

Simon Krek, Iztok Kosem

Odgovor na prispevek "SSKJ danes in jutri, potem pa ..." ¹

Objavljen v: Delo, Sobotna priloga, 21. 9. 2013

Spletna stran: <http://www.delo.si/zgodbe/sobotnapriloga/sskj-danes-in-jutri-potem-pa-hellip.html>

Uvodno pojasnilo:

V nadaljevanju je izvorno besedilo prispevka iz Sobotne priloge označeno z olivnozeleno barvo, posamezni deli, na katere referiramo v odgovorih, pa z različnimi barvami, s katerimi so označene tudi številke pred posameznim odgovorom. Številka, označena z rumeno, torej označuje odgovor na trditev v predhodnem navedku iz prispevka v Sobotni prilogi, ki je prav tako označena z rumeno.

Proti koncu letošnje pomladi so dr. Simon Krek, dr. Iztok Kosem in dr. Polona Gantar na spletnih straneh zasebnega zavoda Trojina objavili Predlog za izdelavo Slovarja sodobnega slovenskega jezika, ki naj bi nasledil naš temeljni enojezični razlagalni Slovar slovenskega knjižnega jezika. Predlog predvideva izdelavo samo spletno in mobilno dostopnega slovarja, narejenega popolnoma na novo s korpusnim pristopom v petih letih, za katerega naj bi davkoplačevalci odšteli le 4,2 milijona evrov.

1. Predlog predvideva izdelavo slovarja, ki je metodološko zasnovan za digitalni medij, kar pomeni, da lahko v obsežni, podatkovno tako rekoč neomejeni slovarski bazi ali več povezanih bazah vsebuje večino informacij o slovenskem jeziku. Slovar na spletu ali v mobilnih napravah je zgolj ena od vizualizacij dela podatkov iz slovarske baze. V predlogu so na straneh 49-52 opisane tudi predvidene možnosti natisa podatkov v primerih, ko je to smiselno. Klasični tisk je opisan na strani 52:

»Knjižni« natis je še bolj strnjen in je pravzaprav namenjen klasičnemu tisku, torej natisu in vezavi slovarja v knjigo, če se bo morda pri določenih uporabnikih kazala takšna potreba. V tem natisu so ohranjene samo nekatere informacije iz gesla, po vsebini in izgledu pa je zelo podoben standardnim knjižnim slovarjem, tudi SSKJ in SNB. Transformacija gesel iz slovarske baze za natis v katerikoli od omenjenih variant je avtomatska.

2. Že od 19. stoletja naprej se vsi resni slovarji izdelujejo "s korpusnim pristopom", kar pomeni, da se slovaropisec pri izboru iztočnic, opisu pomenov, vključevanju zgledov itd. ne zanaša na introspekcijo, temveč ti podatki temeljijo na analizi in konsistentni interpretaciji zbranega gradiva ali "korpusa". Do osemdesetih let 20. stoletja je bilo to večinoma listkovno gradivo, izpisano iz knjig in drugih virov, od 90. let naprej pa izključno obsežni računalniški besedilni korpusi, saj ti omogočajo neprimerno hitrejše in kvalitetnejše analize. Semantični opis sodobnega jezika lahko temelji le na temeljiti analizi čim bolj obsežnega korpusa sodobnega jezika, kar je smiselno in možno izvesti zgolj s podporo računalniške analize besedil.

Leksikografska analiza vsebuje več kot le preprosto statistično analizo korpusa. Po nujnosti vključuje tudi jezikoslovno interpretacijo oz. dodajanje jezikoslovnih metainformacij v slovar, ki izhajajo iz interpretacije jezikovnega sistema, npr. opredelitev iztočnic glede na besednovrstno kategorijo ali njene lastnosti, način vključevanja skladenjskih informacij in podobno. Besedilo prispevka ustvarja

¹ Avtorja se zahvaljujeta vsem strokovnjakom, ki so prebrali in komentirali besedilo.

vtis o razliki med "korpusnim pristopom", pri katerem bi se informacije – denimo sistem strojnega jezikoslovnega označevanja korpusov – slepo prenesle v slovarsko bazo, in "tradicionalnim pristopom", ki naj bi upošteval tradicionalno jezikoslovno interpretacijo jezikovnega sistema. Ta razlika ne obstaja, saj je v obeh primerih tako ali drugače treba izpeljati jezikoslovno interpretacijo jezikovnih dejstev po izbrani jezikoslovni teoriji. V predlogu zagovarjamo sodobne slovaropisne teorije, ki upoštevajo neločljivo povezanost leksikalnih in skladenjskih lastnosti besed, v temelju kontekstualno naravo semantike besed, iz česar izhaja, da šele konkretni kontekst odloča o aktivaciji pomenskega potenciala iztočnice, ter sistem razlag, ki lahko vključuje tudi tipične skladenjske vzorce, v katerih se pojavlja iztočnica.

Pobuda navedenih treh jezikoslovcev je nastala mimo prizadevanj Inštituta za slovenski jezik Frana Ramovša ZRC SAZU, ki od ustanovitve dalje kontinuirano skrbi za opremljenost slovenskega jezika s temeljnimi jezikovnimi priročniki. Na inštitutu so namreč v preteklih desetletjih nastali vsi temeljni in mnogi aplikativni slovarski priročniki slovenskega jezika. Večina teh del je poleg v knjižni obliki prosto dostopna tudi na spletu v digitalni.

3. Pri predlogu gre za slovarski koncept, ki je nastal na podlagi štiriletnega dela na [Leksikalni bazi za slovenščino](#) kot delu projekta »[Sporazumevanje v slovenskem jeziku](#)«, ki sta ga iz strukturnih skladov financirala Evropska unija ter ministrstvo za izobraževanje, znanost in šport in pri katerem sta sodelovala tako zavod Trojina kot tudi Inštitut za slovenski jezik oz. ZRC SAZU. Predstojnik Inštituta je bil od leta 2008 redno obveščan o rezultatih projekta. Poleg tega so bili rezultati, torej leksikografske rešitve in korpusna metodologija, predstavljeni v znanstvenih člankih in na [rednih letnih konferencah](#) projekta. Iz [Specifikacij](#), ki so od junija 2009 na voljo na spletu, prav tako izhaja, da je Leksikalna baza za slovenščino namenjena izdelavi sodobnih jezikovnih priročnikov oz. različnih tipov slovarjev.

4. Termina »temeljni slovar« in »temeljno slovaropisje« v evropski in svetovni leksikografiji ne obstajata. Avtor prispevka ju že nekaj časa uporablja kot protipol prav tako neobstoječim »aplikativnim slovarjem« in skuša Inštitut s tem (samo)opredeliti kot »temeljno institucijo«, ki izdeluje »temeljni slovar«. Izraz, ki se v leksikografiji dejansko uporablja za podobne slovarje, kot je SSKJ, je »akademski« ali »akademijski« slovar, ki je določen tip slovarja in nima »aplikativnih« izpeljank.

5. Leta 2013, ko smo že globoko v digitalni dobi, v tem stavku ne bi smelo biti besed "večina" in "poleg". Vsa dela Inštituta bi morala biti prosto dostopna tako na spletu in kot baze podatkov, in sicer v skladu z [Resolucijo o raziskovalni in inovacijski strategiji Slovenije 2011–2020](#), ki izpostavlja "prost dostop do surovih podatkov iz raziskav, financiranih z javnimi sredstvi", ter [priporočilom Evropske komisije o dostopu in ohranjanju znanstvenih podatkov](#).

Ko govorimo o splošnem slovaropisju, imam v mislih predvsem Slovar slovenskega knjižnega jezika in njegovo dopolnilo Slovar novejšega besedja slovenskega jezika. Ta bo skupaj s preostalimi dodatki in popravki vkomponiran v drugo izdajo Slovarja slovenskega knjižnega jezika, ki jo v knjižni in elektronski obliki načrtujemo prihodnje leto. Ko k temu prištejemo še Slovenski pravopis, terminološke, etimološke in zgodovinske slovarje ter lingvistični atlas, pri čemer so ta dela teoretično podprta s skoraj sto naslovi znanstvenih monografij in več kot tisoč znanstvenih člankov, ugotovimo, da smo glede na opremljenost svojega jezika z opisnimi in normativnimi priročniki primerljivi z večino srednjeevropskih narodov.

6. SSKJ je nastajal od sredine prejšnjega stoletja do l. 1991 na gradivu, ki je bilo še starejše. To pomeni, da imamo v njem trenutno opisano le slovenščino, kakršna je bila v rabi nekje od začetka prejšnjega stoletja do sredine ali malo čez. Če se želimo poučiti o sodobnih pomenih besed, ta slovar danes težko uporabimo, razen kot zgodovinski vir, saj tudi za še aktualne pomene ne moremo biti prepričani, ali ustrezajo sodobni rabi ali ne. Slovar je tudi težko razumljiv, tako za splošno publiko kot za šolsko populacijo. To so znana in (znanstveno) dokumentirana dejstva. Leta 2013 izdani slovar novejšega besedja skoraj v celoti sledi leksikografski metodologiji SSKJ, zato kombinacije obeh omenjenih slovarjev ne bo mogoče vzeti za izhodišče novega slovarja oz. digitalne slovarske baze, kakršno bomo potrebovali v prihodnosti. Ta baza bo morala omogočiti črpanje raznovrstnih informacij o slovenščini in njihovo vizualizacijo za različne tipe uporabnikov, od učencev do znanstvenikov.

7. Vprašanje je, če se dejansko želimo primerjati zgolj s srednjeevropskimi narodi, ne pa tudi z zahodno- in severnoevropskimi. Pri dilemi, kako izdelati nov slovar slovenskega jezika, ne gre za to, kaj je trenutno na voljo, temveč v kaj je smiselno vlagati energijo in sredstva v prihodnje: na novo zasnovano digitalno slovarsko bazo ali nadgradnjo slovarjev s knjižno zasnovano, ki so bili izdelani v času pred digitalno dobo. Po Evropi potekajo številni slovarski projekti, ki jih je kljub zavedanju nove realnosti treba dokončati po stari metodologiji, saj imajo izhodiščno knjižno zasnovano, velik del slovarja pa je že izdelan. Tipična primera sta denimo slovaški² in švedski³ akademski slovar. Na drugi strani pri nemščini,⁴ nizozemščini,⁵ danščini,⁶ poljščini⁷ in drugih jezikih vlagajo v razvoj na novo zasnovanih digitalnih slovarskih baz. Pri slovenščini smo trenutno na točki, ko je knjižno zasnovan slovar (s sodobnejšim dopolnilom) na voljo, v načrtu je izdelava novega slovarja. Smiselno ga je konceptualizirati kot digitalno slovarsko bazo in ga začeti izdelovati po novi, sodobnejši metodologiji.

Slovaropisno delo na Inštitutu za slovenski jezik je tako v preteklosti kot danes sledilo sodobnim jezikoslovnim in slovaropisnim trendom. Glavni gradivni vir za splošno slovaropisje je bil nekdaj v obliki 6-milijonske listkovne kartoteke in ustreznih kartotečnih postopkov, danes so to besedilni korpusi slovenskega jezika in ustrezni jezikovnotehnoški postopki, redakcija pa poteka z uporabo vrhunskega leksikografskega programa. Slovarska vsebina in njen prikaz kakovostno napredujeta skladno z razvojem vseh za slovaropisje pomembnih jezikoslovnih področij in usmeritev, od fonetike prek pomenoslovja, frazeologije in normativistike, vključno s terminologijo, do etimologije.

