

Marko Snoj

Inštitut za slovenski jezik Frana Ramovša ZRC SAZU

Objavne oblike novega slovarja slovenskega jezika

Prispevek izraža stališče Inštituta za slovenski jezik Frana Ramovša ZRC SAZU, po katerem bi moral novi splošni slovar slovenskega jezika iziti v spletni in vsebinsko enaki lokalni elektronski izdaji, knjižna pa bi bila od teh lahko siromašnejša le za vsebine, ki jih papir ne prenese zlahka ali sploh ne. Dotika se tudi vprašanja prioritetenih lastnosti novega slovarja, pri čemer izpostavlja predvsem zanesljivost njegovih vsebin in jezikoslovno neoporečnost.

Summary

The paper presents the views of the Fran Ramovš Institute of the Slovenian Language of the Research Centre of the Slovenian Academy of Sciences and Arts according to which the new general dictionary of the Slovenian language should be published in online electronic edition and in its local version whose contents should be identical. The print edition could only do without those contents that cannot be easily adapted to paper format or cannot be adapted to it at all. It also deals with the question of which features of the new dictionary should be given priority underlining in particular the reliability of its contents and the need for it to be faultless from the point of view of linguistics.

Ključne besede: slovaropisje, elektronske in knjižna izdaja, zanesljivost slovarske vsebine

Keywords: lexicography, electronic and print edition, reliability of the dictionary content

V začetku letošnjega leta, vendar že po 2. januarju, ko smo bili Službi za slovenski jezik obvezani poslati povzetke predavanj za današnji posvet, sem od ministra za kulturo, dr. Uroša Grilca, izvedel, da je eden od petih pogojev za financiranje novega slovarja popolna prilagojenost digitalnemu okolju in izid v fizični obliki. Če to formulacijo, zlasti uporabo pridevnika *fizični*, razumem prav, se je politika o tem vprašanju odločila tako, kot smo se pred nekaj leti odločili na Inštitutu za slovenski jezik: sodobni slovar, ki ga v celoti financiramo davkoplačevalci, mora biti prosto dostopen na spletu, iziti mora v lokalni elektronski izdaji in knjižni obliki, pri čemer velja, da mora biti osnovna vsebina elektronskih izdaj enaka vsebini

knjižne, ob tem pa je vsebina elektronskih lahko bogatejša za tiste prvine, ki jih knjižna iz naravnih ali tradicionalnih razlogov ne prenese, kot sta npr. večpredstavnost in povezave z zunanjimi bazami.

