

UDK 81'374

IZAZOVI LEKSIKOGRAFSKE OBRADE U JEDNOJEZIČNOME MREŽNOM RJEČNIKU (na primjeru *Hrvatskoga mrežnog rječnika – Mrežnika*)¹

Lana Hudeček

*Institut za hrvatski jezik i jezikoslovlje,
Republike Austrije 16, 10 000 Zagreb, Hrvatska,
lhudecek@ihjj.hr*

Na primjeru *Hrvatskoga mrežnog rječnika – Mrežnika*, koji se kao istraživački projekt koji financira Hrvatska zaklada za znanost provodi u Institutu za hrvatski jezik i jezikoslovlje od ožujka 2017. godine, u radu se pokazuju neki od izazova pred kojima se nalaze obrađivači i urednici takva rječnika, odnosno pokazuje se u kojim elementima rad na izradi mrežnoga rječnika uvodi pogled na leksikografsku obradu drukčiji od onoga u tiskanim rječnicima. Velik broj mogućnosti koje takav rječnik otvara (npr. rad na korpusima i sa suvremenima alatima za pretragu korpusa, povezivanje s izvanrječničkim izvorima) uvodi i potrebu za njihovom pažljivom analizom te ograničavanjem na one mogućnosti koje su korisniku zaista potrebne i njemu korisne. Predstavit će se i izazovi s kojima su se urednici i obrađivači susreli tijekom pokusne obrade u prvome projektnom razdoblju te rješenja koja su donijeli oslanjajući se, između ostaloga, i na suvremenu e-leksikografsku teoriju, rješenja primijenjena u drugim velikim nacionalnim e-rječnicima tena svoje leksikografsko iskustvo u izradi tiskanih rječnika.

Ključne riječi: e-rječnici, e-leksikografija, Hrvatski mrežni rječnik, Mrežnik, jednojezični rječnici.

DOI: <http://dx.doi.org/10.30970/vpl.2018.69.9293>

Uvod. 1. ožujka 2017. u ustanovi nositelju projekta Institutu za hrvatski jezik i jezikoslovlje započeo je s provedbom projekt *Hrvatski mrežni rječnik – Mrežnik*. U provedbu projekta, čiji je cilj izrada jednojezičnoga hrvatskoga mrežnog rječnika izrađena u skladu sa suvremenim e-leksikografskim teorijama, spoznajama, istraživanjima i postignućima, uključen je tim najvećim dijelom sastavljen od iskusnih leksikografa koji su već surađivali na *Školskome rječniku hrvatskoga jezika* objavljen u tiskanome izdanju 2012. godine. Osnovni razlog za prijavu projekta *Mrežnik* bio taj što za hrvatski jezik ne postoji suvremeni jednojezični mrežni rječnik izrađen u skladu sa suvremenim spoznajama računalnoga jezikoslovlja² [5: 329] te ocjena da je u hrvatskome jezikoslovlju nužno potaknuti sustavno upoznavanje s teorijom i praksom

¹ Rad je izrađen na istraživačkome projektu *Hrvatski mrežni rječnik – Mrežnik* (voditeljica: Lana Hudeček), koji u cijelosti financira Hrvatska zaklada za znanost.

² Za hrvatski jezik postoji nekoliko općih jednojezičnih e-rječnika: na internetu su javno dostupni *Hrvatski jezični portal* (<http://hjp.novi-liber.hr/>) i *Wječnik* (<https://hr.wiktionary.org>). Prvi je, međutim, samo formalno mrežni rječnik jer je zapravo riječ o tiskanome rječniku (Enciklopedijski rječnik hrvatskoga jezika) postavljenom na mrežu; drugi je Wikipedijin rječnik koji uređuju njegovi korisnici. Veliki rječnik hrvatskoga jezika ima pristup mrežnoj inačici s pomoću zatvorenoga koda, ali i tu je riječ samo o promjeni medija. DVD-om je opremljen prvi školski rječnik hrvatskoga jezika, inovativan po tome što korisnici mogu čuti i izgovor riječi te (rječnik je namijenjen učenicima nižih razreda osnovne škole) učiti hrvatski jezik igrajući se jer se na DVD-u nalaze i poučne igre prilagođene dobnoj skupini korisnika rječnika.

izrade e-rječnika i mogućnostima koje otvaraju nove rječničke platforme, potaknuti povezivanje s ustanovama u kojima se e-rječnici izrađuju te provođenje novih istraživanja u području e-leksikografije.