8. Slediti sodobnim jezikoslovnim in slovaropisnim trendom pomeni biti strokovno aktiven v okviru evropske leksikografije. Na področju sodobnega slovaropisja se v zadnjem desetletju in še posebej v zadnjih petih letih veliko dogaja: slovarji, ki nastajajo danes, niso več to, kar so bili – predvsem po zaslugi prehoda v digitalni medij in upoštevanja potreb uporabnikov. O tem se je mogoče poučiti v strokovni literaturi, predvsem pa na srečanjih z leksikografi na mednarodnih leksikografskih in jezikovnotehnoških konferencah. Najbolj indikativni trendi zadnjih let se kažejo v nastanku in popularnosti serije konferenc *E-lexicography in the 21st century* (2009, 2011, 2013) ter vzpostavitvi vseevropskega projekta COST *European Network of e-Lexicography* (2013-2017), pri katerem

² [Slovník súčasného slovenského jazyka.](#)

³ [Svenska Akademiens ordbok.](#)

⁴ [Das Digitale Wörterbuch der deutschen Sprache.](#)

⁵ [Algemeen Nederlands Woordenboek.](#)

⁶ [Den Danske Ordbog.](#)

⁷ [Wielki słownik języka polskiego.](#)

sodelujejo predstavniki 25 držav, večinoma z inštitutov za nacionalne jezike. Osnovni namen avtorjev predloga je zagotoviti jezikovno infrastrukturo in približati jezikovni opis slovenščine ravni, kot jo imajo naprednejši evropski jeziki. Eden od avtorjev je član glavnega odbora evropske zveze leksikografov Euralex,⁸ dva sta predstavnika Slovenije v izvršnem odboru evropskega projekta COST *European Network of e-Lexicography*,⁹ vsi pa smo soorganizatorji konferenc *E-lexicography in the 21st century* (Bled 2011,¹⁰ Talin 2013¹¹).

9. Glej odgovor pod št. 2 in št. 6.

Poleg tega obstaja še akuten problem računalniške obdelave slovenskega jezika. Sodobna računalniška obdelava jezikov je s svojimi praktičnimi aplikacijami v zadnjih letih prišla do točke, ko se je treba lotiti tudi semantičnega razdvoumljanja besed v sobesedilu. Za uporabo te tehnologije je treba imeti na voljo slovar določenega jezika v obliki, ki je računalniško berljiva in prosto dostopna. Ta tehnologija je pomembna za številne računalniške aplikacije, od črkovalnikov, strojnega prevajanja, razpoznave govora do umetne inteligence na skrajnem koncu. Semantično razdvoumljanje postaja ena od temeljnih tehnologij, ki jo bo treba zagotoviti za vsak jezik, ki bo sploh zmogel prehod v digitalno dobo. Pomembnosti tega prehoda preprosto ni mogoče dovolj poudariti in je po teži verjetno primerljiv z vzpostavljanjem knjižne slovenščine v 16. stoletju. Za omenjeno (računalniško) rabo velja, da je treba imeti slovarsko bazo, kompatibilno s tistimi za druge jezike, strojno berljivo, aktualno, redno obnavljano in prosto dostopno.

Druga izdaja SSKJ bo premostila časovno vrzel do novega temeljnega razlagalnega slovarja, prilagojenega sodobnemu jezikovnemu uporabniku, v katerem bo uporabnik našel vse ali vsaj večino verodostojnih podatkov o našem besedju, torej nabor sodobnih slovenskih besed z izgovarjavo in vsemi (dinamično in tonemsko) onaglašeni stranskimi oblikami pregibnih besed, vključno s slogovno in časovno označenimi ali neoznačenimi dvojnicami, pomenske razlage, sinonime, vezljivostne podatke, slovnične, slogovne, časovne in terminološke kvalifikatorje, sodobno ponazarjalno gradivo, stalne besedne zveze, prikaz in razlago frazemov, besedotvorne in etimološke podatke ter še marsikaj drugega.

10. V leta 2012 izdanem prispevku z naslovom "Podgesla v Novem slovarju slovenskega jezika"¹² Marko Snoj trdi naslednje:

"Na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU potekajo pripravljalna in začetna dela za novi razlagalni slovar slovenskega jezika srednjega obsega, ki bo predvidoma prikazoval 70.000 sodobnih slovenskih nelastnoimenskih besed. Kljub dejstvu, da bodo uporabniki novega slovarja posegali predvsem po njegovi elektronski izdaji, bo ta sestavljen tako, da bo uporaben tudi v knjižni. Naš namen je narediti slovar v takem obsegu, da bomo lahko sodobno besedje slovenskega jezika prikazali tudi v eni ne preveč debeli knjigi slovarskega formata, in sicer s črkami take velikosti, da jih bo brez očal lahko prebral tudi zmerno daljnoviden uporabnik. Ta namen narekuje gospodarno ravnanje s prostorom, saj besedilo celotnega slovarja skoraj ne bo smelo biti daljše od 12 milijonov znakov, kar je približno polovica obsega SSKJ."

⁸ [European Association for Lexicography \(EURALEX\)](#).

⁹ [European Network of e-Lexicography](#).

¹⁰ Konferenca *eLex 2011* – [Electronic lexicography in the 21st century: new applications for new users](#).

¹¹ Konferenca *eLex 2013* – [Electronic lexicography in the 21st century: thinking outside the paper](#).

¹² Marko Snoj, "Podgesla v Novem slovarju slovenskega jezika". Škrabčevi dnevi 7 – Zbornik prispevkov s simpozija 2011, uredila Franc Marušič in Rok Žaucer. Nova Gorica : Univerza, 2012.

...

"Če bi se odločili za samostojni prikaz manjšalnice, bi morali namesto enega geselskega članka napisati dva, ki bi namesto 314 znakov obsegala 378 znakov. K temu moramo prišteti še **povprečno za polovico prazne vrstice daljši seštevek neizkoriščenega prostora za zadnjim znakom geselskega članka in četrtinski medvrstični razmik med dvema člankoma, kar skupaj zneso prostora za vsaj 100 znakov.**"

...

1. Zakaj uvedba podgesel, če se bo slovar uporabljal predvsem v elektronski izdaji? Ali ni vsaka beseda svoja enota, ki si zasluži individualno obravnavo v samostojnem geselskem članku? In ali ne bo sistem podgesel celo otežil iskanja po elektronski izdaji?

Odgovor. Obseg za elektronsko izdajo res ni takega pomena kot za knjižno, pa **vendar bo slovar izšel tudi v knjigi, ki ne bo smela biti preobsežna.** Iskanje v elektronski izdaji bo urejeno tako, da bo **privzeto nastavljeno iskanje po geselskih in podgeselskih besedah, ne pa tudi po geselskih besedah vodilkah.**

Navedki kažejo, da je zasnova za Novi slovar slovenskega jezika izrazito knjižna, čeprav je tudi avtorju prispevka jasno, da bo uporaba slovarja zgolj digitalna. Za razliko od javno objavljenega slovarskega predloga Inštitut svojega koncepta novega slovarja ni objavil, razen sporadično v omenjenem članku, kar pa za kakršnokoli resnejšo oceno ne zadostuje. Dejstvo je, da novi slovar mora biti zasnovan kot slovarska baza podatkov, vprašanja o številu znakov, varčevanju s prostorom na knjižni strani in druge navedbe iz omenjenega članka so konceptualno povsem sekundarne, takorekoč banalne tehnične narave in rešljive s preprostim strojnimi mehanizmom. Objavljena poskusna gesla pa nakazujejo, da bo novi slovar očitno še težje berljiv in manj razumljiv kot SSKJ:

máčka -e ž_{S200g} **1.** domača žival, ki lovi miši: *Mačka prede* **2.** mačja samica: *Mačka ima mladiče* **3.** kot povdk., nav. s pril., ekspr. izraža, da se v osebkcu imenovani osebi ženskega spola pripisuje, da je postavna, privlačna: *Ona je dobra mačka* || s pomenskim dopolnilom kot delovalnik *Pri mizi sta sedeli dve črnolasi mački máčkin -a -o* svoj. prid.^{49c} k pomenoma 1.–2. ↑ **máčica** -e ž_{S200g} manjš. **máčkica** -e ž_{S200g} manjš. **T** žival., mn. družina živali roparic z okroglo glavo, gibčnim trupom in ostrimi kremplji; *Felidae: levi, tigri, gepardi in druge mačke* **E** = hrv., srb. *máčka*, slovaš. *mačka* < slovan. **mačьka* iz vabnega klica *mac(a)*

máčica -e ž_{S200g} **1.** manjš. od mačka 1.–2.: *mačice in kužki* **2.** nav. mn. socvetje v obliki podolgovate kepice: *vrbove mačice* || vejice s takimi socvetji: *šopek mačic* **3.** nav. s pril., ekspr. ljubka, mikavna ženska: *spoznati ljubko mačico* **E** < **mačćica*, manjš. od ↑ **máčka**

pomaránča -e ž_{S200g} **1.** južni sadež z oranžno lupino: *limone in pomaranče* || zveza *grenka pomaranča* sorta pomaranč, primerna za sokove, marmelade **2.** pog. pomarančevc **1.** **pomaránčica** -e ž_{S200g} manjš. **E** < star. it. *pomancia* iz *pomo* 'jabolko, sadež' + *arancia* 'pomaranča' < špan. *naranja* < arab. *nārandž* < perz. *nārandž*

nosoróg -a _{m_{S1k}} velika čokata žival z zelo debelo kožo in enim ali dvema izrastkoma na sprednjem delu glave; *Rhinocerus*: *povodni konji in nosorogi*
nosorógov -a -O svoj. prid. _{p1k} **nosorógec** -gca _{m_{S1k}} manjš. **nosoróginja** -e _{ž_{S200f}}
nosoróginjin -a -O svoj. prid. _{p1k} **E** ↑nós + ↑róg po zgledu lat. *rhīnoceros* iz gr. *rhís* 'nos' + *kéras* 'rog'

dekàn -ána in **dekán** -a _m člov. _{S1a} in _{S1e} **1.** predstojnik fakultete: *Dekan je sklical sejo akademskega zbora* **2.** predstojnik dekanije: *Škof je napovedal pogovore z dekani* **dekánov** -a -O svoj. prid. _{p1a} **dekánica** -e _{ž_{S200a}} k pomenu 1. **dekáničin** -a -O svoj. prid. _{p1a} **dekánja** -e _{ž_{S200a}} k pomenu 1. **dekánjin** -a -O svoj. prid. _{p1a} **dekánka** -e _{ž_{S200a}} k pomenu 1. **dekánkin** -a -O svoj. prid. _{p1a} **E** < nem. *Dekan* < lat. *decānus* 'desetnik' k *decem* 'deset'

lisják -a _{m_{S1e}} **1.** lisičji samec: *V daljavi je zalajal lisjak* **2.** kot povdk., ekspr. izraža, da se v osebkcu imenovani moški osebi pripisuje, da je zvita, prebrisana: *Naš sosed je lisjak* || kot psovka *ti lisjak stari* || s pomenskim dopolnilom kot delovalnik *Temu lisjaku ne zaupaj!* **lisjákov** -a -O svoj. prid. _{p1a} **E** < **lisvjakъ* k ↑lisica

Predlog trojice jezikoslovcev iz prvega odstavka se odlikuje po uporabi nekaterih Inštitutu za slovenski jezik nedostopnih jezikovnotehnoloških orodij, ki naj bi delovni proces pohitrila in rezultate izboljšala, ter po uporabniku prijaznejši vizualizaciji slovarskih vsebin, vse njegove šibke točke pa je težko dokončno oceniti, saj 150 strani dolgo besedilo, ki opredeljuje predvsem slovaropisne postopke, organizacijo dela in finančni načrt, ni slovarski koncept. Iz prikaza osmih večinoma ne v celoti predstavljenih vzorčnih geselskih sestavkov pa lahko sklenemo, da po tem predlogu nastali slovar ne bi bil temeljni.

11. Tako kot vsi drugi uporabniki tudi Inštitut za slovenski jezik lahko uporablja vse vire in orodja, ki so bila razvita v okviru projekta Sporazumevanje v slovenskem jeziku in drugih javno financiranih projektih. Dostop do komercialnih virov, kot je Sketch Engine,¹³ na kar verjetno namiguje avtor prispevka, je za vse stranke tega angleškega podjetja po svetu enak in je odvisen od nakupa oz. dogovora z lastniki. Zato se vsiljuje misel, da morda Inštitut šteje orodja za nedostopna zgolj zato, ker nima ustreznih znanj o njih in njihovi uporabi.

