

Iztok Kosem

Trojina, zavod za uporabno slovenistiko

Slovar v digitalnem mediju

V prispevku se osredotočamo na utemeljevanje digitalnega medija kot izhodiščnega medija za zasnovo novega slovarja sodobnega slovenskega jezika, kar podpremo tudi z relevantnimi primeri iz evropske leksikografske prakse. Predstavimo prednosti medija tako za leksikografe kot uporabnike, pa tudi izzive, ki pri prehodu v ta slovarski format čakajo leksikografijo kot vedo. Zaključimo, da moramo za potrebe jezikoslovcev in računalniškega procesiranja slovenščine razmišljati tudi o odprto dostopni slovarski bazi.

A dictionary of contemporary Slovene in the digital form

The paper focusses on arguing for the digital form as the main form of a new dictionary of contemporary Slovene; the argument is supported by relevant examples from European lexicography. The paper presents the advantages of the digital form for lexicographers as well as dictionary users, and the challenges of conceptualizing born-digital dictionaries for lexicography as a discipline. In conclusion, we point out that open access to the dictionary database must be offered in order to meet the needs of linguists and natural language processing.

Ključne besede: digitalni medij, koncept, predlog za izdelavo slovarja sodobnega slovenskega jezika, uporabniki

Keywords: digital form, concept, a proposal for a dictionary of contemporary Slovene, users

0 Uvod

Izbira medija oz. medijev, v katerih bo slovar ponujen uporabnikom, je postala pomemben del konceptualizacije sodobnih slovarjev. Še do pred desetletjem so bili slovarji vselej zasnovani za knjižno obliko; njihove elektronske verzije so bile zgolj (marketinški) dodatek. V zadnjih 5–10 letih se je zaradi razmaha tehnologij in svetovnega spleta ter s tem povezanih zahtev in potreb slovarskih uporabnikov oblikovala povsem drugačna slika; prevladal je digitalni oz.

spletni medij, pojavili so se pozivi po koncipiranju in izdelavi slovarjev, zasnovanih za elektronski oz. spletni medij (Nesi 2000; de Schryver 2003; Pajsz 2009). Kazalcev digitalne revolucije v leksikografiji je več: historični slovar Oxford (OED) in nedavno slovar angleščine za tuje govorce Macmillan sta se popolnoma odrekla tiskani izdaji, izhodiščno spletne slovarje so začeli izdelovati na Poljskem, Nizozemskem ipd., podobna razmišljanja kažejo leksikografi v mnogih drugih državah. Navsezadnje o zavedanju leksikografske skupnosti, v katero smer morajo iti razmisleki sodobne leksikografije, pričajo tri mednarodne konference eLex na temo elektronske leksikografije in ustanovitev evropske mreže elektronske leksikografije v okviru akcije COST.

Slovenska leksikografija, zlasti enojezična, se na digitalno revolucijo še ni ustrezno odzvala. Spletni slovarji slovenščine ostajajo zgolj na spletu objavljeni knjižni slovarji, pri čemer še vedno uporabljamo uporabniško neprijazne poslovne modele, kot je objavljjanje spletnih izdaj več mesecev ali let po izdaji knjižne (Slovar novejšega besedja slovenskega jezika, 2013¹) ali celo pogojevanje dostopa do spletne verzije z nakupom knjižne izdaje (načrtovano za SSKJ2²).

Prispevek se osredotoča na utemeljitev izbire digitalnega medija kot glavnega medija za slovar sodobnega slovenskega jezika ter na predstavitev segmentov slovarja, ključnih za zadovoljitev potreb leksikografov, slovarskih uporabnikov in jezikoslovne znanosti nasploh.

Koncept: slovar ali slovarska baza?