Mrežnik je zamišljen kao rječnik koji ima tri međusobno neovisna, a ipak dijelom i povezana modula: modul za opće korisnike (10 000 natuknica), modul za učenike nižih razreda osnovne škole (3000 natuknica) te modul za osobe koje uče hrvatski kao drugi i strani jezik (1000 natuknica). Njihova se neovisnost očituje u činjenici da su za svaki modul izrađena posebna pravila obrade te u programu *TLex* uspostavljena drukčija struktura u skladu s potrebama korisnika modula, a povezanost u činjenici da su modul za strance i modul za učenike u pravilu uspostavljeni redukcijom strukture osnovnoga modula te da su sve riječi obrađene u modulu za strance i modulu za učenike obrađene i u osnovnome modulu. Modulna struktura *Mrežnika* ima posljedica i na odabir osnovnoga abecedarija od 10 000 riječi. On je dobiven tako da je ekscerpcijom udžbenika i nastavnih planova i programa za niže razrede osnovne škole dobiven abecedarij modula za učenike od 3000 riječi, a ekscerpcijom udžbenika za osobe koje uče hrvatski jezik kao drugi i strani jezik abecedarij od 1000 riječi. Ta su dva abecedarija preklapljena, pa je tako dobiven abecedarij dopunjen riječima prema popisu dobivenu od preklapljenih i izbalansiranih čestotnih popisa riječi u dvama hrvatskim korpusima: *hrWacu* i *Hrvatskoj mrežnoj riznici* (ne automatski, nego na temelju uredničkoga odabira), korpusima na kojima se temelji *Mrežnik*. On je zamišljen kao korpusno utemeljen, a ne korpusom vođen rječnik, odnosno kao rječnik koji se oslanja na korpusne podatke, ali se pri obradi njegovi obrađivači služe i svim drugim dostupnim izvorima i podacima (iz tiskanih rječnika, s interneta itd.). Korpusi se pretražuju s pomoću alata *SketchEngine*, a za potrebe projekta napisana je te potrebama projekta u više koraka prilagođena gramatika skica (*SketchGrammar*) za hrvatski jezik, s pomoću koje se jednostavno dobivaju najvažniji podatci o gramatičkome i kolokacijskome ponašanju pojedine riječi [2: 25–26]. *Mrežnik* je zamišljen kao normativni rječnik, pa i ta činjenica objašnjava zašto on ne može biti korpusom vođen rječnik – nijedan, naime, od spomenutih korpusa nije u pravome smisli riječi standardnojezični korpus; osobito u *hrWacu* nalazi se i niz tekstova s društvenih mreža, foruma itd., dragocjenih za istraživanje jezika u posebnoj uporabi, ali nerelevantnih (osim kao pokazatelj veoma čestih jedinica koje se približavaju razgovornome funkcionalnom stilu standardnoga jezika) za standardni jezik. [3: 208–210].

Rječnički članak uz natuknicu *zubar* koji donosimo u slici 3., u osnovnome se modulu sastoji od ovih elemenata:

- naglašene natuknice ;
- nenaglašene natuknice (neće se prikazivati u konačnome prikazu rječnika, potrebna je zbog pretraživanja rječnika);
- gramatičkoga bloka s naglašenim oblicima riječi (oblici su razrađeni i dogovoreni za svaku vrstu riječi te uređeni prema morfološko-naglasnim tipovima, što osigurava njihovu potpunu sustavnost);
- stilske oznake (*razg.*);
- definicije značenja;
- primjera preuzetih iz korpusa;
- kolokacijskoga bloka uspostavljena po modelu pitanja (npr. *Kakav je xxx?*) ili rubrika (npr. *Koordinacija*:) i njima pridruženih najčešćih kolokacija u kojima se pojavljuje riječ iz natuknice;

- normativne napomene kojom se objašnjava status riječi u standardnome jeziku;
- podatka o muškome/ženskome parnjaku³;
- podatka o sinonimu;
- tvorbene raščlambe;
- popisa tvorenica.

Sketch Engine hr Q Croatian Web (hrWaC 2.2, RFTagger)

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Corpus info
My jobs
User guide

Save
Change options
Cluster
Sort by freq
Hide grammels
More data
Less data
Sketch grammar

zubar (noun)
Croatian Web (hrWaC 2.2, RFTagger) freq = 7,213 (5.13 per million)

prijedlog	veznik	kakav?	komu-čemu	koordinacija
25.22	20.55	14.99	11.59	10.95
kod + 1,110 4.95 kod zubara	at 5 1.27 ili + 150 0.74 zubara ili	privatnom 5 5.81 kninski 15 5.67 kninski zubar	ići + 172 8.47 otići + 116 8.47 odvesti 26 7.97 boljeti 12 7.97 posjetiti 19 7.75 posjeti zubaru	ginekolog 34 8.66 zubara i ginekologa
k 34 2.47 k zubaru	iako 5 0.49 pa 45 0.46 zubara pa	dežuran 14 5.04 dežurnog zubara	otisati 6 7.68 odlaziti 16 7.50 isti 6 7.23 otisti 5 7.20 bojati 7 6.42 voditi 21 6.16 planirati 5 5.90 voziti 5 5.72	ortodont 9 8.38 frizer 18 7.89 kirurg 14 7.83 okulist 6 7.72 doktor 35 7.20 liječnik 67 6.93 liječnika i zubara
imenica_iza_prijedloga	ali 18 0.11 jer 50 0.06 zubara ali zubara jer	privatan 94 4.22 privatnog zubara socijalan 58 3.41 socijalnog zubara	prijašnji 10 3.14 ugledan 6 2.23 nijedan 8 2.19 tadašnji 8 2.00 izvrstan 6 1.75	specijalist 5 6.14 veterinar 5 6.05 odvjetnik 10 5.29 izvaditi 5 5.10 zub 10 4.87

Slika 1. Skica riječi *zubar*

File Edit View Lemma Dictionary Format Tools Window Help

Lemma: zubar LemmaSign=zubar, Naglaseno=zubár, vrsta_rijeci=im. m., Modified=2018-05-23 13:11:11, Created=2018-02-20 11:47:11, A

Gramaticki_blok
Oblici: naglaseni=GA zubára, DL zubáru, V zúbáru/zúbáre, I zubárom/zubárem; %imn.%i NV zubári, G zubárá, DLI zubárima, A zubárá

Sense: SenseNumber=1, stilska_odrednica=razg.