12. Od 150 strani besedila v slovarskem predlogu SSSJ je zgolj 12 strani namenjenih organizaciji dela in finančnemu načrtu, "slovaropisne postopke" pa štejemo med nujne dele vsakega slovarskega koncepta. V ostalih delih dokumenta v podrobnostih opisujemo vsebinske rešitve za predlagani slovar. Za podrobnejši opis slovaropisnega postopka so na podlagi [predstavitve na konferenci E-lexicography v Talinu](#) denimo že izrazili zanimanje na Inštitutu za nizozemsko leksikologijo v Leidnu in na Oxford University Press.

13. Ta trditev najbrž postavlja pod vprašaj sam pojem "koncepta", ki ga – tudi po SSKJ – poznamo kot *osnovni, temeljni načrt, zasnova*. O objavljenem slovarskem predlogu si (strokovna) javnost vsaj lahko ustvari mnenje, saj je dokument javno dostopen. Inštitut svojega slovarskega koncepta nikjer ni razkril v vsaj približno podobno podrobnem opisu, zato si o njem ne moremo ustvariti mnenja. To je

¹³ <http://www.sketchengine.co.uk/>.

nenavadno, saj je Inštitut financiran iz javnih sredstev in bi njegovo delovanje moralo biti transparentno in javno. Ko je bilo vprašanje dostopnosti shem XML, iz katerih bi bilo mogoče vsaj sklepati, kakšna je struktura bodočega slovarja, javno zastavljeno na eni od jezikovnotehnoloških konferenc v Ljubljani, sta raziskovalca z ISJFR, po naročilu vodstva, kot sta se izrazila, lahko povedala le, da je to interni podatek ISJFR.

Dokler Inštitut javno ne objavi svojega slovarskega koncepta, je predlog za izdelavo Slovarja sodobnega slovenskega jezika edini koncept, o katerem je sploh mogoče razpravljati. O alternativni, torej o naslovnem *Novem slovarju slovenskega jezika* je mogoče delno sklepati zgolj iz precej bolj skromnega prej omenjenega prispevka: (1) vsebinsko bo novi slovar zasnovan za tisk, (2) imel bo zelo zapleten sistem oznak, kazalk in drugih slovarskih sredstev, ki so znana iz sveta tiskanih slovarjev, (3) njegovo zasnovo bodo v veliki meri določali izračuni velikosti fontov, dolžine vrstic in podobno. Razen o tem, da bo vseboval 70.000 iztočnic in o odločitvah, v katerih primerih bodo iztočnice uvrščene kot podgesla, o vsebini slovarja ne vemo nič. Če bo šlo zgolj za dopolnjeni SSKJ in SNB, morda celo bolj kriptičen in neberljiv, kar je mogoče sklepati po nekaj objavljenih geslih, to za prihodnji opis slovenščine predstavlja problem, saj tak slovarski koncept v letu 2014 ne more biti sprejemljiv.

Temeljni slovar je namreč na znanstvenih metodah zasnovani slovar, ki (1) izhaja iz primarnih virov, v primeru splošnega slovarja torej ustreznih besedil, (2) prinaša skladno s stopnjo razvoja različnih jezikoslovnih področij in usmeritev zanesljive podatke in interpretacije in (3) je obsegovno čim bolj popoln. Od naštetih parametrov predlog trojice izpolnjuje le prvega, pa še tega pogojno, saj besedilni korpus Gigafida, ki je njegov poglavitni vir, žanrsko ni uravnotežen. Obsega namreč le dva odstotka leposlovnih in le štiri odstotke (poljudno)znanstvenih besedil, v njem prevladujejo publicistične objave (77 odstotkov), vsebuje pa tudi vse preveč slogovno zares šibkih spletnih stvaritev, namreč 16 odstotkov, med katerimi se najdejo celo besedila ali deli besedil v angleščini, hrvaščini in srbščini.

14. Kot smo že omenili, pojem "temeljni slovar" v leksikografiji ne obstaja. Če pa se osredotočimo na tri pogoje, ki so izpostavljeni pri slovarju, ki bi ga Inštitut želel izdelati v prihodnosti, se je z njimi mogoče povsem strinjati. Prav zato je presenetljiva trditev, da predlog izpolnjuje le prvi pogoj, pa še tega pogojno, zato bomo v nadaljevanju podrobneje analizirali argumente. Kot izhodišče pri tem pa velja upoštevati predvsem to, da nerazumljivost informacij za navadnega uporabnika nikakor ne pomeni, da so informacije znanstveno zanesljive, kot tudi znanstvena zanesljivost ne pogojuje nerazumljivosti. Zanesljivost informacij in znanstvenost je bila pri slovarjih vedno pogojena z analizo velikih količin podatkov, bodisi ročno, kot nekoč, bodisi strojno, kot se dela danes, in njihovo konsistentno interpretacijo. Ključno pa je, da mora biti slovarska baza s pomočjo različnih vizualizacij namenjena vsem uporabnikom, ne le jezikoslovcem.

15. Očitek žanrske neuravnoteženosti in slogovne šibkosti (spletnih) besedil v korpusu Gigafida je presenetljiv, predvsem s stališča zastopnika največjega inštituta v Sloveniji, ki je zadolžen za opis slovenskega jezika. Zakaj? Najprej že zato, ker inštitut sam pripravlja slovarske opise brez žanrsko uravnoteženega, obsežnega, sodobnega in jasno dokumentiranega gradiva – takšnega namreč v zadnjih dveh desetletjih ni zgradil. A vrnimo se k Gigafidi. Pod točko (1) smo ugotovili, da mora ciljni slovar izhajati iz primarnih virov, to pa so zgolj "ustrezna" besedila. Očitno se v pojmu ustreznosti skriva tista past, ki korpus Gigafida iz primarnega slovarskega vira spremeni v neprimeren vir. Iz nadaljevanja argumentacije vidimo, da med ustrezne primarne vire spadajo le besedila, ki niso "slogovno šibka", pri čemer so problematična kar vsa spletna besedila, poleg tega je – verjetno iz

istega razloga – glede ustreznosti vprašljiva tudi večina besedil iz revij in časopisov. Na drugi strani so očitno "ustrezna", vrednejša in zaželena leposlovna in (poljudno)znanstvena besedila. In če to izhodišče izpeljemo do konca, se lahko vprašamo: bo Inštitut tisti, ki bo nase prevzel vlogo *arbitra elegantiarum* za celotno slovenščino? Kdo si bo upal reči: jezik v Dnevniku in Večeru je slabe kakovosti in ne sodi v korpus; roman *Trainspotting* ne sodi v stilno prečiščen nabor; zelo brani, a po vsebini trivialni romani niso vredni, da jih leksikografsko obdelujemo, komentatorji vsebin spletnih časopisov so nepismeni?

Iz tega izhaja predvsem eno vprašanje: ko bomo nekoč pogledali v slovar sodobnega slovenskega jezika – ali v njem želimo videti prikaz slovenščine, kot jo uporablja čim večja množica piscev, v knjigah, revijah, časopisih, na spletu, tviterju, v blogih, skratka kdorkoli, kjerkoli in kakorkoli? Ali po drugi strani želimo dobiti vnaprej samoomejeni idealizirani opis, ki primarno izhaja iz tega, kako slovenščino uporabljajo "dobri avtorji", pisatelji in znanstveniki, prefiltriran skozi jezikoslovčev vrednostno sito? O tem govori teza o neustreznosti korpusa Gigafida – če resno vzamemo nalogo, da želimo dobiti opis sodobnega slovenskega jezika, kakršen ta dejansko je, potem o ustreznosti Gigafide ne more biti nikakršnega dvoma. Če želimo v slovarsko bazo vključiti zgolj "lepo", "visoko" slovenščino, potem korpus Gigafida res ni primeren.

V sodobnem slovaropisju so prakse, pri katerih bi si z izbiro gradiva kar sami zamejili pogled na lastni jezik, izjemno redke, čeprav v vzhodni Evropi dejansko še obstajajo. Standardno sodobno leksikografsko izhodišče je, da je o jeziku najprej treba vedeti čim več, da bi potem z analizo izluščili, kaj je osrednje in obrobno, kaj standardno ali nestandardno, regionalno omejeno, stilno opredeljeno, dovolj stabilno za vključitev, dovolj marginalno za izključitev itd. Zato je treba imeti čim večji in čim bolj raznolik korpus, ki zajema besedila predvsem s stališča recepcije v jezikovni skupnosti – čim več govorcev in govork dejansko bere določena besedila (ne glede na njihovo "slogovno šibkost"), tem večji vpliv imajo ta na jezik in toliko bolj so zanimiva za leksikografsko obravnavo, ki v konsistentno zasnovanem procesu vsebino slovarske baze opremi z relevantnimi informacijami za različne tipe uporabnikov. Ta trenutek je za takšen tip analize pri nas daleč najbolj primeren korpus Gigafida skupaj z uravnoteženim korpusom Kres.

Če prvi del odgovora na očitek o slogovni šibkosti in žanrski uravnoteženosti govori o izbiri leksikografske "ideologije", ki v enem primeru pristopa k primarnim virom samoomejevalno s slogovnimi plašnicami, v drugem pa v njih vidi zrcalo celotnega realnega jezika, je drugi del odgovora bolj tehničen. Prispevek korpus predstavi kot amorfno maso, ki jo uporabimo ali zavržemo kot celoto. V korpusu Gigafida je skoraj 40.000 dokumentov različne velikosti, od celih knjig do posameznih časopisnih prispevkov, vsi pa vsebujejo bibliografske podatke o konkretnem viru (naslov, leto nastanka in vrsta besedila), večinoma pa tudi druge, bolj podrobne informacije. Poleg tega je s strojnimi mehanizmi mogoče prečesavati tudi vsebino besedil do ravni odstavka in posameznim delom pripisovati lastnosti, kot so tematika, o kateri govori besedilo, podobnost z drugimi dokumenti, jezik itd. Če se resno lotimo slovarskega projekta, je treba vse te informacije uporabiti pri leksikografski interpretaciji. Za poljuben jezikovni pojav lahko že s statistično analizo ugotovimo, ali ima razmeroma enakomerno distribucijo po vseh besedilnih žanrih in virih. To pomeni, da ji z leksikografskega stališča lahko pripišemo neko splošnost, ne da bi bilo treba ročno pregledati na tisoče konkordančnih vrstic. Prav tako lahko na podlagi avtomatske analize korpusa ugotovljamo, da se denimo "piškotek" pojavlja predvsem v računalniških in kulinarčnih okoljih, kar kaže na možno interpretacijo semantike. Na drugi strani nas statistična analiza tudi samodejno opozori, da je

določena beseda izrazito podpovprečno zastopana v literarnih virih, npr. beljakovina, brskalnik, bivalen, kar morda kaže na njihovo manj splošno naravo, bodisi zaradi terminološkosti, publicističnosti, spletnosti ali drugih razlogov, ki jih mora leksikograf znati prepoznati in leksikografsko opisati v skladu z dogovorjenim konceptom. Zamislimo si, kaj bi pomenilo, če bi imeli na voljo zgolj omejen nabor besedil, izbran po načelih "dobrega sloga". Vse tovrstne analize bi postale invalidne, saj o tem, česar ni, ni mogoče nič reči.

