

BELEŽENJE IN PRIKAZOVANJE PODATKOV O JEZIKOVNI RABI: OD LEKSIKALNE BAZE DO SPLETNEGA SLOVARJA

Polona Gantar

Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU, Ljubljana

Iztok Kosem

Trojina, zavod za uporabno slovenistiko, Ljubljana

UDK 81'27:81'322:81'374

V prispevku obravnavamo postopek pridobivanja in slovarskega prikazovanja podatkov o rabi besed oz. pomenov v različnih govornih položajih, ki jih narekujejo realne sporočanje situacije. Najprej opišemo način beleženja in prikazovanja tovrstnih podatkov v predkorporusnih slovarjih, nato pa prikažemo spremenjene okoliščine beleženja jezikovne rabe, ki so posledica analize obsežnih besedilnih korpusov, avtomatizacije leksikografskih postopkov, predhodne izdelave slovarskih baz in prenosa slovarske informacije na splet. Zadnji del prispevka je namenjen opisu postopka pridobivanja podatkov iz referenčnega korpusa Gigafida, s poudarkom na avtomatizaciji ter načinu beleženja in strukturiranja tovrstnih podatkov v Leksikalni bazi za slovenščino.

kvalifikatorji, oznake, avtomatizacija leksikografskih postopkov, korpusna leksikografija

This paper deals with the extraction, recording and visualisation of information related to usage of words or their meaning(s) in different communication situations. First, methods of recording and visualising word usage information in non-corpus-based dictionaries are discussed. Then, new trends and approaches to recording word usage are presented, based on analyses of large corpora, automated lexicographic procedures, building of dictionary databases and subsequent transfer of dictionary information to the internet. We then present the process of usage data extraction from the Gigafida corpus for the purposes of the Slovene Lexical Database, focussing on automatic methods used, and ways of recording and structuring usage information.

qualifiers, labels, automating lexicographic work, corpus lexicography

1 Uvod

Pojasnila o tem, kako je besedo, zvezo ali posamezni pomen smiselno rabiti v besedilu glede na to, katere slovnične omejitve predvideva, na katerem strokovnem področju se rabi, kakšen odnos do vsebine ali udeležencev izraža, na katero časovno obdobje, besedilni tip, regionalne ali kulturološke posebnosti je vezana, se v slovarjih navajajo v obliki krajših pojasnil, kot so kvalifikatorji, kvalifikatorska pojasnila ali oznake. Ta pojasnila pa so lahko zajeta tudi v razlage ali druge segmente geselskih člankov, kar pomeni, da se od preostalih delov oblikovno ne ločijo. Za naš prispevek je pomembno, da

ostajajo v slovarjih, ki so zasnovani za spletni medij, ti podatki pričakovani del informacije.

V nadaljevanju najprej opišemo beleženje jezikovne rabe v slovarjih predkorporusnega obdobja, ki niso nastajali na podlagi vnaprej strukturiranih slovarskih baz in so bili pretežno vezani na knjižne izdaje. V slovenskem prostoru sta taka *Slovar slovenskega knjižnega jezika* (SSKJ) in njegov naslednik *Slovar novejšega besedja slovenskega jezika* (2012; SNB). Nato na kratko predstavimo prenovljeno slovarsko situacijo, ki jo določajo obsežne elektronske besedilne zbirke, in tipe oznak, ki smo jih uporabili v Leksikalni bazi za slovenščino (LBS), izdelani pri

projektu Sporazumevanje v slovenskem jeziku.¹ Posebej izpostavimo prednosti avtomatizacije leksikografskih postopkov na prehodu iz korpusa v slovarsko bazo ter različne možnosti posredovanja podatkov o jezikovni rabi uporabniku, ki jih omogoča spletni medij.

2 Beleženje jezikovne rabe

2.1 Kategorizacija oznak v predkorpusnih slovarjih

Predkorpusni slovarji so temeljili na ročni analizi obsežnih listkovnih datotek, saj kartotek ni bilo mogoče računalniško urejati glede na besedilni tip, slovnično omejitve, preferenco določenega skladenjskega vzorca, področje rabe ipd. Zato je bilo določanje rabe besed, zvez in pomenov z vidika omenjenih posebnosti ne le zamudno, ampak tudi bolj ali manj prepuščeno intuiciji leksikografov.