Definicija: definicija=Zubar je liječnik koji se bavi stomatologijom.

primjeri
Primjer: Primjer=Tko želi zdrave zube i čvrste desni, mora zube prati 2 do 3 puta dnevno, svilenim koncem pročistiti prostor između zubi.
Primjer: Primjer=U nekoliko trenutaka možete postići osmijeh iz snova, a da pritom ne idete zubaru na zahvate koji koštaju i oštećuju zube.

kolokacije
Kolokacija: odrednica=Kakav je zubar?, kolokacija=besplatan, dežurni, izvrstan, poznati, privatni, socijalni, ugledan, uspješan, vr
Kolokacija: odrednica=Što zubar može?, kolokacija=brusiti zub, liječiti desni/zub, otvoriti zub, praviti/raditi (navlaku, plombu, prot
Kolokacija: odrednica=Što se sa zubarom može?, kolokacija=bojati ga se, ići k njemu, dolaziti k njemu, imati ga, javljati mu se, n
Kolokacija: odrednica=Koordinacija, kolokacija=kirurzi i zubari, zubar i ortodont, zubari i doktori, zubari i frizeri, zubari i ginekolo
Kolokacija: odrednica=Povezuje se s, kolokacija=preporuka, strah, usluga

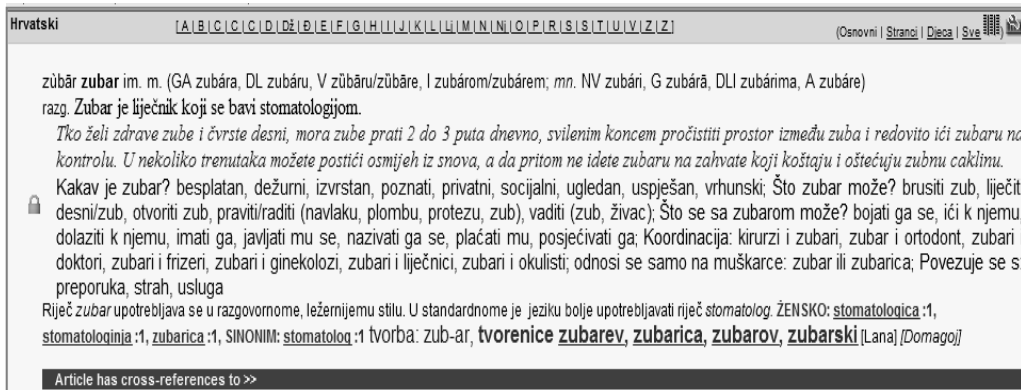
Uporaba: uporaba=Riječ %izubar%i upotrebljava se u razgovornome, ležernijemu stilu. U standardnome je jeziku bolje upotrebljavati %zubar%.

Poveznice
References: ŽENSKO: stomatologica(:1), ŽENSKO: stomatologinja(:1), ŽENSKO: zubarica(:1), sinonim stomatolog(:1)

Tvorbena_raz: tvorba=zub-ar
References: tvorenica zubarev, tvorenica zubarica, tvorenica zubarov, tvorenica zubarski

Slika 2. Prikaz strukture rječničke natuknice *zubar* u osnovnome modulu u programu TLEX

³ U rubrikama *Žensko:Muško*: ne donose se mocijski, nego značenjski parnjaci; naravno da su mocijski parnjaci uključeni u kategoriju muško-ženskih parnjaka, ali se rubrike *Žensko:Muško*: ne odnose isključivo na njih; u toj će se rubrici npr. uz riječ *koza* donijeti podatak *Muško: jarac*.



Slika 3. Prikaz obrade riječi *zubar* u osnovnome modulu u programu *TLex*

Osim navedenih elemenata postoje i drugi elementi, npr. uporabna/pragmatička napomena, podatci o ostalim značenjskim odnosima u koje riječ ulazi, npr. o antonimima, meronimima, hiponimima, kohiponimima – podatci za posljednje tri kategorije donose se samo kad je riječ o manjim i zatvorenim skupinama, o vidskome parnjaku, o gramatičkoj restrikciji povezanoj s određenim značenjem natuknice itd.; u školskome modulu nalazi se i dodatni element, podatci o tome kako se riječ koja se nalazi u natuknici rastavlja na slogove.

