

ISSN 1452-4112

С Р П С К А А К А Д Е М И Ј А Н А У К А И У М Е Т Н О С Т И
О Г Р А Н А К У Н О В О М С А Д У



АНАЛИ
ОГРАНКА САНУ У НОВОМ САДУ

Број 14 за 2018.

НОВИ САД
2019.

СТАРИ РЕЧНИЦИ И НОВЕ ТЕХНОЛОГИЈЕ¹

Снежана Петровић²

Абстракт. У раду је, након краћег прегледа историјата рада на пословима дигитализације у Институту за српски језик САНУ, представљена речничка платформа *Расковник* и речници који се на њој налазе: *Српски речник* Вука Карадића, *Речник косовско-метохиској гујалеке* Г. Елезовића, *Речник јавора јужне Србије* М. Златановића, *Речник јавора Прошићења (ког Мојковца)* М. Вујичића и *Речник дубровачке јавора* М. Бојанића и Р. Трикуницац. У основним цртама приказан је начин на који су речници дигитализовани и главне функције које *Расковник* пружа. Указано је на научни и шире друштвени значај дигитализације речника и језичких ресурса.

Дигитализација у Институту за српски језик САНУ

Поред тога што је Институт за српски језик САНУ централна научна установа у Србији у којој се проучава српски језик из синхроне и дијахроне перспективе, у његову основну делатност, од самог оснивања, спада и израда капиталних научних лексикографских дела – речника и атласа. О научном значају ових подухвата и потреби да се они израђују најбоље сведочи чињеница да је Институт настао на темељима Лексикографског одсека који је 1893. године формиран у Српској краљевској академији, на иницијативу академика Стојана Новаковића, управо са циљем да лексичко благо српског језика буде сакупљено и језик у својој целини проучаван на највишем научном

нивоу. Како су се лингвистичка знања током година продубљивала и ширила, а перспективе истраживања и облици њиховог публиковања увећавале, тако се и делатност Института гранала. Осим првог пројекта, израде *Речника српскохрватске књижевној и народној језику*, данас у Институту постоји још њих четири, као и један потпројекат.³

Ширење делатности Института за српски језик САНУ последица је разумевања тога да проучавање матичног језика представља континуиран процес који мора бити усклађиван са савременим научним токовима како би доносио квалитетне резултате. Уз то, осавремењавање истраживања захтева и примену техника и метода подобних времену у коме она настају. Због тога су у Институту деценијама уназад постојале иницијативе да се у рад, на првом месту лексикографски, уведе примена нових технологија.

¹ Овај рад је настао у оквиру пројекта „Интердисциплинарно истраживања културног и језичког наслеђа Србије и израда мултимедијалног интернет портала Појмовник српске културе“ (47016), који финансира Министарство просвете, науке и технолошког развоја Републике Србије.

² Институт за српски језик САНУ.

³ Више о историјату Института и његовој делатности може се наћи на сајту <http://www.isj.sanu.ac.rs/>

Прве идеје о томе појавиле су се крајем осамдесетих година прошлог века и везане су за рад на *Речнику српскохрватској књижевној и народној језику*. Током деведесетих година, у оквиру пројекта израде *Етимолошкој речнику српској језику* јавила се идеја о потреби за дигитализацијом извора (на првом месту лексикографских) ради бржег и ефикаснијег рада. Међутим, ове идеје морале су да сачекају да прођу деценије и да се стекну различити, објективни и субјективни услови, како би се неке од њих и реализовале.

Први корак у тој реализацији у оквиру научних пројеката који се израђују у Институту десио се 2011. године, када је као део III пројекта *Интардисциплинарно истраживање језичкој и културној наслеђа Србије*. *Израда мултимедијалној интернетне Јордана Појмовнику српске културе* чији је носилац Етнографски институт САНУ, формиран потпројекат *Примена лингвистичких истраживања у изради дигиталној Јордану Појмовнику*. Пројектни тим из Института за српски језик САНУ чинили су сарадници са три постојећа пројекта (дијалектолошког, етимолошког и старословенистичког), а један од главних задатака било је осмишљавање и сарадња у изради дигиталне лексикографске базе дијалекатских речника српског језика. Нажалост, неадекватан начин финансирања пројекта није омогућавао суштинске помаке на овом пољу у оквиру научног пројекта.