Različni mediji, v katerih lahko uporabniku predstavimo slovarsko vsebino, vzbujajo vtis, da morajo leksikografi izbrati za koncept slovarja primarni medij, kateremu bodo ostali mediji podrejeni. Dejstvo je, da je takšna izbira zgolj navidezna, saj jo narekujejo slovarski uporabniki in lastnosti razpoložljivih slovarskih medijev. Slovarski uporabniki danes živijo in delujejo v digitalnem svetu in do vseh informacij, tudi slovarskih, dostopajo prek digitalnih medijev; uporabniki tiskanih slovarjev so v veliki manjšini. Digitalni medij oz. mediji lahko vsebujejo veliko več informacij kot tiskani medij, omogočajo različne oblike vizualizacije, redno posodabljanje in povezovanje z ostalimi referenčnimi in drugimi viri. Oblikovanje

¹ Pri čemer je pri spletni izdaji SNB težko govoriti o elektronskem mediju, saj gre zgolj za digitalizirano verzijo knjige.

² Mag. Aleš Pogačnik, vodja založbe ZRC, na javni razpravi *Slovensko slovaropisje danes*, 11. decembra 2013.

slovarskega koncepta na podlagi tiskane verzije tako ni smiselno, saj se po eni strani odločitev o izdaji tiskane verzije lahko prestavi na čas, ko je slovar dokončan oz. je dokončana zadovoljiva količina gesel,³ po drugi strani pa se že med samo konceptualizacijo, pa tudi izdelavo, izognemo razmišljanju o vseh omejitvah medija in posledično vključevanju manjšega števila informacij (tako že v zgodnjih fazah lahko izbiramo manjše število kolokacij in zgledov, bolj radikalno krajšamo zglede ipd.).

Izbira slovarskega medija torej ni več vprašanje – slovar mora biti zasnovan za digitalni medij. Vendar pa velja razmišljati tudi širše: leksikografi med analizo gradiva lahko beležijo tudi informacije, ki morda ne bodo končale v slovarju, a so dragocene za izdelavo priročnikov (npr. za kolokacijski slovar, slovar sinonimov), jezikoslovne raziskave in jezikovnotehnološke namene. Ravno zato je dejansko bolj smiselno govoriti o izdelavi koncepta za slovarsko bazo, saj bi se na tak način zagotovila uporabnost informacij v njej za različne namene in izognilo podvajanju (leksikografskega) dela in denarja.

V nadaljevanju je predstavljenih nekaj prednosti, ki jih digitalni medij prinaša leksikografom, slovarskim uporabnikom in jezikoslovni znanosti.

Digitalni medij: leksikografove sanje

“[...] jezik je živ organizem, zato se nenehno razvija in spreminja [...]” (Bajec, Kolarič, Rupel 1956)

Ena od večjih težav slovaropisja je sama narava jezika – ker se jezik nenehno spreminja, vsak slovar predstavlja jezik v določenem trenutku oziroma v določenem (preteklem) obdobju, posledično pa to pomeni, da leksikografovega dela nikoli ni konec. Tega so se zavedali že med pripravo SSKJ:

Na rob vsemu pa še ugotovitev, da bi bilo treba pri izbiri načina redakcijske obdelave stvarno posnemati najboljše in najmoderneje tuje obdelave, vendar s samostojno aplikacijo na slovenskem gradivu. Pri našem počasnem tempu in pomanjkanju strokovnjakov se namreč utegne zgoditi, da bo delo v času izdaje zastarelo, pa če bo še tako sodobno zasnovano. Težiti je treba k temu, da bo zaostanek kar se da majhen. (Pogorelec 1964: 242)

Spoznanje je bilo še enkrat potrjeno ob izdaji:

³ Takrat se lahko tudi na podlagi študij med uporabniki odloči, ali je tiskana verzija sploh potrebna.

“Vendar slovar najbrž ne bo ušel očitku, da se je že rodil z nadihom starinskosti; toda temu se v celoti ne more izogniti noben slovar [...]” (Predgovor SSKJ, Franc Jakopin, 1991)

Trditev Franca Jakopina za sodobne slovarje drži v precej manjši meri ravno na račun digitalnega medija, ki je med drugim omogočil prekinitev navezanosti na abecedni vrstni red (obdelava in objava gesel lahko poteka pod drugimi kriteriji, npr. glede na prioritizacijo določenih segmentov besedišča), omogoča pa tudi sprotno objavo rezultatov leksikografskega dela v vnaprej določenih časovnih razponih (npr. vsake tri mesece, pri čemer so posodobitve dokumentirane)⁴ in s tem takojšnji odziv na spremembe v jeziku. Digitalni medij ponuja tudi različne načine komuniciranja med uporabniki in leksikografi (npr. direktno, z omogočanjem povratne informacije prek kontaktnih obrazcev in, po leksikografovi presoji, vključevanje nove vsebine ali popravkov; ali posredno, tj. z analizo uporabniških poizvedb) in posledično sodelovanja pri izboljšavi slovarja.

