

ЛЕКСИКОЛОГИЈА И ЛЕКСИКОГРАФИЈА
У СВЕТЛУ САВРЕМЕНИХ ПРИСТУПА

Зборник научних радова

SERBIAN LANGUAGE INSTITUTE OF SASA

LEXICOLOGY AND LEXICOGRAPHY
IN THE LIGHT OF CONTEMPORARY
APPROACHES

A collection of papers

Edited by:

Stana Ristić, PhD, scientific advisor

Ivana Lazić Konjik, PhD, senior research associate

Nenad Ivanović, PhD, research associate

Belgrade

2016

ISBN 978-86-82873-56-3

ИНСТИТУТ ЗА СРПСКИ ЈЕЗИК САНУ

ЛЕКСИКОЛОГИЈА И ЛЕКСИКОГРАФИЈА У СВЕТЛУ САВРЕМЕНИХ ПРИСТУПА

Зборник научних радова

Уредништво:

др Стана Ристић, научни саветник
др Ивана Лазић Коњик, виши научни сарадник
др Ненад Ивановић, научни сарадник

Београд
2016

Издавач:
Институт за српски језик САНУ

Рецензенти:
проф. др Андреја Желе
проф. др Вера Васић
проф. др Даринка Гортан Премк
др Људмила Ришкова

Коректура: аутори и Уредништво

Припрема за штампу:
Давор Палчић

Штампа:
Чигоја штампа

Тираж: 300

Издавање ове књиге финансијски је помогло Министарство просвете,
науке и технолошког развоја Републике Србије

НАПОМЕНА О АУТОРСКОМ ПРАВУ:

Ниједан део ове публикација не може се прештамповати, репродуковати или употребити у било
ком облику без писменог одобрења аутора као носиоца ауторског права.

COPYRIGHT NOTICE:

No part of this publication may be reprinted, reproduced or utilized in any form without permission in
writing from the author, as the holder of the copyright.

Ненад Б. ИВАНОВИЋ, Милена М. ЈАКИЋ, Стана С. РИСТИЋ
Институт за српски језик САНУ, Београд
nenad.ivanovic@isj.sanu.ac.rs, milena.jakic@isj.sanu.ac.rs,
stana.ristic@isj.sanu.ac.rs

ГРАЂА РЕЧНИКА САНУ – ПОТРЕБЕ И МОГУЋНОСТИ ДИГИТАЛИЗАЦИЈЕ У СВЕТЛУ САВРЕМЕНИХ ПРИСТУПА*

У раду се, у три засебна дела, даје опис структуре грађе Речника САНУ и процеса његове израде у светлу планирања дигитализације и аутоматизације рада на овом речнику. У првом делу говори се о месту Речника САНУ у европској академијској лексикографији, како у смислу његовог обима и темпа израде, тако и у смислу репрезентативности његове грађе за опис савременог српског књижевног језика. Други део рада говори о потребама и оптималним начинима дигитализације различитих типова грађе за Речник САНУ, имајући у виду чињеницу да грађа тренутно постоји у папирном формату. У трећем делу говори се о елементима корисничког окружења за израду Речника САНУ, али и о другим могућностима експлоатације дигитализоване грађе.

Кључне речи: грађа Речника САНУ, метаанализа речника, дигитализација лексикографских ресурса, модернизација и унапређење израде Речника САНУ.

* Рад је настао у оквиру пројекта 178009 *Лингвистичка исцртавања савременог српског књижевног језика и израда Речника српскохрватског књижевног и народног језика САНУ*, који у целини финансира Министарство просвете, науке и технолошког развоја Републике Србије.