Прави почетак реализације пројекта дигитализације десио се тек формирањем три краткорочна пројекта⁴: *Призрен – живот у речима. Дигитализација рукописне збирке речи из Призрена Димитрија Чемерикића (2013–2014)*, *Расковник: српски лексикографски Јордан*

⁴ Сва три пројекта подржало је Министарство културе и информисања Републике Србије у наведеним годинама, док даље стање о њима финансира Институт из средстава постојећих научних пројекта. Три сараднице Института ангажоване су и на пројекту ELEXIS (European Lexicographic Infrastructure) из програма Хоризонт 2020, у оквиру кога се такође ради на дигитализацији наших речника.

⁵ О поступку дигитализације ове збирке, анатацији и обогаћивању података в. Петровић / Тасовац 2016.

⁶ Започета је и дигитализација *Речника САНУ*, а о томе шта је до сада на том пољу урађено в. Стијовић / Станковић 2018.

штал и платформа за упоредна истраживања српске лексике (2015–2016) и *Израда дигиталној каталога листића Речника САНУ* (2017).

Основни послови на пољу дигитализације језичких ресурса у Институту за српски језик САНУ данас се одвијају на три поља:

Дигитализација Јордана, коју углавном чине милиони руком писаних или на писаћим машинама куцаних листића, минимално уобличених у лексичке јединице. Највећи број листића део су картотеке за израду *Речника САНУ*. На овом плану урађено је много у последњих пет година у оквиру два пројекта: а) дигитализована је рукописна збирка речи из Призрена Димитрија Чемерикића, једна од највећих појединачних збирки која је 1950. године предата Институту да послужи као грађа за *Речник САНУ*, али је остала изван његове картотеке. Насловљени и напредно обележени листићи налазе се на институтском платформи *Прејис* (<http://www.prepis.org/items/browse?collection=1>)⁵ и б) комплетна картотека грађе за *Речник САНУ*, око шест милиона листића, биће скенирана до краја 2018. године. Израђена је база података, а у току је насловљавање скенова. Када овај посао буде завршен та драгоценна грађа биће сачувана од евентуалног уништења и приступачнија за лексикографски и научни рад;

Израда речника и атласа у дигиталном окружењу, чији се почетак планира за наредну годину;

Дигитализација извора, чију прву фазу чини дигитализација речника, дијалекатских, историјских, специјалних, терминолошких и других. Дигитализовани речници налазе се на институтској речничкој платформи *Расковник*.⁶

Речничка платформа *Расковник*

Речничка платформа *Расковник* пројекат је Института за српски језик САНУ у оквиру којег се развија напредно окружење за дигитализацију српског речничког наслеђа и развој алата за истраживање српског језика и културе и налази се на адреси <http://raskovnik.org/>. Овај пројекат се реализује у сарадњи са Центром за дигиталне хуманистичке науке из Београда.

Речничка платформа је интернет сајт или низ међусобно повезаних сајтова путем којих се врши приступ већем броју различитих електронских речника. Сваки речник на платформи може се претражити и као самостална јединица. Важна особина речничке платформе јесте и то да је садржина речника обележена по истим, компатибилним принципима и да је структурисана на такав начин да омогућава међусобно повезивање података из свих дигитализованих речника према различитим, унапред одређеним параметрима. На пример, на порталу *Расковник*, све речи које су означене као деминутиви могу се наћи путем јединствене претраге, без обзира на то у ком су од дигитализованих речника посведочене. У дигиталном окружењу, будућност речника лежи у њиховом преласку од засебних, изолованих приручника ка интегрисаним онлајн истраживачким сервисима (De Schryver 2003: 157). Речничка платформа је управо такво место, на коме се одвијају захтевна претраживања и истраживања речника.

Речници на платформи *Расковник*

Српски лексикографски портал заслужио је да прво и на симболичан начин почасно место у њему заузме *Српски речник* Вука Стефановића Караџића – то је био први речник постављен на *Расковник*. Значај овог дела за нашу културу и историју готово да се не може преценити. Поред тога што представља један

од темеља на којима је створен савремени српски језик, ово дело је и права енциклопеђија српске материјалне и духовне културе и ризница описа народних обичаја, умотворина, одевања, исхране, друштвеног живота, породичних односа и др. У њему се огледају најважнија поља Вуковог рада: филологија, фолклористика, етнологија и историографија, а актуелност *Речника* за ове научне области и дан-данас није умањена. Дигитално издање Вуковог *Речника* заснива се на издању из 1898. године, уз додате одреднице из првог издања из 1818. које из каснијих издања изостају.