U ovome će radu dalje biti riječi o izazovima leksikografske obrade za mrežni rječnik s kojima su se susreli njegovi urednici i obrađivači, koji su (većina njih) leksikografsko iskustvo skupljali radom na tiskanim rječnicima, u prvome redu na *Školskome rječniku hrvatskoga jezika*.

Od mnogih izazova s kojima smo se susreli u izradi mrežnoga rječnika, ovdje ćemo opisati samo tri, prvi povezan s unutarrječničkim i izvanrječničkim povezivanjem, drugi povezan s drukčijom perspektivom pri razgraničavanju značenja povezanom s mogućnošću detaljnoga i iscrpnoga uvida u kolokacijsko ponašanje riječi i sveza te treći povezan s dvojbama koje nameće istodobno slijeđenje načela sustavnosti i načela korpusne utemeljenosti.

Izazovi leksikografske obrade u mrežnome rječniku

Unutarrječničko i izvanrječničko povezivanje. Medij za koji se izrađuje mrežni rječnik omogućuje pouzdano povezivanje unutar rječnika s pomoću poveznica. Poveznicima se povezuju riječi koje se nalaze u kojemu značenjskom odnosu (sinonimi, antonimi, meronimi, hiponimi, muško-ženski parnjaci – obostrano povezivanje) ili u kojemu gramatičkom odnosu (tvorenice i osnovna riječ, vidski parnjaci – obostrano povezivanje) ili u kojemu normativnome odnosu (upućivanje s pomoću uputnice vidi v.– jednostrano povezivanje). Iako i u nekim tiskanim rječnicima postoje rubrike za neke od navedenih kategorija, internet omogućuje da se jednim pritiskom na poveznicu pogleda i rječnički članak natuknice koja se nalazi na poveznici ili pojedino njezino značenje (značenjski se odnosi uspostavljanju među značenjima, a ne među natuknicama), a program za pisanje rječnika omogućuje potpun nadzor nad povezivanjem natuknica ili pojedinih njihovih značenja koja stoje u kojemu značenjskome, gramatičkom ili normativnome odnosu. U natuknici na Slici 3. riječ *zubar* povezana je sa ženskim parnjacima *stomatoloģica*, *stomatoloģinja* i *zubarica*, sino-

nimom *stomatolog* te tvorenicama *zubarev*, *zubarica*, *zubarov*, *zubarski*. Kako rječnik u prvoj fazi svoje izradbe, koja traje do 28. veljače 2021., obuhvaća (samo) 10 000 natuknica, bilo je potrebno predvidjeti i mogućnost donošenja tzv. mrtvih sinonima, antonima, tvorenica, muških/ženskih parnjaka itd. jer u prvoj fazi ne mogu sve riječi koje će se nalaziti na poveznicama biti uključene u *Mrežnik*. Njih je ipak važno zabilježiti, u prvome redu kao podatak za korisnika te oživljavanje takvih mrtvih jedinica uvođenjem njihovih rječničkih članaka imati u vidu u sljedećoj fazi obrade *Mrežnika*.

Medij za koji se izrađuje mrežni rječnik omogućuje uklapanje u rječnik različitih podataka s interneta. Predviđeno je uklapanje u *Mrežnik* podataka koji se već nalaze na internetu (na portalima Instituta za hrvatski jezik i jezikoslovlje *Jezični savjetnik* www.jezicni-savjetnik.hr, *Bolje je hrvatski* www.bolje.hr, *Kolokacijska baza hrvatskoga jezika* ihjj.hr/kolokacije, *e-Glava* (baza glagolski 10.5h valencija) valencije.ihjj.hr, *Struna* (baza stručnoga nazivlja) www.struna.hr), a također i izrada dodatnih baza za potrebe *Mrežnika*. Neke su od njih već smještene na mrežne stranice *Hrvatski u školi* (hrvatski.hr), a osim za potrebe modula za učenike dobrošao su izvor jezičnih podataka svim korisnicima toga portala namijenjena učenicima osnovne i srednje škole. One se nalaze u rubrikama *Savjeti* i *Frazemi*. U prvoj se donose jezični savjeti za učenike nižih razreda osnovne škole, a građa za tu rubriku uzeta je prema *Mrežnikovu* abecedariju za učenike. U drugoj se donose zanimljivi podatci o podrijetlu pojedinih frazema koji će također biti obrađeni u modulu za učenike. U okviru projekta *Mrežnik* izrađuje se i baza etnika i kjetika te baza veznika.⁴