Digitalni medij nam tudi omogoča, da jezik opisujemo z drugimi, nejezikovnimi sredstvi, ki so včasih učinkovitejša od razlag (Kiefer in Sterkenburg 2003: 360). Tako si na primer na podlagi razlage "priprava na dolgi palici za beljenje" za besedo *maček* (SNB 2013) uporabnik, ki ni poznavalec pleskanja, zelo težko predstavlja, za kakšno pripravo gre,⁵ ob dodani sliki in/ali videoposnetku pa mu postane takoj jasno ne samo to, kako priprava izgleda, temveč tudi, kako se uporablja. Avdiovizualna gradiva, npr. zvočni posnetek izgovora ali videoposnetek prikaza v znakovnem jeziku, lahko zagotovijo tisto avtentičnost, ki je zapisana oblika nikoli ne more ponuditi v enaki meri ali pa sploh ne.

Vključitev multimedijskih gradiv v slovar je predvidena tudi v Predlogu za izdelavo slovarja sodobnega slovenskega jezika (Krek idr. 2013: str. 31 in 39), v katerem sta predstavljena tudi dva možna načina prikaza razmerij med besedami. Prvi je izbor kolokatorjev za določeno iztočnico, pri čemer pogostost kolokatorja odraža velikost pisave (Slika 1), drugi pa razmerja med sinonimi (Slika 2), ki so prikazana s pomočjo najpogostejših kolokatorjev iztočnice in njenega sinonima, "in sicer na 5-stopenjski lestvici, ki uporabniku pomaga pri odločitvi,

⁴ Angleški historični slovar Oxford na primer objavi nova ali popravljena obstoječa gesla vsake tri mesece.

⁵ Tu ni nezanemarljiva tudi dvoumnost razlage, ali gre za "pripravo na dolgi palici, s katero belimo" ali "pripravo, ki jo najdemo na dolgi palici za beljenje". Takšni tipi razlage so že v SSKJ naleteli na kritike (gl. Kosem 2006: 37–38).

kakšno besedilno okolje je za določen sinonim v razmerju do iztočnice najbolj sprejemljivo". (Krek idr. 2013: 38).