Referenčni korpusi bodo vedno samo vzorci jezika, v tem se Gigafida ne loči od nabora listkovnega gradiva, ki je bil podlaga za SSKJ, ali enakih korpusov drugod po svetu. V Gigafidi je 1,1 milijarde besed, ki prihajajo iz 534 leposlovnih knjig 55 različnih založnikov, 1.082 različnih priročnikov, vodnikov, učbenikov itd., ki so izšli pri 89 različnih založnikih, dalje 51 različnih časopisov, 127 naslovov revij itd. Pleteršnikov slovar in SSKJ sta v slovenskem prostoru uresničila ključno zahtevo po verodostojnosti slovarja, ki je prav v dokumentiranosti njegovega gradiva. Tega se je ekipa, ki je gradila Gigafido in iz nje nastali uravnoteženi 100-milijonski korpus Kres, zelo dobro zavedala, kot tudi še dveh dodatnih, z omenjenima slovarjema uresničenih zahtev: (a) gradivo za slovar naj bo čim bolj obsežno ter (b) izbor (takrat izpisovanih) besedil mora temeljiti na premišljenih merilih in mora biti razviden. Prav zato smo leta 2012 izdali knjigo, kjer je gradnja Gigafide in Kresa natančno opisana, vključno z merili za izbor besedil, natančnim prikazom vsebine ter samokritiko lastnega dela in izdelka, ki je nastal.

16. Noben dokument v Gigafidi ni v celoti v tujem jeziku, vsebujejo pa dele v tujih jezikih, ker so ti sestavni deli različnih vrst dokumentov. Taka je narava besedil nasploh, v Gigafidi pa so vsa besedila, kolikor je mogoče, avtentična, torej vanje nismo posegali s stališča dobrega sloga in kriterija, po katerem ne bi vključili besedil, ki vsebujejo karkoli tujejezičnega. Naloga korpusa je, da pokaže jezikovno realnost, in ena od lastnosti te realnosti je mešanje različnih jezikov. To torej ni površnost ali malomarnost, temveč zavestna odločitev o načinu sestavljanja korpusa sodobnega jezika.

Sklepajoč po Spletnem slovarju slovenskega jezika z 2500 iztočnicami, ki je nastal v okviru projekta Sporazumevanje v slovenskem jeziku, pri katerem je delo vodila soavtorica predloga trojice (sicer zaposlena na Inštitutu za slovenski jezik), se postavlja tudi vprašanje, ali se ugotavljanje pomenov posameznih besed lušči neposredno iz gradiva ali s škiljenjem na SSKJ. V Spletnem slovarju slovenskega jezika je namreč beseda maček pomensko opredeljena kot »domača žival«, »zrel in izkušen človek« in »slabo počutje« (pustimo ob strani dejstvo, da ni vsaka domača žival že maček, da zrela in izkušena ženska tudi ni maček in da slabo počutje zaradi prehlada prav tako ni maček), kar vse in še več, vendar mnogo natančneje, prinaša že SSKJ, manjka pa pomen »pleskarski valjček«, ki ga SSKJ še ni zaznal in smo ga prvič uslovarili v letos objavljenem Slovarju novejšega besedja slovenskega jezika. Pričakovali bi, da bi avtorji Spletnega slovarja slovenskega jezika novi pomen izluščili iz besedilnega korpusa Gigafida, kjer se pojavlja v več primerih, npr. v povedi Na vrhu lojtre pa Matevž z mačkom v eni in kanto z barvo v drugi roki.

17. Sugestija, da bi morale biti tako semantična distribucija kot tudi vse pomenske razlage v kateremkoli drugem slovarju ali slovarskem predlogu za vse obravnavane besede za vsako ceno drugačne, je z leksikografskega stališča nevzdržna, dokler opis ustreza realnosti sodobne slovenščine. Če torej nadpomenka iz SSKJ-jevske razlage lepo deluje kot t. i. pomenski indikator v slovarski bazi, je edino primerno, da se jo uporabi, ne pa da zaradi apriornega strahu pred plagiatorstvom na silo iščemo drugačno rešitev. Govorimo pa o sovpadanju zgolj delčka razlage iz SSKJ, prepisovanje

celotnih razlag z vidika predlagane metodologije oz. ciljnega uporabnika tako ali tako ne bi bilo sprejemljivo, če je bilo ugotovljeno, da so te za splošno publiko in za šolsko populacijo težko razumljive.

18. V tem delu so rešitve v slovarskem predlogu navedene selektivno. Pomenski opisi besed in besednih zvez so ključna informacija, ki jo uporabniki pričakujejo od slovarja (in ključna informacija za računalniško obdelavo jezika), hkrati pa najzahtevnejša leksikografska naloga. Tej problematiki smo zato namenili veliko pozornosti že pri snovanju leksikalne baze v okviru projekta Sporazumevanje v slovenskem jeziku. Pomembna spoznanja, ki so jih prinesle analize in različni pristopi k oblikovanju slovarskih definicij, so med drugim: (1) da je zapisovanje pomenskih razlag metaleksikografsko delo, ki nima veliko skupnega s tem, kako govorci v vsakdanji komunikaciji prepoznavamo pomene besed, da pa se je smiselno pri tem čim bolj približati diskurzu, ki simulira pojasnjevanje besed v naravni situaciji; (2) da ne obstaja ena sama možna in edina ustrezna razlaga posameznega pomena besede (slovenski uporabniki bi se imeli o tem možnost prepričati, če bi imeli poleg splošnega (oz. znanstvenega) na voljo vsaj še slovar, prilagojen šolski rabi); (3) da uporaba enega samega načina definiranja, npr. zgolj klasičnoslovarskega pomenskosestavinskega ali zgolj stavčnega z vidika pomenske obvestilnosti ni smiselna, pri čemer pri izbiri definicijskega načina ne odloča le besedna vrsta besede, ki jo razlagamo, pač pa predvsem narava njenega pomena, njene tipične upovedovalne možnosti, delež pragmatičnih pomenskih elementov itd.; in (4) kar je najpomembneje, da besedni pomeni sami na sebi, tj. v izolaciji, ne obstajajo, obstajajo zgolj njihove pomenske tendence, da pa se besedni pomeni vedno realizirajo v sobesedilu, ki je v veliki meri predvidljivo. Zadnje je tudi razlog, da v kontinuiranem diskurzu do zamenjav tipa vile – razkošne hiše in vile – kmečko orodje prihaja zelo redko.

Zato smo v slovarskem predlogu, predhodno pa že v slovarski bazi, predlagali pomenski opis, ki je sestavljen iz treh medsebojno povezanih elementov, od katerih ima vsak tudi svojo zelo specifično nalogo. Prvi je t. i. pomenski meni, interaktivni prikaz pomenske zgradbe večpomenske besede, ki daje uporabniku osnovni pregled nad tem, kako pomensko razvejana je beseda, v kakšnem medsebojnem razmerju so posamezni pomeni in kako si sledijo glede na svojo (relativno) pogostost oz. aktualnost v sodobnem jeziku. Drugi podatek so t. i. pomenski indikatorji, katerih naloga je najprej oblikovati pomenski meni, sicer pa so, zlasti kadar so navedeni v obliki sinonimov ali nadpomenk, namenjeni vzpostavljanju pomenskih povezav ali t. i. semantičnih mrež, ki omogočajo združevanje pomensko sorodnih besed bodisi znotraj slovarske baze bodisi v povezavi z drugimi za slovenščino obstoječimi semantičnimi bazami, kot je denimo slovenski Wordnet. Indikatorsko oznako »počutje« oz. »slabo počutje« ima v slovarski bazi denimo še beseda brezvoljnost, oznako »domača žival« pa denimo še mačka, muc in pes. Bistveni del pomenske informacije je v slovarskem predlogu namenjen t. i. pomenskim shemam, ki so primarno namenjene formaliziranju skladenjskega vzorca, v katerem se realizira določen pomen besede, sicer pa predstavljajo izhodišče za oblikovanje stavčnih (ali nestavčnih) definicij, ki prikazujejo pomensko rabo besede v njenem tipičnem besedilnem okolju na način, ki je govorcem blizu iz vsakdanje komunikacije. Avtor je v svojem prispevku izpostavil le pomenske indikatorje, ki sami na sebi torej ne predstavljajo zaključene pomenske informacije.

19. Beseda maček ima v korpusu Gigafida, ki ima 1,2 milijarde besed, 25.526 pojavitev, kar jo uvršča med relativno pogoste slovenske besede, pri čemer velik delež, približno 43 %, predstavljajo lastna imena. Zveza *pleskarski maček* v pomenu »pleskarski valjček« ima 2 pojavitvi v enem samem viru, poleg tega se maček v tem pomenu pojavi samostojno v nekaj zgledih, med drugim v tistem, ki je

naveden v prispevku. Maček kot "priprava na dolgi palici za beljenje" (SNB) glede na korpus Gigafida predstavlja torej povsem zanemarljivo pomensko rabo, ki je v pisni slovenščini izjemno redka. Po podatkih iz korpusa v pisni slovenščini za ta pojem uporabljamo besedno zvezo *pleskarski valjček* (14) ali *pleskarski valj* (11). V slovaropisni literaturi se za ustaljeni del besedišča nekega jezika štejejo besede in pomeni, ki izkazujejo relativno pogosto rabo skozi daljše obdobje in so v tem času izpričane v različnih besedilnih virih. Za dani pomen besede maček to gotovo ne velja. Naloga splošnega slovarja je beležiti to, kar je v jeziku tipično in osrednje. Predstavljanje enkratnih ali redkih pomenskih rab kot ustaljenega pomenskega fonda je z vidika legitimne slovarske informacije prav tako vprašljivo kot zanemarjanje očitnih pomenov ali besed. Kljub vsemu pa je vendarle to stvar slovarskega koncepta, ki mora vsebovati načelne odločitve o stopnji pomenske razčlenjenosti in pogostnostnem statusu vključenih pomenov.

Oglejmo si spornega "mačka" še malo podrobneje: v uvodu v Slovar novejšega besedja je izvor slovarskih podatkov pojasnjen takole:

Pri sestavi geslovnika smo se oprli na zbirko novejšega besedja, ki je nastala v okviru projekta Novejša slovenska leksika (v povezavi s spletnimi jezikovnimi viri). Projekt, ki je potekal v Leksikološki sekciji Inštituta za slovenski jezik v letih 2007-2009, je bil zasnovan kot predstavitev nove leksike in leksikalnega stanja v slovenščini za dopolnitev zbirk in geslovnikov novejšega besedja za različne tipe slovarjev standardne slovenščine. Najpomembnejše pa je, da je bil za pripravo zbirke novejšega besedja kot temeljni gradivni vir uporabljen besedilni korpus Nova beseda (javno dostopen na naslovu http://bos.zrc-sazu.si/s_beseda.html), ki je nastajal na Inštitutu za slovenski jezik od leta 1994. Ta korpusno zgrajeni geslovník je bil primerno izhodišče za slovarsko obravnavo, ki korpusno orientacijo v veliki meri razširja na slovarsko besedilo.

V okviru projekta Novejša slovenska leksika (v povezavi s spletnimi jezikovnimi viri) je bila objavljena znanstvena monografija z istim naslovom, kjer dobimo nadaljnja pojasnila glede gradivnih potrditev za posamezne iztočnice:

1. razdelek zbirke NSLSJV: gre za t. i. reprezentativni razdelek zbirke, v katerega se uvrščajo rezultati projekta NSLSJV, namenjeni in pripravljeni tudi za objavo; razdelek ima dve enoti: a) slovaropisni razpravni elaborat k leksikografskemu seznamu NSLSJV z naslovom Leksikografska pojasnila in ureditvena načela za leksikografski seznam NSLSJV, b) leksikografski seznam NSLSJV.

2. razdelek zbirke NSLSJV: gre za t. i. informativni delovni razdelek zbirke: v njem so zbrane novejše leksemske enote (zbiranje in evidentiranje se nadaljuje) obenem s pripisano navedbo gradivnih virov in (orientacijskim) podatkom o frekvenci (gre torej za evidenco mest besedilnih potrditev za leksemske enote, ki so nato - večinoma - urejene v leksikografski seznam NSLSJV in tako preidejo v enoto b) 1. razdelka zbirke NSLSJV).