Gradivo za izdelavo SSKJ predstavlja listkovna kartoteka, ki je ob koncu izdelave vsebovala štiri milijone izpisov (Müller 2009: 17, 22). Kvalifikatorjev in kvalifikatorskih pojasnil, ki določajo normo knjižnega jezika oz. opozarjajo na razmerje do nje, je več vrst in so podrobno opisani v Uvodu (str. XVIII–XXII). V pregledu kritičnih odzivov na SSKJ (Kržišnik 2004; Logar Berginc 2000/01; Müller 2009: 18) je v zvezi s kvalifikatorji izpostavljeno, da so potrebni »sistemske in distribucijske preureditve in prevetritve« (Müller 2009: 18), v zvezi s slovničnimi kvalifikatorji, zlasti pri preverjanju njihove razumljivosti pri šolski populaciji (Rozman 2010), pa je bilo ugotovljeno, da je sistem prikazovanja prezahteven in zato pogostokrat za uporabnika nerazumljiv ali celo zavajajoč.²

Približno 20 let pozneje je z namenom preseči obdobje, ko slovenska leksika ni bila

popisana v slovarju, izšel SNB. Čeprav temelji na korpusu Nova beseda,³ pregled nekaterih razmeroma pogostih besed (npr. *biatlonka*, *bluetooth*, *bedno*)⁴ v korpusu Gigafida⁵ in njihova odsotnost v SNB pokažeta neprimernost uporabljenega korpusa za izdelavo sodobnega slovarja (prim. tudi Logar Berginc 2009: 157). Kvalifikatorji so v SNB glede na SSKJ deloma preurejeni in prinašajo nekatere premike v pravo smer. Taka je na primer kombinacija kvalifikatorja »žargonsko« s terminološkim kvalifikatorjem, npr. film. žarg. – filmski žargon, in premik v kvalifikatorska pojasnila tipa »zlasti v sproščenem ožjem krogu«, ki dopolnjuje pogosto nejasen ali neustrezen kvalifikator »pogovorno«, npr. za *hostel*, *bar koda*, *nacionalka*, ki glede na kontekst in besedilno okolje take zaznamovanosti ne kažejo. Med pomanjkljivosti sodi tudi uporaba kvalifikatorja »ekspresivno«, zlasti tedaj, ko bi bila potrebna jasnejša opredelitev »čustvene zaznamovanosti«, npr. pri *balkanec*, *goveja glasba*, oz. odsotnost kakršnekoli opredelitve pri besedah, kot so: *googlati*, *googlanje* ipd.

2.2 Beleženje jezikovne rabe v obdobju pospešenega razvoja jezikovnih tehnologij

V obdobju zadnjih nekaj let so se slovarske informacije preselile na splet, kjer postajajo del celovite jezikovne podpore za različne profile uporabnikov. Sodobni slovar ima glede na slovarske priročnike v knjižni obliki več prednosti. Kot končni izdelek je predviden za pregledovanje na spletu oz. na različnih mobilnih napravah (Lorentzen, Theilgaard 2012). Da je kot spletni priročnik zasnovan že v izhodišču, je pomembno z več vidikov, izpostaviti pa je mogoče neomejenost s prostorom, vključevanje večje količine informacij, različne iskalne možnosti,

¹ Projekt financira Evropska unija iz Evropskega socialnega sklada ter Ministrstvo za izobraževanje, znanost in šport, <http://projekt.slovenscina.eu/Vsebine/Sl/Domov/Domov.aspx>.

² Npr. oznaka »nedov.« za 'nedovršno', razumljena tudi kot 'nedovoljeno' (Rozman idr. 2010: 208).

³ Sestava geslovnika temelji na zbirki novejšega besedja, ki je nastala pri projektu Novejša slovenska leksika (v povezavi s spletnimi jezikovnimi viri) in na korpusu Nova beseda (SNB Uvod: 9).

⁴ V Gigafidi imajo vse tri besede več kot 2100 pojavitev.

⁵ <http://demo.gigafida.net>

hiperpovezave, vključitev multimedije, redno posodabljanje ter prilagodljivost prikaza informacij.

