

ISSN 0027–8084
УДК 808.61./62

ИНСТИТУТ ЗА СРПСКИ ЈЕЗИК САНУ

НАШ ЈЕЗИК

XXXVIII/3-4

БЕОГРАД
2017

УДК 811.163.4'374
930.25:004

РАДА Р. СТИЈОВИЋ*
(Институт за српски језик САНУ,
Београд)**

Оригинални научни рад
Примљен 11. децембра 2017. г.
Прихваћен 22. децембра 2017. г.

ГРАЂА ЗА РЕЧНИК САНУ – БЛАГО КОЈЕ ТРЕБА САЧУВАТИ (о дигитализацији листића)

Грађа за Речник САНУ прикупљана је више од 120 година, садржи преко 5.000 извора и око 300 збирки речи из народних говора. Забележена је, у највећој мери, на око 6.000.000 листића. Дигитализација грађе и њена организација у дигитални репозиторијум омогућиће да се ови у највећем броју врло трошни листићи сачувају, да се учине подесним за аутоматизацију рада на Речнику, што води знатном убрзању његове израде, као и да се грађа записана на њима учини доступном научној, стручној и широј културној јавности.

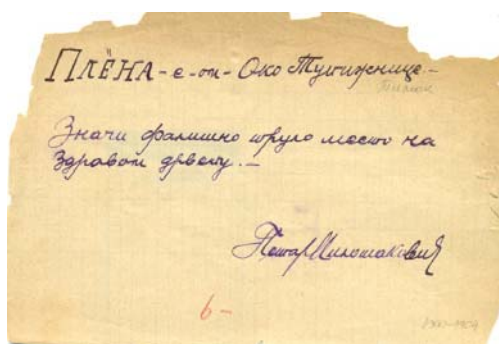
Кључне речи: Речник САНУ, грађа за Речник САНУ, листићи, дигитализација, анотација.

Речник српскохрватској књижевној и народној језика САНУ представља, како је више пута истицано, монументално дело српске и словенске лексикографије и несумњиво је једно од најзначајнијих дела савремене српске науке и културе. Предвиђа се да ће, када буде завршен, садржати око 500.000 речи, које припадају књижевном језику у току протекла два века (од Доситеја и Вука па до данас) и народним говорима са целокупног српског језичког простора. Свака од

* rada.stijovic@isj.sanu.ac.rs

** Овај рад је настао у оквиру пројекта *Линџивистичка испитивања савременој српској књижевној језика и израда Речника српскохрватској књижевној и народној језика САНУ* (бр. 178009), који финансира Министарство просвете, науке и технолошког развоја Републике Србије.

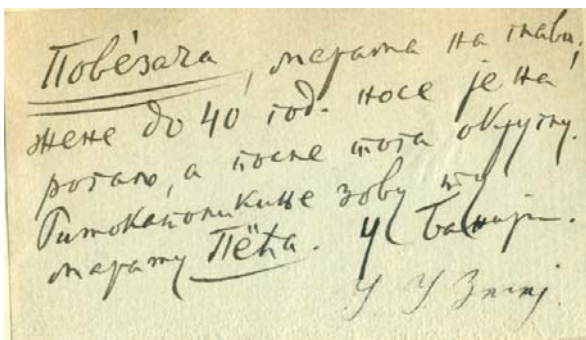
ових речи углавном је потврђена бројним примерима (изузимајући хапаксе, поједине покрајинске речи и сл.), који су забележени на листиће и, узбучени, смештени у одговарајуће кутије. Овако богат лексички материјал, који се прикупља већ више од 120 година, ексерпиран је из око 5.000 извора и око 300 рукописних збирки речи са целокупног српског језичког простора.¹ На њему су радила многа знеменита имена наше културне историје. Грађу на терену бележили су песник и лекар Јован Јовановић Змај (највише у Војводини, али и у другим српским крајевима), историчар, а тада „гимназист“ Владимир Ђоровић (у Херцеговини), проф. Велике школе и председник СКА Сима Лозанић (у Рипњу и Параћину), учитељица Марица Косановић, рођена Тесла (по Хрватској), сликар Новак Радонић (у Бачкој), књижевница Јелена Билбија (у Босни и Херцеговини) и многи други. Међу онима који су подвлачили речи у књигама из којих ће се ексерпирати грађа и онима који су те речи исписивали на картице били су: филолог Милан Решетар, филолог, историчар књижевности, дипломата Стојан Новаковић, књижевни критичар Јован Скерлић, књижевници Радоје Домановић, Јаша Продановић, Исидора Секулић и други.²



Из збирке свештеника Петра Милошаковића из тимчког краја (1898–1902)

¹ О грађи за Речник САНУ и збиркама речи прикупљеним за Речник – њиховој структури, ауторима, времену настанка и сл. писали су Н. Ивановић (в. Ивановић 2013: 151–180), затим Н. Ивановић, М. Јакић и С. Ристић (Ивановић и др. 2016: 141–142).