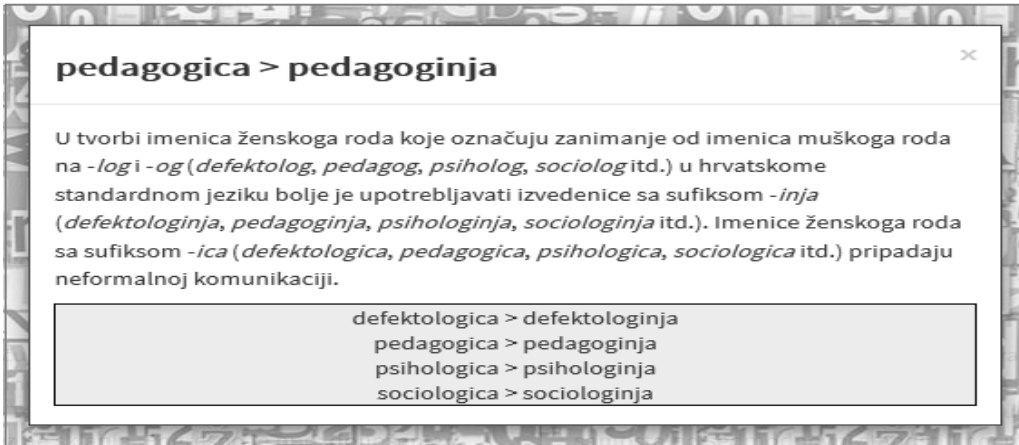
1. Primjer uklapanja savjeta s portala *Jezični savjetnik* u *Mrežnik*. Neki se podatci s već postojećih portala Instituta za hrvatski jezik i jezikoslovlje pri uklapanju u *Mrežnik* prilagođuju određenoj natuknici. Stoga neće s natuknicom u *Mrežniku* biti povezani poveznicom, nego će njihov prilagođen sadržaj biti kopiran u odgovarajuću rubriku u *Mrežniku*. Donosimo primjer takva preuzimanja savjeta s portala *Jezični savjetnik*, odnosno prilagodbe savjeta koji se odnosi na normativni status ženskih mocijskih parnjaka (koji se tvore sufiksima *-ica* i *-inja*) tvorenih od muških mocijskih parnjaka koji završavaju na *-og* na određenu natuknicu u *Mrežniku* – za primjer uzimamo natuknicu *stomatologinja/stomatologica*.

Iz primjera koji donosimo na Slici 4. i Slici 5. vidljivo je da se savjet koji uzimamo sa stranica *Jezičnoga savjetnika* prilagođuje specifičnoj natuknici; tako je, primjerice, u savjet koji se donosi uz natuknice *stomatologinja* i *stomatologica* u nizu primjera koji se nalazi na tim stranicama dodan i taj primjer te uvršten u abecedni niz primjera.

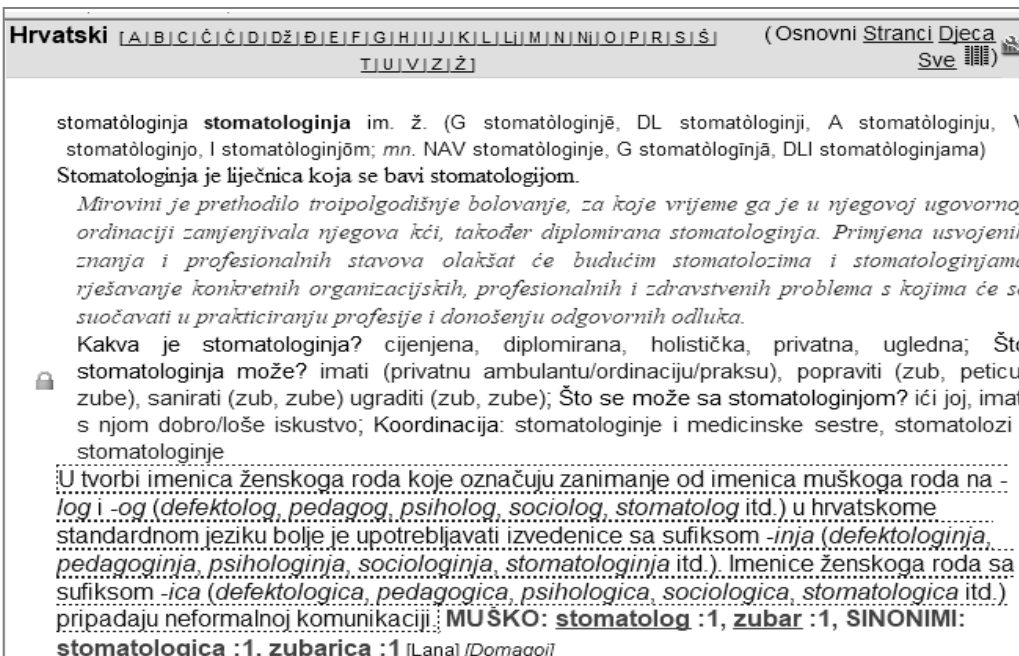
Uz velike mogućnosti okupljanja podataka s različitih mrežnih mjesta, ta mogućnost krije i zamku da se u rječnik uvrštavaju i podatci samo zbog toga što su dostupni, a ne zbog toga što su u njemu zaista potrebni. Urednice rječnika vrlo su rano postale svjesne te zamke te su svi izvori s kojima se podatci u *Mrežniku* povezuju vrlo pažljivo odabrani te su izrađene točne upute za povezivanje sa svakim od njih; u ovome ih radu ilustriramo primjerom povezivanja s bazom jezičnih savjeta na stranicama jezicni-savjetnik.hr. Osobito je pažljivo trebalo pristupiti preuzimanju podataka iz već postojećih baza u *Mrežnik* jer one, naravno, ne odgovaraju u potpunosti njegovim potrebama. Baze koje nastaju u okviru projekta osmišljene su tako da se podatci mogu iz njih preuzimati bez prilagodbe. Među načelima rječničke obrade tako načelo