Poleg omenjenih lastnosti je spletnim slovarjem skupno tudi to, da temeljijo na elektronskih slovarskih podatkovnih bazah, ki so strukturirane tako, da je podatke mogoče v čim večji meri pridobivati avtomatsko, jih urejati in povezovati z drugimi podatkovnimi bazami in uporabljati za nadaljnje jezikoslovne analize, hkrati pa jih izrabljati tudi v jezikovnotehnološke namene. Kot primere dobre prakse lahko omenimo DANTE, leksikalno bazo za angleščino,⁶ Splošni nizozemski slovar (Algemeen Nederlands Woordenboek),⁷ leksikalno podatkovno bazo za francoščino (Base lexicale du français)⁸ in Veliki slovar poljskega jezika (Wielki słownik języka polskiego).⁹

Ena od značilnosti takih slovarskih baz je, da vsebujejo več informacij, kot se jih dejansko prikazuje v slovarju. Sem sodi tudi beleženje jezikovne rabe v obliki oznak, kjer je med prednostmi, ki jih leksikografom ponuja slovarska baza, mogoče izpostaviti zlasti fleksibilnost in izčrpnost pri njihovem opisu in kategorizaciji. Fleksibilnost se kaže v možnosti razširitve prvotnega nabora oznak

in relativne svobode pri njihovi ubeseditvi, saj se leksikografu ni treba tega držati seznama, če realna jezikovna raba sugerira določene posebnosti. To sicer ne velja v enaki meri za vse tipe oznak, ampak predvsem za tiste, kjer jezikovna raba niha in kaže različne pomenske, stilne, pragmatične in druge omejitve, ki jih je težko ustrezno zajeti z vnaprej določenimi kategorijami.

Ko govorimo o izčrpnosti, imamo v mislih težnjo, da se v fazi izdelave slovarske baze jezikovna raba opiše čim bolj podrobno, in sicer ne glede na to, kako in če sploh bodo ti opisi predstavljeni v slovarju. Na ta način leksikograf opozori na posebnost v jezikovni rabi, odločitev o tem, ali in v kakšni obliki se bo podatek prenesel v slovar, pa je stvar končne redakcije in načina njihove predstavitve. Primer tega je slovarska baza, na kateri nastaja Splošni nizozemski slovar, kjer je beleženje področnih oznak zelo podrobno in poteka na več ravneh: geslo *qualifier* ('kvalifikant'), kot prikazuje Slika 1, ima oznako *šport*, ki je lahko še podrobneje členjena, npr. *badminton*, *biljard*, hkrati pa so navedene tudi nadkategorije, npr. *šport z žogo*, *dvo-ranski šport* ipd.



Slika 1: Del gesla *qualifier* v slovarski bazi ANW

Podrobno beleženje oznak ne ponuja prednosti le leksikografom, temveč tudi slovarskim uporabnikom, saj lahko na podlagi podrobne kvalifikacije leksike v slovarski

bazi, ki ni nujno prikazana tudi v slovarju, omogočimo različna napredna iskanja. Kot primer navajamo spletni enojezični angleški slovar Oxford,¹⁰ ki omogoča napredno

⁶ <http://www.webdante.com>

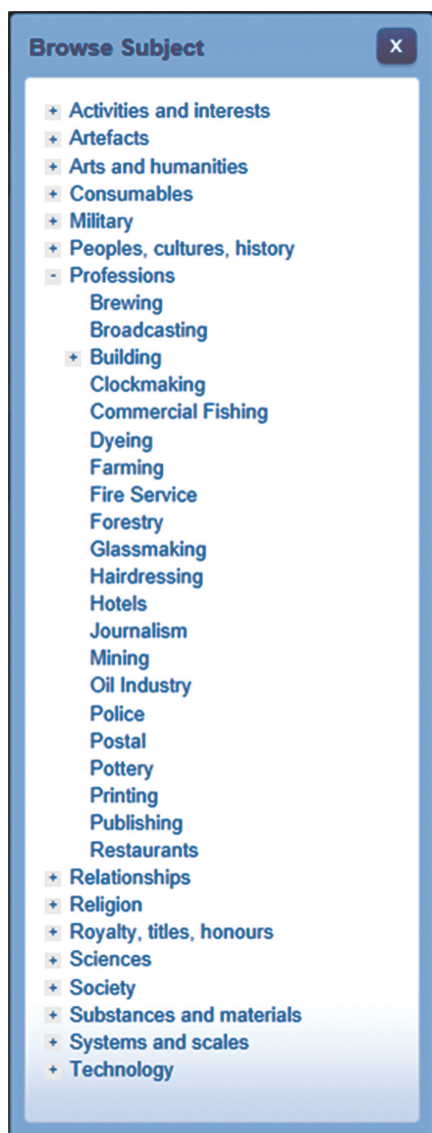
⁷ <http://anw.inl.nl>

⁸ <http://ilt.kuleuven.be/blf>

⁹ <http://www.wsjp.pl>

¹⁰ Slovar je dostopen na spletni strani: <http://oxforddictionaries.com>.