Поред овог речника, на *Расковнику* се тренутно налазе још четири дијалекатска речника:

Речник косовско-међахиској дијалекти Г. Елезовића (Елезовић 1932, Елезовић 1935) први је модеран српски дијалекатски речник, састављен на основу грађе прикупљане у периоду од 1902. до 1928. године, са простора чији говори спадају у косовско-ресавски дијалекат. Уз семантичке дефиниције и примере говора, речник садржи и ауторове етимолошке коментаре, првенствено код речи турског и албанског порекла, као и бројна лична имена, топониме и фразеологизме.

Речник јовора јужне Србије (Златановић 1998) плод је вишедеценијског рада Момчила Златановића, врсног познаваоца народног стваралаштва јужне и југоисточне Србије. Грађу за речник чине речи из усменог песништва и живог народног говора. Лексички материјал сакупљан је на ширем терену – у Врању и околини, Пчињи, Пољаницама, Иногошту, Прешевској Моравици и Прешевској Црној гори, Грделичкој клисури, Власини и Црној Трави. Речник садржи велики број личних и породичних имена, микротопонима и примера из народних песама, пословица и загонетки.

Речник јовора Прошћења (код Мојковац) Милоша Вујичића (Вујичић 1995) доноси грађу из живог говора становника Прошћења, села у општини Мојковац. Поред дијалектизма, аутор је бележио и речи које спадају у основни лексички фонд српског језика, као

и лексику страног порекла, од чега највише турцизама.

Рјечник дубровачкој јовора Михаила Бојанића и Растилаве Тривунац (Бојанић / Тривунац 2002) бележи грађу из живог говора становника Дубровника и његових предграђа и садржи неколико хиљада речи и израза забележених у првој половини XX века. Настао је као резултат личног интересовања Михаила Бојанића, по струци правника, за говор свога града. Збирку је за штампу припремила његова сестра Растилава Тривунац, доктор књижевних наука, у чему су јој помогли врсни дијалектологи.

Српски рјечник Вука Каракића има 46.210 одредница, *Речник косовско-међиохиској дигиталној лекцији* Глише Елезовића 12.836 одредница, *Речник јовора јужне Србије* Момчила Златановића 9.178 одредница, *Рјечник јовора Прошићева (код Мојковца)* Милоша Вујићића 5.213 одредница, а *Рјечник дубровачкој јовори* М. Бојанића и Р. Тривунац 9.904 одреднице. Уз могућност интегралног претраживања на *Расковнику* добија се „нови рјечник“ који броји 84.041 одредницу.

Начин дигитализације рјечника

За дигитализацију лексичке грађе и штампаних рјечника постоје четири приступа која се могу комбиновати: захватање слике (image capture), захватање текста (text capture), лексикографско моделирање података (lexicographic data modeling) и лексикографско обогаћивање података (lexicographic data enrichment) (Tasovac / Petrović 2015: 386–387). За потребе објављивања на *Расковнику*, сви речници су најпре прекуцани, а потом су урађене три коректуре текста. Покушаји скенирања и коректуре OCR-ованог текста показали су се мање ефикасним од прекуцавања, због бројних акцентованих словних знакова у насловима одредница и примерима и комбиновања различитих писама (ћирилица, латиница, ста-

рословенски, готица и др.). За сваки рјечник урађен је модел микроструктуре, на основу којег су они автоматски обележени у проширивом језику за означавање (Extensible Markup Language, XML), а у складу са смерницама Иницијативе за кодирање текста (Text Encoding Initiative, TEI). Пре објављивања на платформи, урађене су исправке у XML-у сваког рјечника. У појединим случајевима додаване су уредникове допуне, увек на посебан начин маркиране. На тај начин, у дигитализованим речницима урађено је лексикографско моделирање, а започето је и обогаћивање података.