3. razdelek zbirke NSLSJV: gre za t. i. pomožni, interni gradivni razdelek: vanj se zgolj delovno ilustrativno uvrščajo (posamezni) dobri, učinkoviti primeri besedilnih potrditev (iz spletnih jezikovnih virov) za (vsaj nekatere) posebno zanimive ali kako drugače specifične leksemske enote, uvrščene v leksikografski NSLSJV; razdelek ima dva podrazdelka: a) primeri besedilnih potrditev (iz spletnih jezikovnih virov) v elektronski obliki, b) natisi posebnih, specifičnih besedilnih potrditev (iz spletnih jezikovnih virov) za (nekatere) posamezne leksemske enote.

4. razdelek zbirke NSLSJV: gre za pomožni, interni pomenski orientacijski razdelek: vanj se uvrstijo morebitne ob zbiranju oz. identifikaciji novejšega besedja sproti oblikovane približne pomensko

orientacijske usmeritve oz. pridobljene stvarne pomenske informacije vsaj za nekatere, iz kakega razloga posebno zanimive leksemske enote.

V pričujoči publikaciji je v celoti objavljen prvi razdelek, glede ostalih treh pa je podana le osnovna informacija o njih z nekaj konkretnimi primeri za ilustrativni vpogled v sicer interne delovne materiale.

V čem je torej leksikografska zadrega s pleskarskim mačkom? Če resno vzamemo načelo, da mora biti za uvrstitev pomena v slovar na voljo zadostno število zgledov, ki izhajajo iz različnih, med sabo neodvisnih gradivnih virov, so z mačkom težave. Namreč:

- v korpusu Nova beseda, ki naj bi bil prvi in najpomembnejši vir za Slovar novejšega besedja, ni niti enega zgleda, ki bi potrjeval ta pomen besede maček
- v "reprezentativnem razdelku zbirke NSLSJV" je navedeno zgolj naslednje: ■ maček (priprava za beljenje)
- do "informativnega delovnega razdelka zbirke NSLSJV", v katerem "so zbrane novejše leksemske enote /.../ obenem s pripisano navedbo gradivnih virov in (orientacijskim) podatkom o frekvenci", nimamo dostopa in torej ne vemo, iz katerih konkretnih gradivnih virov izhajajo morebitni zgledi za ta pomen
- Gigafida, skoraj štirikrat večji in žanrsko bistveno bolj raznolik korpus kot Nova beseda, izkazuje izjemno redko, marginalno rabo mačka v tem pomenu
- na spletu (Google, Najdi.si) najdemo zelo redke zglede, izključno na spletnih klepetalnicah in forumih (npr. *Ali je najboljša varianta "na pištolo", ali bi se lahko malo "pomatral" z mačkom? | Jaz bi pa fotru za rojstni dan kupil en set penzlov pa mačka in ostale opreme za beljenje. | Se pravi lepo raztegovati barvo nanest potem pa barvo z mačkom lepo raztegnit, da ni neke debela plast pa tistih črt od mačka.*)

Iztočnica maček je v Slovarju novejšega besedja opremljena s kvalifikatorjem *teh. žarg.*, kar pomeni, da se uporablja v žargonu določenega področja tehnike, v tem primeru verjetno slikopleskarstva. Iz tega sledi logično vprašanje: če pomislimo, koliko žargonskih izrazov se uporablja po spletnih klepetalnicah in forumih (o katerih kritiki predloga večino menijo, da tako ali tako niso ustrezen primarni vir), zakaj je ravno maček dobil ta privilegij, da je bil "uslovarjen"? Ne pa še na tisoče podobnih, bistveno pogosteje rabljenih žargonizmov, ki so vsekakor legitimna in zanimiva plast jezika, vendarle pa ne prva pri beleženju standardnega jezika, še posebej, če je konkretni žargonski izraz razmeroma redek in po vsej verjetnosti omejen na govorjeni jezik.

To seveda pomeni, da obstaja dober razlog za to, da maček v pomenu pleskarskega valjčka v Leksikalno bazo za slovenščino ni bil uvrščen. Na drugi strani pa denimo v SNB za razliko od LBS ni zabeležena dejansko obstoječa in relativno pogosta zveza *mišični maček*, ki ima zaznavno pojavnost v Gigafidi (47) in celo v Novi besedi (4), nadomešča pa pogovorni "musklfiber".

V geslu maček v SNB lahko opazimo še dve stvari, ki v povezavi z zgoraj navedenimi trditvami nista pomembni, sta pa zanimivi za nadaljevanje prispevka. Maček je namreč definiran kot *priprava na dolgi palici za beljenje*, kar je absolutna trditev, ki se ji slovanske razlage navadno izogibajo. Maček oz. pleskarski valjček je namreč sestavljen iz valja in ročaja, ki sta nujna dela priprave, dolga palica kot podaljšek ročaja pa je nenujni dodatek. Verjetno bi bila torej ustrežnejša razlaga *priprava za beljenje, sestavljena iz valja in ročaja, navadno pritrjena na dolgo palico* ali kaj podobnega. Razlaga v SNB je rahlo problematična tudi zaradi uporabe dvoumne skladišne strukture, pri kateri ni povsem jasno,

kaj so jedra zvez in kaj prilastki. Razlago je zlahka mogoče brati, kot da gre za pripravo, ki je na *palici za beljenje*, ki je pač določena vrsta palice, ena od njenih lastnosti pa je dolgost. Drugi problem je zgled "beliti z mačkom", ki ga ni mogoče najti niti v enem viru, ne na spletu, nikjer. Če ga ne bomo našli v zbirki NSLSJV, in sicer v realnem besedilu iz konkretnega gradiva, to pomeni, da je bil izmišljen, napisan na podlagi introspekcije kot leksikalno-skladenjska možnost, ne pa kot navedek realne in neodvisne rabe. Ena od osnovnih maksim sodobnega slovaropisja pa je ta, da si zgledov rabe nikoli ne smemo kar izmisliti, ker so sistemsko možni.

V vzorčnih geslih, ki jih prikazuje predlog trojice, najdemo tudi vse **preveč faktičnih napak**, tako da slovar, izdelan po tem predlogu, ne bi bil zanesljiv – zanesljivost informacij pa je verjetno najpomembnejše pričakovanje uporabnika temeljnega slovarja. Naj omenim le eno izmed več takih napak. V vzorčnem geslu žena **je navedeno, da je ta beseda v slovenščini prvič zapisana v Dalmatinovi Bibliji leta 1584**, čeprav je bila natisnjena že v Trubarjevem Katekizmu iz leta 1550, prvič pa je bila zapisana že v Rateškem rokopisu iz druge polovice 14. stoletja. Ta napaka ni naključna, temveč jo narekuje temeljna usmeritev predloga trojice, ki je **upoštevanje le izbranih digitalnih virov**. Ker sestavljavec gesla žena v spletni zbirki Jezikovni viri starejše slovenščine, ki obsega le nekaj odstotkov naših starejših besedil, ni našel nobenega starejšega zapisa obravnavane besede, je **preprosto povzel, da je navedeni najstarejši**. Točne podatke o starejših zapisih te besede **bi sestavljavec lahko prepisal iz ustreznih knjig, vendar listanja po knjigah predlog trojice ne predvideva**, zato po tem predlogu zasnovani slovar **ne more vsebovati podatkov, ki so dostopni le v tiskani in ne tudi digitalni obliki**.

20. V slovarskem predlogu zagotovo nikjer ni trditve, da bi bila beseda žena prvič v slovenščini napisana v Dalmatinovi Bibliji leta 1584. Kar v predlogu v resnici piše, je naslednje:

»Pod zavihkom Izvor najde uporabnik dva tipa historičnih podatkov o iztočnici oz. njenem pomenu /.../, in sicer: (1) podatek o etimološkem izvoru besede in njenih sorodnih oblikah v različnih jezikih ter (2) podatek o starinskih oblikah ali zapisih besede v slovenskem jeziku glede na časovno umeščenost besedila, v katerem se je konkretna oblika pojavila. /.../ Podatke o starejših oblikah za besedo v iztočnici je mogoče dobiti avtomatsko na podlagi Jezikovnih virov starejše slovenščine, ki vsebujejo Slovar starejšega slovenskega knjižnega jezika in starejše oblike, oboje izdelano v okviru projekta IMP. Temu se nato v modri fazi izdelave gesla dodajo še opombe o izvoru besede in njeni etimologiji, ki jih iztočnici pripiše etimolog.«

Podatek o "starinskih oblikah ali zapisih besede" izhaja iz integracije zbirke starejših slovenskih besedil v spletni prikaz na slovarskem portalu v okviru enega od zavihkov. Tako kot v drugih primerih, kjer so v portal integrirane druge, od slovarske baze neodvisne podatkovne zbirke, se prikaz prilagaja spremembam vsebine same zbirke, namen tega dela portala pa seveda ni ta, da bi bralca informiral o prvih pojavitvah neke besede v slovenskem jeziku, temveč da bi mu omogočil dostop do konkretne vsebine oz. konkordanc starejših besedil, v katerih se pojavlja izbrana iztočnica, hkrati pa vizualiziral, kakšno obliko iztočnice lahko pričakuje v besedilu. Tega namena tako iz grafičnega prikaza kot iz opisa v slovarskem predlogu ni tako težko razumeti, zato ni jasno, od kod interpretacija, da zavihkek izvor prikazuje prve pojavitve besede v slovenščini. Prikazuje pač prvo in vse ostale pojavitve besede v zbirki besedil, ki jo vizualiziramo. Ali konkretnije: takoj ko se bosta Trubarjev Katekizem iz leta 1550 in Rateški rokopis iz druge polovice 14. stoletja pojavila v korpusu starejših besedil, bo informacija o pojavitvi in letnici izdaje na voljo tudi na časovnici, še vedno pa iz prikaza ne bo implicirano, da je ena ali druga prva pojavitve besede v slovenščini, temveč le to, da je glede na leto izida kronološko prva v dani zbirki.

21. Ta očitek slovarskemu predlogu je zanimiv na več ravneh. Prvič, treba je ločiti med podatkom, vnesenim v slovarsko bazo s strani leksikografa, etimologa ali kakega drugega strokovnjaka, npr. o letnici prve pojavitve neke besede v slovenskih besedilih (če bi se denimo za vnos te vrste informacije odločili sestavjalci slovarske baze) in prikazom oz. vizualizacijo zbirke, ki vsebuje večjo količino besedil, ki so jih napisali različni avtorji – korpusom starejših besedil. To smo že ugotovili. Teh besedil bo s procesom digitalizacije zgodovinskega tiskanega fonda predvidoma na voljo vedno več in od tod tudi motivacija za njihovo vključitev v zavihek Izvor. Ključna zanimivost pri tem očitku pa je, da je piscu sploh prišlo na misel, da bi pri sestavljanju slovarja sodobnega jezika želeli brskati po knjigah, ugotavljati, kdaj se je neka beseda prvič pojavila v slovenščini in ta podatek eksplicitno vnesti v slovarsko bazo, in to kar za vse slovarske iztočnice. Taka naloga je že na prvi pogled neizvedljiva, vsaj v okviru predlaganega projekta. Po drugi strani si to nalogo navadno zadajajo historični slovarji, ki sistematično opisujejo zgodovinski razvoj besed. Takega slovarja slovenščina še nima in ga verjetno tudi ne bo imela, razen kot kombinacije slovarjev slovenščine različnih obdobj. Prva pojavitve in historični razvoj besed kot leksikografska informacija ni del ne SSKJ, ne SNB ali kateregakoli drugega slovarja slovenskega jezika. Morebitna izjema bo slovar protestantskih piscev 16. stoletja, ki je že več deset let v procesu izdelave na Inštitutu in naj bi bil podprt tudi z namenskim korpusom, ki pa izven ISJ ni dostopen. Podatke iz tega slovarja bo vedno mogoče vključiti v slovarski portal, skupaj z informacijami o sodobnem jeziku, seveda če bo za to izkazan interes in bodo viri prosto dostopni.