iskanje po 15 različnih strokovnih področjih s številnimi podpodročji, kot prikazuje Slika 2. Uporabnik lahko na ta način poišče vsa gesla, ki vsebujejo pomene oz. podpomene z določeno področno oznako.



Slika 2: Napredno iskanje po strokovnem področju v slovarju Oxford

Pri posameznem (pod)podpomeni se v posebnem okenčku izpišejo tudi druge v bazi obstoječe kategorije, ki uporabnika opozarjajo na (pod)pomene ali besede sorodnih kategorij, kot prikazuje Slika 3.

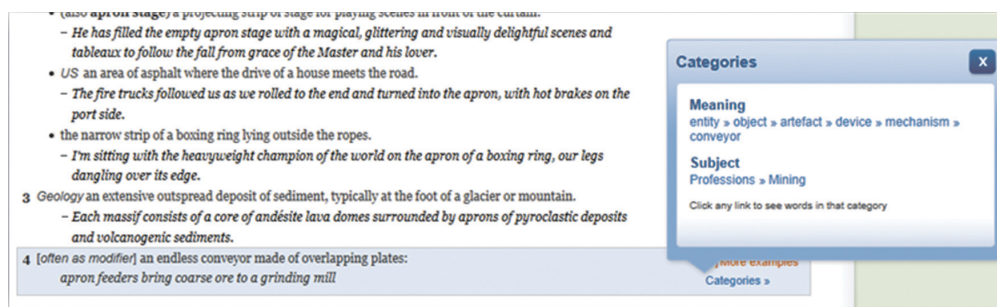
3 Oznake v leksikalni bazi za slovenščino (LBS)

V LBS smo jezikovno rabo beležili v obliki oznak, za katere smo predvidevali, da jih bo v spletni verziji slovarja mogoče prikazovati v obliki krajših, uporabniku čim bolj razumljivih opozoril, in kot del stavčnih razlag. Tako prakso poznajo zlasti slovarji, namenjeni učenju tujega jezika, ki večino pragmatičnih in skladenjskih elementov pomena vključujejo neposredno v razlago,¹¹ zaradi česar uporabnik lažje uzavesti rabo besede ali njenega pomena (Barnbrook 2002: 178). Podobne ugotovitve veljajo za navajanje slovnčnih podatkov, saj ti pogosto zaradi prezahtevnega metajezika ali krajšanj ne dosežejo uporabnikov oz. jih uporabniki največkrat ignorirajo (Cowie 1981: 206; Summers 1987: F8). Pri beleženju podatkov o jezikovni rabi v slovarski bazi smo zato sledili načelu sprotnega oblikovanja in ubesedenja oznak, kategorizacijo in poenotenje pa smo, kjer je bilo to smiselno, opravili ob zaključni redakciji.

Največ besedišča je v LBS označenega glede na strokovno področje rabe, kjer smo predvideli seznam področij na podlagi obstoječih taksonomij, hkrati pa smo področja dodali, če se je ob pomenski analizi pokazala taka potreba.

Poleg področnih smo na ravni baze predvideli tudi t. i. kontekstualne oznake, ki so namenjene označevanju omejitve rabe besede, zveze ali pomena na določen diskurz ali besedilni tip, vendar se je pokazalo, da je na nekaterih področjih, zlasti športnem, političnem, v računalništvu ipd., težko ločevati med strokovno rabo besede ali pomena in

¹¹ V zvezi s tem je bila dana pobuda za stavčno razlaganje členkov (Müller 2009: 105) v novem slovarju slovenskega jezika, kjer naj se razlag ne bi več ločevalo od kvalifikatorskih pojasnil, ampak bi se vsa pojasnila oz. t. i. situativna semantika vključila v razlago.



