Mediji slovarja

VSMS bo izhodiščno zasnovan za digitalne medije, od spleta do tabličnih in mobilnih naprav, pri čemer študije (Müller-Spitzer 2014; za dvojezični prostor gl. Kosem in Kovács 2017) kažejo na prevlado spletne oblike kot preferenčnega medija za slovarje med uporabniki. Če hoče doseči čim večje število uporabnikov, mora biti slovar dostopen v čim več oblikah, k čemur v digitalnem okolju lahko najbolj pripomore prosta dostopnost slovarske baze in fleksibilna organiziranost podatkov. Konceptualizacija slovarja zato ne sme biti omejena zgolj na en format oz. medij, temveč mora vključevati izdelavo digitalne slovarske baze, ki vključuje informacije, na podlagi katerih je mogoče izdelati slovarje z različnimi kombinacijami informacij v različnih formatih, s čimer lahko posledično zadovoljimo potrebe različnih tipov uporabnikov.

Kot kažejo raziskave med ciljnim uporabniki slovensko-madžarskih jezikovnih virov (Kosem in Kovács 2017), mnogi še vedno posegajo po tiskanih slovarjih, vendar pa zaradi dejstva, da je veliko slovensko-madžarskih virov na voljo samo v tiskani obliki, ni jasno, ali uporabniki tako obliko slovarja dejansko uporabljajo zaradi lastnih preferenc ali zaradi nerazpoložljivosti oz. neobstoju digitalnih izdaj slovarjev. Pri konceptu VSMS to ni težava, saj je digitalna oblika primarna in je tiskano obliko vedno mogoče izpeljati iz digitalne, medtem ko pri obratnem postopku marsikateri elementi zaradi omejitev tiskane verzije izostanejo (gl. Gorjanc idr. 2015; Krek idr. 2013). O pripravi tiskane oblike se bomo odločali sproti, tudi na podlagi nadaljnjih študij med ciljnim uporabniki.

VSMS bo primarno na voljo kot samostojen slovarski vir, saj je to preferenčna oblika dostopanja do dvojezičnih slovensko-madžarskih virov med ciljnim uporabniki (gl. Kosem in Kovács 2017), vključen pa bo tudi v (večjezične) portale. Hkrati že koncept slovarja oz. slovarskih gesel predvideva vključevanje povezav z ostalimi viri, kot so slovarji sinonimov in korpusi, pa tudi leksikoni besednih oblik. Tako je predvidena vključitev v portal jezikovnih virov Centra za jezikovne vire in tehnologije Univerze v

Ljubljani, ki bo omogočal hkratno zadovoljitev obeh zahtev, torej omogočanje samostojnosti vira in njegovo povezljivost z ostalimi relevantnimi viri.

1.4 Obseg slovarja in geslovník

Slovar bo obsegal vsaj 70.000 iztočnic, ki ponujajo slovnične, pomenske, pragmatične in druge informacije o leksikalnih enotah, tako eno- kot večbesednih. Slovar bo pokrival obsežno slovensko splošno besedišče in strokovno izrazje, pa tudi lastna imena.

Izhodiščni geslovník temelji na frekvenčnem seznamu, pridobljenem iz referenčnih korpusov, npr. korpus Gigafida⁴ za pisno slovenščino in korpus Gos⁵ za govorjeno slovenščino, uporabljamo pa tudi korpus Kres⁶ in specializirane korpusse. Seznam bo ustrezno prečiščen in redno preverjan, sploh ob pomembnejših spremembah oz. dopolnitvah korpusnega gradiva (npr. objavi nove verzije korpusa). V ta namen se bodo tudi razvila oz. uporabila orodja, ki omogočajo sprotno beleženje jezikovnih sprememb, tako na makrostrukturni kot mikrostrukturni ravni.

Slovar se bo v skladu s sodobnimi leksikografskimi praksami v spletni izdaji objavljaj postopoma, najbrž enkrat letno ali vsakih šest mesecev. Tako spletnim uporabnikom ne bo treba čakati do dokončanja slovarja, sproti bodo že lahko dostopali do dokončanih gesel. Na ta način bomo tudi sproti naslavljalí pomanjkljivosti obstoječih slovensko-madžarskih slovarjev, kot sta zastarelost in skromna pokritost. Poleg tega bodo uporabljeni pristopi, ki jih omenja Predlog o izdelavi Slovarja sodobnega slovenskega jezika (Krek idr., 2013;) in na njem temelječi Slovar sodobnega slovenskega jezika (Gorjanc idr. 2015), tj. prioritiziranje izdelave gesel (tj. najprej izdelati gesla, ki so za uporabnike bolj relevantna) in vnaprejšnji objavi delno izdelanih gesel. Pri tem v našem primeru delno izdelano geslo vsebuje preverjene leksikografske informacije, od katerih pa določeni mikrostrukturni elementi še niso dokončani, npr. vsi zgledi še niso prevedeni oz. je preveden samo del kolokacij (kontrastivno najbolj zanimive), skladenjskih zvez ipd. S tem k izdelavi slovarja pristopamo v plasteh, tako da določenim (obveznim) mikrostrukturnim elementom dajemo prioriteto.

1.5 Slovarski uporabniki

Kot kažejo raziskave (Müller-Spitzer 2014; za pregled situacije v Sloveniji gl. Arhar Holdt 2015), uporabniki od slovarjev pričakujejo predvsem zanesljive in ažurne

⁴ <http://www.gigafida.net/>

⁵ <http://www.korpus-gos.net/>

⁶ <http://www.korpus-kres.net/>

informacije, ki so predstavljene na uporabniku prijazen način in so prosto dostopne.⁷ VSMS je namenjen najširšemu krogu ciljnih uporabnikov, od prevajalcev, učiteljev, študentov in učencev, pa tudi poslovnežem in politikom oz. zaposlenim v javni upravi. Iz tega razloga so tudi podatki v bazi opremljeni z ustreznimi metapodatki (npr. frekvenca, kontrastivna relevantnost), saj to omogoča prilagajanje prikaza različnim ciljnim skupinam. Pripravljene bodo tudi rešitve za osebe s posebnimi potrebami, kot tudi veleba Resolucija. Predvideva se, da bo večina uporabnikov uporabljala digitalne oblike slovarja, predvidena pa je tudi tiskana oblika, sploh če študije to potrdijo.

Posebna pozornost priprave podatkov v slovarski bazi bo namenjena udeležencem dvojezičnega modela izobraževanja. V ta namen se bo opravila analiza učbenikov in šolskih gradiv, s katero bomo prepoznali leksiko, ki je bolj relevantna v izobraževanju. Opremljenost z ustreznimi jezikovnimi viri je namreč eden pomembnih elementov za izboljšanje učinkovitosti dvojezičnega izobraževanja, v katerem morajo imeti strokovni delavci ustrezno razvito sporazumevalno zmožnost v slovenskem in madžarskem jeziku na taki ravni, da ju lahko uporabljajo kot učna jezika. Zasnova dvojezičnega slovarja, ki upošteva potrebe dvojezičnega izobraževanja, tako pomeni prvi korak k razvoju in izboljšanju kompetenc učiteljev, posledično pa tudi boljšega znanja učencev in s tem tudi njihovo konkurenčno prednost na trgu dela ter pri delu boljšo delovno uspešnost.

Pomembna lastnosti VSMS je tudi ta, da čeprav je enosmerni dvojezični vir, je namenjen tako slovenskim uporabnikom (tistim, ki jim je madžarščina drugi/tuji jezik) kot madžarskim uporabnikom (tistim, ki jim je slovenščina drugi/tuji jezik), saj upravičeno lahko pričakujemo, da bodo glavni uporabniki slovarja iz dvojezičnega območja na obeh straneh meje, torej v Prekmurju in v Porabju. Madžarskemu uporabniku tako VSMS omogoča pasivno rabo oz. dekodiranje, tj. razumevanje pisane in govorne slovenščine. Osnovna naloga je torej pomagati madžarskemu uporabniku pri razumevanju slovenskih besedil in pri prevajanju iz slovenščine v madžarščino. Po drugi strani bo slovar v pomoč tudi slovenskemu uporabniku pri aktivni rabi oz. enkodiranju, tj. tvorbi pisane in govorne madžarščine. Posledično so geselski članki podatkovno bogatejši kot pri dvojezičnih slovarjih, ki so namenjeni zgolj dekodiranju, in vključujejo elemente, ki običajno v tovrstne slovarje ne bi bili vključeni oz. bi bili vključeni v manjši meri. Na ta način slovar prispeva tudi k izpolnjevanju ciljev Resolucije, povezanih z jezikom madžarske manjšine v Republiki Sloveniji (Resolucija: 21-22).

⁷ Podobne ugotovitve ponujajo tudi prvi izsledki nedavno opravljene evropske raziskave o rabi slovarjev, pri kateri je kot član osrednje skupine sodeloval tudi I. Kosem (članek bo objavljen v drugi polovici leta 2018).

1.6 Gradivo

Zaradi obstoječega stanja na področju enojezičnih slovarjev, za katero je značilno pomanjkanje slovarja, ki bi odražal sodobni slovenski jezik (Krek idr. 2013; Gorjanc idr. 2015b; Stabej 2015; Gantar 2016), se pri pripravi slovenskega dela geslovnika opiramo predvsem na referenčni korpus slovenskega jezika Gigafida (gl. Logar idr. 2012; Logar idr. 2013; Logar 2015), ki trenutno obsega 1,2 milijarde besed in je obsežna zbirka slovenskih besedil najrazličnejših zvrsti, nastalih med letoma 1990 in 2011, od dnevnih časopisov, revij do knjižnih publikacij vseh vrst (leposlovje, učbeniki, stvarna literatura), spletnih besedil, prepisov parlamentarnih govorov in podobno.⁸

Pri pripravi slovenskega geslovnika uporabljamo tudi odprto dostopne jezikovne vire, kot so *Leksikalna baza za slovenščino* (Gantar in Krek 2011; Gantar idr. 2012), *Slovar sopomenk sodobne slovenščine*,⁹ *Kolokacijski slovar sodobne slovenščine* (Kosem idr. 2017; Krek idr. 2016), *Slovar slovenskih frazemov* (Keber 2011), *Slovar tviterščine* (Gantar idr. 2016b)¹⁰ in podobne. Posvetujemo se tudi z obstoječimi slovarskimi viri, in sicer od splošnih slovarjev oz. slovarskih baz uporabljamo *Slovenski pravopis* (Toporišič ur. 2001), *Slovar slovenskega knjižnega jezika* (prva in druga izdaja), od specializiranih slovarjev pa *Sinonimni slovar slovenskega jezika* (Snoj idr. 2016). Uporabljamo tudi *Slovenski oblikoslovni leksikon Sloleks* (Dobrovoljc idr. 2015a, 2015b) ter portala Fran¹¹ in Termania.¹²

Enojezični madžarski viri so uporabljeni za preverjanje prevodov in iskanje kandidatov za prevode. Med glavnimi viri je korpus madžarskega jezika huTenTen12 (Jakubiček idr. 2013), 2,5-milijarde besed obsegajoč korpus spletnih besedil,¹³ ki naj bi ga v kratkem nadomestil večji in sodobnejši korpus huTenTen18. Uporabljamo tudi razlagalni slovar madžarskega jezika (*A Magyar Nyelv Értelmező Szótára I–VII. Kötet*)¹⁴ in madžarski pravopis,¹⁵ ter slovarje sopomenk, kot je npr. spletni sinonimni slovar madžarskega jezika.¹⁶

⁸ Predvidoma konec leta 2018 bo objavljena nova verzija korpusa, tj. Gigafida 2.0, ki bo vključevala tudi besedila, nastala v letih 2012-2018 (<https://www.cjvt.si/gigafida/>).

⁹ <http://viri.cjvt.si/sopomenke/slv/>

¹⁰ <http://lexonomy.cjvt.si/slovar-tvitterscine/>

¹¹ <http://www.fran.si/>

¹² <http://www.termania.net/>

¹³ Madžarski referenčni korpus Magyar Nemzeti Szövegtár (MNSZ, zadnja verzija 2.0.4) je velik malo več kot milijardo besed, a je dostop do korpusa omejen, poleg tega pa ni na voljo v orodju Sketch Engine, ki omogoča napredna iskanja, potrebna za leksikografske namene.

¹⁴ <http://mek.oszk.hu/adatbazis/magyar-nyelv-ertelmezo-szotara/elolap.php>

¹⁵ <http://helyesiras.mta.hu/> Od 1. maja 2017 je v veljavi novi pravopis (12. izdaja).

¹⁶ <https://szinonimaszotar.hu>

Kot izhodišče pri pripravi prevodnih kandidatov za leksikografe uporabljamo dva dvojezična jezikovna vira. Prvi so e-zbirke izrazov¹⁷ za 15 področij, ki se poučujejo na osnovnih in srednjih šolah. E-zbirke izrazov temeljijo predvsem na učbenikih, ki se uporabljajo v dvojezičnih šolah, vsebujejo pa 29.685 enobesednih in večbesednih slovenskih izrazov in njihove prevode v madžarščini. Gre za pomemben jezikovni vir, saj so po eni strani najsodobnejši dvojezični vir za par slovenščina-madžarščina, po drugi strani pa dober pokazatelj predvsem terminoloških, pa tudi splošnih, potreb uporabnikov v izobraževanju na dvojezičnem območju.

Drugi vir je avtomatsko izdelana slovensko-madžarska baza prevodov založbe KDictionaries,¹⁸ ki vsebuje več kot 24.000 slovenskih iztočnic z več kot 41.000 pomeni in njihovimi prevodi v madžarščino.

Za namene posvetovanja uporabljamo še obstoječe dvojezične slovarje, kot so Slovensko-madžarski in madžarsko-slovenski navzkrižni slovar (Hradil 2012), ki vsebuje 37.908 gesel in je na voljo tudi v elektronski različici, Slovensko-madžarski slovar (Hradil 1996) s 40.000 gesli (in okrog 60.000 drugimi slovarskimi podatki ter izrazi) in slovensko-madžarski del slovarja Elizabete Bernjak iz leta 1995 (Bernjak 1995), ki po grobi oceni vsebuje približno 22.000 geselskih člankov (gl. Bálint Čeh in Kosem 2017).

Pri delu uporabljamo tudi ostale dvojezične vire, kot sta npr. prosto dostopni bazi prevodov Linguee¹⁹ in pa Glosbe.²⁰ Pri iskanju v Linguee²¹ mora uporabnik vnesti besedo ali več besed, lahko tudi samo del besede, ali kolokacije v narekovajih. Podobno Glosbe ponuja orodje za iskanje ustreznice, tako za enobesedne iztočnice, kot tudi za kolokacije v drugem jeziku, in pokriva tudi par slovenščina-madžarščina. Za slovensko-madžarski jezikovni par je trenutno na voljo 31.527, za madžarsko-slovenski par pa 31.204 prevedenih stavčnih zgledov.

Pri iskanju prevedkov za slovenske frazeme je na voljo tudi paremiološka platforma za učenje pregovorov SprichWort.²² V slovenski podatkovni bazi omenjene platforme je 212 slovenskih in 210 madžarskih pregovorov, ki so razporejeni po abecednem

¹⁷ <http://eslovar.datadev.si/>. Zbirke so nastale v okviru projekta *E-kompetence učiteljev v dvojezičnih šolah*, ki sta ga financirala Evropska unija iz Evropskega socialnega sklada in Ministrstvo za izobraževanje, znanost in šport Republike Slovenije in je potekal od oktobra 2011 do avgusta 2013 pod vodstvom Zavoda za kulturo madžarske narodnosti. V projektu so sodelovali učitelji dvojezičnih osnovnih šol in srednje šole v Prekmurju, Inštitut za narodnostna vprašanja in Filozofska fakulteta Univerze v Mariboru.