Оваквим обележавањем преводи се нама разумљив текст рјечника на језик који разумеју компјутери. На Слици 1 приказана је једна одредница из *Речника јовора јужне Србије* М. Златановића, онако како је ми читамо и означене су информације које имплицитно сазнајемо из начина на који су у лексикографском тексту приказане. (Слика 1)

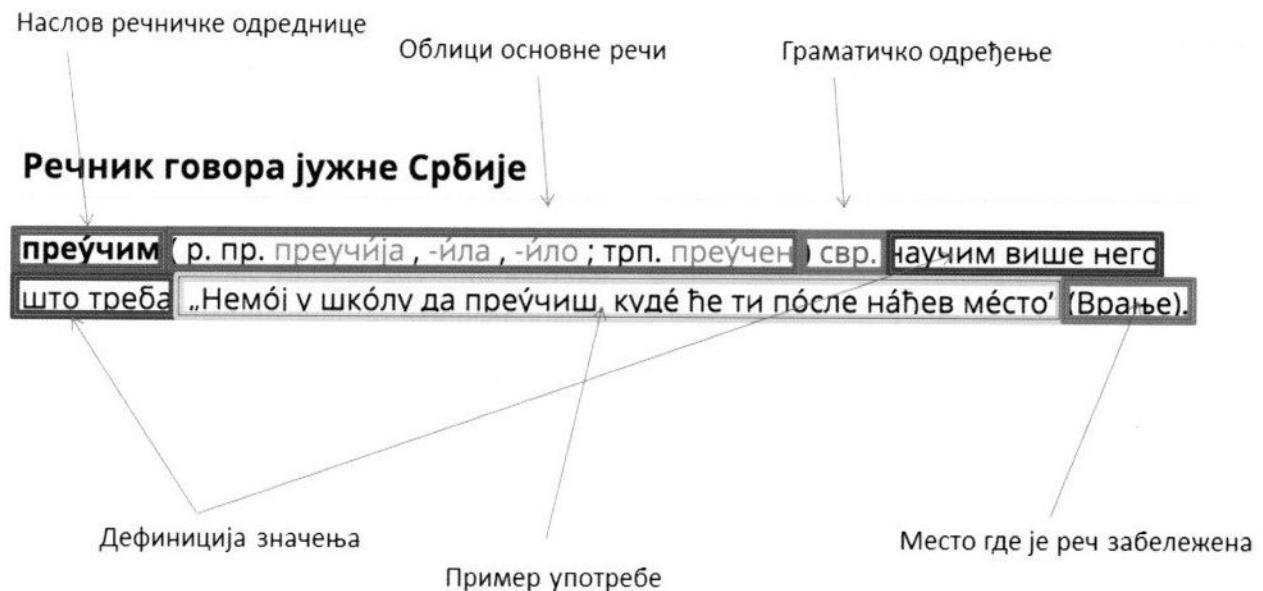
Те имплицитне информације су приликом дигитализације експлицитно обележене и на Слици 2 приказана је иста одредница у језику XML-у.

Ознаке којима су речници обележени служе томе да оно што ми, људи, разумемо постане разумљиво и компјутерима, и на тај начин међусобно повезано, као и да представља основу за структурисано претраживање преко свих параметара експлицитно обележених у тексту.