⁴ Na hrcak.srce.hr/193900 može se preuzeti rad objavljen u časopisu *Hrvatski jezik* u kojemu se na str. 6–7 donosi slikovni prikaz povezivanja *Mrežnika* s drugim izvorima te iscrpniji opis baza s kojima se povezuje. [4: 6–7].

svrhovitosti u Mrežniku, pa i u svakome drugom mrežnom rječniku dobiva dodatnu dimenziju povezanu s medijem koji otvara iznimno velike mogućnosti povezivanja podataka.



Slika 4. Tipiski savjet koji se odnosi na sinonimne ženske mocijske parnjake sa sufiksima *-inja* i *-ica* u *Jezičnome savjetniku* (jezicni-savjetnik.hr)



Slika 5. Primjer prilagodbe tipškoga savjeta pojedinoj riječi (na primjeru natuknice *stomatologinja*); isti se savjet ponavlja i uz riječ *stomatologica*

2. Kolokacijski odnosi i razgraničavanje značenja. Činjenica da je riječ o rječniku utemeljenu na korpusu na koji je primijenjena gramatika skica te su obrađivačima rječnika dostupne skice riječi (jasan pregled gramatičkoga i kolokacijskoga ponašanja riječi) za svaku natuknicu pred obrađivače rječnika stavila je i problem određivanja značenja i njegova opsega, odnosno prototipnosti značenja. Pred cijelim pregledom kolokacijskoga ponašanja pojedine riječi dolazi se do mnogo većega broja značenja nego što smo ih, primjerice, imali uz pojedinu natuknicu u *Školskome rječniku hrvatskoga jezika* [1]. Uzmimo za primjer riječ *list*: jedno je od značenja te riječi u *Školskome rječniku hrvatskoga jezika*: *što tanko, što izgledom podsjeća na list* te se kao primjeri navode kolokacije *list papira, list tijesta* [1: 290]. U *Mrežniku*, u naprijed opisanome sustavu donošenja kolokacija, takva bi definicija značila da ćemo, uz to značenje, u kolokacijskome bloku uz pitanje *Što se s listom može?* imati odgovore *ispisati ga, isprintati ga, nadjenuti ga, napisati što na njega, pisati po njemu, poderati ga, premazati ga (jajem, vrhnjem), presaviti ga, razvući ga, sviti ga, zgužvati ga, odrezati ga*. Međutim, problem koji su uočili obrađivači *Mrežnika* u tome je što se dio tih odgovora vezuje samo uz izraz *list papira (ispisati ga, isprintati ga, napisati što na njega, pisati po njemu)*, a dio samo uz izraz *list tijesta (nadjenuti ga, premazati ga (jajem, vrhnjem), razvući ga)*, dok se određen broj odgovora može odnositi na oba izraza (*poderati ga, presaviti ga, sviti ga, zgužvati ga, odrezati ga*). Iz toga proizlazi da ono što smo letimičnijim uvidom u korpus (uvidom u korpus bez skica riječi) pri izradi *Školskoga rječnika hrvatskoga jezika* smatrali jednim značenjem u *Mrežniku* razdvajamo kako bi u kolokacijskome bloku sve kolokacije bile primjenjive za jedno značenje. U *Mrežniku* je to riješno tako da je pod definicijom *List je tanak komad čega, ono što izgledom podsjeća na istoimeni dio biljke*. uvedena sveza *list papira*. Tako se broj značenja znatno povećava u odnosu na broj značenja u *Školskome rječniku hrvatskoga jezika*.

U *Hrvatskome školskome rječniku* riječi *vozač* pridruženo je tvorbena značenje *osoba koja vozi* [1: 872], međutim, uvid u širi kolokacijski opseg te korpusne potvrde doveo je do toga da u *Mrežniku* ta riječ ima znatno veći broj značenja s pridruženim primjerima, kolokacijama i svezama; ovdje navodimo samo značenja te svezu pridruženu 4. značenju, a opravdanost uvođenja tih značenja objašnjavamo samo po jednom koordiniranom kolokacijom u zgradama koja jasno upućuje na to da je riječ o posebnome značenju):

1. Vozač je odrasla osoba bez obzira na spol ili muškarac koji zna upravljati vozilom (*vozači i nevozači*);

2. Vozač je odrasla osoba bez obzira na spol ili muškarac koji upravlja osobnim vozilom (*vozač i suvozač*);

3. Vozač je odrasla osoba bez obzira na spol ili muškarac koji se profesionalno bavi upravljanjem vozilom (*vozač i tjelohranitelj*);

4. *sp.* Vozač je odrasla osoba bez obzira na spol ili muškarac koji vozi utrke (*vozači i mehaničari*);

5. testni vozač (*testni vozač i kaskader*).