Vsebino mojega prispevka napovedujeta predvsem druga (digitalna slovarska baza ali slovar v knjižni obliki) in deseta alineja (sprotna objava slovarja na spletu) v vabilu predlaganih tem, dotakne pa se tudi sedemnajste (organizacija dela in usposobljenost kadrov). Uvodoma bi rad povedal, da izraz iz druge alineje *digitalna slovarska baza ali slovar v knjižni obliki* ne odseva prave dileme, saj si verjetno nihče ne predstavlja, da bi bilo treba narediti samo digitalno slovarsko bazo ali izdelati samo slovar v knjižni obliki. Digitalna slovarska baza je namreč eden od pogojev, ki omogoča objavo slovarja v kakršni koli obliki, od katerih je po mnenju mnogih, ne samo mene, ena spletna (z mobilno različico), druga elektronska lokalna z zmogljivejšim iskalnikom, tretja pa knjižna. Slovarska baza mora zadoščati trem pogojem, in sicer (1) tehničnemu, (2) obsegovnemu in (3) jezikoslovnemu. K 1: sestavljena mora biti v formatu xml, ki omogoča najrazličnejša procesiranja vsebovanih podatkov, povezljivost z izbranimi zunanjimi bazami kot spremljajočimi deli slovarja in dolgoročno uporabnost, vključno z možnostjo dopolnjevanja in spreminjanja. Vprašanje standardiziranosti formata je vprašanje, ki ga morata rešiti v sodelovanju jezikoslovec in računalničar. K 2: vsebovati mora vsaj jedrne slovarske sestavine tistega besedja, ki ga določa ustrezni del slovarskega koncepta, pri čemer je pričakovati, da je število enobesednih nelastnoimenskih lem okrog 100.000. Izraz *jedrne slovarske sestavine* obsega dinamično in tonemsko naglašeno lemo, njeno zborna izgovarjavo (v knjižni izdaji je to fonetični zapis /kakšen, je stvar dogovora, jaz bi zagovarjal način, kot je v SSKJ, saj je domačemu uporabniku prijaznejši kot IPA ali kak drug/ izgovorno nepredvidljivih lem ali stranske oblike, v elektronskih je smiselni tudi zvočni posnetek), pregibnost (ki v e-izdajah sestoji iz navajanja vseh njenih dinamično in tonemsko onaglašanih oblik, v knjižni pa iz tradicionalne druge oblike in sklica na ustrezni pregibnostno-naglasni vzorec v uvodnem delu), pomenske indikatorje, opis kategorialnega, denotativnih, konotativnih in po potrebi pragmatičnih pomenov ter zglede rabe, tudi v obliki tipičnih besednih zvez (ki jim danes rečemo kolokacije) in iztržkov. Temu bi bilo že v sami bazi treba priključiti še frazeološke podatke in navedbo ter opise stalnih besednih zvez ter vsaj do določene stopnje tudi splošnoslovarsko obravnavo širše poznanih terminov. Priporočljivo je, da posamezno geslo v slovarski bazi zaokroži etimološka osvetlitev, ki je zares dobra le, če jo vpeljejo podatki o zgodovini obravnavane besede v slovenskem pisanem jeziku. K 3: vsebina slovarske baze mora biti prek navodil za sestavo in slovarskega koncepta skladna s sodobno jezikoslovno teorijo. Tu bi izmed mnogih izpostavil le dve možni težavi. (1) Če skušamo

doseči, da bo slovarski uporabnik intuitivno razumel slovarsko vsebino, ki na korekten način opisuje jezikovno stvarnost, mu moramo v eni lemi podati le opis ene besede in zvez z njo. To v prvi vrsti pomeni, da se pod eno lemo ne sme opisovati dveh ali več enakoglasnic (*atlas*) ali celo enakopisnic (*kila*), ki jih moramo znati ločevati od večpomenskih besed (*sejati*), in da se pod eno lemo ne smejo pojavljati oblike, ki spadajo h kaki drugi besedi (tožilniška oblika *hčero* npr. spada k lemi *hčera*, če bo uvrščena, in ne k lemi *hči*, kamor je napačno uvrščena npr. v bazi Sloleks). (2) Od slovarja pričakujemo natančno besednovrstno uvrstitev, kot ta pojem razumemo v okviru jezikovnosistemskega pristopa, ki ga je smiselno obogatiti z besedilnofunkcijskim. Način prikaza besednovrstne uvrstitve, zlasti pri konvertitih, je stvar doslednega upoštevanja izbrane teoretske podlage. Sistemskojezikovni pristop upošteva tako zunajbesedilno kakor tudi znotrajbesedilno merilo in je v srednjeevropskem prostoru tradicionalno priznan tudi v šolski praksi, zato mu kaže dati prednost pred zgolj besedilnofunkcijskim pristopom, ki ga je treba smiselno vgraditi. Beseda *moški* je v svoji osnovni funkciji torej vrstni pridevnik, ki v moškem spolu konvertira v samostalnik s pomenom »človek moškega spola«. Način slovarskega prikaza takih konvertitov, ki se oblikovno ne razlikujejo od dela oblik motivirajoče besede, je sicer stvar dogovora, vendar se zdi mogoče zaradi oblikovnega merila obe besedi ohraniti v isti makrostrukturni enoti (SSKJ ju loči). Argument za obravnavo v isti makrostrukturni enoti je zunajbesedilna besednovrstna opredelitev besede *moški* in dejstvo, da je takih primerov vse več, npr. *maturantski* v pomenu »maturantski izlet«, *tretji* v pomenu »tretji dan v mesecu«. Nasprotno pa npr. samostalnika *ženska* ni mogoče obravnavati znotraj iste makrostrukturne enote kot pridevnik *ženski*, saj se pridevniške in samostalniške oblike razlikujejo v rodilniku dvojine in množine. Zunajbesedilna besednovrstna opredelitev besede *ženska* torej pogojuje samostojni slovarski sestavek, tako kot dva samostojna sestavka opredeljuje v primeru zaimka *jaz mene* in samostalnika *jaz -a*.