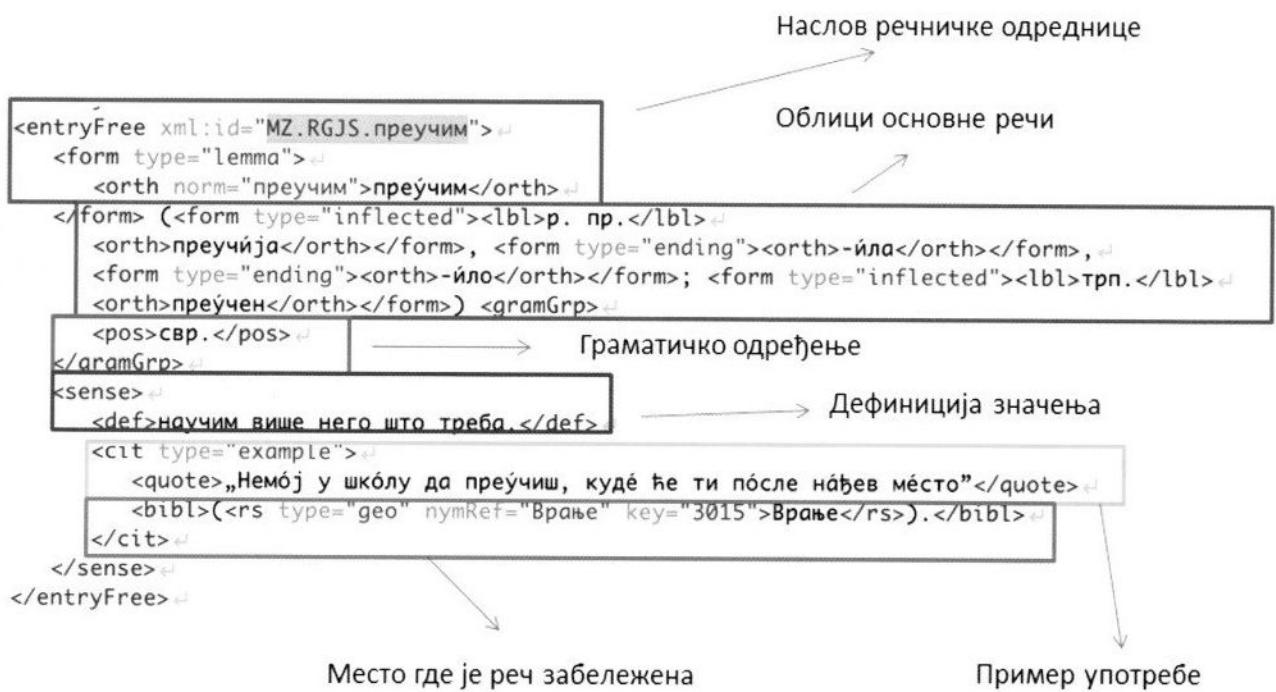
Функције на *Расковнику*

Детаљним обележавањем рјечника омогућене су три основне функције на *Расковнику*: претраживање, исечци и мапе.

Претраживање. Постоје четири врсте претраге на платформи: претрага у главном менију, претрага из одредница, претрага по речницима и напредна претрага.



Слика 1



Слика 2

Претрага у главном менију, или инстант-претрага, налази се на почетној страници *Расковника* (Слика 3). Укуцавањем речи у поље за претрагу добијају се резултати у свим или у изабраним речницима.

Слика 3

Претрага из одредница значи да се долажење до нових речи може наставити и на страницама самих одредница, и то у сваком

речнику понаособ (Слика 4). Кликом на знак „+” поред сваког резултата, нова реч додаје се испод претходног налаза.

Претрага по речницима, азбучним редом, почиње одабиром појединачног речника са почетне стране *Расковника*. Након тога се изабре жељено слово са списка на левој страни екрана, као и начин приказивања, који може бити у виду списка речи или целокупних одредница. (Слика 5).

Напредна претрага служи томе да на платформи претражујемо целокупан текст речника или одређене његове делове, а не само насловне речи одредница. За напредну претрагу могу се одабрати следећи делови: лема, целокупна одредница, семантичка дефиниција или пример употребе (Слика 6). Такође, у овој врсти претраге могу се користити посебни знаци којима се у упиту замењује један или више словних знакова. На пример, знак „..” у некој речи замењује тачно један словни знак, знак „..?” замењује један или ниједан словни знак, док ознака „.*” један или више словних знакова. Ово је само део могућности које нам

Српски рјечник

бàсма, f. оне ријечи скупа што бајалица или бајач г Zauberspruch, carmen.
бàсма* f. 1. gedruckte Leinwand, linteum pictum. 2. pyxidis ad pulverem nitratum [cf. везма].

← баскјати

= 10 × 1

Речник косовско-метохиског дијалекта

бàсма, e ж. циц. У В. 2. У В. 2 бàсма ж. 1) без оног др њега наведено. У RJA. 2. *bàsma* ж. 1).