² О изворима за Речник, нарочито о листићима – њиховом изгледу, ауторима, начину ексерпирања примера забележених на њима и др. в. Павковић 2014.



Из збирке Ј. Ј. Змаја из разних српских крајева
(1900–1904)

Грађа за Речник повремено се допуњава и осавременује. Ексерпиреју се дела савремених писаца, консултују нови речници (књижевног језика и дијалеката), енциклопедије, уџбеници, часописи и прилажу нове збирке речи. Међутим, наш циљ у овом раду јесте да укажемо на трошност старе и у многоме оштећене грађе, њену ненадокнадиву вредност, као и на потребу да се она сачува, учини поделеном за аутоматизацију рада на Речнику и доступном научној, стручној и широј културној јавности.

Вредност грађе је вишеструка. Не само што на основу ње треба да буде написано још најмање 15 томова Речника, већ и зато што она пружа обиље других података. Она може бити основа за разна језичка истраживања у области фонетике, морфологије и синтаксе српског језика, на основу ње се може пратити развој језика током протекла два века, могу се сазнати бројне дијалекатске црте из времена када су збирке настале (већина примера једини су подаци о говору неког краја тога времена), могу се правити етимолошке студије итд. Грађа пружа податке и за ванјезичка истраживања – о материјалној и духовној култури српског народа током протекла два века, о његовим обичајима, веровањима, занимањима, начинима размишљања, етичким начелима, научној мисли и др. Могу се на основу ње изводити и разноврсна статистичка истраживања, формирати синонимски низови итд. И, што је веома важно, она може бити најшира полазна основа за формирање српског језичког корпуса. Ипак, за потребе Института, због којих је и прикупљана, најважније је сачувати је како би се написали преостали томови Речника и дати јој форму која ће омогућити убрзавање рада на Речнику.

С тим циљем се у Институту 2016. године почело прво са скенирањем целокупне грађе³, које се успешно реализује.⁴ Године 2017. приступило се процесу њене дигитализације.⁵ Имајући у виду то да је остало да се за Речник обради око 3.000,000 листића од слова П па до слова Ш, почело се прво са том грађом, а потом ће се наставити са грађом која је већ искоришћена у Речнику (у мери у којој је за Речник било потребно), али је драгоцену за друга истраживања.

Израдом дигиталног каталога листића са грађом за Речник САНУ, на системски, квалитетан и уједначен начин обезбедиће се организовање ове обимне грађе, као и њена лакша доступност и прегледност. Имплементацијом система за управљање дигиталним каталогом скенираних листића, њиховим обележавањем на одговарајући начин и похрањивањем у базу података биће обезбеђено лакше руковање листићима и прегледање грађе по обележеним подацима.

Информатички тим са Рударско-геолошког факултета у Београду осмислио је софтверско решење, у виду веб-апликације, према спецификацији и потребама корисника који су запослени у Институту за српски језик САНУ и који у свакодневном раду користе листиће, а које је израдио лингвистички тим из Института.

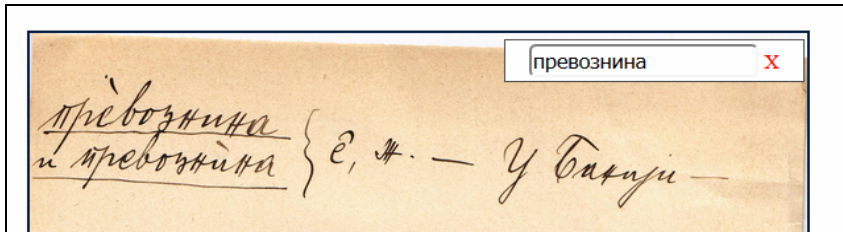
Апликација је заснована на комбинацији PHP и MySQL технологија и омогућује везу између корисника различитих профила, апликативног сервера, репозиторијума скенираних листића⁶ и базе података у коју се анотације похрањују. Овлашћени корисници могу јој приступити путем веб читача, анотирати листиће коришћењем форми апликације, чиме се допуњује база података. Постоје два сценарија анотирања. Прво је скраћена анотација, која подразумева обележавање одредница.

³ Треба рећи да је скенирање, односно каснија дигитализација целокупне речничке грађе, и оне која је употребљена за већ објављене томове речника, неопходна, јер се у Речник уноси само њен мали део – можда око 20%.

⁴ Драгоцену помоћ у овоме, нарочито у првим корацима, Институту је пружила Универзитетска библиотека из Београда.