¹⁸ <http://kdictionaries.com/>

¹⁹ <https://sl.linguee.com/>

²⁰ <https://sl.glosbe.com/>

²¹ Podatek o obsegu slovensko-madžarske v Linguee ni dostopen.

²² http://www.sprichwort-plattform.org/sp/Sprichwort_si

vrstnem redu. Platforma s pomočjo korpusno pridobljenih podatkov prikazuje tudi tipično rabo pogovorov.

V slovarska gesla vključujemo tudi posnetke izgovorjav delov slovenskih gesel in madžarskih prevodov. Za slovenščino uporabljamo sintetizator govora eBralec (<http://ebralec.si>), ki so ga izdelali podjetji Alpineon d.o.o. in Amebis d.o.o. ter Institut "Jožef Stefan". Za madžarščino uporabljamo sintetizator govora Profivox HMM TTS Univerze za tehnologijo v Budimpešti (Oddelek za Telekomunikacije in Medijsko informatiko).²³

1.7 Metodologija

1.7.1 Orodje za analizo korpusa

Za analizo podatkov in preverjanje prevodov se pri pripravi gesel uporablja predvsem korpusno orodje Sketch Engine²⁴ (Kilgarriff idr. 2004; Kilgarriff idr. 2014), s pomočjo katerega je mogoče pridobivati relevantne leksikalne, slovnične in pomenske podatke iz korpusa z uporabo aplikacij, kot so:

- Konkordančnik za analizo in urejanje konkordančnega niza za določeno besedo, besedno obliko, zvezo ipd.
- Besedne skice (gl. tudi Krek in Kilgarriff 2006; Kilgarriff idr. 2010; Krek 2012, 2015) – avtomatsko profiliranje leksikalnega obnašanja besed in besednih zvez na podlagi oblikoskladenjsko označenega korpusa
- Primerjalne skice – avtomatska primerjava besedilnega okolja za različne besede, besedilne žanre ipd.
- Tezaver – izdelava seznamov pomensko sorodnih besed idr.

Sketch Engine uporabljamo tako pri pripravi slovenskega dela gesel, kjer je pomembno zagotoviti predvsem širok nabor kolokacij, skladenjskih zvez in ostalih informacij, ki so potem na voljo leksikografu pri izbiri relevantnega gradiva in pripravi madžarskih prevodov, kot tudi pri iskanju in preverjanju prevodov, saj nam ponuja tudi sežetek okolice prevodov in njihovih kolokacij, kar je koristno za prevajanje večbesednih elementov v geslih.

²³ <http://smartlab.tmit.bme.hu/>.

²⁴ <http://www.sketchengine.co.uk/>

1.7.2 GDEX za luščenje kandidatov za slovarske zgledе

Pri pridobivanju slovenskih in madžarskih slovarskih zgledov uporabljamo orodje za prepoznavo dobrih zgledov GDEX (ang. *Good Dictionary Examples*; Kilgarrieff et al. 2008), ki nam ponudi nabor kandidatov za dobre slovarske zgledе, med katerimi leksikografi potem izberejo ustrezne. GDEX razvršča zgledе glede na njihovo kakovost pri značilnostih, kot so dolžina zgleda, celostavčna oblika, preprosta ali manj kompleksna skladenjska zgradba povedi, prisotnost ali odsotnost redkih besed, spletnih in elektronskih naslovov ipd. Mnoge od teh značilnosti so posredno povezane s tipičnostjo, informativnostjo in razumljivostjo, torej lastnostmi dobrega zgleda. Značilnosti lahko razdelimo v dve skupini: v prvi so tiste, ki jih zgled mora vsebovati, npr. celostavčnost, odsotnost spletnih naslovov, odsotnost izredno dolgih ali redkih besed ipd. Če zgled ne ustreza vsaj eni od teh značilnosti, dobi toliko kazenskih točk, da se takoj znajde na dnu vseh zadetkov. V drugi skupini so značilnosti, ki so bodisi zaželenе bodisi nezaželenе (stopnjo (ne)zaželenosti določimo s težo, ki jo pripišemo posamezni značilnosti, in višino dodatnih/odbitih točk), a je pomemben predvsem kumulativni seštevek vrednosti vseh značilnosti v konfiguraciji.

Za namene VSMS smo izdelali dve novi konfiguraciji, po eno za vsak jezik. Pri GDEX konfiguraciji za slovenske zgledе smo izhajali iz najnovejše GDEX konfiguracije za slovenščino (Kosem 2015a; gl. tudi Kosem idr. 2011, 2013), ki vsebuje sledeče klasifikatorje:

- Cela poved. Na ta način prioritiziramo zgledе, ki ustrezajo načelu celostavčnosti.
- Ne vsebuje pojavnic s frekvenco manj kot 3. Iščemo zgledе, ki ne vsebujejo zelo redkih besed, napak in korpusnega šuma.
- Minimalna dovoljena dolžina povedi.
- Maksimalna dovoljena dolžina povedi.
- Poved ne sme vsebovati ponovitve iztočnice. Gre za pomemben klasifikator, kajti večkratno ponavljanje iztočnice zgledu jemlje razumljivost in informativnost.
- Vsebuje elektronski ali spletni naslov. Zgledi, ki ustrezajo temu kriteriju, prejmejo visok kazenski pribitek.
- Optimalna dolžina (med X in Y pojavnic). Medtem ko s klasifikatorjema za minimalno in maksimalno dolžino povedi izločamo prekratke in predolge povedi (jih potiskamo na dno seznama), z optimalno dolžino nagrajujemo povedi z dolžino znotraj danega razpona.
- Vsebuje redke leme. Klasifikator dodeli točkovni odbitek povedi za vsako redko lemo, ki jo vsebuje. Frekvenčna meja, ki opredeljuje redkost, je odvisna od velikosti korpusa.
- Vsebuje pojavnice, daljše od X znakov. Klasifikator kaznuje vsako pojavnico, ki izpolnjuje omenjeni kriterij.

- Število ločil v zgledu (brez vejic). Klasifikator točkovno kaznuje poved, v kolikor je preseženo določeno število ločil v njej, pri čemer se vejice ne upoštevajo.
- Število vejic v povedi. Klasifikator točkovno kaznuje povedi v več kot tremi vejicami, saj je bilo ugotovljeno, da so takšne povedi pogosto kompleksnejše in posledično slabši kandidati za dobre zglede.
- Pojavnice z velikimi začetnicami. Klasifikator točkovno kaznuje povedi, ki vsebujejo pojavnice z velikimi začetnicami, in je namenjen predvsem kot dopolnilo klasifikatorju za lastna imena.
- Pojavnice z mešanimi simboli (npr. črke in številke). Klasifikator točkovno kaznuje nebesede in korpusni šum.
- Lastna imena. Klasifikator točkovno kaznuje povedi s pojavnicami, ki so v korpusu označene kot lastna imena. Če je takšnih pojavnic v povedi več, je kaznovana vsaka posamezna pojavitev.
- Zaimki. Klasifikator z odbitkom kaznuje vsako pojavitev zaimka v povedi. Klasifikator je koristen predvsem, ko je zaimkov v povedi več, saj so takšne povedi ponavadi manj razumljive oz. potrebujejo dodaten kontekst.
- Položaj leme v povedi. Klasifikator točkovno kaznuje povedi, kjer se lema pojavlja izven določenega razpona v povedi. Tako je bilo za glagolske leme ugotovljeno, da so boljši kandidati za dobre zglede tiste povedi, v katerih se glagol ne pojavlja na začetku, tj. v prvih 40 odstotkih pojavnicah povedi.
- Seznam prepovedanih besed na začetku povedi. Pri izdelavi konfiguracij se je izkazalo, da so določene besede na začetku povedi že dober indikator, da ne gre za dobrega kandidata za slovarski zgled. Za namene klasifikatorja je bil na podlagi analize in opažanj pri evalvaciji konfiguracij izdelan seznam takšnih besed. Klasifikator tako točkovno kaznuje povedi, ki se začenjajo s katero od besed na seznamu.
- Seznam prepovedanih besednih zvez na začetku povedi. Podoben klasifikator kot klasifikator za prepovedane besede, s tem da kaznuje pojavitev določenih večbesednih nizov na začetku povedi.
- Tretji kolokator. Eden najpomembnejših klasifikatorjev, ki točkovno nagrajuje zglede, ki vsebujejo najbolj tipične kolokatorje določene kolokacije, in s tem posredno upošteva merilo koligacijske tipičnosti. Npr. pri kolokaciji *klavrn + podoba* klasifikator dodeli dodatne točke zgledom s statistično pomembnim tretjim kolokatorjem *kazati*, pri čemer se izkaže, da tako identificirani zgledi pa vsebujejo tudi tipično širšo strukturo kolokabilne okolice: *kazati klavarno podobo česa*.

- Jaccardov indeks. Gre za algoritem,²⁵ ki meri podobnost med nizi, v našem primeru povedmi. Če klasifikator najde dve podobni ali celo enaki povedi, tisto z nižjim točkovanjem vrže na dno seznama zadetkov.

Pri pripravi konfiguracije GDEX za VSMS smo obdržali vse zgoraj našete klasifikatorje, smo pa prilagodili nekatere nastavitve, zlasti tiste, povezane z dolžino povedi in pogostostjo pojavnic v povedih. Zglede pridobivamo iz korpusa Gigafida, po potrebi pa tudi iz drugih relevantnih, referenčnih ali specializiranih, korpusov slovenščine.

Pri pripravi konfiguracije GDEX za madžarščino smo morali izdelati povsem novo konfiguracijo, kjer pa smo se tudi oprli na slovensko različico in uporabili predvsem klasifikatorje, ki so se v preteklosti že izkazali za jezikovno neodvisne (Kosem idr., v pripravi).²⁶ Tudi pri pripravi madžarske konfiguracije smo največ pozornosti posvečali dolžini povedi in pogostosti pojavnic v povedih, saj iščemo predvsem relativno kratke in razumljive zglede. Zglede pridobivamo iz korpusa huTenTen12 oz. v danem trenutku najsodobnejšega in prosto dostopnega korpusa madžarščine.

Za vsak izluščeni zgled izvažamo tudi metapodatke o besedilu, iz katerega zgled izvira, npr. leto, vir, avtor, naslov ipd. To nam zagotavlja sledljivost zgleda, ponuja pa tudi možnost kasnejšega prikaza tovrstnih informacij v slovarju.

1.7.3 Slovarsko orodje

Izdelava slovarskih gesel poteka v dveh slovarskih orodjih. Prvo orodje je iLex (Erlandsen 2004), ki ga je izdelalo podjetje Erlandsen Media Publishing in je dobro poznano v mednarodnem in slovenskem prostoru. iLex ponuja številne napredne funkcionalnosti, ki so za leksikografsko delo nepogrešljive, npr. napredno iskanje po slovarskih geslih, vzporeden pogled dveh gesel ali več, enostavno in hitro urejanje ipd. Pomembno je tudi, da orodje omogoča dokaj enostavno izvažanje in uvažanje podatkov, kar olajša morebitne potrebe po sistematičnem dodajanju podatkov v obstoječa gesla, kot so na primer zgledi za madžarske prevodne ustreznice. S tem orodjem delajo predvsem uredniki pri oblikovanju končnih gesel.

Drugo slovarsko orodje, ki ga uporabljamo pri izdelavi slovarja, je Lexonomy (Měchura 2017).²⁷ Orodje je bilo v zadnjem letu snovanja koncepta precej nadgrajeno tako z vidika ponujenih funkcionalnosti kot z vidika uporabniške prijaznosti. Orodje je spletno dostopno, tako da za uporabo ni potrebno namestiti nobene programske opreme.

²⁵ https://en.wikipedia.org/wiki/Jaccard_index

²⁶ V času pisanja koncepta je bil prispevek že odobren za objavo v reviji International Journal of Lexicography.

²⁷ <https://www.lexonomy.eu/>

Lexonomy omogoča uvažanje in izvažanje datotek, prilagajanje prikaza slovarskih gesel in tudi takojšnjo vizualizacijo slovarskih gesel na spletu. Lexonomy uporablja osrednja redaktorska ekipa pri izdelavi gesel, zaradi enostavnega dostopa in uporabe pa tudi zunanji sodelavci (npr. lektorji, terminologi) pri pregledovanju gesel in potrjevanju rešitev.

2 Sestava slovarskega sestavka

2.1 Iztočnice

Vsako geslo v VSMS se začne z iztočnico. Podiztočnic ne uporabljamo. V VSMS imamo enobesedne in večbesedne iztočnice, so pa v primeru večbesednosti odločitve na ravni iztočnice sprejete na podlagi pomenskih in formalnih meril.²⁸ Posledično je večbesedna leksikalna enota lahko samostojna iztočnica ali pa samostojni segment gesla enobesedne iztočnice (pod stalnimi zvezami ali frazeološkimi enotami), ki se nahaja v večbesedni enoti. Prednost tega pristopa je tudi v tem, da nam nudi fleksibilnost pri izdelavi slovarja, kar je sploh koristno pri sprotnem objavljanju gesel – tako lahko večbesedne leksikalne enote, ki jih npr. dobimo iz obstoječih baz, prvotno obravnavamo kot samostojne iztočnice (npr. Bela hiša → *Fehér Ház*), ko pa izdelamo geslo oz. gesla za njene sestavne elemente (enobesedne iztočnice), pa jo v bazi lahko umestimo pod relevantne enobesedne iztočnice (*bel* in *hiša*). Hkrati lahko večbesedne leksikalne enote, ki jih imamo zabeležene pod enobesednimi iztočnicami, premaknemo na raven iztočnice, če se izkaže, da se enobesedna iztočnica pojavlja izključno kot sestavina stalne zveze. Nekoliko drugače je v primeru, ko se beseda pojavi zgolj v frazemih – takrat obdržimo enobesedno iztočnico in ponudimo samo frazeološki del.

Kot enobesedne iztočnice obravnavamo tudi glagole s prostim glagolskim morfemom *si/se*, v primeru ko je morphem sestavni del glagola v vseh njegovih pomenih, npr. *smejati se*, *bati se*, *svitati se*. Če je morphem sestavni del glagola v samo določenem njegovem pomenu oz. pomenih, ga navedemo samo tam (kot ustaljeno obliko) in ne že v zapisu iztočnice. Podoben pristop uporabljamo tudi pri glagolih z morfemom *ga* oz. *jo* (gl. tudi Gantar 2016) in podobnih kolokacijskih tipih (npr. *na tešče* pri prislovu *tešče*).

Večbesedne iztočnice so tudi prevzeti leksemi, kot so *ad hoc*, *alma mater*, *de facto* in podobni, ne pa tudi besedotvorni sklopi, ki se pišejo tako skupaj in narazen (npr. *po navadi*, *na svidenje*), saj kot iztočnice obravnavamo le zapis skupaj (npr. *nasvidenje*), zapis narazen pa pri relevantni enobesedni iztočnici (npr. *svidenje*).

Kot iztočnice so v VSMS vključene tudi črke, kratice in okrajšave. Kot samostojnih iztočnic pa nismo obravnavali predpon, npr. anti-, ampak le celotne besede, na katere se pripenjajo.

²⁸ Na ta način sledimo pristopu, ki ga uporabljata Leksikalna baza za slovenščino (Gantar in Krek 2011; Gantar 2016) in Slovar sodobnega slovenskega jezika (Gorjanc idr. 2015) in se nekoliko razlikuje od pristopa SSKJ (obe izdaji) ter NSSKJ (Gliha Komac idr. 2015).