← баскјин

= 10 × 1

бая	
бàја	+
бáја	+
бàјалица	+
бàјало	+
бàјам	+
бàјат, бàјатан	+
бàјати	+
Бàјац	+
бàјач	+
бàјање	+
Бáјко	+
Бàјнà Лóка	+

Слика 4

Српски речник

А	Прикази:	списак речи	•	целокупне одреднице	укупно на б.														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Б																			
1.	баба , f. (gen. pl. баба) 1. очина или материна мајка, die Grossmutter, avia. [cf. стара 2]. 2. стара жена, altes Weib, verula. [cf. стара 1, старица]. 3. у добровњику се свака дојкиња и дадиља зове баба; макар била и девојка од 12 година, die Amme, Kindsfrau, nutrix: ancilla curans parvulos. 4. многи муж зове своју жену [1] бобом, ако и није стара. 5. гдекоји и за пуничу [1] реку: моја боба, die Schwiegermutter, socrus. 6. [бобо] планина у Србији, између Ниша и Парагина [близу Мораве] изнад села Лешја, урезу парагинском округа моравског: ein Berg in Serbien, mons Serbieae. 7. [Бобо] ein Berg in der Herzegowina, mons Herzegovinae. 8. превести бобу, каку дјеца кад баце малу плочицу од камена преко воде, али је тако баце да одскоче поврх воде.																		
2.	баба [*], m. (ист.) vide бабо.																		
бабаје , f. pl. (у Мостару) некаке трешње, које се зову и роковице, Art Kirschen, cerasi genus. [vide рокавац]. cf. руцтеви.																			
бабајин , a, o, adj. у пјесми vide бабов: И изведе ђогу бабајин.																			
бабајко , m. (понајвише у пјесмама) [hyp. v. бабо]: [cf. балко, тајко, чале 1, ћако, ћаћко].] vide бабо [1]: Мој бабајко, не слушам те стари — Камо, Раде, мој мили бабајко — Каж бабајку утрос на уранку.																			
бабајко *, a, o, adj. vide бабов: Видиш, брате, токе бабајкове.																			

Слика 5

пружа напредна претрага, а детаљна упутства о њеној употреби налазе се на *Расковнику*.

Основна претрага Напредна претрага

Преко прече, наоколо ближе...

Напредне могућности претраге

Речници:

- ✓ Сви речници
- ✓ Вук Стеф. Карадић, Српски речник
- ✓ Глиша Елевовић, Речник косовско-међохиској дијалекто
- ✓ Момчило Златановић, Речник говора јужне Србије
- ✓ Милош Вујићић, Речник говора Прошићења
- ✓ М. Бојанић и Р. Тривунац, Речник дубровачког говора

Опсег претраге:

- одредница
- лема
- дефиниција
- пример

Слика 6

Ова функција омогућава да се брже и прецизније дође до нових информација, на првом

месту корисних за научноистраживачки рад и да се урони у дубинску структуру текста и сагледају речници из нове перспективе. Као илустрацију могућности које нам ова напредна функција пружа навешћемо пример речи *йосле* у дијалекатским речницима. Наиме, ова врста приручника углавном је писана диференцијално у односу на стандардни, књижевни језик. Тако овај прилог у пет речника у основној претразији налазимо само у Вуковом *Речнику*. Међутим, када га уз помоћ напредне претраге потражимо у примерима осталих речника, добијамо другачију слику – *йосле* је у примерима у *Речнику Јовора Прошићења* посведочено 1 пут, у *Речнику Јовора јужне Србије* 8 пута, а у *Речнику косовско-међохиској дијалекто* чак 107 пута. Само једна од предности функције напредне претраге примера јесте то што нам пружа могућност да речнике користимо и као врсту корпуса.