Slika 3: Informacija o kategorijah (*Categories*) pri določenem pomenu v slovarju Oxford

vezanostjo na besedila z določeno tematiko, npr. *kloniti* – v športu ‘izgubiti tekmo’; *bobnati* – v političnem kontekstu ‘veliko govoriti’.

Stilne oznake smo v LBS ločili na (a) registrske oznake za označevanje rabe besede oz. pomena, glede na vrsto sporočanje situacije, npr. v neformalnih govornih situacijah (*kasirati*), ali besedišča, ki ga uporabljajo samo določene skupine govorcev, npr. mladostniki (*keš*), ali ki se uporablja v govoru z otroki (*ritka*), in na (b) konotacijske oznake, s katerimi smo označevali rabo besede oz. pomena glede na konotativne pomenske lastnosti, kot je odnos govorca do sogovorca, vsebine besedila ipd. Skupni oznaki, kot je npr. »ekspresivno« v SSKJ in SNB, smo se zaradi premajhne obvestilnosti odpovedali in raje uporabljali krajše opise, s katerimi smo skušali natančneje zajeti vrsto konotacije, npr. *klišé* – z negativnim vrednotenjem ‘vzorec; ponavljanje’.

Pri glagolskih pomenih, zlasti pa pri frazeoloških enotah smo uporabljali t. i. pragmatične oznake, s katerimi smo opozorili na rabo besede v posebni vlogi, vezani na sporočanje situacijo, kot je npr. pozdrav, opozorilo, izraz odobravanja ipd., sicer pa smo tovrstne podatke vključevali v t. i. pomenske sheme ali razlage. Prav tako smo v pomenski opis vključevali podatek o tipični skladišnji rabi posameznega pomena, npr. *aktiven* – ‘dejavnen’: *človek je aktiven na nekem področju ali v neki dejavnosti, če se vanjo vključuje*

in pri njej redno sodeluje. Z implicitnim navajanjem tovrstnih podatkov smo želeli pri uporabniku uzavestiti tipično skladišnji rabo na način, ki najbolje odraža dejansko stanje, kar je zlasti pomembno pri učenju slovenščine kot tujega jezika in pri šolskih uporabnikih.

3.1 Avtomatizacija in objektivizacija beleženja jezikovne rabe

Novi trendi v leksikografiji, ki jih določajo velike količine jezikovnih podatkov in prenos slovarskih informacij na splet, v veliki meri spreminjajo način izdelave in podobo slovarjev. Leksikografi, ki svoje delo opravljajo v digitalnem okolju, iščejo rešitve, ki bi v enaki meri upoštevale relevantnost in kredibilnost leksikalnih informacij ter možnosti za optimizacijo in racionalizacijo leksikografskega dela. Ena od rešitev je čim večja avtomatizacija rutinskih leksikografskih postopkov, ki je neposredno povezana s prihrankom časa in denarja. Rundell in Kilgariff (2011) pri razmišljanju o možnostih avtomatizacije različnih segmentov leksikografskega dela med drugim izpostavita, da je o avtomatskem beleženju oznak treba razmišljati že v času gradnje korpusa (prim. Kosem 2011). Eden od načinov avtomatskega luščenja oznak je namreč izdelava podkorpusov besedil, ki se uvrščajo pod obseg določene oznake, npr. za določeno strokovno področje, regijo ali časovno obdobje. Leksika, ki se pojavlja večinoma ali izključno v besedilih

katerega od podkorpusov, pridobi tudi z njim povezano oznako.

Nekoliko drugačen pristop zahteva avtomatsko luščenje slovničnih oznak, ki smo ga preizkusili v zaključni fazi izdelave LBS. Postopek je bil primarno namenjen luščenju slovničnih relacij, kolokacij in z njimi povezanih zgledov iz korpusa neposredno v programski vmesnik LBS (Kosem idr. 2011), hkrati pa smo na ta način pridobili tudi podatke o jezikovni rabi. Na eni strani gre za informacije, ki leksikografa opozarjajo na tipično sopojavljanje dane besede s številom (*finalist*), količinskimi izrazi (*denar*) ali lastnimi imeni (*mesto*), tipično pojavljanje v povratnih se/si-stavčnih konstrukcijah (*izvoliti*), v nedoločniku (*blažiti*) in v zvezi z glagolom *biti* (*všeč*). Na drugi strani pa za informacije, ki jih je mogoče neposredno prepisati v slovarske oznake, npr. tipično v 3. os. (*boleti*), navadno v množini (*katakomba*), pogosto zanikano (*nakloniti*) ipd.