Ker govorimo o slovarju, ki bo v splošni rabi nadomestil SSKJ, moramo zagotoviti njegovo vsebinsko in referenčno zadostnost in zanesljivost. Za to je v prvi vrsti potrebno visoko strokovno oziroma znanstveno tehnološko in slovaropisno delo, ki ga opravljajo čim bolj usposobljeni strokovnjaki posameznih področij. Čim višja je usposobljenost sodelujočih, tem boljšo kakovost izdelka pričakujemo. Vlaganje sredstev v slovar, ki mu ne bi zastavil jasnega cilja, da mora biti vsebinsko in jezikoslovno zanesljiv, bi bilo zlasti pri nas, ki si ne moremo privoščiti več splošnih slovarjev, nesmotrno in neodgovorno. V Nemčiji izpeljana anketna raziskava je pokazala, da uporabniki od spletnih slovarjev v prvi vrsti pričakujejo zanesljivost

slovarskih vsebin, tej sledi pričakovanje po jasnosti in aktualnosti. Ista raziskava je pokazala, da bi uporabniki najlažje pogrešali tiste vsebine in oblike, ki so uresničljive le v spletnem mediju, namreč večpredstavnost, namigi za nadaljnje brskanje, prilagodljivost vmesnika in povezave do drugih slovarjev ter korpusov (Müller-Spitzer, Kopleinig in Töpel 2011: 203–208).¹

Kakovostnega in dovolj obsežnega slovarja ni mogoče sestaviti v kratkem času,² zato o tem ni treba izgubljeni besed. Jezikovni uporabniki bodo slovar sprejeli kot zanesljiv in referenčen priročnik le, če bodo objavljeni podatki ustrezali jezikovni stvarnosti in bodo ostali nespremenjeni v celotnem času njegove primarne uporabe, tj. do naslednje izdaje istega slovarja ali do sestave novega. Če se odločimo za sprotno spletno objavljane že izdelanih slovarskih sestavkov, se torej njihova vsebina po predhodni objavi ne sme več spreminjati, dopolnjuje pa se lahko samo v obliki povezav s tistimi zunanji bazami, ki v času objave posameznih slovarskih sestavkov še ne bodo pridružene. Ko bo slovar dokončan, mora poleg v elektronskih oblikah, predvidoma spletni, mobilni in lokalni, iziti tudi v tiskani, knjižni obliki, za kar govorijo vsaj štirje razlogi. (1) Pričakovati je, da bodo tudi v prihodnjih letih obstajali uporabniki, ki bodo dajali prednost tej obliki. Poleg tega obstajamo uporabniki, ki za določene potrebe uporabljamo elektronsko, za kako drugo, da npr. dobimo vpogled v slovarski koncept, pa knjižno obliko. (2) Knjižna oblika je edini porok, da se slovarske vsebine v spletnih in mobilnih objavah ne bodo po tistem spreminjale, s čimer se vzpostavi referenčnost. Pričakovanje, da se slovar v času primarne uporabe ne spreminja, je nujno iz povsem praktičnih razlogov. Navajam enega. S sodišč ali odvetniških družb na Inštitut za slovenski jezik naslavlajo zaprosila po ekspertnem mnenju, v katerih pričakujejo opredelitev, ali je uporaba določene besede ali besedne zveze žaljiva in kaj je v resnici sporočilo obravnavane izjave. Podajanje mnenja o konotativnih in denotativnih pomenih določene besede ali besedne zveze v prvi vrsti temelji na njenem opisu v SSKJ, ki se od prve objave do danes ni spreminjal, v drugi vrsti seveda tudi od znotrajbesedilnih okoliščin. Če bi se pomeni in oznake v splošnem slovarju lahko spreminjale čez noč, uporabniki ne bi imeli nobenega