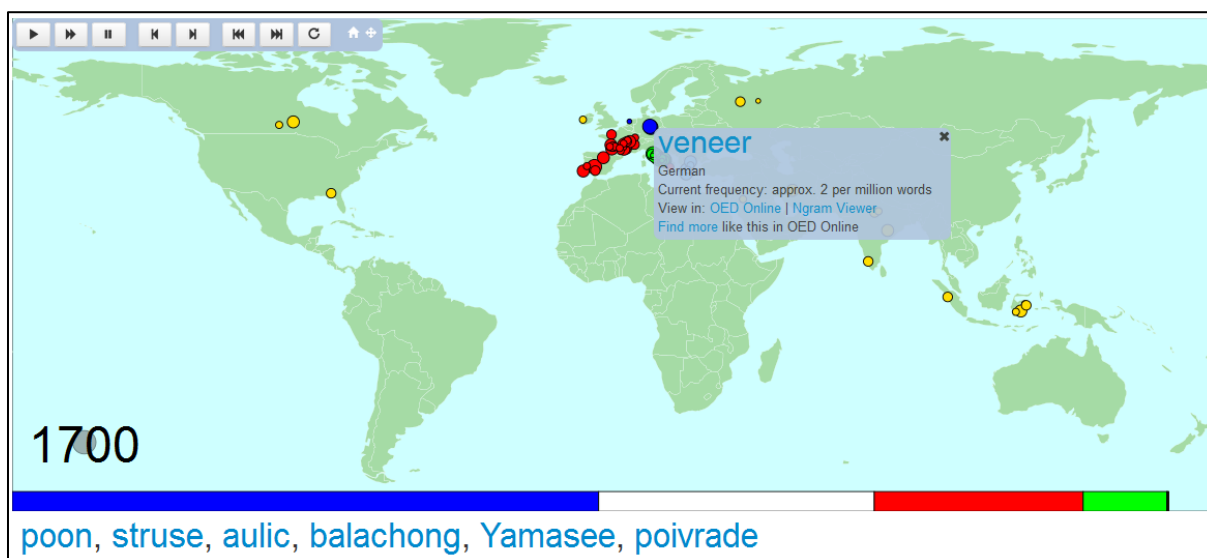


Slika 1: Izbor kolokatorjev



Slika 2: Prikaz razmerja med sinonimi

Slovarske informacije lahko povežemo z informacijami iz ostalih virov (slovarjev, korpusov ipd.) in na ta način ponudimo uporabniku vse informacije o jeziku na enem mestu. Navsezadnje pa lahko informacije iz slovarske baze prikažemo na drugačen, interaktiven, a manj klasičnoslovarski način. Dober primer je prikaz izvora angleških besed glede na geografsko območje, ki temelji na etimoloških podatkih v historičnem slovarju založbe Oxford. Gre za interaktivno karto sveta, na kateri si lahko ogledamo, kako so po posameznih letih (od sredine 12. stoletja naprej) besede prodirale v angleški jezik. Vsak krogec predstavlja besedo, ko se nanj premaknemo z miško, se nam prikažejo relevantni podatki in tudi povezava na slovar.



Slika 3: Izvor angleških besed glede na območje in leto izvora

Digitalni medij: uporabnikova realnost

Seveda doslej našteje prednosti digitalnega medija veljajo tako za leksikografe kot tudi za slovarske uporabnike. Ravno ciljne uporabnike morajo snovalci slovarja imeti še kako v mislih, saj naj bi slovar reševal potrebe in ugodil zahtevam uporabnikov. Kot pravita Atkins in Rundell (2008: 5), je eden najpomembnejših nasvetov, ki jih lahko dasta izdelovalcem slovarjev, ravno ta, da morajo dobro poznati svojega uporabnika oz. uporabnike.

Ciljne uporabnike slovarja sodobnega slovenskega jezika bi lahko razdelili v tri skupine: šolski, splošni in specializirani uporabniki (jezikoslovci, lektorji, prevajalci ipd.). Specializirani uporabniki so od vseh treh skupin najzahtevnejši in najbolj jezikovno podkovani, potrebujejo pa osnovne slovarske informacije precej manj pogosto kot dodatne informacije, npr. lektorji normativne, prevajalci sinonime ipd. Pa vendar se slovaropisem zelo rado zgodi, da prilagodijo rešitve zahtevnejšemu in ne povprečnemu uporabniku. To je bilo pogosto razbrati na slovarskem posvetu leta 2008 (Perdih 2009) v kritikah SSKJ in iz predlaganih rešitev za novi slovar. Pomembno je torej opredeliti t. i. povprečnega uporabnika (gl. Vidovič-Muha v Perdih 2009: 24) in njegove lastnosti ter potrebe, mu prilagoditi rešitve, po drugi strani pa razmišljati o zadovoljitvi potreb vseh tipov ciljnih uporabnikov, npr. s stopenjskim podajanjem slovarskih informacij.

Na področju, ki se ukvarja z analiziranjem uporabe slovarjev in študijami slovarskih uporabnikov, je slovenska leksikografija zelo slabo podprta (ena redkih študij je npr. Rozman 2010), hkrati pa tudi stroka nikoli ni pokazala velikega interesa po vključitvi uporabniških študij v konceptualizacijo in izdelavo slovarja. Še več, neredka je praksa, da strokovnjaki določeno predlagano slovarsko rešitev utemeljujejo z lastnimi izkušnjami uporabe slovarjev (ki pa seveda nikakor ni odraz izkušnje povprečnega uporabnika), npr.:

Sheme so v Uvodu v SSKJ, ampak meni kot uporabnici je težko iskati po shemah, čeprav so prikazane zelo sistematično. (Jožica Škofic v Perdih 2009: 142)

in

Na hitro sem pogledala angleški slovar, ampak se mi na prvi pogled ni zdelo tako zelo strogo strokovno, kot je tukaj videti. (Erika Kržišnik v Perdih 2009: 89)

Dejstvo, da nimamo veliko podatkov o tem, kako uporabniki uporabljajo slovarje, lahko deloma nadomestimo z upoštevanjem splošnih značilnosti sodobnih slovarskih uporabnikov: močno prevladujejo pripadniki generacij, rojenih v digitalnem svetu, ki večji del svoje receptivne in produktivne rabe jezike opravijo na digitalnih medijih. Hkrati v Sloveniji, tako med mladimi kot odraslimi, vlada vse slabša (bralna) pismenost, slabšajo se bralne navade; kot je bilo med drugim izpostavljeno na posvetu o bralni pismenosti, ki je potekal 7. februarja 2014 na Ministrstvu za izobraževanje, znanost in šport, igrajo pomembno vlogo v bralni pismenosti tudi ustrezni in kvalitetni jezikovni viri in orodja. Za uporabnike je tako ključno, da jim čim prej ponudimo slovar v digitalni obliki oz. oblikah, s katerim bodo lahko hitro in na enostaven način našli raznorazne informacije za reševanje svojih jezikovnih zagat.

Digitalni medij: znanstveni izziv

Ni dvoma, da so slovarji zelo pomembni za razvoj slovenske jezikoslovne znanosti, ki pa se že od izida SSKJ ne more ponašati z večjim leksikografskim dosežkom. Ker je v teh 23 letih prišlo do pomembnih premikov v leksikografski teoriji in praksi (npr. uveljavitev korpusne leksikografije in elektronske leksikografije), bo načrtovani slovar, ki naj bi oz. mora vključevati sodobne pristope, tudi odraz stanja slovenske leksikografske in jezikoslovne stroke. Gledano v prihodnost pa za slovenske jezikoslovce, pa tudi za strojno obdelavo

slovenščine in ostale discipline, ki potrebujejo jezikoslovne podatke, načrtovani slovar ne bo tako dragocen kot slovarska baza, ki bo omogočala različne raziskave slovenskega jezika; ravno zato je nujno zagotoviti njeno odprto dostopnost (gl. Erjavec 2014).

Načini izdelave digitalnih slovarjev, prikaz in povezovanje slovarskih informacij, uporaba sodobnih metod in orodij pa so pomemben izziv tudi za leksikografijo kot vedo, o čemer med drugim priča nedavna ustanovitev evropske leksikografske mreže ENeL, financirane s strani EU v okviru akcije COST, ki se osredotoča na elektronsko leksikografijo in združuje leksikografe, jezikoslovce, jezikovne tehnologe iz 26 evropskih držav, tudi Slovenije. Z digitalno zasnovanim slovarjem bi se pridružili majhnemu naboru držav, ki so že začele izdelovati tovrstne slovarje, in prispevali k uveljavitvi slovenske znanosti; tako bi se lahko zgodilo, da bi za spremembo leksikografij iz tujine navajali dosežke slovenske leksikografije kot primer sodobne leksikografske prakse. Da to ni nemogoče, dokazuje tudi zanimanje založbe Oxford in Inštituta za nizozemski jezik za predstavitev metodologije, opisane v Predlogu (Krek idr. 2013), in sicer zlasti avtomatsko luščenje podatkov iz korpusa in uporabo moči množic.

Zaključek

V času od posveta o novem slovarju slovenskega jezika (Perdih 2009) žal še nismo dobili novega slovarja; dejansko nismo niti začeli z njegovo izdelavo. V vmesnem obdobju se je zgodilo veliko pomembnega: v okviru projekta Sporazumevanje v slovenskem jeziku (2008–2013), financiranem z evropskim in državnim denarjem v širšem konzorciju, so bili izdelani viri in orodja, ki so ključnega pomena za izdelavo slovarjev: 1,18-milijardni referenčni korpus Gigafida, uravnoteženi korpus Kres, korpus govornjene slovenščine Gos, leksikon besednih oblik Sloleks, oblikoslovni označevalnik, skladijski razčlenjevalnik in leksikalna baza za slovenščino. Pri izdelavi leksikalne baze so bile izdelane in preizkušene sodobne in jezikovnotehnološko podprte leksikografske metode, med pomembnimi rezultati, ki izhajajo iz dela na projektu, pa je tudi Predlog za izdelavo slovarja sodobnega slovenskega jezika (Krek idr. 2013), ki je pomagal oživiti in poživiti razprave o novem slovarju.