Tovrstne informacije smo iz korpusa pridobili prek gramatičnih relacij v slovnici besednih skic v orodju Sketch Engine, kjer posamezno relacijo v osnovi sestavljajo trije elementi: (procesna) direktiva, ki določa tip odnosa v relaciji, ime in poizvedba (Krek 2010). Znotraj poizvedbe je mogoče skladenjske odnose definirati glede na en sam element, ki izpostavlja en pojav v odnosu do vseh ostalih elementov v korpusu. Za namene pridobivanja slovničnih oznak je glede na poizvedbo – tj. besedo v iztočnici – mogoče tako pridobiti podatek o tem, da je nek pojav, npr. množinska oblika, sopojavljanje s količinskimi izrazi ipd., statistično izstopajoč. Pomembna odločitev pri luščenju slovničnih oznak je vezana na statistično mejo, pri kateri se posamezni iztočnici pripiše oznaka. Kot ponazarjata Rundell in Kilgariff (2011), je odstotkovno mejo mogoče določiti glede na razmerje pojavitev v določeni obliki do vseh pojavitev, npr. iztočnica se mora pojaviti v množinski obliki v X odstotkih, da se ji pripiše oznaka *navadno v množini*, temu pa je po potrebi mogoče dodati še merilo, ki določa-

nemu odstotku tako izbranih iztočnic pripiše oznako v slovarju.

4 Zaključek

Podatki, ki besede, njihove variante ali pomene opredeljujejo z vidika slovničnih, stilističnih, pragmatičnih, področnih, časovnih, regionalnih, registrskih in drugih omejitev, ostajajo ena ključnih informacij v slovarjih tudi v času, ko se slovarske vsebine selijo na splet in postajajo sestavni del celostnih jezikovnopriročniških portalov. Obsežni korpusi in napredna orodja za njihovo analizo na eni strani ter oblikovanje predhodnih slovarskih baz z možnostjo avtomatiziranja postopkov pridobivanja podatkov, njihovega urejanja ter procesiranja na drugi so v temelju spremenili načine beleženja in prikazovanja informacij o jezikovni rabi. Ročno pregledovanje velike količine podatkov, sledenje intuiciji in vnaprej določenim seznamom kategorij so nadomestili postopki avtomatizacije zlasti pri prepoznavanju slovničnih omejitev besed in pri prepoznavanju njihovega tipičnega pojavljanja na določenem strokovnem področju. Beleženje rabe, ki se nanaša na stilistične, pragmatične in druge pomensko specifične posebnosti sicer ostaja v pristojnosti leksikografske analize, vendar pa elektronske leksikalne baze, iz katerih izhajajo končni slovarski produkti, ponujajo vrsto prednosti tako v načinu beleženja tovrstnih podatkov kot v možnostih njihovega urejanja. Podatke je v bazi mogoče kadarkoli poenotiti, opustiti ali pa uporabiti za združevanje besedišča s sorodnimi stilističnimi, pragmatičnimi in drugimi pomenskimi lastnostmi. Podlaga, ki jo imajo spletni slovarji v slovarskih bazah, omogoča tudi različne možnosti končnega, slovarskega prikazovanja podatkov o jezikovni rabi. Ker vsebujejo podrobne, za posamezne besede in pomene specifične opise, jih je ob prenosu v slovar mogoče vizualizirati na različne načine: eksplicitno v obliki kratkih in razumljivih pojasnil, npr. koketiranje – v političnem diskurzu *medsebojno povezovanje*, in implicitno

znotraj stavčno strukturiranih slovarskih razlag, ki uporabniku omogočajo uzaveščanje informacije na način, ki ga pozna iz vsakodnevne komunikacije oz. ki je za pripadnike določene jezikovne skupnosti najbolj naraven, npr. *komentar – če da ČLOVEK komentar na nek DOGODEK, SITUACIJO ali TEMO, izrazi o tem svoje mnenje*.