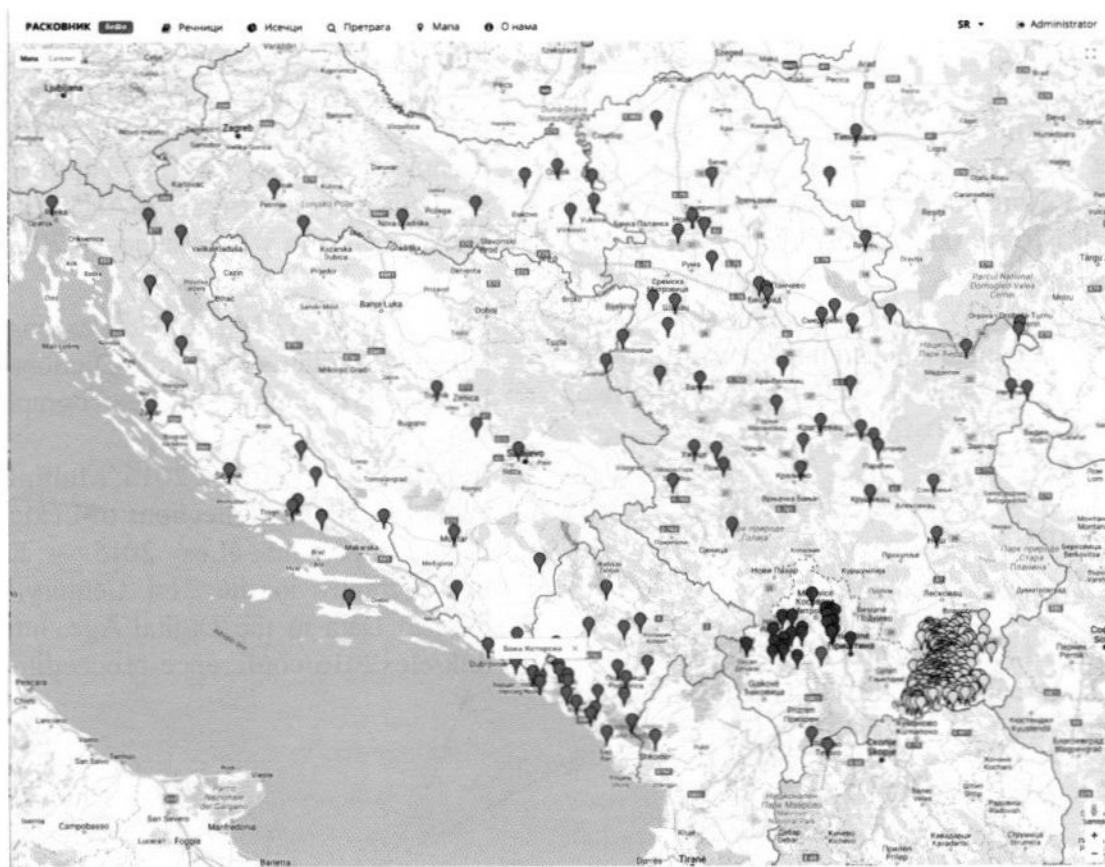
Исечци. Исечци су делови речника који садрже одређену, посебно одабрану врсту лексике и који су издвојени из речничке целине

захваљујући томе што су делови одредница прецизно обележени. Навешћемо неке исечке који се налазе на *Расковнику*: *Og абе до шућура* (речи турског порекла у Вуковом *Рјечнику*), *Грчко наслеђе* (речи из Вуковог *Рјечника* у којима се грчке речи појављују као етимони или служе за појашњење значења), *Изјубљени у преводу* (речи које у Вуковом *Рјечнику* немају адекватан превод на латински и немачки), *Без цензуре* (опсцене речи као одреднице или примери у којима се оне јављају, а који се налазе у првом издању Вуковог *Рјечника*, док су из каснијих изостављени), *Врањанске пословице* (одреднице које садрже примере из пословица у *Речнику шовора јужне Србије*) и др.

Мапе. Називи места у којима су речи из речника на *Расковнику* посведочене повезани су, геолоцирани, на Гугл-мапи, а сваки речник обележен је посебном бојом (Слика 7). Кликом на пункт на мапи, појављује се прозор са именом места, а кликом на име обележеног

места отвара се нови прозор са списком свих речи у којима се име тог места јавља. На овај начин *Расковник* пружа још једну могућност „малих речника“ – према локацији. Као примере навешћемо Боку Которску, чији „мали речник“ у Вуковом *Рјечнику* садржи 673 речи и Вучитрн, из кога је 112 речи из *Речника косовско-мейохиског дијалекта*.