V primeru različnih zapisov leme (recimo na ravni domače – prevzeto), npr. *džez* – *jazz* itd., smo v VSMS izdelali dve oz. več samostojnih gesel, če so imele posamezne oblike zadostno število pojavitev v korpusu. To odločitev smo sprejeli zato, ker predvidevamo, lahko v povezavi z različnimi zapisi pričakujemo tudi drugačne leksikalno-gramatične podatke na domala vseh ravneh leksikografske obravnave (npr. različne kolokacije, izbira registra, besedilnega tipa, stilne vrednosti, tvorjenje stalnih zvez, frazeologije ipd.).

Primere konverznosti, tj. ko beseda v stavku spremeni svoj skladenjski položaj in s tem prevzame tudi skladenjske in pomenske lastnosti, ki so za ta položaj značilne, smo reševali znotraj izhodiščne iztočnice. Na primer pri samostalnikih, ki pred drugimi samostalniki dobijo vlogo nesklonljivega prilastka, npr. **latino**, se ravnamo glede na prevladujočo vlogo; če npr. prevladuje pridevniška vloga, postanejo pridevniške iztočnice, konverzija v samostalnika pa je na ravni pomena označena z oznako »v samostalniški rabi.« V primeru, ko pa gre za samostalnike, ki se pred samostalnikom obnašajo kot pridevniki, ki se sklanjajo z ničto končnico (npr. **jeans**), pa pri relevantnem pomenu uporabimo oznako »v pridevniški rabi«.

Posamostaljene pridevnike tipa **dežurni** zaradi spremembe v kategoriji in paradigmi obravnavamo drugače, in sicer kot samostojne iztočnice. To velja tudi za primere, ko oblika posamostaljenega pridevnika v samostojnem pomenu ni predvidljiva oz. prekrivna z določno obliko pridevnika, kot je npr. **bolniška** v pomenu bolniškega dopusta. Na ta način odstopamo od pristopov, opisanih v Krek idr. (2013), Gorjanc idr. (2015), Gantar (2016) in tudi Gliha Komac idr. (2015), razlog pa je predvsem v optimizaciji beleženja podatkov v bazi in omogočanje enostavnejšega povezovanja z drugimi jezikovnimi viri.

2.1.1 Lastna imena

Za razliko od enojezičnih slovarjev so lastna imena za uporabnike dvojezičnih slovarjev bolj relevantna, ker so pogosto kontrastivno zanimiva že samo na ravni zapisa; to še sploh velja za madžarščino in slovenščino, kjer se pravila zapisa lastnih imen močno razlikujejo. Kot kažejo analize (npr. Bálint Čeh in Kosem 2017), obstoječi slovensko-madžarski slovarji v vključujejo skromno število lastnih imen in kulturološko zanimivih iztočnic, čeprav so le-ta za uporabnike relevantna.²⁹

V VSMS so tako vključena in prevedena vsa lastna imena (osebna, zemljepisna in stvarna), ki se kakorkoli ločijo od zapisa v madžarščini, npr.:

²⁹ Nenazadnje o tem priča že analiza e-zbirk izrazov, v katerih je od približno 12.000 enobesednih izrazov 6 % lastnoimenskih.

Jeruzalem → *Jeruzsálem*

Sicilija → *Szicília*

Madžarska → *Magyarország*

Poleg tega so vključena tudi lastna imena, ki se v zapisu ne ločijo, so pa v rabi pogosta in so zlasti relevantna za ciljne uporabnike ali pa so del zaprtega niza leksemov (npr. **Afrika** → *Afrika* kot ena izmed celin, **Jupiter** → *Jupiter* kot eden izmed planetov osončja ipd.).

Najpogostejša lastna imena so s področij geografije, zgodovine, astronomije, umetnosti in kulture. V slovar so umeščene tudi pomembnejše kulturološko zanimive iztočnice, tudi če zanje prevedek ne obstaja – v tovrstnih primerih ponudimo samo razlagalno ustreznico.

Razlago uporabljamo tudi v kombinaciji s prevedkom, in sicer pri manj znanih lastnih imenih, tj. večinoma pri tistih, ki presegajo srednješolsko znanje. Takšno razlago imenujemo razlagalno dopolnilo.

Abesinija → ZGODOVINA *Abesszínia (Etiópia egykori elnevezése)*

Panonija → ZGODOVINA *Pannónia (római tartomány)*

Sfinga → MITOLOGIJA *Szfinx, Sphinx (görög mitológiai lény)*

2.1.2 Homonimija in homografija

Pri obravnavi homonimije VSMS sledi Leksikalni bazi za slovenščino (Gantar in Krek 2011; Gantar idr. 2012; Gantar 2016) in Predlogu za izdelavo Slovarja sodobnega slovenskega jezika (Krek idr. 2013) in na njem temelječem Slovarju sodobnega slovenskega jezika (Gorjanc idr. 2015). Tako so kot homonimi oz. samostojne leme obravnavane »samo tiste besede, ki so izrazno prekrivne znotraj istih oblikoskladenjskih kategorij, in sicer v celotni sklanjatveni, spregatveni ter naglasni paradigmi« (Gantar 2016: 112), in sicer v kolikor se razlike pojavljajo na ravni besedne vrste, oblikoslovne in naglasne paradigme, ali samo naglasne paradigme. Vse druge prekrivne oblike na ravni leme so obravnavane kot večpomenske, torej znotraj enega gesla pod skupno iztočnico. Etimološki kriterij pri opredeljevanju homonimije ni

upoštevati, saj od uporabnika zahteva preveč predznanja in odločitev že pred samim iskanjem besede v slovarju (prim. Moon 1987: 89).³⁰

Kadar prihaja pri enakopisnih (ne pa tudi enakoglasnih) prekrivnih oblikah znotraj iste besednovrstne kategorije do spremembe paradigme zgolj na ravni naglasa, iztočnico v slovarju zapišemo z naglasom, npr. *částiti* in *častíti*. Enako velja za primere, ko prihaja do pisne, ne pa tudi naglasne prekrivnosti, npr. *poročèn* – *poróčen*; *omrežèn* – *omréžen*, pri osnovni obliki, ne pa tudi v celotni sklanjatveni ali spregatveni paradigmi.

Iztočnic, ki imajo prekrivno osnovno obliko in različno oblikoslovno paradigmo ter hkrati nimajo prekrivnih slovničnih kategorij, kot sta besedna vrsta ali slovnični spol, npr. *plesen* – samostalnik, *plesen* – pridevnik; *prst* – samostalnik (ženski spol) in *prst* – samostalnik (moški spol), nismo obravnavali znotraj večpomenskosti, ampak smo izdelali samostojni iztočnici.

2.1.3 Razvrstitev iztočnic

Razvrstitev iztočnic v digitalnem mediju ni več relevantna, velja pa razmišljati o razvrstitvi za morebitno tiskano različico VSMS. V tem primeru bi upoštevali abecedni vrstni red, in sicer po principu črke, torej velja dano zaporedje črk, ne glede na presledke med besedami.

Kadar imamo naglašene iztočnice, npr. pri homografih, je iztočnica z naglasom na prvem zlogu pred iztočnico z naglasom na drugem zlogu, naglašena iztočnica pred nenaglašeno ipd.

2.2 Podatki v geselski glavi

Podatki, ki jih lahko pripišemo iztočnici v vseh njenih pomenih in se načeloma ne spreminjajo (če pa se, je to izpostavljeno pri vsakem konkretnem pomenu oz. leksemu, bodisi enobesednem ali večbesednem), so predstavljeni v glavi oz. na začetku gesla (v tiskanih slovarjih se ta del imenuje zaglavje).

Iztočnice so prikazane v osnovni obliki, npr. samostalnik v imenovalniku ednine, pridevnik v imenovalniku moškega spola ednine, glagol v nedoločniku ipd. Praviloma so iztočnice navedene brez jakostnega naglasa, izjeme so le homografi, pri katerih je to glavni razločevalni kriterij. Tako imenovanih prvih stranskih oblik (gl. Koncept NSSKJ, Gliha Komac idr. 2015: 11) ne navajamo, saj bodo gesla ponujala direktno

³⁰ Za podrobno analizo obravnave homonimije v obeh izdajah SSKJ, v predlogu NSSKJ (Gliha Komac idr. 2015) in, primerjalno, Slovarju sodobnega slovenskega jezika (Gorjanc idr. 2015), glej Gantar (2015).

povezavo do Slovenskega oblikoslovnega leksikona ali podobnega vira s podatki o vseh oblikah besed.

V glavi so ponujeni tudi podatki o besedni vrsti iztočnice in pa pogostosti iztočnice v referenčnem korpusu. V bazi je podatek o pogostosti zabeležen v obliki absolutne ali relativne pogostosti, pri prikazu gesel v slovarju pa bo omenjena informacija prikazana na uporabniku bolj prijazen in predvsem informativen način, npr. v obliki zvezdic ali podobnih ikon, ki bodo nakazovale mesto iztočnice na frekvenčnem seznamu in predvsem v odnosu do drugih iztočnic v slovarju.

2.3 Geselski del

2.3.1 Pomen

Pomenski nivo je v VSMS hierarhično nadrejen vsem podatkom, ki jih beležimo v zvezi s posamezno besedo v iztočnici. Izhodišče za beleženje jezikovnih podatkov na pomenskem nivoju je vedno konkretni, na podlagi korpusne analize zabeleženi pomen besede. Pri tem izhajamo iz stališča do koncepta pomena, kot ga je prevzela Leksikalna baza za slovenščino, namreč da besede same na sebi nimajo pomenov, pač pa le pomenske tendence (Hanks 2009: 11), ki vplivajo na izbire v besedilnem okolju, to pa se odraža v stavčnih vzorcih, skladenjskih strukturah in kolokacijah, ter tudi v prevzemanju nekaterih slovničnih kategorij, kot je npr. (ne)števnost.

Pomenska členitev v VSMS je podana zgolj na enem nivoju, torej uporabljamo pomene, ne pa tudi podpomenov, ki jih sicer najdemo v enojezičnih slovarjih in leksikalnih bazah. Manjše pomenske odmike, ki bi jih drugače lahko ponazorili s podpomeni, v tem primeru ponazorimo z uporabo podpičja med prevodnimi ustreznici, uporabo oznake ali razlagalnih oz. prevedkovih dopolnil.

Pri določanju pomenske zgradbe besed smo uporabili tudi t. i. formalne kazalce večpomenskosti. Tako je bila upoštevana besednovrstna opredelitev besede, kjer se kot pomenskorazločevalne kažejo za posamezno besedno vrsto značilne slovnične kategorije (Vidovič Muha 2000: 32). Na primer pri samostalnikih so to predvsem spol in podkategoriji živost in človeškost, poleg tega števnost ter število, pri glagolih spremembe na ravni prehodnosti in v udeleženski zgradbi pomena ter glagolski vid, pri pridevniki in prislovih pa zlasti možnost stopnjevanja. Pomembni in hkrati najbolj očitni formalni pokazatelji večpomenskosti so kolokatorji, pri glagolih pa zlasti udeleženska zgradba, ki se spreminja glede na pomenske tendence posameznega glagola.

Z vsebinskega vidika je pri pomenski členitvi ključen kriterij predvsem prevodni del. Tako imamo lahko v slovenščini več pomenov, ki pa imajo v madžarščini vsi enake prevedke. To posledično pomeni, da se prvotna pomenska členitev, narejena za slovenski del gesla, lahko na predlog leksikografa tudi spremeni; v večini primerov gre za združevanje več pomenov v enega.

ananas (samostalnik)

PRVOTNA POMENSKA ČLENITEV:

1. (rastlina) *ananász*

nasad ananasa → *ananászültetvény*

2. (sadež) *ananász*

svež ananas → *friss ananász*

zrel ananas → *érett ananász*

KONČNA VERZIJA:

1. (rastlina in sadež) *ananász*

Poudariti je treba, da ponavljanje prevedkov še ne pomeni avtomatično potrebe po združevanju pomenov; ločene pomene npr. ohranimo, ko imamo zraven še neprekrivne prevedke, pomena kažeta povsem različne strukture, kolokacije ipd., je področje rabe precej različno (tudi v primeru, ko ni potrebe po eksplicitno uporabljeni področni oznaki) ali pa obstaja razlika v zaznamovanosti.

jajce (samostalnik)

1. (pri pticah in plazilcih) *tojás*

kokošje jajce → *tyúktojás*

prepeličje jajce → *fürjtojás*

gnezdo z jajci → *fészek tojásokkal*

2. (v kulinariki) *tojás*

kuhano jajce → *főtt tojás*

stepeno jajce → *felvert tojás*

3. (predmet ovalne oblike) *tojásszerű, tojásalakú, tojás formájú*

4. (testisi; moda) **V MNOŽINI VULGARNO** *here*

brcniti koga v jajca → *tökön rúg valakit*

5. (nekaj slabega) **SLABŠALNO** *nyavalya*

benigen (pridevnik)

1. (v medicini) *jóindulatú*

benigni tumor → *jóindulatú daganat*

benigni izrastek → *jóindulatú kinövés*

2. (neškodljiv) *jóindulatú*

benigna razlaga → *jóindulatú magyarázat*

2.3.1.1 Razvrščanje pomenov

Pri razvrščanju pomenov slovenskih iztočnic sledimo obstoječi praksi dvojezičnih slovarjev, zlasti pristopih korpusno temelječega Velikega angleško-slovenskega slovarja Oxford-DZS (Krek ur. 2005/2006). Tako so praviloma najprej ponujeni pomeni, ki na podlagi analize korpusnega gradiva predstavljajo osrednje in hkrati najpogostejše rabe. Sledijo jim stilno in časovno označeni pomeni, nato pa še področno označeni pomeni. Vseeno načelo pogostosti prevlada nad načelom označenosti, torej je zelo pogost področni pomen lahko podan celo pred neoznačenim pomenom, kot pri spodnjem primeru iztočnice **satelit**, kjer je pomen iz astronomije precej pogostejši kot ostali trije splošni pomeni za njim.

satelit (samostalnik)

1. (vesoljska naprava) *mesterséges hold, műhold, szatellit*
2. (nebesno telo) **ASTRONOMIJA** *mellékbolygó, hold*
3. (zvočnik) *szatellit hangfal, szatellit hangszóró*
4. (manjša država ali organizacija) *csatlós, csatlós állam*
5. (teniško tekmovanje) *szatellit (teniszmérkőzés)*

2.3.1.2 Pomenski indikatorji

Naloga pomenskih indikatorjev je na kratko in na razumljiv način opisati ali bolje določiti pomenski obseg ali pomensko področje obravnavane besede, pri večpomenskih besedah pa poleg tega vzpostaviti še razliko glede na druge pomene. Poleg tega so indikatorji sestavni del pomenskega menija, za katerega velja, da je samostojna slovarska informacija in mora biti razumljiv sam na sebi. Pri enopomenskih geslih indikatorjev praviloma ne uporabljamo, razen če gre za redkejša in predvidoma uporabnikom manj znane besede ali za homonimne pare.

Najpogosteje so uporabljeni sinonimni indikatorji, pri čemer težimo k temu, da se enobesedna sopomenka v čim večji meri, tj. glede na kolokacijske, stilne, zvrstne, pogostnostne in druge lastnosti, približa besedi v iztočnici, hkrati pa smo pozorni na to, da sinonimni indikator ne zajame pomenskega polja iztočnice preširoko ali preozko. Možna je uporaba tudi dveh sinonimnih indikatorjev za en pomen, v kolikor to pripomore k lažji prepoznavi pomena (se pa nizanju sinonimov načeloma izogibamo).

jedek (pridevnik)

1. (ki razjeda)
2. (intenziven; oster)
3. (zajedljiv, piker)

Druga skupina indikatorjev so nadpomenke, ki jih uporabljamo predvsem pri pomenskih opisih samostalniških iztočnic.