Literatura

- BARNBROOK, Geoff, 2002: *Defining language: A local grammar of definition sentences*. Studies in Corpus Linguistics. Amsterdam: John Benjamins Publishing Company.
- BIZJAK, Aleksandra, SNOJ, Marko (ur.), 2012: *Slovar novejšega besedja slovenskega jezika*. Ljubljana: Založba ZRC, ZRC SAZU.
- COWIE, Anthony, 1981: Introduction to special issue on pedagogical dictionaries. *Applied Linguistics* 2/3. Oxford: Oxford University Press.
- KOSEM, Iztok, 2011: Prihodnost leksikografije: dinamični slovar. Marko Jesenšek (ur.): *Izzivi sodobnega slovenskega slovaropisja*. Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti, Filozofska fakulteta. 38–48.
- KOSEM, Iztok, GANTAR, Polona, KREK, Simon, 2012: Avtomatsko luščenje leksikalnih podatkov iz korpusa. Tomaž Erjavec, Jerneja Žganec Gros (ur.): *Zbornik osme konference Jezikovne tehnologije, 8. do 12. oktober 2012, Ljubljana, Slovenija: zbornik 15. mednarodne multikonference Informacijska družba – IS 2012, zvezek*. Ljubljana: Institut Jožef Stefan. 117–122.
- KREK, Simon, 2010: *Pridobivanje jezikovnih podatkov iz besedilnih korpusov za namen izdelave enojezičnih slovarjev in slovnice. Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- KRŽIŠNIK, Erika, 2004: Določanje jezikovno-in/ali besedilnozvrstne zaznamovanosti v frazeologiji. Erika Kržišnik (ur.): *Aktualizacija jezikovnozvrstne teorije na Slovenskem: členitev jezikovne resničnosti. Obdobja 22*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete. 103–120.
- LOGAR BERGINČ, Nataša, 2000/01: Kvalifikator ekspr. v Slovarju slovenskega knjižnega jezika na ravni frazeologije. *Jezik in slovstvo* 46/4. 137–148.
- LOGAR BERGINČ, Nataša, 2009: O dveh znanstvenomonografskih leksikalnih seznamih: Alenka Gložančev, Primož Jakopin, Mija Michelizza, Lučka Uršič, Andreja Žele: *Novejša slovenska leksika (v povezavi s spletnimi jezikovnimi viri)*. Ljubljana: Založba ZRC, ZRC SAZU, 2009. *Jezik in slovstvo* 54/3–4. 153–159.
- LORENTZEN, Henrik, THEILGAARD, Liisa, 2012: Online dictionaries – how do users find them and what do they do once they have? R. Vatvedt Fjeld, J. M. Torjusen (ur.): *Proceedings of the 15th EURALEX International Congress, Oslo, 7–11 August 2012*. Oslo: University of Oslo, Department of Linguistics and Scandinavian Studies. 654–660.
- MÜLLER, Jakob, 2009: Kritične misli in zamisli o SSKJ. Andrej Perdih (ur.): *Strokovni posvet o slovarju slovenskega jezika, 23. in 24. oktober 2008*. Ljubljana: Založba ZRC, ZRC SAZU. 17–20 (25).
- ROZMAN, Tadeja, 2010: *Vloga enojezičnega razlagalnega slovarja slovenščine pri razvoju jezikovne zmožnosti. Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- ROZMAN, Tadeja, STRITAR, Mojca, KRAPŠ VODOPIVEC, Irena, KOSEM, Iztok, KREK, Simon, 2010: *Nova didaktika poučevanja slovenskega jezika: sporazumevanje v slovenskem jeziku*. Ljubljana: Ministrstvo za šolstvo in šport, Amebis. www.slovenscina.eu/Media/Kazalniki/Kazalnik15/Nova_didaktika_Sporazumevanje.pdf.
- RUNDELL, Michael, KILGARRIFF, Adam, 2011: Automating the creation of dictionaries: where will it all end? F. Meunier, S. De Cock, G. Gilquin, M. Paquot (ur.): *A Taste for Corpora. A tribute to Professor Sylviane Granger*. Benjamins. 257–281.
- Slovar slovenskega knjižnega jezika (1970–1991/ spletna različica 2000)*: <http://bos.zrc-sazu.si/sskj.html>. Ljubljana: Založba ZRC, ZRC SAZU.
- SUMMERS, D., 1987: *Introduction to LDOCE2*. Harlow: Longman.