Након овог кратког представљања дигитализације у Институту за српски језик САНУ поставља се питање због чега је важна дигитализација старих речника? На првом месту, то је начин да се традиција учини живом, да речници не буду само предмети у библиотекама и архивима, на располагању малом броју стручњака и ентузијаста, већ да буду, путем интернета, доступни великим броју корисника у форми која је разумљива и новим генерацијама. Такође, она је важна за научна истраживања. Дигитализација речничког наслеђа, уз поштовање највиших академ-



Слика 7

ских и технолошких стандарда омогућује научницима различитих струка (лингвисти- ма, историчарима, етнолозима и другима) олакшани приступ извornoј грађи. Штавише, лексичка грађа на *Расковнику*, структурисана на описани начин, као и алати које користимо, пружају научницима могућност да уоче проблеме који линеарним читањем понекад остају замагљени, чак невидљиви. Затим, дигитализовани речници могу се користити и у образовне сврхе, од основношколског до постдипломског нивоа, а могу послужити и као приручници лекторатима за српски језик у иностранству. Због своје доступности, они су намењени и широј јавности – свима које занима лексичко богатство српског језика.

На крају, треба имати у виду и стратешки значај дигитализације речника, али и свеукупних ресурса српског језика. У параметре којима се мери статус једног језика, поред броја говорника и броја и квалитета производа на том језику уведен је и критеријум присутности у дигиталном окружењу. Данас, када се велики део комуникације одвија управо у том медију, они језици који не улажу у дигитализацију и језичке технологије најпре ће у тој комуникацији заостајати, а потом из ње и бити избрисани. Ниједна технологија не може заменити резултате квалитетног научног рада, што је основни циљ делатности Института за српски језик САНУ, али је и осавремењивање њиховог објављивања и начина израде посао који морамо урадити ми у Институту, јер нико други то неће учинити уместо нас.

ЛИТЕРАТУРА

1. Бојанић М., Р. Тривунац 2002: Речник дубровачког говора, Српски дијалектолошки зборник 49.
2. Вујићић М. 1995: Рјечник говора прошћења (код Мојковца), Посебна издања ЦАНУ књ. 29, Одјељење умјетности књ. 6, Подгорица.
3. Елезовић Г. 1932: Речник косовско-метохијског дијалекта, Српски дијалектолошки зборник 4.
4. Елезовић Г. 1935: Речник косовско-метохијског дијалекта, Српски дијалектолошки зборник 6.
5. Златановић М. 1998: Речник говора јужне Србије (провинцијализми, дијалектизми, варваризми и др.), Учитељски факултет, Врање.
6. Ивановић Н., М. Јакић, С. Ристић 2016: Грађа Речника САНУ – потребе и могућности дигитализације у светлу савремених приступа, Лексикологија и лексикографија у светлу савремених приступа, Институт за српски језик САНУ, 133–154.
7. Петровић С., Т. Тасовац 2016: Лексикографска анатација нестандардних облика у служби претраживости електронских речника, Лексикологија и лексикографија у светлу савремених приступа, Институт за српски језик САНУ, 155–168.
8. Стијовић Р., Р. Станковић 2018: Дигитално издање Речника САНУ: формални опис микроструктуре Речника САНУ, Научни састанак слависта у Вукове дане 47/1, 427–440.
- *
9. De Schryver, G-M., 2003: Lexicographer's Dreams in the Electronic-Dictionary Age, International Journal of Lexicography, 16/2, 143–199.
10. Tasovac, T., S. Petrović 2015: Multiple Access Paths for Digital Collections of Lexicographic Paper Slips. In: eLex 2015 – Electronic Lexicography in the 21st Century: Linking Lexical Data in the Digital Age. <https://elex.link/elex2015/conference-proceedings/paper-25/>

OLD DICTIONARIES AND NEW TECHNOLOGIES

Abstract

After a brief review of history of digitization at the Institute for the Serbian Language of SASA, this paper presents the Institute's dictionary platform *Raskovnik* and five dictionaries edited on this platform: *Srpski rječnik* by V. Karadžić, *Rečnik kosovsko-metohiskog dijalekta* by G. Elezović, *Rečnik govora južne Srbije* by M. Zlatanović, *Rječnik govora Prošćenja (kod Mojkovca)* by M. Vujičić and *Rječnik dubrovačkog govora* by M. Bojanović i R. Trivunac. The paper, in basic terms, shows how dictionaries have been digitized and marked and describes the main functions *Raskovnik* provides. It also points at the scientific and broader social significance of digitization of dictionaries and language resources.