čili (samostalnik)

1. (začimba)
2. (jed)

Tretjo skupino predstavljajo nanašalni indikatorji, ki se uporabljajo predvsem pri pridevniških iztočnicah. Pogosto vlogo indikatorja lahko prevzame tudi slovarska oznaka, zlasti področna (npr. v športu).

Pri indikatorjih sta ključni razumljivost in pomenska razlikovalnost, zato smo, če je bilo to najbolj optimalno, za različne pomene posamezne iztočnice uporabili različne tipe indikatorjev. V spodnjem primeru za **arhivski** imamo tako uporabljena dva nanašalna indikatorja, indikator-oznako in sinonimni indikator.

arhivski (pridevnik)

1. (o ustanovi)
2. (o zbirki)

3. (V RAČUNALNIŠTVU)

4. (star)

V redkih primerih se poslužujemo tudi kombiniranja različnih tipov indikatorjev znotraj posameznega pomena, je pa zaradi teženja h kratkosti indikatorjev to manj zaželeno. Na splošno se pri ubeseditvi indikatorjev izogibamo rabi večpomenskih besed in besed v katerem od obrobni ali atipičnih pomenov, manj zastopanih besednih oblik in besed, ki so redko zastopane v običajni rabi. Izogibamo se tudi navajanju indikatorjev v nikalni obliki, pa tudi indikatorjem, ki izražajo frazeološki ali metaforični pomen. Glavno vodilo so čim preprostejše ubeseditve, kar skušamo doseči z izbiro indikatorjev, ki najbolje asociirajo ključno pomensko lastnost iztočnice, izražajo ustrezno besedno vrsto in glagolski vid iztočnice (izjema so nanašalni indikatorji, ki te možnosti nimajo) in so hkrati čim bolj kratki in razumljivi.

Indikatorji so namenjeni predvsem uporabnikom, ki uporabljajo VSMS pri tvorjenju madžarskih besedil, tj. maternim govorcem slovenščine. Pomagali naj bi pri ločevanju med pomeni in predvsem pri identifikaciji relevantnega pomena. Iz tega razloga so indikatorji ponujeni samo v slovenskem jeziku, izjema so le indikatorji-oznake, ki so tako kot vse oznake ponujeni tako v slovenščini kot madžarščini.

2.3.2 Prevedki iztočnice

V VSMS uporabljamo dva tipa prevedkov oz. prevodnih ustreznice: neposredne prevedke in razlagalne ustreznice. Neposredni prevedek ima vedno prednost pred razlagalno ustreznico.

2.3.2.1 Neposredni prevedki

Znotraj posameznega pomena je najprej ponujen niz prevodnih ustreznice. Ključno pravilo pri oblikovanju niza prevedkov je, da je na prvem mestu ponujen najbolj tipičen in pogost prevedek iztočnice, ki pomensko in stilno ustreza iztočnici. Raba madžarskih prevedkov preverjamo v korpusih, pa tudi v obstoječih enojezičnih slovarjih. Prvi ponujeni prevedek mora zadovoljiti večino uporabnikovih potreb po prevedkih oziroma mora biti uporaben v večini besedil.

Pri navajanju prevedkov upoštevamo tudi posebnosti posameznih besednih vrst, pri čemer izhajamo iz prakse obstoječih dvojezičnih slovarjev. Tako npr. pri glagolskih iztočnicah madžarske prevodne ustreznice podajamo v 3. osebi ednine. npr. **hiteti** → **siet**, **igyekszik**; **cveteti** → **virágzik**.

Prvemu oz. osnovnemu prevedku sledijo prevedki, ki se pokažejo v posebnih kontekstih, njihova raba pa mora biti vedno potrjena v kolokacijah in/ali zgledih.

šumeč (pridevnik)

1. (ki povzroča šumeč zvok) *suhogó; zúgó, zuhogó*

šumeč potok → *zúgó patak*

šumeč slap → *zuhogó vízesés*

šumeč gozd → *suhogó erdő*

šumeče listje → *suhogó levelek*

Sopomenske prevedke ločimo z vejico, pomensko (zlasti stilno ali konotativno) različne prevedke znotraj istega pomena pa ločujemo s podpičji. V zgornjem primeru za **šumeč** je za prevedkom *suhogó* podpičje, ker se prevedek uporablja v drugem sobesedilu kot prevedka *zúgó* in *zuhogó*. Uporaba prevedkov se potrjuje tudi s prevajanjem slovenskih zgledov v madžarski jezik in s pregledom rabe prevedka v madžarskih korpusih.

Dodajanje prevedkov je koristno tudi pri razločevanju pomenov, ki imajo en skupen prevedek. Če se v enem od naslednjih pomenov ponovi glavni prevedek iz prvega oz. prejšnjega pomena, je praviloma na prvem mestu tisti prevedek, ki je ločevalen.

arhitektura (samostalnik)

1. (veda in dejavnost) *építészet, architektúra*

...

3. (struktura, ureditev) *felépítés, architektúra*

Včasih je zato, da je pomen prevedka jasen, dodana tudi mini razlaga, ki lahko funkcionira tudi kot možen prevedek. Na primer pri področno označenem pomenu, katerega prevod je tujka, je pogosto navedena tudi madžarska ustreznica.

aorta → (ANATOMIJA) *aorta, főútőér, főverőér*

anoreksija → (MEDICINA) *anorexia, kóros étvágytalanság*

Če je tujka pogostejša ali je pomen vezan na določeno področje, **ars poetica** → *ars poetica, költői hitvallás* ali **agraren** → *agrár-, mezőgazdasági*, najprej navedemo tujko, nato pa še madžarski domači izraz. Če je domači izraz precej pogostejši, je naveden na prvem mestu, nato sledi tujka, npr. **aplikacija** → *alkalmazás, applikáció*. Včasih je ustaljen domači madžarski prevedek za iztočnico, ki je v slovenščini tujka, medtem ko

je tujka v rabi zelo redka oz. se sploh ne uporablja (podatek preverjamo v korpusu); v takšnih primerih tujke med prevedki sploh ne navedemo, npr. **ablacija** → *gleccserkopás* ali **aranžma** → *turisztikai kínálat*.

V nekaterih primerih je prevedku lahko dodano tudi t. i. razlagalno dopolnilo, ki je ponujeno v oklepaju, recimo pri manj znanih lastnih imenih.

Klavdij → ZGODOVINA Claudius (*római császár*)

Lombardija → GEOGRAFIJA Lombardia (*olaszországi táj*)

Nekoliko drugačno vlogo od razlagalnega dopolnila ima prevedkovo dopolnilo, pri katerem gre za sestavni del prevedka, ki pa ni nujen oz. se v jezikovni rabi vedno ne realizira. Prevedkovo dopolnilo se lahko pojavlja pred ali za prevedkom. Primer rabe prevedkovskega dopolnila so samostalniške iztočnice, ki se nanašajo npr. na pripadnike narodnosti in poklic:

Madžar → magyar (férfi), **Slovenec** → szlovén (férfi)

Madžarka → magyar (nő), **Slovenka** → szlovén (nő)

Madžarščina namreč nima slovnične kategorije za spol (isti prevedek se uporablja za moško in žensko obliko), zato pri madžarskih prevedkih dodamo prevedkovo dopolnilo v oklepaju. V določenih primerih, npr. pri nekaterih poklicih, pa je prevedkovo dopolnilo že sestavni del enobesednega prevedka:

učiteljica → tanárnő

vzgojiteljica → óvónő

Po drugi strani vedno pazimo, da prevedki ne vsebujejo nepotrebnih dodatnih besed oz. podvajanja. Prevedek namreč ni prevedena slovenska razlaga oz. njena interpretacija, kar je razvidno v sledečih primerih:

- prevedek za **brizgalko** (kot napravo za vbrizganje tekočine v žilo) ni *injekciós fecskendő*, temveč samo *fecskendő*;
- prevod **Severnice** je *Sarkcsillag*, ne pa *Északi Sarkcsillag*.
- prevedka za **senat** (v starem Rimu) sta *szenátus* in *senatus*, ne pa tudi Római Szenátus (dobesedno prevedeno »rimski senat«).

2.3.2.2 Razlagalne ustreznice

Kadar ni ustrezne prevodne ustreznice v madžarščini, bodisi da gre za kulturološko ali splošno leksikalno praznino, je navedena razlagalna ustreznica. V slovarski bazi je razlagalna ustreznica beležena drugače kot neposredni prevedki, podobno je tudi v

slovarju prikaz razlagalnih ustreznice jasno razlikovan od prikaza neposrednih prevedkov (zaenkrat je predvidena uporaba kurzive).

mandrač → *belső halászkikötő*

Slovenska beseda *mandrač* pomeni majhno, notranje ribiško pristanišče. Leksem v madžarščini nima prevodne ustreznice, zato uporabimo razlagalno ustreznico. Pri kolokacijah pa lahko ohranimo citatno prevzeto besedo in dodamo razlagalno ustreznico v oklepaju, npr. **Piranski mandrač** → *Pirani mandrač (belső kikötő)*.

Če ima en pomen dva pomenska odtenka (pogosto je prevedek enak), ki zahtevata pojasnilo, je med razlagama podpičje, podobno kot pri neposrednih prevedkih, npr. Saturn kot planet in rimski bog:

Saturn → *Szturnusz, Saturnus (bolygó; római istenség)*

Kadar imamo samostojno razlagalno ustreznico, je ta lahko tudi nekoliko daljša, vendar pa se izogibamo rabi podrednih stavkov in deležnikov. V primeru, ko imamo kombinacijo neposrednega prevedka in razlagalne ustreznice, razlagalna ustreznica dobi vlogo razlagalnega dopolnila, ki vedno sledi neposrednemu prevedku, ni pa od njega ločena z ločilom, temveč je navedena v oklepaju. Razlagalno dopolnilo je praviloma kratko – ima le bistvene pomenske elemente. To še zlasti velja za razlago v funkciji identifikacije (npr. pri lastnih imenih). Težimo k jedrnatosti, torej k nominalizaciji, neosebni oblikam in predložnim zvezam.

Pri določenih geslih, večinoma pri kulturoloških pojmi, je kot prevedek uporabljen približek iz madžarske kulture, zraven pa je navedeno razlagalno dopolnilo, na primer:

okrajno sodišče → **PRAVO** járásbírótság (*elsőfokú bíróság*)³¹

okrožno sodišče → **PRAVO** körzeti bíróság (*másodfokú bíróság*)³²

šola v naravi → **ŠOLSTVO** erdei iskola (*többszapos kirándulás az oktatás folyamatában*)³³

³¹ Prevod razlagalnega dopolnila v slovenščini je »prvostopenjsko sodišče«.

³² Prevod razlagalnega dopolnila v slovenščini je »drugostopenjsko sodišče«.

³³ Prevod razlagalnega dopolnila v slovenščini je »večdnevni izlet v okviru pouka«.

2.3.3 Kolokacije in skladenjske zveze

Pod vsakim pomenom najdemo kolokacije, včasih pa tudi razširjene kolokacije in skladenjske zveze. Kolokacije so kombinacije dveh ali treh besed (pri tribesednih kolokacijah je ena od besed predlog, npr. **brskati po smeteh** → a szemétben turkál). Razširjene kolokacije so kolokacije, pri katerih je dodan ustaljen leksikalni element, npr. **organizirati okroglo mizo** → kerekasztal-beszélgetést szervez, **govoriti (slovenski, angleški...) jezik** → beszél (szlovén, angol...) nyelven. V primeru, ko ima razširjena kolokacija niz leksikalnih elementov, za VSMS izberemo kontrastivno najbolj zanimivega ali, če ni kontrastivno zanimivih, najpogostejšega v nizu.

Kolokacije in razširjene kolokacije so v slovarski bazi grupirane glede na skladenjsko strukturo, npr. pridevnik + samostalnik, glagol + samostalnik v tožilniku, v slovarskem prikazu pa te informacije ne podajamo.

Skladenjske zveze³⁴ so večbesedni nizi, ki nimajo statusa leksikalne enote in za katere veljajo naslednja načela (Krek idr. 2013):

- so (relativno) pomensko prozorne besedne kombinacije, kar pomeni, da je njihov pomen (načeloma) kombinacija pomenov posameznih sestavin.
- hkrati so dovolj strukturno trdne (ustaljene), kar pomeni, da so posamezni elementi leksikalizirani in jih ni mogoče nadomestiti z drugimi oz. da je niz nadomestljivih elementov omejen, npr. *temperatura pade pod ničlo/ledišče*.
- imajo pogosto tudi semantično in/ali oblikovno predvidljivo prosto mesto znotraj stavčne zgradbe, v kateri se tipično pojavljajo.
- imajo lahko za razliko od kolokacij in razširjenih kolokacij znotraj svoje zgradbe več kolokabilnih nizov, npr. *otroci v starosti od [x] do [x] let; odigrati [tekmo, dvoboj] [x.] kroga končnice*.

Pri skladenjskih zvezah s prostim mestom zaradi delne abstraktnosti zveze navadno poleg prevedka ponudimo tudi zgled, v primeru kolokabilnih nizov pa že v izhodiščni zvezi (in prevodu) namesto x lahko ponudimo konkretno vrednost iz enega izmed zgledov.

alkohol (samostalnik)

2. (kemična spojina) *alkohol*

5-odstotni alkohol → 5 *száزالékos alkohol*

1,6 promila alkohola v krvi → 1,6 *ezrelék véralkoholszint*

³⁴ Kot ugotavlja Gantar (2016: 330), so skladenjske zveze razmeroma pogost jezikovni pojav, ki se mu v slovarskih opisih ni mogoče izogniti, so pa zaradi svoje skladenjske raznolikosti in relativne transparentnosti v literaturi različno obravnavane in imajo tudi različna poimenovanja.

Če del skladijske zveze predstavlja niz elementov, podobno kot pri razširjeni kolokaciji vzamemo s prevodnega vidika kontrastivno najbolj zanimivega oz. v rabi najpogostejšega. Na primer pri iztočnici **jajce** je bila zabeležena skladijska zveza [vonj, smrad, zadah] po gnilih jajcih. Podrobnejša analiza pokaže, da ima zveza z *vonj* 131 zadetkov v korpusu Gigafida, ostali dve pa samo sedem (*smrad*) oz. štiri (*zadah*). Posledično smo prevedli skladijsko zvezo **vonj po gnilih jajcih** → záptojásszag, smo pa hkrati ponudili tudi podobno, na smrad navezano glagolsko zvezo, in sicer **smrdeti po gnilih jajcih** → búzlik, mint a záptojás.

Pri izbiranju kolokacij in skladijskih zvez upoštevamo predvsem kontrastivni vidik (prevod besedne zveze se razlikuje ali pravopisno ali drugače od strukture izhodiščne besedne zveze, zahtevnost prevoda besedne zveze, pravopisno težji pari), zaradi vloge slovarja pri enkodiranju v madžarščino pa tudi pogostost oz. tipičnost večbesednih enot v izhodiščnem jeziku, tj. slovenščini. Kot primer navajamo tri pogoste kolokacije pri iztočnici **brisača** (pripomoček za brisanje):

kopalna brisača → fürdőlepedő

papirnata brisača → papírtörölköző

mokra brisača → nedves törölköző

Prva kolokacija je kontrastivno zanimiva, saj se prevodna ustreznica popolnoma razlikuje od prevedkov *törölköző* in *törölköző*, ki je ponujen za ta pomen iztočnice. Druga kolokacija je zanimiva zaradi zapisa, saj je madžarski prevod dvobesedne slovenske kolokacije ena beseda. Tretja kolokacija je prevodno manj zanimiva, je pa zaradi svoje pogostosti v slovenskem jeziku relevantna predvsem za uporabnike, ki jim je madžarščina drugi/tuji jezik.