Teoretično vprašljivo je tudi **enačenje enakopisnic in večpomenskih besed**, kot ga predvideva predlog. Tako bo npr. dve besedi z izrazno podobo bor (prva pomeni »drevo«, druga »kemijski element«) slovar prikazoval na enak način kot večpomenski samostalnik maček. Ker se avtorji predloga sklicujejo na intuitivnost pri uporabi slovarja, iz takega prikaza intuitivno sklepam, da je **bor ena beseda s prvim pomenom »drevo« in drugim »kemijski element«**, kar **seveda ni res**. Slovenščina ima dve besedi bor, od katerih prva pomeni »drevo«, druga pa »kemijski element«. Najbolj površinski dokaz tej trditvi je, da besedo bor s pomenom »kemijski element« nekateri še izgovarjajo z ozkim ó, torej drugače kot poimenovanje drevesa, ki ga izgovarjamo s širokim. Sledeč temu načelu, bo tudi geslo **bučen v slovarju samo eno in bo imelo dva pomena: »hrupen« in »nanašajoč se na bučo«**. Po jezikoslovnih spoznanjih, ki jih poučujejo tudi v naših osnovnih in srednjih šolah, gre tu za **dve besedi, ki se pri tonemskih govornih celo razlikujeta v naglasu**, prva ima nizkega, torej akutiranega, druga pa visokega, torej cirkumflektiranega.

22. Pri problemu *bor* gre za obravnavo homonimije nasproti večpomenskosti, ki ima v slovarjih različne rešitve. V skladu z obravnavo homonimije v SSKJ so tudi v SNB nekatere besede obravnavane v več samostojnih slovarskih geslih, pri katerih so iztočnice navedene zaporedoma z nadpisanim indeksom, torej: bor¹, bor², in ne kot ena beseda oz. iztočnica z več pomeni. Taka rešitev je v veliki meri prilagojena tiskanemu mediju, kjer je abecedna ureditev pomembna za uporabnikovo uspešnost pri iskanju slovarske informacije. Digitalni medij v marsičem spreminja način predstavitve jezikovnih podatkov, saj se uporabniku ni treba ukvarjati z iskanjem besed v abecedni razvrstitvi. Niti s tem, ali bo zvezo čarterski polet našel pod čarterski, pod polet ali kot samostojno geslo, ker svojo poizvedbo enostavno vtipka v iskalno okence.

Kar je z vidika obravnave homonimov kot več samostojnih besed sporno, je predvidevanje, da uporabnik pozna zgodovinski razvoj besede, še preden ve, kaj ta dejansko pomeni (saj je sicer ne bi iskal v slovarju). Uporabnik bi moral torej vnaprej vedeti, da npr. beseda *pero* v slovenščini lahko pomeni (razlage so prirejene po SSKJ): 1. kožna tvorba, ki pokriva telo ptic, 2. kovinska priprava za

pisanje na črnilo, 3. pisatelj, 4. vzmet in 5. rastlinski list, medtem ko v ločenih geslih denimo mula¹ pomeni "domačo žival", mula² pa "kazanje jeze", atlas¹ "zbirko zemljevidov", atlas² "svileno tkanino", atlas³ "prvo vratno vretence". Čeprav je z jezikoslovnega vidika tako rešitev mogoče utemeljiti na podlagi razvojne (ne)povezanosti pomenov, se je mogoče tudi vprašati, ali gre dejansko za informacijo, ki je za uporabnika relevantna na prvi ravni, torej do te mere, da so ji podrejene vse druge informacije v slovarski bazi, zlasti torej to, kaj beseda pomeni danes, kako jo učinkovito rabiti v kontekstu in kakšne so njene slogovne in gramatične posebnosti. Novi veliki poljski slovar Inštituta za poljski jezik, ki je zasnovan za digitalni medij, tako "pero" kot "atlas" obravnava na enak način, podobno kot je predvideno v Slovarju sodobnega slovenskega jezika.

WIELKI SŁOWNIK JEZYKA POLSKIEGO

Praca naukowa finansowana ze środków na naukę w latach 2007-2012 jako projekt rozwojowy.

CHRONOLOGIZACJA FRAZEOLIZMY POKAŻ WSZYSTKO

A B C D E F G H I J K
L M N N O O P R S S
T U V Y Z Z Z <<

pióro

1. ptaka
2. rzecz o takim kształcie
3. do pisania
4. do ekranu dotykowego
5. pisarstwo
6. sposób pisania
7. autorstwo
8. pisarz

peró -ésa s (ô ê)

1. koźna tvorba iz roźevinastega tulca s pahljačastimi izrastki, ki v velikem številu pokriva telo ptic:
2. majhna kovinska priprava s priostrenim koncem za pisanje, risanje: // taka priprava s peresnikom: // priprava za pisanje, ki se polni s črnilom:
3. pisatelj, knjiźevnik:
4. vzmet:
5. (rastlinski) list:

ksiąźk. pisarz

DEFINICJA
WARIANTY
KWALIFIKACJA TEMATYCZNA
POŁĄCZENIA
CYTATY
ODMIANA

WIELKI SŁOWNIK JEZYKA POLSKIEGO

Praca naukowa finansowana ze środków na naukę w latach 2007-2012 jako projekt rozwojowy.

CHRONOLOGIZACJA POKAŻ WSZYSTKO

A B C D E F G H I J K
L M N N O O P R S S
T U V Y Z Z Z <<

atlas

1. geograficzny
2. grzybów
3. kość
4. podpora
5. gimnastyczny
6. rakietą

átlas¹ -a m (ā) zbirka zemljevidov, slik iz določene stroke v obliki knjige:
átlas² -a m (ā) svilena tkanina, ki se blešči zlasti na pravi strani:
átlas³ -a m (ā) prvo vratno vretence

anał. pierwszy krąg u człowieka i zwierząt podtrzymujący czaszkę

DEFINICJA
KWALIFIKACJA TEMATYCZNA
RELACJE ZNACZENIOWE
CYTATY
ODMIANA
POCHODZENIE

Večina leksikografov se je še v času tiskanih slovarjev zavedala, da so meje med polisemijo in homonimijo zelo tanke. Oglejmo si dve od teh mnenj: "Homonimija se začne na točki, ko govorniki jezika ne morejo več dojeti različnih pomenov kot povezanih med sabo..." (Ladislav Zgusta, *Manual of*

Lexicography: 1971), "Čeprav se etimologija pogosto omenja kot kriterij, predstavlja negotovo vodilo, saj so etimološko različne besede včasih razvile podobne pomene, besede, ki jim je mogoče določiti isti izvor, pa so razvile različne pomene, tako da jih govorec sodobnega jezika ne dojema kot sorodne" (Sidney I. Landau, *The Arts and Crafts of Lexicography*: 2001). Če je etimološki kriterij nezadosten in nezanesljiv ter so praktični razlogi razvrščanja besed po abecedi v digitalnem mediju izginili, je vsekakor treba ponovno premisliti logiko določanja homonimije in polisemije, ki ima v ozadju še bolj filozofsko in usodno vprašanje: kaj pravzaprav sploh je "beseda", ki vodi do vnosa "iztočnice" v slovarsko bazo, v kateri so predstavljeni pomembnejši "leksemi" nekega jezika? Za potrebe izdelave slovarske baze bo vse detajle morda res treba natančneje določiti, zanesljivo izhodišče pa je dano z rešitvami iz Predloga.

Pri različnih zapisih/oblikah iztočnice na ravni leme smo se zato odločali za samostojno obravnavo (tj. dve samostojni iztočnici) le v primeru spremembe kategorije spola, npr. prst – ž. sp. : prst – m. sp., spremembe besedne vrste, npr. naglas – prislov : naglas – samostalnik, ali spremembe, tudi naglasne, v oblikoskladenski (tj. spregatveni ali sklanjatveni) paradigmi, npr. molíti – mólim : molíti - molím.

Če bi upoštevali tudi razlike v tonemskem naglasu, bi bilo to verjetno že bolj problematično, saj nujna empirična potrditev teoretskih predpostavk o tonemskem naglasu zahteva ogromen vložek, ob tem pa je ta podatek po vsej verjetnosti nerelevanten za veliko večino uporabnikov.

Skratka, če upoštevamo navedene argumente, je samostalnik bor, katerega obravnavo v predlogu problematizira avtor prispevka, v resnici poseben problem zaradi svoje domnevno različne naglasne podobe pri posameznem pomenu, kar osmišlja obravnavo dveh ločenih besed že na ravni iztočnice. Tako rešitev predvideva tudi slovarski predlog, ko predlaga obravnavo enakoizraznih besed kot samostojnih iztočnic, kadar prihaja do sprememb v oblikoslovni in naglasni paradigmi, vendar pa je – prav zanesljivosti slovarske informacije na ljubo – potrebno najprej ugotoviti (in to ne na podlagi zastarelih podatkov ali zapisa lastnega jezikovnega občutka, ampak na podlagi velike količine podatkov govorjenega jezika), ali je naglasna različnost besede bor za pomen drevo in za pomen kemijski element v sodobnem jeziku tudi dejansko izpričana.

23. O tem, kako bi bilo v slovarski bazi rešeno geslo "bučen", je avtor prispevka zgolj ugibal, ker tega gesla v Leksikalni bazi za slovenščino še ni. Kot smo omenili, so kriteriji za uvedbo homonimnih gesel v slovarski bazi različni, smiselno pa so vezani predvsem na oblikoslovno-glasoslovne razlike, ne pa na etimološki izvor. V tem smislu je eden od kriterijev tudi razlika v oblikoslovni paradigmi pri lastnostnih (bučen = glasen) in izsamostalniških vrstnih pridevnikih (bučen = iz buče). Tipično drugi namreč nimajo primerniških in presežniških oblik, kar je zadostni razlog za izkazovanje razlike v leksikonu besednih oblik in posledično v slovarski bazi. Kot je bilo omenjeno, razlika v tonemskem naglasu po vsej verjetnosti ne bi zadostovala za tovrstno razlikovanje.

Zatrjevanje, da gre v primeru prikazane rešitve za napako, in postavljanje SSKJ-jevske obravnave kot edine strokovno sprejemljive pomeni bodisi nepoznavanje drugačnih jezikoslovnih obravnav problema v sodobnih slovarjih bodisi namenoma zavajajoče razglašanje ene same, čeprav sicer možne jezikoslovne interpretacije, kot edine znanstveno utemeljene.

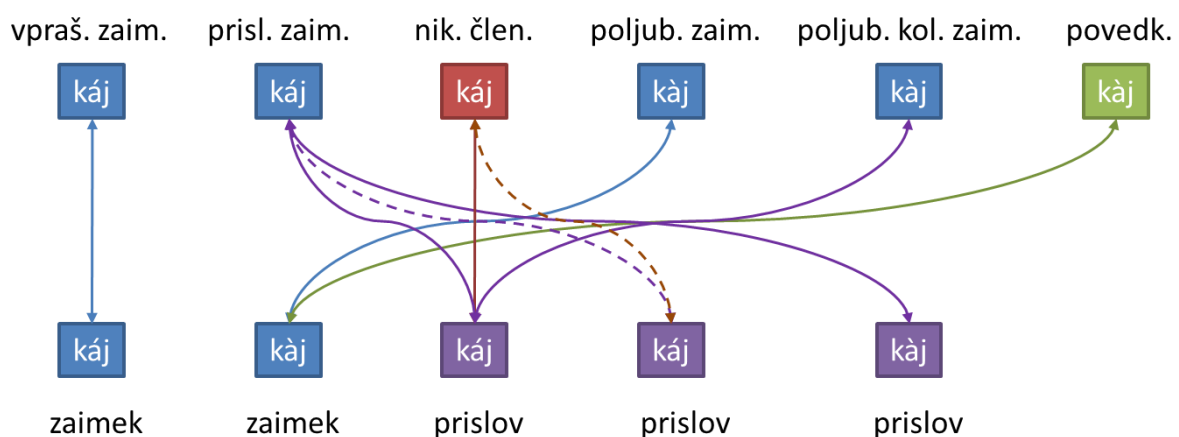
Morda se bodo komu zdele te podrobnosti za vsakdanjo rabo odveč in morda je res tako, vendar pa sta zanesljivost in globina opisa teh in še mnogo takih podrobnosti za temeljni slovar neizogibni. Ignoriranje teh in takih znanstvenih resnic si lahko privoščimo kvečjemu kak specialni, za določene

namene napisani slovar, temeljni pa nikakor ne. Temeljni slovar slovenskega jezika si prav tako ne more privoščiti popolnega ignoriranja tonemskega naglasa, ki je skoraj tako močna posebnost slovenščine kot naša dvojina, samo manj govorimo o njem. Tretja značilnost temeljnega slovarja je njegova težnja k popolnosti. Druga izdaja SSKJ bo imela približno 100.000 geselskih sestavkov, torej toliko, kot jih predvideva predlog trojice. Vendar predlog trojice v to število prišteva tudi nezanemarljivo število večbesednih iztočnic tipa državni zbor, ki so v SSKJ obravnavane na mikrostrukturni ravni pod enobesednimi iztočnicami, poleg teh pa še nekaj lastnoimenskih, ki zaradi pomenske izpraznjenosti ne spadajo v splošni razlagalni slovar. Če v drugi izdaji SSKJ preštejemo tudi podiztočnice in razlage, ki pomensko opredeljujejo večbesedne leksikalne enote, dobimo številko 175.000, kar krepko presega zastavljeni cilj trojice, v katerem je celo zapisano, da bo načrtovani slovar obravnaval samo polnopomenske besede (samostalnike, glagole, pridevnike in prislove – ob tem avtorji pozabljajo, da so nosilci pomena in že zato potrebni slovarske obravnave tudi medmeti, členki in povedkovniki ter da so lastna imena pomensko prazna), zavestno pa izpuščajo števnike, zaimke, veznike in predloge, ki so najpogostejše rabljene besede slovenskega jezika.