¹ <http://www.trojina.si/elex2011/Vsebine/proceedings/eLex2011-27.pdf> (dostop 20. 2. 2014).

² Za samo spletni *Wielki słownik języka polskiego* (februarja 2014 je dostopen na spletnem naslovu <http://www.wsjp.pl/>) je bilo v petih letih izdelano 15.000 slovarskih sestavkov. Pri nastajanju je sodelovalo 39 stalnih in 13 občasni avtorjev, ob tem pa še štirje informatiki, štirje sestavljalci pregibnostnih oblik in šest dodatnih sodelavcev pri konceptu. Pri prav tako samo spletnem nizozemskem *Algemeen Nederlands Woordenboek* (februarja 2014 je dostopen na spletnem naslovu <http://anw.inl.nl/search>) je 24 sodelavcev doslej (s spletne strani ni razvidno, koliko časa poteka delo) izdelalo 12.600 geselskih sestavkov. Pri slovaškem slovarju *Slovník súčasného slovenského jazyka*, ki je doslej izšel do črke L samo v knjižni obliki, sodeluje 15 urednikov. Ob ustreznih jezikovnotehnološki podpori so v 13 letih izdelali skoraj 55.000 slovarskih sestavkov.

zagotovila, da so v svojem delovanju uporabili besedo v želenem (denotativnem in konotativnem) pomenu, sodišča pa nobenega trdnega kriterija za presojo o tem. (3) Edino knjižna oblika omogoča uporabo slovarja tudi v neidealnih okoliščinah, kot sta npr. izpad elektrike in uporaba na krajih, kjer ni računalnika. Lokalna izdaja omogoča uporabo slovarja tam, kjer ni povezave z internetom. K temu je treba dodati, da ima v Sloveniji po podatkih Statističnega urada Republike Slovenije dostop do interneta 76 odstotkov gospodinjstev.³ Knjižna in elektronska lokalna izdaja torej omogočata uporabo slovarja vsem, ne le 76 odstotkom gospodinjstev. (4) Knjižna oblika od vseh danes uporabljenih medijev najučinkoviteje kljubuje izgubi podatkov, še zlasti če je izdana v visoki nakladi. Tako sta se nam npr. kljub uničenju, sežigu večine izvodov kmalu po izidu vse do danes ohranila dva skoraj pol tisočletja stara izvoda Trubarjeve *Cerkovne ordninge*, medtem ko je marsikatera datoteka z začetka devetdesetih let prejšnjega stoletja že za vedno izgubljena.⁴

Sklenem s povzetkom. Načrtovani slovar naj izide v vsebinsko enaki spletni, lokalni in knjižni izdaji, pri čemer je prvi in deloma drugi smiselno pridružiti le tiste zanesljive in kakovostno izdelane podatkovne zbirke, ki jih knjižni medij ne prenese zlahka ali sploh ne. Morebitne predhodne objave slovarskih sestavkov morajo biti tako domišljene, da bodo lahko ostale nespremenjene ves čas primarne uporabe slovarja.

³ https://www.stat.si/novica_prikazi.aspx?id=5795 (dostop 20. 2. 2014).

⁴ Pripomba prof. dr. Marka Stabeja, predsedujočega sklopu referatov, v katerem je bil predstavljen ta prispevek, češ da ima knjižna izdaja tudi kalorično vrednost, je vredna objave, komentarja pa ne potrebuje.