Količina kolokacij, razširjenih kolokacij in skladijskih zvez, ki jih prevedemo, je odvisna od:

- Pogostosti iztočnice. Pri iztočnicah, ki so v korpusu bolj pogoste, ponudimo več kolokacij in/ali skladijskih zvez.
- Večpomenskosti iztočnice. Pri enopomenskih besedah praviloma prevedemo manj kolokacij kot pri večpomenskih, kjer želimo zaradi lažjega pomenskega razlikovanja ponuditi vsaj nekaj kolokacij na pomen.
- Splošnosti oz. terminološkosti iztočnice. Pri splošnejšem besedju potrebujemo nekoliko več informacij o okolici in rabi, medtem ko je pri terminoloških iztočnicah oz. pomenov to manj potrebno.
- Števila prevedkov. Če smo za iztočnico oz. njen pomen ponudili več prevedkov, je vedno zaželeno pokazati kontekste rabe posameznih prevedkov.

2.3.4 Večbesedne leksikalne enote

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

V bazi VSMS stalne zveze in frazeologijo beležimo pod enobesednimi iztočnicami, kar tudi pomeni »podvajanje« informacij v različnih geslih, katerih (polnopomenske) iztočnice so sestavni del večbesednih leksikalnih enot. Iz tega razloga se vse identificirane stalne zveze in fraze sproti beleži in skupaj s prevodi avtomatsko prenaša v gesla ostalih sestavnih delov, s čimer se tudi izogibamo podvajanju leksikografskega dela.

V digitalnih verzijah VSMS bo uporabnikom omogočeno tudi iskanje po večbesednih iztočnicah, pri čemer bodo imeli možnost izbire med gesli enobesednih iztočnic, ki sestavljajo stalno zvezo ali frazo. V tiskani različici VSMS pa je pri stalnih zvezah predviden premik na raven iztočnice, pri čemer razvrstitev sledi abecednemu redu, pri frazah pa umestitev pod iztočnico, ali več iztočnic, ki predstavlja njen glavni sestavni element oz. v kateri bi uporabnik frazo najverjetneje iskal.

2.3.4.1 Stalne zveze

V sodobni leksikografski praksi je navajanje stalnih besednih zvez, ki sugerirajo samostojni pomen ali pa izstopajo po pogostosti rabe, kot samostojnih iztočnic ali enot iztočnic, ki imajo vse sestavine pomenov samostojnih iztočnic (npr. pomenski indikator, oznako ipd.), že ustaljena praksa.

Kot stalne besedne zveze so tako v VSMS v ločenem razdelku navedene večbesedne enote (gl. tudi Krek idr. 2013), za katere je bilo na podlagi analize korpusnih podatkov mogoče ugotoviti:

- da so strukturno in pomensko ustaljene oz. da njihove variantne sestavine pripadajo istemu/sorodnemu pomenskemu polju (**morilska čebela / čebela ubijalka** → *gyilkos méh*; **prikriti kurikulum / skriti kurikulum** → *rejtett tanterv*);
- da so vsaj delno pomensko neprozorne – tj. imajo samostojni pomen (*spalna srajca, glasbeni stolp, osnovna šola; veliki petek*), zaradi česar jih je smiselno obravnavati na enak način kot enobesedno leksiko;
- da njihov celostni pomen (ne pa nujno tudi motivirajoči pomen oz. način pomenskega združevanja sestavnih elementov) ni metaforičen, po čemer se ločujejo od frazeoloških enot (*tiskarski skrat, rumeni tisk*);

- da so lahko, kot tudi sicer leksikalne enote, večpomenske (*topla greda* – 1 ‘del vrta’; 2 ‘proces v zemeljskem ozračju’);
- da je njihova raba pogosto omejena na določeno terminološko oz. specializirano področje (**kisli dež** → *savas eső* – ekologija; **pomožni glagol** → *segédige* – jezikoslovje; **notranji glas** → *belső hang* – psihologija);

Stalne zveze imajo pogosto ob dobesečnem (besednozveznem ali kolokacijskem) še osamosvojeni besednozvezni pomen (*orjaški dežnik* – ‘velik dežnik’ : ‘goba’) oz. konkretnega referenta (**rdeči križ** → *vöröskereszt*). V takšnih primerih, zlasti ko gre za pogosto kolokacijo, prevedemo tako kolokacijo kot stalno zvezo, saj s tem uporabniku pomagamo razlikovati med obema rabama.

Pri stalnih zvezah tako kot pri pomenih enobesednih iztočnic lahko ponudimo tudi slovenski indikator, pri čemer je pri terminoloških stalnih zvezah kot indikator ponujena področna oznaka. Prevodni ustreznici po potrebi dodamo tudi razlagalno dopolnilo v oklepaju za prevedkom, npr.

zlati cepin → ALPINIZEM *arany jégcsákány (nemzetközi díj)*

Stalne zveze imajo kot samostojne leksikalne enote pogosto tudi lastno kolokabilno okolje (**učinek tople grede** → *űvegházhatás*, **povzročati toplo gredo** → *űvegházhatást okoz*); v VSMS so podane in prevedene najpogostejše kolokacije oz. je takšno kolokabilno okolje ponazorjeno z zgledi.

Znotraj geselske strukture VSMS so stalne zveze beležene na dva načina. V primeru izkazane pomenske povezave med pomenom zveze kot celote in katerim od pomenov besede v iztočnici, smo zvezo navedli v okviru danega pomena. V primeru, kjer take pomenske povezave na podlagi analize jezikovne rabe ni bilo mogoče ugotoviti, pa smo stalne zveze navajali kot samostojne, od posameznih registriranih pomenov iztočnice, neodvisne enote, npr. **datelj** (sadež) in **morski datelj** (školjka).

V digitalnih različicah slovarja, zlasti spletni, se tak način obravnave kaže v možnosti, da se ob odprtih pomenih, ki vsebujejo stalne zveze, ustrezno filtrira tudi seznam vseh stalnih zvez: tako so na ravni celotnega gesla lahko prikazane vse stalne besedne zveze, na ravni posameznih pomenov, ki jih uporabnik odpre, pa le tiste, ki se nanj nanašajo. Pri stalnih zvezah znotraj registriranih pomenov na ta način v slovarju (in slovarski bazi) ohranjamo tudi informacijo o pomenu iztočnice, ki je z besedno zvezo v pomenski soodvisnosti.

2.3.4.2 Frazologija

Za razliko od stalnih zvez so frazeološke enote vedno navedene v ločenem razdelku na koncu gesla, neodvisno od posameznih pomenov besede v iztočnici. Pri identifikaciji frazeoloških enot si poleg korpusne analize pomagamo tudi z obstoječimi

slovarji, kot so npr. *SSKJ1 in SSKJ2, Slovar slovenskih frazemov* ipd., vendar pa v VSMS vključujemo samo frazeološke enote, ki so izpričane v korpusnem gradivu.

Kot vir za iskanje prevodov frazemov uporabljamo obstoječe slovensko-madžarske slovarje, madžarske enojezične slovarje, slovar madžarskih frazemov, pa tudi korpuse.

Frazeološke enote pojmujeemo v skladu s frazeološko teorijo, ki mednje šteje tiste enote besedišča, ki glede na svoje sestavne elemente izkazujejo samostojni, od sestavnih elementov neodvisni – t. i. frazeološki pomen (Čermak 1985).³⁵ Ker pa ta pogoj velja tudi za prepoznavanje stalnih besednih zvez, smo za frazeološke enote upoštevali še metaforičnost (v nekaterih teorijah pojmovano kot ekspresivnost, slikovitost ipd.) pomena enote kot celote (gl. npr. Vidovič Muha 2000; Gantar 2016). Med frazeološkimi enotami, ki ustrezajo zgornjim merilom, lahko izpostavimo dva osnovna tipa:

- frazeološke enote, ki opravljajo v stavku vlogo stavčnega člena, tj. pretežno, a ne nujno, z zgradbo nestavčne besedne zveze, npr. **rešilna bilka** → *megmentő szalmaszál*, **trojanski konj** → *trójai faló*, **princ na belem konju** → *herceg fehér lovon* ipd.;
- frazeološke enote, ki so same na sebi »zaključena« besedila in imajo navadno strukturo stavka. Sem sodijo zlasti pregovori in reki (**dobro blago se samo hvali** → *jó bornak nem kell cégér*; **ljubezen je slepa** → *a szerelem vak* itd.) ter izrazi s pretežno pragmatično vlogo, ki se kaže v nanašanju na vsebino sporočnega, npr. **če me spomin ne vara** → *ha nem csal az emlékezetem*, ali v vrednotenju udeležencev, okoliščin, splošnih nazorov in prepričanj, vezanih na govorca oz. udeležence v govorni situaciji, npr. **kaj iščeš/iščete tukaj?!** → *mit keresel/mit kerestek itt?!*, itd. V zadnjo skupino prištevamo tudi ustaljene konverzacijske formule in izraze, kot so pozdravi, opozorila, kletvice ipd.: **čestitamo** → *gratulálunk*, **hvala bogu** → *hála istennek!* ipd.

Tipična značilnost frazeoloških enot je, kljub »pregovorni« pomenski in strukturni trdnosti, variantnost, in sicer tako pri posameznih sestavinah kot pri besednih oblikah in »zunanjih« udeležencih:

luč na koncu tunela/predora → *fény az alagút végén*

pasti v kremplje komu/čemu/koga → *valakinek/valaminek a karmai közé kerül*

Variante v VSMS beležimo znotraj ene same frazeološke enote. V tem primeru govorimo o variantnosti besednih oblik, ki so posledica oblikoslovnega prilagajanja

³⁵ Za več o frazeologiji, gl. Kržišnik (1994, 1996) in Gantar (2007).

znotraj konkretne skladenjske kombinacije. Tipični primeri so, ko katera od sestavin frazeološke enote, npr. variantni predlog, glagol ipd., zahteva drugačno sklonsko obliko določila, npr. **stisniti koga v kot** → *sarokba szorít valakit* in **stisniti koga ob zid** → *falhoz szorít valakit*, **zaiti v slepo ulico** → *zsácutcába jut* in **znajti se v slepi ulici** → *zsácutcában találja magát*, ali ko variantnost oblike vpliva na razporeditev udeleženskih vlog, ki jih predvideva frazeološka enota: **srce se stisne komu** → *összeszorul valakinek a szíve* in **kaj stisne srce komu** → *valami összeszorítja a szívét*.

Nekoliko drugače v VSMS obravnavamo pretvorbene možnosti (npr. različne skladenjske oz. stavčnočlenske funkcije frazeoloških enot; različne realizacijske možnosti, zlasti posamostaljenje; različno število in razporeditve udeležencev; tipične realizacije v kateri od upovedovalnih oz. naklonskih možnosti, kot so veelnost, zanikanje ipd.; fakultativne sestavine). V takšnih primerih se pri prevajanju nagibali k izbiri najbolj tipične oziroma pogoste variante oz. v izjemnih primerih več variant (vse morajo biti pogoste in tipične).

Za frazeme vedno skušamo ponuditi prevodno ustreznico, če pa ne obstaja, podamo približek in kratko razlago. Če ima frazem več pomenov (npr. **v lastnem aranžmaju** → *saját szervezésben*; *saját rezsiben*), prevedke ločimo s podpičji ter skušamo ponuditi zgled za vsakega od prevedkov oz. nizov.

Pri frazeoloških enotah, ki so same na sebi zaključena besedila (npr. pregovori), ne ponujamo prevedenih zgledov (lahko pa ponudimo samo slovenske ali samo madžarske zglede).

Sreča je opoteča. → *A szerencse forgandó.*

Kri ni voda. → *A vér nem válik vízzé.*

Zdrav duh v zdravem telesu. → *Ép testben ép lélek.*

Ni vse zlato, kar se sveti. → *Nem mind arany, ami fénylik.*

Obleka ne naredi človeka. → *Nem a ruha teszi az embert.*

2.3.5 Zgledi

Z zgledi ponazarjamo rabo vseh elementov slovarskega gesla v konkretnem kontekstu oz. naravnem okolju in hkrati tudi potrjujemo podatke, ki so na voljo na višjih nivojih gesla: pomen, kolokacije, skladenjske zveze, pa tudi stalne zveze in frazeološke enote. Pri prevajanju zgledov imajo prednost kontrastivno zanimivi zgledi rabe, zlasti takšni,

kjer iztočnice ni mogoče prevesti z ustreznico iz prevodnega niza. Pri navajanju prevedenih zgledov sicer sledimo sledeči hierarhiji:

- najprej so navedeni zgledi, katerih prevod je sorazmerno enostaven in ga lahko izberemo med prevodnimi ustreznici v nizu oz. je edini prevod (npr. pri kolokacijah);
- sledijo zgledi, kjer sobesedilo opazneje vpliva na prevodno ustreznico;
- nazadnje so ponujeni zgledi, kjer iztočnice ni mogoče prevesti s podano prevodno ustreznico.

Pri pripravi dobrih zgledov in njihovih prevodov sledimo ustaljenim načelom leksikografske prakse (Atkins in Rundell 2008; Fox 1987; Kosem 2015a):

- Naravnost oz. avtentičnost, kar zagotovimo z izbiranjem zgledov iz korpusov.
- Tipičnost. Načelo veleva, da morajo slovarski zgledi pokazati tipično rabo iztočnice z vidika konteksta, skladnje, frazeologije in kolokacij. S sodobnimi orodji, kot sta besedne skice in GDEX v Sketch Enginu, si lahko pri iskanju tipičnih zgledov precej pomagamo, saj z njihovo pomočjo poiščemo tipične slovnične strukture, kolokacije in koligacijske lastnosti iztočnice (npr. v kateri obliki ali sklonu se beseda z določenim kolokatorjem najpogosteje pojavlja).
- Informativnost. Izogibamo se kopičenju prevelikega števila zgledov, ki ilustrirajo enake ali podobne informacije, kar velja predvsem za prevodne ustreznice. Tako imamo lahko za kolokacije in podobne elemente več zgledov, če se izkaže, da je v njih raba prevodne ustreznice različna.
- Razumljivost dosežemo z izogibanjem kompleksnim strukturam, redki in zahtevni leksiki, zaimkom, deiktom in ostalim elementom, ki zahtevajo seznanjenost s širšim kontekstom.
- Celostavčnost. Izbiramo celostavčne zglede, daljšim nizom besed oz. iztržkom se izogibamo, saj to obliko predstavljajo že večbesedni elementi v geslu.
- Pri izbiri zgledov smo biti družbeno občutljivi in odgovorni, npr. praviloma ne vključujemo lastnih imen (v nekaterih primerih so lahko izjema znane javne osebnosti) ipd.
- Modificiranje zgledov praviloma ni zaželeno, a v odsotnosti dobrih krajših zgledov se poslužujemo predvsem krajšanja daljših zgledov z izpuščanjem nerelevantnih oz. za razumevanje nebitvenih delov, kot so odvisniki ali vrinjeni stavki. Krajšanje je dejansko velikokrat povsem legitimno z vidika informativnosti, saj stavki večkrat vsebujejo dele, ki so nepotrebni oz. nerelevantni, če niso podani v širšem kontekstu besedila. Če se le da, se izogibamo poenostavljanja kompleksnih struktur in zamenjave besed ali besednih zvez s pogostejšimi ali ustrežnejšimi poimenovanji (več o tem gl. Kosem 2015a).