24. Iz tega dela besedila lahko do neke mere sklepamo, kaj ima avtor prispevka v mislih, ko govori o "aplikativnih" oz. "specialnih" slovarjih v primerjavi s "temeljnimi". Kot kaže, "temeljni" slovarji nimajo specifičnega namena oz. naslovnika, obstajajo zgolj zato, da sporočajo znanstvene resnice o jeziku – za koga in komu, pa ni jasno. Če je tako, je tudi jasno, zakaj je ta distinkcija v leksikografiji nepoznana – po tej premisi so namreč vsi obstoječi slovarji "aplikativni", saj so tako ali drugače namenjeni specifičnim naslovnikom ali prilagojeni specifičnim namenom.

Status znanstvenih resnic v slovarjih Inštituta za slovenski jezik sicer zasluži nadaljnjo eksplikacijo. Po navedbah avtorja prispevka na Lingvističnem krožku 28. oktobra 2013, se v prenovo obravnave funkcijskih besednih vrst v drugi izdaji SSKJ Inštitut ne bo spuščal, ker "za to ni časa". Zgolj droben primer težav, ki si jih v prihodnje lahko ometamo uporabniki slovarja, ko bomo soočeni z vprašanjem, kako določiti besedno vrsto besedice "kaj", je opisan v nadaljevanju.

Druga knjiga SSKJ, v kateri je pet geselskih člankov z iztočnico *kaj*, je izšla leta 1975. Leta 2001 je izšel Slovenski pravopis, v katerem je šest geselskih člankov z iztočnico *kaj*. V SSKJ je teh pet iztočnic opredeljenih bodisi kot zaimek in kot prislov, v SP pa so štirje zaimki, členek in povedkovnik. Vsebinsko je preslikava med njimi izjemno kompleksna, saj je zgolj ena iztočnica od petih oziroma šestih neposredno in enoznačno primerljiva (vprašalni zaimek). Preslikavo grafično prikazujemo na spodnji sliki:



Predstavljajmo si torej zmedo, ko bo leta 2014 izšla druga izdaja SSKJ, ki bo ponovno uvedla skoraj štirideset let star besednovrstni sistem iz SSKJ. Poglejmo si komentar Jožeta Toporišiča iz leta 1974 (*Besednovrstna vprašanja slovenskega knjižnega jezika*, Jezik in slovstvo, letnik 20, 2/3):

Preden se poslovimo od pridevniške besede, si oglejmo še zaimenske vrste (v SSKJ so kvalificirane samo z zaim., kar nikakor ne zadošča, saj jih še SP 1962 večinoma natančno določa, prim. čigar, svoj, tvoj, kateri, neki, nekaj ipd. proti moj, nekateri ipd.). Po našem bi bilo treba te kvalifikacije v slovarju (in naši zavesti) napraviti še natančnejše, ne pa da so praktično odpravljene. Mislim, da se je to zgodilo zopet tudi zaradi nezadostne teoretične ozaveščenosti (ali iz nestrokovnega ozira).

Spričo teh in podobnih drugih vprašanj, ki sem si nanje moral kot predavatelj slovenskega jezika neprestano odgovarjati, sem prišel do naslednjega pojmovanja besednih vrst: V slovenskem jeziku obstajajo 4 velike besedne (makro-)vrste: samostalniška in pridevniška beseda (obe sklonljivi), glagol (spregljiv) in prislov (nespregljiv, nesklonljiv, pač pa deloma pregiben); poleg tega je še 5 besednih vrst: členek, predlog, veznik, medmet in predikativ (zadnje je nepregibna besedna vrsta, eventualno z obliko za srednji spol ednine, npr. tiho v zvezah kot tiho biti/sem).

V Slovenskem pravopisu iz 2001 je torej Jože Toporišič uvedel omenjeni besednovrstni sistem, ki pa ni kompatibilen ne s SSKJ niti ne z obravnavo v učbenikih za slovenščino, kjer večinoma izpuščajo povedkovnik oz. predikativ. Zgodovinski razvoj in obravnavo besednih vrst v trenutno veljavnih učbenikih in učnih načrtih navajamo spodaj. Druga izdaja SSKJ z letnico 2014 bo verjetno ustvarila še večjo metodološko zmedo s potrditvijo sistema iz SSKJ.

SLOVNICE

	Kopitar (1808/9) Janežič (1854) Breznik (1912) Bajec, Kolarič, Rupel (1956)	Toporišič (2000)
1	samostalnik	samostalniška beseda
2	pridevnik	pridevniška beseda
3	zaimek	povedkovnik
4	števnik	členek
5	glagol	glagol
6	prislov	prislov
7	predlog	predlog
8	veznik	veznik
9	medmet	medmet

UČBENIKI ZA SLOVENŠČINO

• Učbeniki:

- Na pragu besedila, Rokus
- Z besedo do besede, MK
- Govorica jezika, Modrijan
- Besede, DZS
- ...

• Učni načrti:

- Program osnovnošolskega izobraževanja: SLOVENŠČINA (1998)

	besedna vrsta
1	samostalniška beseda
2	pridevniška beseda
3	
4	členek
5	glagol
6	prislov
7	predlog
8	veznik
9	medmet

25. Podatke o tonemskosti je vedno mogoče dodati iztočnicam in jih pripeti ne le na izrazno podobo besede kot take, pač pa tudi na ustrezni posamezni pomen, podobno kot je to pri Velikem poljskem slovarju. V zvezi s tonemskostjo si je ob tem, da trenutno ni ustreznega gradiva, na katerem bi bilo mogoče preučiti in določiti tonemskost sodobne slovenščine, potrebno zastaviti vsaj še eno vprašanje: Ali so bile opravljene kakršnekoli analize slovarskih uporabniških potreb glede tega, koliko je za splošnega uporabnika ta informacija dejansko zanimiva/potrebna? Tako kot pri večini projektov, kjer je obseg naloge izjemno velik, se je treba odločiti za prioritete. Podatke o tonemskosti je v slovarsko bazo (oz. leksikon) vedno mogoče dodati tudi kasneje, če je ustrezno organizirana.

26. Pri številki 175.000 večbesednih zvez je treba upoštevati dejstvo, da je SSKJ kot tak že na voljo. Ko torej štejemo iztočnice, podiztočnice in razlage stalnih besednih zvez iz tega slovarja, s tem operiramo z že obstoječim gradivom, ki je za opis slovenskega jezika relevantno z zgodovinskega stališča. Na drugi strani Inštitut po redkih znanih podatkih v Novem slovarju slovenskega jezika ponuja v osmih letih 70.000 gesel. Izračun iz predloga kaže, da je mogoče v petih letih narediti

100.000 gesel na novo, torej z novo pomensko interpretacijo sodobnega jezika in novim, uporabniško bolj prijaznim slovarskim opisom. Glede na to, da izračun potrošenih sredstev za SNB kaže, da bi bil za ekvivalenten slovar s 100.000 gesli potreben približno sedemkratnik vsote za predlagani slovar, je seveda mogoče ustrezno povečati tudi obseg predlaganega Slovarja sodobnega slovenskega jezika. V osmih letih bi denimo po naši metodologiji mogoče narediti tudi 175.000 gesel, pri čemer bi bila potrebna vsota še vedno nekajkrat nižja od tiste za "klasični" slovar. Poleg tega se je treba zavedati, da so v predlagani slovarski bazi večbesedne iztočnice tipa "državni zbor" opremljene z bistveno bogatejšimi podatki kot zgolj z "razlagami, ki pomensko opredeljujejo večbesedne leksikalne enote", zato je z njimi nekaj več dela, so pa bolj informativne. Za to, da dobijo status samostojne iztočnice, morajo večbesedne iztočnice izkazovati lastno besedilno okolje, samostojen, od sestavnih delov (relativno) neodvisen pomen itd. Neposredna številčna primerjava zato ni ustrezna.

27. Glede izločitve funkcijskih in slovničnih besed je v predlogu zapisano sledeče: »Funkcijske in slovnične besede, kot so vezniki, predlogi itd., bodo obravnavane drugače od ostalih gesel, poudarek bo na njihovi vlogi v besedilu oz. slovničnem opisu,« kar ne predvideva njihove vnaprejšnje izločitve, kot želi prikazati avtor prispevka. Hkrati pa je – skladno z razvojem jezikoslovne teorije – treba opozoriti na to, da besede, kot so predlogi, vezniki in zaimki v jeziku ne predstavljajo pomenonosnega leksikalnega fonda. Če pogledamo zadevo z vidika korpusa, lahko ugotovimo, da zavzemajo te besede npr. v enomilijonskem korpusu daleč najopaznejši delež, da se ta delež ohranja tudi v petmilijonskem korpusu, vendar pa to besedje za razliko od ostalega besedišča ostaja izrazno vedno enako. Jezik torej ne producira novih predlogov, veznikov in zaimkov itd. Te besede so seveda v jeziku najpogostejše, vendar predstavljajo njegov statičen slovnični, in ne leksikalni potencial. Iz tega sledi, da tudi njihova obravnava v slovarju potrebuje drugačen pristop, pri čemer je izpostavljena predvsem slovnična, skladišna vloga. Glede na to, da ostaja vloga teh besednih vrst v jeziku bolj ali manj nespremenjena, nenazadnje zadošča tudi obstoječi opis njihove slovnične funkcije v SSKJ in slovenskih slovnica.

Bomo torej naš temeljni enojezični slovar in vse prednosti, ki jih pri pisanju novega prinašajo **več kot polstoletne institucionalne izkušnje in pridobljeno znanje**, zamenjali z manj zanesljivim in manj obsežnim, posledično torej manj uporabnim slovarjem, ki ga bo treba tudi **po zaključku redakcije nenehno dopolnjevati, nadgrajevati in vse to plačevati iz javnega denarja**? Zdi se, da je ob ustreznem piaru danes mogoče prav vse. Če se bo tehtnica nagnila v to stran, bomo v prihodnjem obdobju obsojeni na slovar z nezanesljivimi in nepopolnimi podatki, ko pa bo šlo zares, npr. pri prevajanju, v šolah, pri branju naših literarnih klasikov in besedil 20. stoletja, v znanstvenem jeziku, na sodiščih, bomo natančnejše pomenske razlage in vse drugo še vedno iskali v časovno vedno bolj odmaknjenem Slovarju slovenskega knjižnega jezika, ki ga ne bo nihče več nadgrajeval in izpopolnjeval, saj država ne bo financirala dveh slovarjev, ki naj bi služila istemu namenu.