Če še leta 2008 slovarski medij ni bil med osrednjimi temami posveta, je danes to ena ključnih odločitev snovalcev slovarskega koncepta. V tem prispevku smo pokazali, kako je digitalni medij ne samo prava, temveč tudi nujna izbira za izdelavo sodobnega slovarja za

sodobne uporabnike, pri čemer moramo za potrebe jezikoslovcev in računalniškega procesiranja slovenščine razmišljati tudi o odprto dostopni slovarski bazi. Vso potrebno infrastrukturo in znanje za izdelavo takšnega slovarja imamo, zdaj pa je naša odgovornost do uporabnikov in slovenske znanosti, da upravičimo njihov dosedanji vložek v nas in jim čim prej ponudimo slovar, kot si ga slovenski govorniki in tudi slovenski jezik zaslužimo.

Reference

- ATKINS, Sue B. T.; Michael RUNDELL, 2008: *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- BAJEC, Anton; Rudolf KOLARIČ, Mirko RUPEL, 1956: *Slovenska slovnica*. Ljubljana: Državna založba Slovenije.
- DE SCHRYVER, Gilles-Maurice, 2003: *Lexicographers' Dreams in the Electronic-Dictionary Age*. *International Journal of Lexicography* 16(2), 143–199.
- ERJAVEC, Tomaž, 2014: *Odprt dostop do podatkovne baze slovarja*. *Posvet o novem slovarju sodobnega slovenskega jezika*, 12. 2. 2014. Ministrstvo za kulturo Republike Slovenije.
- KIEFER, Ferenc; Piet VAN STERKENBURG, 2003: *Design and production of monolingual dictionaries*. V Piet van Sterkenburg, ur.: *A Practical Guide to Lexicography*. Amsterdam: John Benjamins, 350–365.
- KOSEM, Iztok, 2006: *Definicijski jezik v Slovarju slovenskega knjižnega jezika s stališča sodobnih leksikografskih načel*. *Jezik in slovstvo* 51(5), 25–45.
- KREK, Simon; Iztok KOSEM, Polona GANTAR, 2013: *Predlog za izdelavo Slovarja sodobnega slovenskega jezika*. Verzija 1.1., 22. 5. 2013, www.sssj.si.
- NESI, Hilary, 2000: *The Use and Abuse of EFL Dictionaries. How learners of English as a foreign language read and interpret dictionary entries*. Tübingen: Max Niemeyer Verlag.
- PAJZS, Julia, 2009. *On the Possibility of Creating Multifunctional Lexicographical Databases*. V Henning Bergenholtz, Sandro Nielsen in Sven Tarp, ur.: *Lexicography at a Crossroads*. Bern: Peter Lang. 327–354.
- PERDIH, Andrej, ur., 2009: *Strokovni posvet o novem slovarju slovenskega jezika*. 23. in 24. 10. 2008. Ljubljana: Založba ZRC, ZRC SAZU.
- POGORELEC, Breda, 1964: *Ob poskusnem snopiču Slovarja slovenskega knjižnega jezika*. *Jezik in slovstvo* 9(7–8), 232–242.
- ROZMAN, Tadeja, 2010: *Vloga enojezičnega razlagalnega slovarja slovenščine pri razvoju jezikovne zmožnosti: doktorska disertacija*. Ljubljana: Filozofska fakulteta UL.

Slovar novejšega besedja slovenskega jezika, 2013. Ljubljana: Založba ZRC.

Slovar slovenskega knjižnega jezika, 1991. Ljubljana: SAZU, ZRC SAZU, Inštitut za slovenski jezik, DZS.