V digitalnih izdajah slovarja, predvsem spletni, kjer je prostorskih omejitev manj, bodo poleg prevedenih slovenskih zgledov ponujeni še neprevedene slovenske zglede in avtentične madžarske zglede. Neprevedeni slovenski zgledi vključujejo predvsem nekontrastivno zanimive zglede za kolokacije, skladijske zveze ipd., tj. zglede, v katerih se prevodna ustreznic kolokacije, skladijske zveze ipd. tudi v zgledu pojavlja v eni svojih oblik. Madžarski zgledi, pridobljeni iz referenčnega korpusa madžarščine z orodje GDEX, ponazarjajo avtentično rabo prevedkov pomenov, kolokacij, skladijskih zvez, stalnih zvez in frazeoloških enot.

Dodatni zgledi za slovenski in madžarski del slovarskih gesel omogočajo nadaljnjo kontekstualizacijo rabe slovenskih iztočnic oz. madžarskih prevodnih ustreznice, kar je zlasti nematernim govorcem v pomoč pri bogatenju znanja jezika (npr. nematernim govorcem madžarščine pri učenju madžarščine). Poleg tega dodatni zgledi za oba dela istega elementa slovarskega gesla, npr. kolokacije in njene prevodne ustreznice, lahko ilustrirajo zelo podoben oz. primerljiv kontekst, kar omogoča vsaj delno prevedljivost sobesedila in povečuje didaktično oz. obvestilno vrednost tovrstnih zgledov.

2.4 Slovarske oznake

Oznake so različni tipi slovarskih pojasnil, ki uporabnike opozarjajo, da ima beseda, besedna zveza, njen pomen itd. določene slovnične omejitve, da se nanaša na določeno časovno obdobje, tip besedila, regionalne posebnosti, da se uporablja na določenem strokovnem področju, da izraža določen odnos do vsebine ali udeležencev ipd. (Gantar in Kosem 2013; Kosem 2015a; Gantar 2016).

Pri VSMS z oznakami označujemo samo izhodiščni del, torej v našem primeru slovenska gesla, zglede in pomene, ne pa madžarskih prevodov. Obstajajo sicer določene izjeme, recimo ko želimo opozoriti na razlike med samimi prevedki, npr. pri ločevanju različne stilne zaznamovanosti.

cenen (pridevnik)

1. (slabše kakovosti) *olcsó, silány*

cenen izdelek → *silány portéka*; NENEUTRALNO *bóvli*

Pri iztočnici **cenen** smo tako pri drugem pomenu pri kolokaciji **cenen izdelek** ponudili tudi prevedek *bóvli* in ga s podpičjem ločili od prevedka *silány portéka*. Oba prevedka sta sicer kontrastivno zanimiva, vendar pa se pomensko razlikujeta; raba prevedka *bóvli* je namreč bolj ozka, kar potrjuje tudi Razlagalni slovar madžarskega jezika,³⁶ v katerem je *bóvli* označen s stilno oznako (*bizalmas*, kar se dobesedno prevede z

³⁶ <http://mek.oszk.hu/adatbazis/magyar-nyelv-ertelmezo-szotara/elolap.php>

'zaupno').³⁷ Takšno razliko med prevedkoma v VSMS vzpostavljamo z oznako 'nenevtrarno' (gl. tudi poglavje 2.4.2).

Zaradi dvojne naravnosti VSMS, torej tako za dekodiranje kot enkodiranje, so oznake beležene v obeh jezikih. V digitalnih izdajah slovarja bi bilo uporabnikom mogoče tudi ponuditi možnost enostavnega preklapljanja med jezikom, kar pa bi pomenilo dodatno klikanje, tako da bomo takšne dodatne funkcionalnosti pred vpeljavo najprej preverili z uporabniškimi študijami.

Pri leksikografski analizi in izdelavi slovenskega dela gesel beležimo čim več z oznakami povezanih informacij. Sem sodi tudi prvotno avtomatsko luščenje potencialnih oznak, ki je opravljeno že pri izvozu podatkov iz korpusov, kot je npr. učbeniški korpus. Tovrstne oznake so pri izdelave končne verzije slovenskega dela gesel preverjene in potrjene ali odstranjene, lahko pa so tudi prekategorizirane v skrite oznake (gl. spodaj), kar pa velja predvsem za področne oznake. Upoštevati moramo namreč raznolike potrebe ciljnih uporabnikov VSMS ter možnosti nadaljnje izrabe slovarske baze.

Domet posamezne oznake določa njena umestitev znotraj geselskega članka. Če je oznaka navedena pred vsemi pomeni, tj. za iztočnico oz. v glavi, zajema tudi vse njene pomene; če je navedena znotraj pomena pred prevedki, velja do naslednje oznake oz. do konca določenega niza prevedkov. Pri označevanju stalnih zvez velja oznaka vedno le za določeno stalno zvezo, zato so oznake pri več zaporednih stalnih zvezah, ki so značilne npr. za isto strokovno področje, vedno ponovljene. To v spletnem slovarju omogoča ločen prikaz vsake stalne zveze kot samostojne leksikalne enote z vsemi pripadajočimi informacijami, v tiskanem slovarju pa enostaven prenos stalne zveze na raven večbesedne iztočnice, razvrščene glede na abecedni vrstni red.

Uporabljamo več tipov oznak: slovnične oz. skladenjske, stilne, področne, diskurzne oz. kontekstualne, pragmatične in časovne. Vsak tip oznake ima znotraj slovarske baze svoj atribut. Večina tipov oznak je natančno opredeljenih in ima vnaprej predviden nabor, izjema so le kontekstualne in do neke mere pragmatične oznake, katerih ubeseditev je odvisna od okoliščin rabe leksema. Posledično se nabor kontekstualnih in pragmatičnih oznak pri redakciji sproti usklajuje in dopolnjuje. Na splošno za oznake velja, da se njihova konkretna ubeseditev lahko prilagodi, če se tekom izdelave slovarja pojavi smiselni boljši predlog ali pa se izkaže, da za uporabnike niso dovolj razumljive (to preverjamo z uporabniškimi študijami).

Leksikalne enote imajo lahko več oznak. V tem primeru velja praviloma naslednja hierarhija: slovnične, stilne, pragmatične, časovne, kontekstualne in področne. Možna je tudi uporaba več oznak istega tipa, npr. dveh področnih oznak, pa tudi odkrite in

³⁷ Podana je tudi področna oznaka (ekonomija), vendar pa analiza rabe v korpusu, pa tudi obravnava v drugih slovarjih, tega ne potrjujejo oz. kažejo splošne kontekste rabe.

skrite oznake. V primeru nizanja oznak istega tipa so oznake razvrščene po abecednem vrstnem redu.

Oznake, zlasti kontekstualne in področne, pogosto prevzemajo tudi pomensko-razločevalno vlogo in so posledično umeščene v pomenski indikator. Tako se tudi izognemo pretiranemu ponavljanju informacij. Tako smo pri 1. pomenu pridevniške iztočnice **benigen** namesto kombinacije indikatorja 'o tvorbi ali bolezni' in področne oznake 'medicina' raje uporabili samo indikator-oznako 'v medicini'.

2.4.1 Slovnične oznake

S slovnimi podatki opredeljujemo informacije o leksemih, kot so besedna vrsta, konverznost in nekatere druge slovnične posebnosti in omejitve. Velik del slovničnih informacij je iz korpusa mogoče pridobiti avtomatsko, npr. *tipično v 3. osebi, navadno v množini, pogosto zanikano* ipd. Pri geslih tako označujemo omejitve besede, zveze ali pomena glede na (a) skladenjsko vlogo, npr. *v pridevniški rabi (latino)*, *v povedni rabi (balzam, tema)*; (b) določeno slovnično obliko, npr. *navadno v množini (brzica, ikra)*, *neštevno (grozdje, tangice)* in (c) določeno vrsto zapisa, npr. *navadno v velelniku (klikniti, paziti)*, ustaljena oblika (**tešče** – *na tešče*).

Pomembno je poudariti, da čeprav tovrstne slovnične oznake v slovarski bazi VSMS beležimo, saj so koristne za leksikografe pri pripravi prevodnih ustreznice, pogosto ostanejo skrite, saj so za uporabnika dvojezičnega slovarja lahko zahtevne oz. premalo informativne. Tako določene slovnične posebnosti, kot so npr. neštevnost, povedna raba ipd. raje ponazorimo s kolokacijami in zgledi. Številni podatki o iztočnicah (besedna vrsta, skloni, primernik in presežnik pri pridevniki ipd.) bodo uporabnikom na voljo tudi prek direktne povezave do Slovenskega oblikoslovnega leksikona.

2.4.2 Stilne oznake

Stilne oznake so razdeljene na registrske in konotacijske oznake.

Registrske oznake uporabljamo za označevanje rabe besede oz. pomena, ki izkazuje v odnosu do nevtralnega besedišča posebnosti glede na vrsto komunikacijske situacije, npr. besedišča, ki ga uporabljajo samo določene skupine govorcev, kot so mladostniki (**keš**) ali se uporablja v govoru z otroki (**ritka**). Uvajamo tudi oznako *nenevtraln* (*nem semleges*), ki jo uporabljamo pri leksemih, katerih raba je na nek način zaznamovana v primerjavi z nevtralnimi sinonimnimi leksemi, npr. **buča** (v pomenu glava) in **čik** (v pomenu cigareta).

S konotacijskimi oznakami označimo rabo besede oz. pomena glede na konotativne pomenske lastnosti, npr. odnos govorca do predmetnosti, sogovorca, vsebine besedila ipd. (npr. *ljubkovalno, vulgarno*). Oznake *ekspresivno*, ki je bila zaradi prevelike

splošnosti ali premajhne obvestilnosti v literaturi pogosto deležna kritik (gl. npr. Müller 2009; Rozman 2010; Kalin Golob in Gantar 2015; Gantar 2016), ne uporabljamo.

2.4.3 Področne oznake

S področnimi oznakami je treba označiti vse iztočnice in pomene, ki spadajo v terminologijo določenega področja. Če geslo ali posamezni pomen nista le ozko strokovna, ampak tudi poljudna, uporabimo t. i. skrito oznako (gre za oznake, ki so vidne samo v slovarski bazi; za več gl. Kosem 2015b; Gantar in Kosem 2013; Atkins in Rundell 2008). Skrite področne oznake so koristne zato, da bo mogoče dalo izraze, ki niso strogo specializirani termini, naknadno sistemsko preveriti, uporabili pa jih bomo tudi pri omogočanju naprednejših iskanj po digitalnih različicah slovarja.

Klasifikacija področnih oznak je hierarhično zasnovana, kar leksikografom v primeru več sorodnih področij omogoča uporabo strokovno najširše oznake, npr. če se izraz uporablja v več različnih športih, uporabimo oznako *šport*. Po drugi strani se poslužujemo podkategorij oznak, v kolikor je uporaba širše oznake premalo obvestilna, npr.

kristalna čelada → v MOTOCIKLIZMU *kristálysisak* (*díj országos bajnokságon*)

2.4.4 Kontekstualne oznake

Diskurzne ali kontekstualne oznake za razliko od vezanosti na določeno strokovno področje označujejo omejitev rabe besede ali pogostejše katerega od njenih pomenov na tipične komunikacijske situacije, besedilne tipe ali diskurze. Kontekstualne oznake so sicer pogosto zelo primerne za umestitev v indikatorski del pomena, a na splošno težimo k temu, da namesto uporabe kontekstualne oznake omejitev rabe leksema raje ponazorimo s kolokacijami in/ali zgledi.

2.4.5 Pragmatične oznake

Pragmatične oznake uporabljamo predvsem pri stalnih zvezah in frazeoloških enotah, in sicer v kombinaciji z indikatorji oz. namesto indikatorjev. S pragmatičnimi oznakami želimo v zvezi z rabo določene besede želeli opozoriti na njeno posebno vlogo, vezano na komunikacijsko situacijo, kot je npr. raba enobesedne ali večbesedne leksikalne enote, kolokacije ipd. kot pozdrav, pregovor, opozorilo, izraz odobravanja, grožnje ipd.

jajce (samostalnik)

FRAZE

odrezati komu jajca → KOT GROŽNJA *kiherél valakit*

obesiti koga za jajca → KOT GROŽNJA *tökénél fogva fellógat valakit*

čednost (samostalnik)

FRAZE

skromnost je lepa čednost → PREGOVOR *a szerénység szép erény*

2.4.6 Časovne oznake

S časovno oznako *starinsko* označujemo primere, ko beseda v razmerju do drugih, sinonimnih besed v sodobnem jeziku ohranja časovno distanco. Gre torej za besede, ki danes delujejo arhaično. Oznake *zastarelo* ne uporabljamo, saj je bilo v literaturi (Vidovič-Muha 2009; Kalin Golob in Gantar 2015) pogosto izpostavljeno težko razločevanje med *starinsko* in *zastarelo*.

2.5 Zvočno gradivo

Raziskave med ciljnimi uporabniki (npr. Kosem in Kovács 2017), pa tudi primeri dobre prakse dvojezičnih slovarjev založnikov, kot sta Oxford University Press in HarperCollins, potrjujejo, da mora digitalni medij izkoristiti čim več prednosti v smislu približanja zvočne oblike slovarskih iztočnic uporabniku. Mednje sodijo tudi zvočni posnetki izgovorjave besed, ki so veliko učinkovitejši od kakršne koli transkripcije izgovora besed. Poleg tega ne smemo pozabiti njihove vrednosti za skupine uporabnikov, kot so slepi in slabovidni. V VSMS, ki je naravnano tako na madžarske kot slovenske govorce, vključno oz. še sploh na uporabnike z dvojezičnega območja, bomo ponudili posnetke izgovorjave slovenskih iztočnic in madžarskih prevodov. Posnetki pa ne bodo ponujeni samo za iztočnice in njihove prevedke, temveč tudi za vse ostale mikrostrukturne elemente, od kolokacij in skladenjskih zvez, pa do frazeoloških enot in zgledov.

Priprava takšnega zvočnega gradiva v snemalnem studiu z naravnimi govorci predstavlja velik časovni in finančni zalogaj, zato smo se odločili uporabiti govorne tehnologije, tj. sintetizatorje govora, ki so za slovenščino in madžarščino že zelo dobro razviti. Ne izključujemo pa možnosti, da bi kasneje računalniško generirane zvočne posnetke nadomestili s studijskimi posnetki, predvsem za slovenske iztočnice in madžarske prevode na ravni pomenov, stalnih zvez in frazeoloških enot, ne pa najbrž za kolokacije, skladenjske zveze in zglede. Obstaja tudi možnost vključitve tovrstnih posnetkov iz obstoječih enojezičnih slovarjev, v kolikor bi bili seveda na voljo.

2.5.1 Slovensko zvočno gradivo

Za pripravo zvočni posnetkov slovenskega dela gesel uporabljamo sintetizator govora eBralec, ki so ga izdelali podjetji Alpineon d.o.o. in Amebis d.o.o. ter Institut "Jožef Stefan" (za več gl. Žganec Gros idr. 2016). Kakovost sinteze govora je zelo dobra, zelo blizu naravnemu govoru. eBralec ponuja moški in ženski glas, za VSMS smo po obsežni analizi testnih posnetkov izbrali moškega.

Da bi zagotovili čim večjo kakovost posnetkov, jih med pripravo vedno poslušamo in prilagajamo pisano vnosno besedilo, npr. dodajamo naglasna znamenja v primeru homografov.

2.5.2 Madžarsko zvočno gradivo

Za madžarščino uporabljamo sintetizator govora Profivox HMM TTS Univerze za tehnologijo v Budimpešti (Oddelek za Telekomunikacije in Medijsko informatiko), ki ga na Madžarskem uporabljajo različne ustanove, npr. madžarske železnice, vodilni ponudnik mobilnih storitev za aplikacijo avtomatskega branja elektronskih sporočil (za več gl. Olaszy idr. 2000). Tudi pri madžarskem sintetizatorju govora je kakovost sinteze govora zelo dobra. Na voljo je več različnih verzij, npr. korpusno temelječi TTS in HMM TTS, pa tudi moški in ženski glas – po analizi vzorčnih posnetkov smo izbrali moški glas HMM TTS.