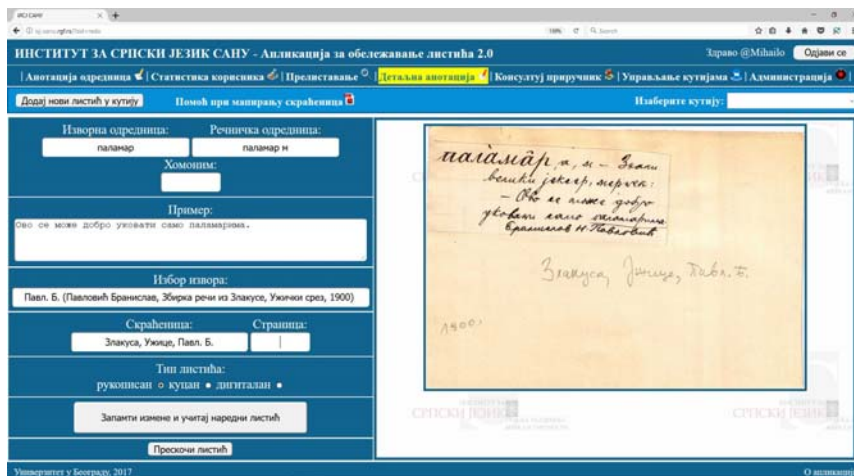
⁵ Овај посао је почет у оквиру пројекта Израда дигиталног каталога листића Речника САНУ, који је суфинансирало Министарство културе и информисања Републике Србије, а који ради информатичко-лингвистички тим на челу са Ранком Станковић и Душком Витасом, односно Радом Стијовић.

⁶ Сви скенирани листићи налазе се у облику *jpg* датотека организованих у каталоге који одговарају кутијама, односно секцијама листића. Методаци о сликама листића, подељене по каталозима према кутији у којој се налазе њихове физичке копије аутоматски су пресликане у базу података, чиме је значајно убрзана анотација листића.



Скраћена анотација: листић анотиран само одредницом

Друго је пуно анотирање, које омогућава унос речничке одреднице тј. одреднице у облику који ће имати у Речнику (што подразумева њен књижевни лик, без обзира на то у ком је лику забележен у извору – дијалекатском, некњижевном или сл.) и одреднице у изворном лику (забележеном у примеру на листићу), затим ознаке за хомоним, текста примера, извора из кога је пример узет са скраћеницом под којом се доноси у Речнику, типа листића (рукопис, куцан текст) и додатне напомене. Различити статистички прикази помажу у праћењу процеса анотације листића.



Пуна анотација

Исти информатичко-лингвистички тим отпочео је 2016. године припрему за дигитализовану верзију РСАНУ. Засада су креиране процедуре за анализу дигитализованог текста 1. и 19. тома, сегментирани речнички чланци и информационе целине у оквиру речничког

чланка.⁷ Дигитализовањем речника, картица са примерима и библиотечких извора, као и осавремењавањем начина израде речника знатно ће се убрзати његова израда.

ЛИТЕРАТУРА И ИЗВОРИ

- Ивановић 2013: Ненад Ивановић, *Речник САНУ и његова улога у лексичкој стандардизацији српског језика (са историјској и лексикографској аспектима)*, необјављен докторски рад, одбрањен на Филолошком факултету Универзитета у Београду.
- Ивановић и др. 2016: Ненад Ивановић, Милена Јакић, Стана Ристић, Грађа Речника САНУ – потребе и могућности дигитализације у светлу савремених приступа, у: Стана Ристић, Ивана Лазић Коњик, Ненад Ивановић (ур.), *Лексикологија и лексикографија у светлу савремених теорија*, Београд: Институт за српски језик САНУ.
- Павковић 2014: Васа Павковић, Грађа за Речник САНУ, у: Рајна Драгићевић (ур.), *Савремена српска лексикографија у теорији и пракси*, Београд: Филолошки факултет, 55–59.
- РСАНУ: *Речник српкохрватској књижевној и народној језика САНУ*, I–XIX, Београд: Српска академија наука и уметности, Институт за српски језик САНУ, 1959–.
- СТИЈОВИЋ–СТАНКОВИЋ 2018: Рада Стијовић, Ранка Станковић, Дигитално издање Речника САНУ: формални опис микроструктуре Речника САНУ, *Научни саставак славистија у Вукове дане 47/1*, Београд, у штампи.

Рада Р. Стијовић

МАТЕРИАЛ ДЛЈА СЛОВАРЈА САНИ – СОКРОВИЦЕ КОТОРОЕ СЛЕДУЕТ БЕРЕЧЬ

(к вопросу о дигитализации листиков)

Резюме

Процессом сканирования листиков с материалом для Словаря САНИ и их перевода в электронную форму, начатым в Институте сербского языка САНИ, сохранится ценный материал, а в то же вре-

⁷ Више о томе у Стијовић–Станковић 2018.

мя, составлением дигитального каталога листиков обеспечится систематичная, качественная и унифицированная организация этого объемистого материала, а также его более легкая доступность и наглядность. Современными методами и технологиями, используемыми на мировой лексикографической практике, работа над Словарем САНИ будет усовершенствована и ускорена.

Ключевые слова: Словарь САНИ, материал для Словаря САНИ, дигитализация листиков, аннотация.