Osim broja značenja, znatno se, iz istih naprijed opisanih razloga, u *Mrežniku* povećava i broj sveza, koje imaju sve elemente koje ima i obrada riječi u natuknici te se ukupna veličina rječničkoga članka bitno povećava. S obzirom na to da u mrežnome rječniku zbog višestrukih mogućnosti organizacije prikaza rječničkoga članka ne postoji problem ograničenosti prostora koji postoji u tiskanome rječniku, u kojemu bi velik broj značenja bitno smanjio preglednost rječničkoga članka te korisniku otežao korištenje rječnikom (kao, primjerice, u Akademijinu rječniku (*Rječnik hrvatskoga ili srpskoga jezika* Jugoslavenske akademije

znanosti i umjetnosti u 23 sveska, koji je izlazio od 1880. do 1976. godine), u kojemu pojedini rječnički članci, tiskani sitnim tiskom, obasežu i nekoliko stranica velikoga formata te je taj rječnik iznimno vrijedan za znanstvene svrhe, ali nepristupačan općemu korisniku.

3. Između korpusa i sustava. U *Školskome rječniku hrvatskoga jezika* veliku smo pozornost poklanjali sustavnosti, odnosno sustavnost je bila jedno od najvažnijih leksikografskih načela kojih smo se u tome rječniku pridržavali. Ona se očitovala na nekoliko razina: u donošenju sustavnih definicija koje su se uspostavljale uvijek kad je to bilo moguće (dakle, kad god je bila riječ o određenoj gramatičkoj – imenice na –ost, glagolske imenice, različite skupine zamjenica i odnosnih priloga, odnosni pridjevi, posvojni pridjevi, ili semantičkoj – godišnja doba, dani u tjednu, mjeseci u godini, horoskopski znakovi, mjerne jedinice, uređaji, biljke i njihovi plodovi, biljke i njihovi cvjetovi itd., skupini

- u donošenju sustavnih tvorbenih gnijezda uz pojedine riječi (npr. sustavnome donošenju ženskoga mocijskoga parnjaka uz muški, nesvršenoga glagola uz svršeni).

Prvo načelo primjenjujemo i u *Mrežniku*, odnosno rječnička se obrada provodi po zatvorenim skupinama od kojih se svaka dodjeljuje pojedinome obrađivaču. Pri izradbi *Školskoga rječnika* takav se model obrade pokazao iznimno dobrim zato što:

- jamči sustavnost definicija i usklađenost obrade jedinica koje pripadaju zatvorenim skupinama

- ubrzava obradu jer jedan obrađivač primjenjuje istu tipsku definiciju koju mijenja tek u elementu koji razlikuje jedinice opisane tom tipskom definicijom, na određen se način “specijalizira” za obradu određene skupine.

Pri primjeni drugoga načela u izradi *Mrežnika* naišli smo na poteškoće prouzročene činjenicom da je *Mrežnik* rječnik koji je utemeljen na korpusu, pa iako to znači da ga korpusni podatci ne obvezuju (nije, naime, riječ o korpusom vođenom rječniku koji bi korpusni podatci obvezivali u smislu da opisuje stanje u korpusu), ipak smo ostali zatečeni činjenicom da neke riječi (uglavnom je riječ o tvorenicama od riječi koje ulaze i u dječji vokabular i čija je tvorba posve sustavna u korpusu (pa i šire, na internetu) nemaju potvrda. Razmjerno se često to događalo za ženske parnjake ili za posvojne i odnosne pridjeve izvedene od muških ili ženskih parnjaka te ih zasad ipak unosimo uz veoma česte muške parnjake uz uporabnu napomenu da u korpusu i na internetu nismo za njih našli potvrda. To radimo samo za ženske parnjake i pridjeve koje možemo pridružiti muškima u temeljnome korpusu od 3000 riječi; za muške parnjake koji se nalaze u ostalih 7000 natuknica nećemo uvoditi ženske parnjake temeljem načela sustavnosti, nego će biti zabilježeni samo ako se nalaze u abecedariju dobivenom balansiranjem i preklapanjem dvaju čestotnih popisa pojavnica u *hrWacu* i *Hrvatskoj jezičnoj riznici*.

Zaključak. U radu na *Hrvatskome mrežnom rječniku – Mrežniku* susreli smo se s mnogim izazovima leksikografske obrade, odnosno morali smo donijeti niz odluka i rješenja koje nije bilo potrebno donositi u rječnicima pisanim za tiskani medij s kojima su suradnici na projektu dosad imali iskustva. Izazov je bio i dopuna gramatike skica koju je izradio vanjski član projekta Nikola Ljubešić kako bi na najbolji način odgovorila na potrebe pisanja *Mrežnika*. Izazov je također bio i pisanje rječnika u programu za leksikografsku obradu *TLexu*, čemu je prethodila intenzivna suradnja projektne skupine za tehnološka rješenja s članom projekta zaduženim za projektiranje polja za unos rječničkih podataka u sva tri modula i intenzivna svakotjedna obuka članova projektne skupine za rad u tome programu namijenjena i upoznavanju svih mogućnosti toga programa. Oba ta izazova projektne je