Madžarske posnetke med pripravo poslušamo in prilagajamo vnosno besedilo, da je rezultat čim bolj kakovosten.

3 Organizacija dela in opis leksikografskega procesa

3.1 Splošni opis

Slovar bo izdelala redakcijska skupina Centra za jezikovne vire in tehnologije Univerze v Ljubljani (CJVT UL),³⁸ v sodelovanju z dvema glavnima zunanjima partnerjema, in sicer Inštitutom za narodnostna vprašanja in Oddelkom za madžarski jezik in književnost na Filozofski fakulteti Univerze v Mariboru. Pri nastajanju slovarja bodo sodelovale še druge javne in zasebne ustanove in posamezniki iz Slovenije in Madžarske.

Glavni uredniški odbor, ki ga sestavljajo strokovnjaki za dvojezično leksikografijo, enojezično leksikografijo in pravopisje, ter za različna druga področja, pomaga pri reševanju težjih vsebinskih in tehnoloških vprašanj, spremlja pa tudi potek dela in predlaga morebitne izboljšave leksikografskega procesa.

Ključni del slovarske ekipe predstavljajo strokovnjaki za jezikovne tehnologije in informatiko, ki sodelujejo pri vseh fazah redakcije, pri čemer skrbijo za pripravo in obdelavo jezikoslovnih podatkov za leksikografe, nemoteno delovanje korpusnih in slovarskih orodij, povezovanje podatkov v različnih delih slovarske baze in pretvorbo podatkov v ustrezen format, npr. za prikaz slovarja v digitalnem okolju.

3.2 Izdelava slovarskih gesel

V tem razdelku predstavljamo posamezne faze izdelave gesel,³⁹ ki temeljijo na dveh ključnih načelih: prioritizaciji iztočnic in plastni izdelavi geselskih sestavkov. Pri prioritizaciji iztočnic se zaradi sprotnega objavljanja slovarskih gesel določenim iztočnicam s frekvenčnega seznama nameni prioriteta obravnava – kriteriji izbiranja so predvsem njihova pogostost v vsakdanji rabi, sodobnost in pa odsotnost v obstoječih priročnikih. Ta pristop dopolnjuje plastna izdelava gesel, kar pomeni, da se pri vsakem geslu najprej izdela osnovni del, torej prevedki pomenov, stalnih, frazemov ter kontrastivno zanimivih kolokacij in zgledov, potem pa se postopoma, tudi glede na časovno razpoložljivost leksikografov, dodajajo dodatni prevedeni zgledi ali samo slovenski oz. madžarski zgledi.

³⁸ CJVT UL (<https://www.cjvt.si/>) je enota Univerze v Ljubljane, ki so jo ustanovile Fakulteta za družbene vede, Filozofska fakulteta, Pedagoška fakulteta, Fakulteta za elektrotehniko in Fakulteta za računalništvo in informatiko.

³⁹ Za več o večfaznem procesu nastajanja sodobnega slovarja, gl. Klosa (2013), pa tudi Krek idr. (2013) in Gorjanc idr. (2015).

3.2.1 Priprava slovenskega dela gesel

Kot izhodišče se najprej pripravi geslovník oz. nabor iztočnic za uvrstitev v VSMS. Zraven se že pripišejo vsi podatki, ki se jih lahko avtomatsko pridobi iz korpusov oz. ostalih na korpusih temelječih virih (npr. povezava do gesla v Slovenskem oblikoslovnem leksikonu), npr. zapis iztočnice v osnovni obliki, podatek o besedni vrsti, pogostost leme v referenčnem korpusu (in morebitnih drugih specializiranih korpusih) in določena slovnična opozorila, ki se v slovarskem geslu generirajo kot oznake, npr. *navadno v množini* ipd. Kot je bilo že omenjeno, se bo ob morebitnih posodobitvah korpusnega gradiva na podlagi analiz dodajalo morebitne nove iztočnice ali pomene obstoječih iztočnic.

Poleg tega se iz korpusov avtomatsko izluščijo tudi slovnične relacije ter pripadajoče kolokacije in njihovi zgledi.⁴⁰ Izvoženi so tudi spremljajoči podatki o pogostosti in jakosti posamezne slovnične relacije in kolokacije, pa tudi metapodatki o korpusnih zgledih.

Leksikografi nato na podlagi vseh informacij pripravijo slovarska gesla, tj. opravijo pomensko členitev, pripravijo indikatorje, vstavijo ustrezne slovarske oznake, izbere se relevanten nabor kolokacij in skladenjskih zvez, identificirajo se stalne zveze in frazeološke enote, za vse elemente mikrostrukture pa se tudi že pripravijo slovarski zgledi. Pri tem se leksikografi ravna po načelih tipičnosti in pogostosti, torej v gesla ne vključujejo zelo redkih rab ali variant (npr. pri frazemih). Še nekoliko strožja merila veljajo pri kolokacijah, kjer se v gesla vključi samo bolj tipične oz. pogoste, pa praviloma tudi pomensko bolj zaokrožene kolokacije.⁴¹

Pri izdelavi gesel v posvetovalne namene uporabljamo tudi obstoječe enojezične vire, kot so *Slovar slovenskega knjižnega jezika* (prva in druga izdaja), *Sinonimni slovar slovenskega jezika*, *Slovar sopomenk sodobne slovenščine*, *Slovar tujk*, *Slovar slovenskih frazemov*, razne terminološke slovarje ipd. Glede obstoja določenih specializiranih izrazov lahko iščemo podatke tudi na spletu, vendar smo pri uporabi teh podatkov posebno pozorni na njihov vir.

⁴⁰ Postopek avtomatizacije je bil za slovenščino že preizkušen pri izdelavi Leksikalne baze (Kosem idr. 2012; Kosem idr. 2013a; Gantar idr. 2015; Gantar idr. 2016a), kjer smo s pomočjo v ta namen izdelane slovnice besednih skic (Krek 2012, 2015), ki deluje v orodju Sketch Engine, s pomočjo posebej prilagojene API skripte, ki vsebuje opise vseh relevantnih slovničnih relacij za luščenje kolokacij, in s pomočjo t. i. GDEX konfiguracije (Kilgariff idr. 2008; Kosem idr. 2011, 2013), ki opredeli lastnosti dobrih zgledov, iz korpusa Gigafida izluščili zgoraj navedene podatke in jih avtomatsko prenesli v slovarsko bazo, kjer so bili pripravljeni za nadaljnjo obdelavo.

⁴¹ Tako se npr. precej redko vključuje kolokacije tipa osebek + glagol, razen če gre za zelo tipično oz. samostojno enoto (npr. **burja brije** → *súvít a bóra*).

3.2.2 Priprava madžarskega dela gesel

Preden gredo slovenska gesla v nadaljnjo redakcijo, se vanje avtomatsko pripišejo prevodne ustreznice, z našega vidika prevodni kandidati, najdene v obstoječih bazah, kot sta npr. slovensko-madžarska baza založbe KDictionaries in baza e-zbirk izrazov. Ker ti prevodni kandidati niso opremljeni s pomensko informacijo, jih vse za posamezno iztočnico pripišemo vsakemu pomenu. S tem leksikografom olajšamo delo in pospešimo izdelavo gesel, saj je namreč lažje brisati nerelevantne prevedke kot vstavljati nove.

Leksikografi, ki so strokovnjaki za slovensko-madžarsko jezikovno kombinacijo, nato prevzamejo slovenska gesla z vstavljenimi kandidati za prevedki in ob analizi predloženih podatkov pripravijo prevodne ustreznice za pomen, stalne zveze in frazeme, pri vsakem pomenu pa prevedejo tudi izbor kolokacij in skladenjskih zvez, ki ga opravijo na podlagi vnaprej določenih meril (gl. poglavje 2.3.3).

Izdelava gesel poteka do določene mere modularno, torej predvidevamo možnost, da leksikograf prevod določenih segmentov prepusti višjim ali specializiranim članom ekipe. Pri pripravi poskusnih gesel se je namreč pokazalo, da se leksikografi razlikujejo glede na usposobljenost in tudi motiviranost izdelave posameznih delov gesel, pri čemer imajo največje težave pri frazeoloških enotah in zgledih, deloma pa tudi pri stalnih zvezah. Tako so v redakcijski ekipi za te naloge posebej določeni strokovnjaki, ki sicer lahko tudi sami izdelujejo celotna slovarska gesla.

Z modularnim pristopom poskrbimo za optimizacijo slovarskega dela, saj naše izkušnje iz podobnih preteklih projektov kažejo, da ravno deli, ki leksikografom povzročajo največ preglavic, lahko jemljejo največ motivacije in močno upočasnijo izdelavo gesel. Ima pa modularna izdelava gesel tudi to prednost, da pri izdelavi določenega dela gesel, npr. prevajanju zgledov, (drugi) leksikograf hkrati preverja in potrjuje prevodne ustreznice pomenov, kolokacij in ostalih delov geselske mikrostrukture, ki jih zgledi ponazarjajo. Na ta način se opravlja že prvi vsebinski pregled gesel.

V vsakem primeru izdelano geslo pregleda vsaj še en leksikograf, ki vnese popravke oz. pusti komentarje za avtorja gesla. Avtor potem popravke pregleda (slovarsko orodje omogoča primerjavo dveh verzij geselskega sestavka), vnese morebitne popravke glede na komentarje in geslo odda. Na tej točki se v primeru modularnega pristopa vnesejo tudi podatki drugih leksikografov.

Oddana gesla nato najprej pregleda glavna redaktorska ekipa, ki jo sestavljajo najbolj izkušeni leksikografi v slovarski ekipi. Glavni redaktorji vnese morebitne popravke in dopolnitve, poskrbijo za tehnično ustreznost gesla (pri tem si lahko precej pomagajo s slovarskimi in podobnimi orodji, npr. pri validaciji XML strukture gesla), hkrati pa tudi

za sistematičnost obravnave povezanih slovarskih sestavkov.⁴² Vsa gesla pregleda tudi madžarski lektor.

3.2.3 Priprava dodatnega gradiva

Po dokončni potrditvi vsebine geselskih sestavkov se pripravijo še dodatne vsebine, ki so relevantne za digitalne različice slovarja, in sicer avtentični madžarski zgledi za prevodne ustreznice ter zvočni posnetki izgovorjav za slovenski del gesel in madžarske prevode. Za pripravo in kasnejše povezovanje zvočnih posnetkov se tudi ustrezno tehnično pripravi slovarska baza.

Kandidati za madžarske zgledse se avtomatično izvozijo iz referenčnega korpusa za madžarščino, nato pa leksikografi pripravijo izbor, pri tem pa upoštevajo merila dobrega slovarskega zgleda in tudi pogostost iztočnice, njeno pomensko razvejanost in število že razpoložljivih v madžarščino prevedenih slovenskih zgledov.

Posnetki izgovorjav se za slovenske iztočnice, kolokacije ipd. in njihove prevodne ustreznice izdelajo s sintetizatorjema govora za slovenščino in madžarščino. Vsak posnetek posluša član ekipe, materni govorec slovenščine oz. madžarščine, ki potrdi njegovo ustreznost in v primeru večjih težav (npr. napačno naglašene besede) ustrezno popravi oz. prilagodi vhodno besedilo. V primeru popravkov se potem pripravijo novi zvočni posnetki.

3.3 Objava slovarja

Slovar se bo objavljial sproti, najbrž enkrat letno. Prvotno bodo gesla ponujena prek slovarskega orodja Lexonomy, ki omogoča tudi spletno objavo slovarjev, hkrati pa se bo izdelal tudi posebej za slovar izdelan vmesnik, ki bo omogočal povezovanje z ostalimi relevantnimi viri in tudi korpusi.

V času izdelave slovarja se načrtuje tudi povezovanje z izdelovalci madžarsko-slovenskega slovarja, ki nastaja na univerzi ELTE v Budimpešti in bo objavljen leta 2018 oz. 2019. Na ta način bo mogoče že sproti vzpostavljati povezave med obema slovarjema in načrtovati objavo skupnega slovarja, tako digitalnih izdaj kot tiskane izdaje, ki bo pripravljena ob dokončanju VSMS.

⁴² Za delno sistematičnost obravnave je že deloma poskrbljeno pri dodeljevanju gesel leksikografom, saj se vsa povezana gesla lahko dodelijo istemu leksikografu oz. leksikografom.

4 Literatura

- Ahlin, Martin, Branka Lazar, Zvonka Praznik in Jerica Snoj, 2014: Slovar slovenskega knjižnega jezika. Druga, dopolnjena in deloma prenovljena izdaja. *Jezik in slovstvo* 59/4. 121–127.
- Arhar Holdt, Špela, 2015: Uporabniške raziskave za potrebe slovenskega slovaropisja: prvi koraki. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): Slovar sodobne slovenščine: problemi in rešitve. Ljubljana: Znanstvena založba Filozofske fakultete UL. 136–149.
- Atkins, B. T. Sue in Michael Rundell, 2008: The Oxford Guide to Practical Lexicography. Oxford: Oxford University Press.
- Bálint Čeh, Júlia in Iztok Kosem, 2017: Prvi koraki do novega velikega slovensko-madžarskega slovarja: analiza relevantnih dvojezičnih virov. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave* 5/2. 113-150. http://slovenscina2.0.trojina.si/arhiv/2017/2/Slo2.0_2017_2_06.pdf
- Bernjak, E. (1995): *Madžarsko-slovenski, Slovensko-madžarski slovar*. Ljubljana. Cankarjeva Založba.
- Čermák, František, 1985: Frazeologie a idiomatika. František Čermák, Josef Filipec (ur.): *Česká lexikologie*. Praha: Academia. 166–248.
- Dobrovoljc, Kaja, Simon Krek in Tomaž Erjavec, 2015: Leksikon besednih oblik Sloleks in smernice njegovega razvoja. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 80-105.
- Dobrovoljc, Kaja, Simon Krek, Peter Holozan, Tomaž Erjavec, Miro Romih, 2015: Sloleks: slovenski oblikoslovni leksikon. Ljubljana: Trojina, Zavod za uporabno slovenistiko. <http://hdl.handle.net/11356/1039>.
- Erlandsen, Jens, 2004: iLex – new DWS. *Third International Workshop on Dictionary Writing systems (DWS 2004)*. Brno, 6. – 7. september 2004.
- Fox, Gwyneth, 1987: The Case for Examples. Sinclair, John McH.(ur.): Looking up: An Account of the COBUILD Project in Lexical Computing. London: Collins. 137–149.
- Gantar, Polona in Iztok Kosem, 2013: Beleženje in prikazovanje podatkov o jezikovni rabi: od leksikalne baze do spletnega slovarja. Žele, Andreja (ur.): *Družbena funkcijskost jezika: (vidiki, merila, opredelitve)*. Konferenca Obdobja. Ljubljana: Znanstvena založba Filozofske fakultete. 133-139.