28. Ob polstoletnih izkušnjah Inštituta obstajata dve šibki točki. Prvič, te izkušnje se nanašajo predvsem na izdelavo slovarja SSKJ s pomočjo "6-milijonske listkovne kartoteke in ustreznih kartotečnih postopkov", nima pa inštitut izkušenj na področju obvladovanja obsežnih digitalnih korpusnih gradivnih virov in (avtomatiziranega) prenašanja leksikografskih (torej semantičnih, skladišnih, slogovnih in drugih) informacij iz teh virov v slovarske baze podatkov, ki morajo biti organizirane popolnoma drugače kot slovarji, namenjeni za tisk. Druga šibka točka je ta, da vsebinska preslikava sistema SSKJ v novi slovar, tako na ravni sistema razlag, uporabe oznak in drugih slovarskih sredstev kot nasploh načina predstavljanja informacij o jeziku, za prihodnji slovar slovenskega jezika

ni sprejemljiva. To pomeni, da je potreben nov, drugačen razmislek o slovarjih, ki na Inštitutu še ni bil opravljen, razen v okviru projekta Sporazumevanje v slovenskem jeziku.

Kljub polstoletnim izkušnjam Inštituta, ki jih avtor poudarja, je dejstvo, da je SNB večinoma izdelek mladih raziskovalcev, torej raziskovalcev z razmeroma omejenimi izkušnjami (in financiranih za študij ter pisanje doktorata), poleg tega so bili usmerjeni v sistem sestavljanja slovarjev, ki je že presežen.

29. Trditev o »večno nedokončanem slovarju« kaže na nerazumevanje potreb sodobnih uporabnikov slovarjev in nasploh tega, kaj je sodobna leksikografija. Ker se jezik spreminja ves čas, noben slovarski projekt, katerega namen je opis sodobnega jezika, ne more biti končan enkrat za vselej, temveč mora ta razvoj spremljati in sproti opisovati. Razmišljanje o »končanju slovarja« je možno zgolj v primeru, če njegov sestavljalac razmišlja le s stališča tiskanega slovarja, ki pred tiskom ne bo na voljo, po tisku pa se ne bo spreminjal. Sodobni digitalni (spletni) slovarji so zasnovani tako, da se lahko vsebinsko spreminjajo sproti glede na zaznane spremembe – bodisi zato, ker so na voljo nova leksikografska dejstva, bodisi zato, ker se je jezik spremenil. Sodobni mediji omogočajo, da na informacije o spremembah ni treba čakati do naslednje tiskane izdaje, ampak so takoj na voljo vsem.

Metodološka nezadostnost za izdelavo temeljnega slovarja se kaže še v eni usmeritvi predloga trojice, tj. v **izključni uporabi korpusnega pristopa**. Korpusni pristop je seveda koristen in uporaben, vendar (še) ni dorasel vsem slovaropisnim opravilom. Predlog omeji opis besedja na to, **kar omogočajo izbrani digitalni viri in današnja tehnologija**. S tem **izloči mnogo podatkov in jezikoslovnih spoznanj zadnjih dvesto let in prekine s slovensko (in srednjeevropsko) slovaropisno prakso**, to pa je ravnanje, ki **ni v skladu z razvojem znanosti**. **Znanost se namreč razvija z dograjevanjem že spoznanega**. Izločanje je ena od značilnosti poslovnega in predvsem komercialnega sveta. To je samo nekaj najbolj površinskih pomanjkljivosti predloga trojice, ki jih razberemo iz prikazanih fragmentov osmih neproblematičnih geselskih sestavkov in spremnega besedila. Resen slovarski koncept bi moral biti preizkušen na nekaj sto dokončno izdelanih slovarskih sestavkih z najrazličnejšo problematiko. Podpirati tako **pomanjkljivo zastavljeni projekt** lahko pomeni samo eno, namreč **vero vanj**. **Znanstvenoraziskovalna misel pa ne verjame, je kritična in dvomi**.

29. Glej odgovor 2.

30. Glej odgovor 24.

Morda je tu na mestu še misel Thomasa Kuhna, ki se je v znanem delu *Struktura znanstvenih revolucij* ukvarjal s fenomenom znanstvenih revolucij: "Ne dvomim, na primer, da je kot orodje za razreševanje ugank Newtonova mehanika boljša od Aristotelove in Einsteinova boljša od Newtonove. Vendar v njihovem zaporedju ne vidim nobene koherentne smeri ontološkega razvoja. Nasprotno, v nekaterih pomembnih vidikih je Einsteinova splošna teorija relativnosti bližja Aristotelovi, kot je katerakoli od obeh Newtonovi." Znanost se ne razvija linearno, temveč kot niz "znanstvenih revolucij", kjer sta dva kompetitivna modela lahko tudi v antagonističnem odnosu in tipično soobstajata, dokler ni jasno, kateri model bolje pojasnjuje izbrane pojave.

Razumljivo je, da državo v vlogi financerja zanima predvsem višina stroškov. Dobri štirje milijoni so za izdelavo temeljnega slovarja malo, zato je predlog trojice lahko mamljiv. Kot je bilo prikazano v prejšnjih odstavkih, bi bil po tem predlogu napisani slovar sicer uporaben za določene namene, vendar **ne bi zadoščal merilom za uvrstitev med temeljne niti po obsegu, še manj pa po kakovosti in**

zanesljivosti vsebine. Slovenci rečemo, da se za malo denarja dobi malo muzike. Poceni naročene strehe na javnih zgradbah v Vipavski dolini odnaša burja, v Ljubljani in še kje začnejo zamakati in uničevati milijonsko računalniško opremo, po vseh predpisih naročeno vodovodno napeljavo na Onkološkem inštitutu bodo morali v celoti zamenjati itd., itd. Ko govorimo o temeljnem splošnem slovarju slovenskega jezika, pa ne govorimo o strehi, ne govorimo o računalnikih ali nepotrebni neoporečni vodi, govorimo o opisu jezika, ki določa nas in našo kulturno raven ter zaradi katerega smo, kar smo. Govorimo o temeljni vrednoti slovenskega naroda in o značilnosti, ki skupaj z ozemljem opredeljuje našo državo. **Neustrezen premislek pri financiranju temeljnega slovarja nas zato lahko stane mnogo več.** Pa s tem ne mislim »samo« na izgubo jezikovne in jezikoslovne tradicije, ki je noben kultiviran narod ne zanemarja, temveč tudi na to, da bo treba prej ali slej vendarle napisati sodoben temeljni slovar z zanesljivimi podatki. Slovar, ki ga brez 4,2-milijonskega obvoda lahko začnemo izdelovati že zdaj.

31. S to trditvijo se je mogoče povsem strinjati. Dejansko se je treba vprašati, kaj potrebujemo za uspešen razvoj slovenščine v prihodnosti in kaj nas bo na dolgi rok stalo več. Sprememba, zaradi katere sploh moramo razpravljati o tem, kako se nacionalnega slovarskega projekta lotiti na pravi način, je sledila s spoznanjem pred nekaj leti, da bomo slovarje kmalu prenehali tiskati, razvoj na tem področju pa bodo določale nove možnosti računalniške obdelave informacij in elektronskih medijev. Kdaj je v slovarskem svetu spoznanje o radikalnem prehodu v digitalni medij dokončno prevladalo, je težko reči, lahko pa rečemo, da se evropski leksikografski svet tega zaveda v celoti in intenzivno išče pravi odgovor na to vprašanje. Ta trenutek univerzalnega odgovora še ni, že vsaj od omenjenega leta pa pri drugih jezikih potekajo slovarski projekti, ki so zapustili knjižno formo in uporabnikom skušajo ponuditi več informacij o jeziku na povsem drugačen način. Poleg same količine je bistveno, da so te informacije lahko prilagojene raznovrstnim potrebam različnih uporabnikov. S svojo brezkompromisno stavo na pretečeni koncept SSKJ-SNB je ISJFR v zadnjih letih zamudil dobro priložnost za vključitev v nove leksikografske tokove z drugačnimi potrebnimi znanji.

Ljudska modrost pravi: Ko ne veš, za kaj gre, gre za denar. Po pisanju Igorja Bratoža so »predlog podprli leksikografsko kompetentni strokovnjaki iz različnih institucij, od Filozofske fakultete do Instituta Jožef Stefan« (Delo, 9. 8. 2013, str. 4). Podporo predlogu je sopodpisalo 15 posameznikov in trije med njimi so res v službi na Filozofski fakulteti, dva pa na Institutu Jožef Stefan (kar seveda še ne pomeni, da ti dve instituciji predlog tudi podpirata), večina preostalih je v službi na Fakulteti za družbene vede ali v zasebnih podjetjih, vendar je med podpisniki le pet takih, ki imajo nekaj manjših slovaropisnih izkušenj, nihče pa doslej še ni sodeloval pri pisanju kakega temeljnega slovarskega dela. Po podatkih iz Cobissa je večina podpornikov predloga pisala članke o slovaropisni dejavnosti, trije so se podpisali med več kot petdeset soavtorjev Velikega angleško-slovenskega slovarja Oxford–DZS, ena podpisnica je vodila delo pri Turističnem terminološkem slovarju, glede katerega je Agencija za raziskovalno dejavnost izjemoma ustanovila komisijo za oceno zaključnega poročila projekta (komisija rektorja Univerze na Primorskem je namreč ugotovila vrsto vsebinskih pomanjkljivosti tega slovarja), še ena sopodpisnica pa je kot uslužbenka pri DZS sodelovala pri urejanju več dvojezičnih slovarjev. Zanimivo ob tem pa je, da domala vseh 15 podpisanih spada v interesni krog zasebnega zavoda Trojina, financiranega izključno z javnim denarjem, s katerim so povezani kot poslovni partnerji, zdajšnji ali nekdanji zaposleni, zdajšnji ali nekdanji honorarni sodelavci ali v smislu solastništva. Bolj kot o leksikografskih kompetencah sopodpisanih bi bilo torej smiselno govoriti o njihovih poslovnih interesih v obsegu (za začetek) 4,2 milijona evrov davkoplačevalskega denarja.

32. Brez komentarja.

Pa vendar predlog trojice vsebuje nekaj dobrih zamisli, predvsem v uporabi nekaterih jezikovnotehnoloških orodij, slovaropisnih postopkov, prikazu kolokacij in sistemu zavihkov, ki bi jih veljalo vključiti v izdelavo novega temeljnega razlagalnega slovarja. Inštitut za slovenski jezik Frana Ramovša ZRC SAZU je v pogovorih že izrazil pripravljenost na sodelovanje s trojico, pri katerem bi vsaka od sodelujočih strani prevzela tista opravila, ki jih dokazljivo obvlada in za katera ima ustrezne reference. S takim sodelovanjem bi Slovenci lahko kmalu dobili nov, zanesljiv in vsestransko uporaben temeljni slovar sodobnega jezika brez izključevanja in brez nepotrebnih razprtij, ki jih Inštitut za slovenski jezik ne želi poglobljati. Predlog je še vedno na mizi.

33. Dejansko je osnovno vprašanje kompetentnost izdelovalcev bodočega slovarja. Od dvanajstih avtorjev, ki so navedeni v kolofonu Slovarja novejšega besedja, je natanko polovica mladih raziskovalcev, ki razen izdelave 7.400 iztočnic po modelu SSKJ s slovaropisjem nimajo večjih izkušenj. Tri sodelavke, ki so sodelovale še pri izdelavi SSKJ, nimajo ustreznih izkušenj pri izdelavi drugače zasnovanih digitalnih slovarjev. Ena od avtoric ni več na Inštitutu, eden je predstojnik Inštituta, torej ne bo imel časa za izdelavo slovarja. Kot kompetentna leksikografka z dolgoletnim stažem in uredniškimi izkušnjami tako ostane ena oseba.

Ljubljana, 30. 12. 2013