skupina uspješno svladala u prvome redu nizom probnih obrada s kojima se započelo već u prvoj projektnoj godini i koja je bila usmjerena utvrđivanju mjesta na kojima se gramatika skica za hrvatski jezik može doraditi te utvrđivanju elemenata u kojima se obrada u *TLexu* može učiniti za obrađivače jednostavnijom. Napominjemo da je u *TLexu* riječ o iznimno složenome okviru za unos podataka koji je takav u prvome redu zbog *Mrežnikove* tromodulne strukture. Osim tih tehnoloških izazova, susreli smo se i s mnogim izazovima same leksikografske obrade, od kojih smo u ovome radu pozornost posvetili trima: untarrječničkome i izvanrječničkome povezivanju, kolokacijskim odnosima i razgraničivanju značenja te problemima koji se odnose na niz odluka o tome u kojoj mjeri i na koji je način korpus relevantan za donošenje podataka u rječniku, odnosno kako postupiti u slučajevima kad korpusni podatci (svojim nepostojanjem) ne podržavaju jedno od temeljnih načela koje smo željeli provesti u *Mrežniku*, načelo sustavnosti. Tri izlučena problema samo su malen prikaz svih izazova s kojima se leksikograf susreće pri izradi mrežnoga rječnika.

Literatura

1. *Birtić M.* Školski rječnik hrvatskoga jezika / M. Birtić i dr. – Zagreb : Školska knjiga i Institut za hrvatski jezik i jezikoslovlje, 2012. – 968 s.
2. *Blagus Bartolec G.* Kad nam korpus ispunjava želje / G. Blagus Bartolec, I. Matas Ivanković // Hrvatski jezik. – Zagreb : Institut za hrvatski jezik i jezikoslovlje, 2017. – T. 4. – N° 3. – S. 25–28.
3. *Hudeček L.* A New Project – Croatian Web Dictionary MREŽNIK / L. Hudeček, M. Mihaljević // The Future of Information Sciences. INFUTURE 2017, Integrating ICT in Society. – Zagreb : Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, 2017. – S. 205–213.
4. *Hudeček L.* Hrvatski mrežni rječnik – Mrežnik / L. Hudeček, M. Mihaljević // Hrvatski jezik. – Zagreb : Institut za hrvatski jezik i jezikoslovlje, 2017. – T. 41. N° 2. – S. 1–8.
5. *Štrkalj Despot K.* A perspective on e-lexicography / K. Štrkalj Despot, C. Möhrs. – Zagreb : Rasprave : Časopis instituta za hrvatski jezik i jezikoslovlje, 2015. – T. 41. – N° 2 – S. 329–353.

Elektronički izvor

6. *Hudeček L.* The Croatian Web Dictionary Project – Mrežnik / L. Hudeček, M. Mihaljević. Dostupno na adresi : <https://elex.link/elex2017/wp-content/uploads/2017/09/paper10.pdf>. Datum pristupa : 10.06.2018.

Стаття: надійшла до редакції 30.07.2018
прийнята до друку 20.08.2018

THE CHALLENGES OF COMPILING A MONOLINGUAL WEB DICTIONARY (on the example of *Croatian Web Dictionary – Mrežnik*)

Lana Hudeček

*Institute of Croatian Language and Linguistics,
Republike Austrije 16, 10 000 Zagreb, Croatia,
e-mail: lhudecek@ihjj.hr*

On the example of Croatian Web Dictionary – Mrežnik, a research project financed by the Croatian Science Foundation which is compiled in the Institute of Croatian Language and Linguistics from March 2017, the author analyzes challenges facing the lexicographers and editors of such dictionaries, i.e. in which elements such work introduces new insights into lexicographic work and differs from the work on printed dictionaries. The di

ctionary entries will contain links to repositories which will be created as a part of this project and made parallel with the dictionary (Linguistic Advice Repository, Conjunction Repository, The Idiom Etymology Repository, The Repository of Ethnic and Ktetics) and repositories which are already being constructed as a part of the projects conducted at the Institute of Croatian Language and Linguistics. Apart from the standard definitions, the dictionary will include definitions for school children (3000) and definitions for learners of Croatian as a foreign language (1000). The dictionary will be corpus based (Croatian Web Repository, corpus of the Institute of the Croatian Language and Linguistics, at: <http://nlp.ffzg.hr/resources/corpora/hrwac/>, and the Croatian Web Corpus, at: <http://nlp.ffzg.hr/resources/corpora/hrwac>). Apart from the aforementioned corpora, other online and printed sources will be taken into account. A number of possibilities that such a dictionary offers (e.g. the work on corpora and with computer tools for corpus search, links to other resources) introduces the need for their careful analysis and limitation to those that are useful for and to the user. The challenges that the editors and lexicographers met during the trial period in the first phase of the project as well as ways of meeting these challenges by means of modern e-lexicographic theory, solutions used in other big international e-dictionaries, and their own lexicographic experience gained while working on printed dictionaries will be shown.

Key words: e-dictionaries, e-lexicography, Croatian Web Dictionary, Mrežnik, monolingual dictionaries.