- Gantar, Polona in Simon Krek, 2011: Slovene lexical database. Majchráková, Daniela in Radovan Garabík (ur.): *Natural language processing, multilinguality*. Brno: Tribun EU. 72–80.
- Gantar, Polona, 2016: *Leksikografski opis slovenščine v digitalnem okolju*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Gantar, Polona, 2015: Homonimija in večpomenskost: od teorije do slovarja. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 340-357.
- Gantar, Polona, 2007: *Stalne besedne zveze v slovenščini: korpusni pristop*. Lingua Slovenica. Ljubljana: Založba ZRC, ZRC SAZU.
- Gantar, Polona, Iza Škrjanec, Darja Fišer in Tomaž Erjavec, 2016b: Slovar tviterščine. Erjavec, Tomaž in Darja Fišer (ur.): *Zbornik konference Jezikovne tehnologije in digitalna humanistika, 29. september - 1. oktober 2016, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana, Slovenija = Proceedings of the Conference on Language Technologies & Digital Humanities, September 29th - October 1st, 2016 Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia*. 1. izd. V Ljubljani: Znanstvena založba Filozofske fakultete: = Ljubljana University Press, Faculty of Arts. 71-76.
- Gantar, Polona, Iztok Kosem in Simon Krek, 2016a: Discovering Automated Lexicography: The Case of the Slovene Lexical Database. *International Journal of Lexicography*, 29/2. 200–225.
- Gantar, Polona, Iztok Kosem in Simon Krek, 2015: Leksikografski proces pri izdelavi spletnega slovarja sodobnega slovenskega jezika. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 280-297.
- Gantar, Polona, Simon Krek, Iztok Kosem, Mojca Šorli, Katja Grabnar, Olga Pobirk, Petra Zaranšek in Nina Drstvenšek, 2012: *Leksikalna baza za slovenščino*. Ljubljana: Ministrstvo za izobraževanje, znanost, kulturo in šport. <http://www.slovenscina.eu/spletni-slovar/leksikalna-baza>, <http://www.slovenscina.eu/> (dostop 9. 1. 2018).
- Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.), 2015: *Slovar sodobne slovenščine: problemi in rešitve*, (Zbirka Prevodoslovje in uporabno jezikoslovje). Ljubljana: Znanstvena založba Filozofske fakultete.

- Gorjanc, Vojko, Simon Krek in Damjan Popič, 2015: Med ideologijo knjižnega in standardnega jezika. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 32-48.
- Hanks, Patrick, 2009: Sestavljanje enojezičnega slovarja za domače govorce. *Jezik in slovstvo* 54/ 3–4. 7–24.
- Hradil, J. (1996): *Slovensko-madžarski slovar*. Ljubljana. Državna založba Slovenije.
- Hradil, J. (2012): *Slovensko-madžarski in madžarsko slovenski navzkrižni slovar*. <http://www.krog-znak-sp.si/slovensko.php/slo-hun-slovar-szotar>.
- Jakubíček, Miloš, Adam Kilgariff, Vojtech Kovář, Pavel Rychlý in Vít Suchomel, 2013: *The TenTen corpus family*. 7th International Corpus Linguistics Conference CL. 125-127.
- Kalin Golob, Monika in Polona Gantar, 2015: Stilistika in enojezični slovar: označevanje jezikovne variantnosti. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 446–465.
- Keber, Janez, 2011: *Slovar slovenskih frazemov*. Ljubljana: Založba ZRC, ZRC SAZU (Zbirka Slovarji).
- Kilgariff, Adam, Miloš Husák, Katy McAdam, Michael Rundell in Pavel Rychly, 2008: GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, Elisenda in Janet DeCesaris. (ur.): *Proceedings of the Thirteenth EURALEX International Congress*. Barcelona, Spain: Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra. 425–432.
- Kilgariff, Adam, Pavel Rychlý, Pavel Smrz in David Tugwell, 2004: The Sketch Engine. Williams, Geoffrey in Sandra Vessier (ur.): *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004 Lorient, France July 6–10, 2004*. Lorient: Université de Bretagne-sud. 105–116.
- Kilgariff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý in Vít Suchomel, 2014: The Sketch Engine: ten years on. *Lexicography* 1/1. 7–36.
- Kilgariff, Adam, Vojtěch Kovář, Simon Krek, Irena Srdanović in Carole Tiberius, 2010: A quantitative evaluation of word sketches. Dykstra, Anne in Tanneke Schoonheim (ur.): *Proceedings of the XIV EURALEX International Congress (Leeuwarden, 6-10 July 2010)*. Ljouwert: Fryske Akademy. 372-379.

- Klosa, Anette, 2013: The lexicographical process (with special focus on online dictionaries). Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard in Herberst Ernst Wiegand (ur.): *Dictionaries. An international Encyclopedia of Lexicography*. Supplement Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin in Boston: de Gruyter. 517–524.
- Kosem, Iztok in Attila Kovács: 2017. Raba jezikovnih virov med učitelji na dvojezičnih šolah v Prekmurju. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave* 5/2. 179–214.
http://slovenscina2.0.trojina.si/arhiv/2017/2/Slo2.0_2017_2_08.pdf
- Kosem, Iztok, 2015a: Slovarski zgledi. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 320–338.
- Kosem, Iztok, 2015b: Oznake: slovarska baza in slovar. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 482–494.
- Kosem, Iztok, Milos Husak in Diana McCarthy, 2011: GDEX for Slovene. Iztok Kosem in Karmen Kosem (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. 151–159.
- Kosem, Iztok, Polona Gantar in Simon Krek, 2012: Avtomatsko luščenje leksikalnih podatkov iz korpusa. Erjavec, Tomaž in Jerneja Žganec Gros (ur.): *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012*. Institut Jožef Stefan. 117–122.
- Kosem, Iztok, Polona Gantar in Simon Krek, 2013a: Avtomatizacija leksikografskih postopkov. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave* 1/2. 139–164.
http://www.trojina.org/slovenscina2.0/arhiv/2013/2/Slo2.0_2013_2_07.pdf
(dostop 8. 11. 2017).
- Kosem, Iztok, Polona Gantar in Simon Krek, 2013b: Automation of lexicographic work: an opportunity for both lexicographers and crowd-sourcing. Kosem, Iztok, Jelena Kallas, Polona Gantar, Simon Krek, Margit Langemets in Maria Tuulik (ur.): *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*. Ljubljana: Trojina, Institute for Applied Slovene Studies in Tallinn: Eesti Keele Instituut. 32–48.

- Kosem, Iztok, Polona Gantar in Simon Krek, 2017: Sense menus in collocations dictionary of Slovene. *Electronic lexicography in the 21st century : lexicography from scratch*. Leiden: Dutch Language Institut; Brno: Lexical Computing; Ljubljana: Trojina Institute for Applied Slovene Studies. 43.
- Kosem, Iztok, Kristina Koppel, Tanara Kuhn, Jan Michelfeit in Carole Tiberius, v pripravi: Identification and automatic extraction of good dictionary examples: the case(s) of GDEX. *International Journal of Lexicography*.
- Krek, Simon (ur.), 2005/2006: Veliki angleško-slovenski slovar OXFORD®-DZS. Ljubljana: DZS.
- Krek, Simon in Adam Kilgarriff, 2006: Slovene Word Sketches. Erjavec, Tomaž in Jerneja Žganec Gros (ur.), *Jezikovne tehnologije: zbornik 9. mednarodne multikonference Informacijska družba IS 2006, 9. do 10. oktober 2006 = Language technologies : proceedings of the 9th International Multiconference Information Society IS 2006, 9th-10th October 2006, Ljubljana, Slovenia*. Ljubljana: Institut "Jožef Stefan". 62–65.
- Krek, Simon, 2012: New Slovene sketch grammar for automatic extraction of lexical data. SKEW3, tretja mednarodna delavnica orodja Sketch Engine. Brno, Češka, 21.–22. marec 2012.
- Krek, Simon, 2014. Prva in druga izdaja SSKJ. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave* 2/2. 114–158. <http://www.trojina.org/slovenscina2.0/si/arhiv/2014-2/2014-2-08/> (dostop 15. 1. 2018).
- Krek, Simon, 2015: Leksikografska orodja za slovenščino: slovnica besednih skic. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 358-378.
- Krek, Simon, Iztok Kosem in Polona Gantar, 2013: *Predlog za izdelavo Slovarja sodobnega slovenskega jezika*. Verzija 1.1. http://sssj.si/datoteke/Predlog_SSSJ_v1.1.pdf (dostop 12. 12. 2017).
- Krek, Simon, Polona Gantar, Iztok Kosem, Vojko Gorjanc in Cyprian Laskowski, 2016: Baza kolokacijskega slovarja slovenskega jezika. Erjavec, Tomaž in Darja Fišer (ur.): *Zbornik konference Jezikovne tehnologije in digitalna humanistika, 29. september - 1. oktober 2016, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana, Slovenija = Proceedings of the Conference on Language Technologies & Digital Humanities, September 29th - October 1st, 2016 Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia*. Ljubljana: Znanstvena

- založba Filozofske fakultete: = Ljubljana University Press, Faculty of Arts. 101-105.
- 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.
- Logar Berginc, Nataša, Miha Grčar, Marko Brakus, Tomaž Erjavec, Špela Arhar Holdt in Simon Krek, 2012: *Korpusi slovenskega jezika Gigafida, Kres, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko in Fakulteta za družbene vede.
- Logar, Nataša, 2015: Gradnja referenčnih korpusov na novo: nadgradnja Gigafide. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 218-240.
- Logar, Nataša, Tomaž Erjavec, Simon Krek, Miha Grčar, Peter Holozan, 2013: *Written corpus ccGigafida 1.0*. Ljubljana: Slovenian Language Resource Repository CLARIN.SI. <http://hdl.handle.net/11356/1035>.
- Madžarski razlagalni slovar - A Magyar Nyelv Értelmező Szótára I–VII. Kötet - <http://mek.oszk.hu/adatbazis/magyar-nyelv-ertelmezo-szotara/elolap.php>.
- Měchura, Michal, (2017): Introducing Lexonomy: an open-source dictionary writing and publishing system. Iztok Kosem, Carole Tiberius, Miloš Jakubiček, Jelena Kallas, Simon Krek in Vít Baisa (ur.): *Electronic lexicography in the 21st century: lexicography from scratch*. Leiden: Dutch Language Institut; Brno: Lexical Computing; Ljubljana: Trojina Institute for Applied Slovene Studies. 662-679.
- Moon, Rosamund, 1987: *The Analysis of Meaning*. Sinclair, John M. (ur.): Looking up: An account of the COBUILD Project in Lexical Computing. London in Glasgow: Collins ELT. 86–103.
- Müller, Jakob, 2009: Kritične misli in zamisli o SSKJ. Perdih, Andrej (ur.): *Strokovni posvet o slovarju slovenskega jezika*. Ljubljana: Založba ZRC, ZRC SAZU. 17–21, 25.
- Müller-Spitzer, Carolin (ur.), 2014: *Using Online Dictionaries*. Berlin in Boston: de Gruyter.
- NSSKJ: Gliha Komac, Nataša, Nataša Jakop, Janoš Ježovnik, Simona Klemenčič, Domen Krvina, Nina Ledinek, Tanja Mirtič, Andrej Perdih, Špela Petric, Marko

- Snoj, Andreja Žele, 2015: Koncept novega razlagalnega slovarja slovenskega knjižnega jezika. Ljubljana: Inštitut za slovenski jezik Frana Ramovša; Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti, 2015. http://www.fran.si/179/novi-slovar-slovenskega-knjiznega-jezika/datoteke/Potrjeni_koncept_NoviSSKJ.pdf.
- Olaszy, Gábor, Géza Németh, Peter Olaszi, Géza Kiss, Csaba Zainkó in Géza Gordos, 2000: Profivox—A Hungarian Text-to-Speech System for Telecommunications Applications. *International Journal of Speech Technology* 3/3-4. 201-215.
- Resolucija o nacionalnem programu za jezikovno politiko 2014–2018. Dostopno prek: <http://pisrs.si/Pis.web/pregledPredpisa?id=RESO91>.
- Rozman, Tadeja, 2010: *Vloga enojezičnega razlagalnega slovarja slovenščine pri razvoju jezikovne zmožnosti*. Doktorska disertacija. Ljubljana: Filozofska fakulteta UL.
- Rudaš, Jutka in Anna Kolláth, 2017: Model dvojezičnega pouka v Prekmurju in slovar kot didaktični pripomoček. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave* 5/2. 64-84. http://slovenscina2.0.trojina.si/arhiv/2017/2/Slo2.0_2017_2_04.pdf.
- Rundell, Michael in Adam Kilgarriff, 2011: Automating the creation of dictionaries: where will it all end? Meunier, Fanny, Sylvie De Cock, Gaëtanelle Gilquin in Magali Paquot (ur.): *A Taste for Corpora. A tribute to Professor Sylviane Granger*. Amsterdam in Philadelphia: John Benjamins. 257–281.
- Snoj, Jerica, Martin Ahlin, Branka Lazar in Zvonka Praznik, 2016: Sinonimni slovar slovenskega jezika. Ljubljana: Založba ZRC, Inštitut za slovenski jezik Frana Ramovša.
- Sopomenke 1.0: Slovar sopomenk sodobne slovenščine, viri.cjvt.si/sopomenke, dostop 18. 03. 2018.
- SSKJ1: Slovar slovenskega knjižnega jezika: prva knjiga A–H (1970); druga knjiga I–Na (1975); tretja knjiga Ne–Pren (1979); četrta knjiga Preo–Š (1985); peta knjiga TŽ (1991) z dodatki od A do Š, Ljubljana: SAZU – Državna založba Slovenije, 1970–1991.
- SSKJ2: Slovar slovenskega knjižnega jezika. Druga, dopolnjena in deloma prenovljena izdaja. Ljubljana ZRC SAZU, DZS, 2014.
- Stabej, Marko, 2015: Daj mi slovar in spremenim ti (jezikovno) skupnost. Gorjanc, Vojko, Polona Gantar, Iztok Kosem in Simon Krek (ur.): Slovar sodobne

slovenščine: problemi in rešitve. Ljubljana: Znanstvena založba Filozofske fakultete UL. 16–31.

Toporišič, Jože (ur.), 2001: Slovenski pravopis. Ljubljana: Založba ZRC, ZRC SAZU.

Vidovič Muha, Ada, 2000: Slovensko leksikalno pomenoslovje. Govorica slovarja. Ljubljana: Znanstveni inštitut Filozofske fakultete Univerze v Ljubljani..

Vidovič Muha, Ada, 2009: Poskus določitve meril slovarskega pomena. Perdih, Andrej (ur.): *Strokovni posvet o novem slovarju slovenskega jezika*. Ljubljana: Založba ZRC, ZRC SAZU. 27–36.

Žganec Gros, Jerneja, Boštjan Vesnicher, Simon Rozman, Peter Holozan in Tomaž Šef, 2016: Sintetizator govora za slovenščino eBralec = The eBralec speech synthesis system for Slovenian. Erjavec, Tomaž in Darja Fišer (ur.). Zbornik konference Jezikovne tehnologije in digitalna humanistika, 29. september - 1. oktober 2016, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana, Slovenija = Proceedings of the Conference on Language Technologies & Digital Humanities, September 29th - October 1st, 2016 Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia. 1. izd. V Ljubljani: Znanstvena založba Filozofske fakultete: = Ljubljana University Press, Faculty of Arts. 180-185. http://www.sdit.si/wp/wp-content/uploads/2016/09/JTDH-2016_Gros-et-al_Sintetizator-govora-za-slovenscino-eBralec.pdf.

4.1 Spletni viri:

Portal Fran - <http://www.fran.si/>

Portal Termania - <http://www.termania.net/>

Slovar tviterščine - <http://lexonomy.cjvt.si/slovar-tviteriscine/>

Sloleks - <http://www.slovenscina.eu/sloleks>

e-zbirke izrazov - <http://eslovar.datadev.si/>

Linguee - <https://sl.linguee.com/>

Glosbe - <https://sl.glosbe.com/>

SPrichwort - http://www.sprichwort-plattform.org/sp/Sprichwort_si

Sinonimni slovar madžarskega jezika - <https://szinonimaszotar.hu>

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

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

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