0. Увод

Циљ овог рада јесте да представи потребе и могућности дигитализације¹ грађе Речника САНУ из перспективе рачунарски оријентисаних приступа модернизацији великих речничких пројеката у савременом добу. Узета су у обзир сазнања из области дигитализације описних речника и примене рачунарских технологија у лексикографији (Meijs 1996; Schryver 2003; Weiner 2009; Abel 2012; Витас и Крстев 2015 и др.), као и досадашњи резултати остварени при успостављању теоријског оквира за дигитализацију и модернизацију Речника САНУ (Сабо и Витас 1988, Ристић и Ивановић 2011, Ристић и др. 2012). На основу тога аутори рада ће настојати да укажу на актуелне културне и научне потребе за осавремењивањем израде овог речника и побољшањем квалитета његове израде (па и убрзањем динамике излажења), уз истовремено очување његове структуре и методологије.² Један од резултата процеса дигитализације представља и повећање доступности грађе, као и самог речничког текста. Поред тога постоје и друге предности рада са дигитализованим речничким материјалом (лакша претраживост, могућност експлицирања и анализе информација које често постоје у виду имплицитних лексикографских знања и сл.). Крајњи циљ, са становишта лексикографа, одређен је као визија новог окружења за израду Речника САНУ, које, уз употребу рачунарских технологија, испуњава две врсте захтева: (1) обезбеђивање доступности различитих типова лексикографски релевантних информација и (2) укључивање тих информација у процес израде самог речника.

¹ У литератури термин *дигитализација* користи се вишезначно, и са неодређеним границама појма. У овом раду под термином *дигитализација* подразумева се обезбеђивање електронске верзије документа (књиге, рукописи и сл.) идентичне са папирном. За кораке који подразумевају даљи рад са електронском верзијом документа (изградња базе података, анотација и сл.) требало би увести други термин.

² Сматрамо да се Речник САНУ, заснован на грађи описаној у Ристић и др. 2012, као и у овом раду, с обзиром на његову микро- и макроструктуру, не може заменити другим типовима речника (нпр. Wiki), како се предлаже у радовима заговорника таквог приступа (Meuер and Gurevuch 2012; Витас и Крстев 2015). Без претензије на информатичко знање, у раду се са лингвистичког, а пре свега са лексикографског аспекта, представљају потребе и могућности дигитализације речничке грађе, при чему се подразумева заједнички рад са информатичким стручњацима, од којих се очекује организовање елемената радног окружења лексикографа у циљу унапређења и убрзања рада на Речнику САНУ.

1. Речник САНУ и његово место у европској тезаурусној лексикографији

1.1. *Речник САНУ као комплексно лексикографско дело.* Једна од чешће цитираних, општих дефиниција речника јесте да је то „систематски уређен списак подруштвљених језичких облика скупљених из говорних навика одређене говорне заједнице и прокоментарисаних од аутора на тај начин да квалификован читалац разуме значење ... сваког појединог облика, и да је обавештен о релевантним чињеницама о функцији тог облика у својој заједници” (Berg, у: Згуста 1991: 189). На основу различитих критеријума у класификацији типова речника (Згуста 1991: 189–210), Речник САНУ спада у *једнојезичне речнике дескриптивног* типа, а по обиму обрађене лексике и броју информација о лексемама спада у *тезаурусе*, односно *академијске речнике*.³ Грађа за овај речник обухвата изворе од 80-тих година 18. века па све до савременог доба, односно дуги (двовековни) период у развоју новоустановљеног српског књижевног и народног језика. Зато он представља речник савременог језика са дијахроним карактеристикама. Начин представљања лексичког материјала у Речнику САНУ омогућава синхроне пресеке појединих раздобља у развоју савременог српског језика, на основу којих се може утврђивати степен њихове стабилности у развоју, и то како српског књижевног тако и народног језика.⁴ На основу наведених општих карактеристика, Речник САНУ није лако дефинисати према строжим класификацијама, јер су параметри, који су обично валидни за типологију речника, мање или више подложни променама током његове израде. У питању су следећи параметри: профилисање циљева и циљне групе корисника речника, методологија израде, временски, територијални и функционалностилски обим грађе, теоријске поставке и др. Динамика њихове промене посебно долази до изражаја при дуготрајној изради овог типа речника, што представља универзални проблем у академијској лексикографији.⁵

³ Термин *тезаурус* у англистичкој лексикографији означава речник синонима, а у словенској велике академијске речнике. Овде се мисли на славистичко значење термина (Згуста 1991: 199, у фусноти 43).

⁴ Ову чињеницу приметили су и други истраживачи, нпр. М. Р. Тешић која наводи слично запажање: „Због дуге временске димензије израде, овај синхрониски заснован речник [РСАНУ] ... нужно добија и битна својства историјски информативног речника, што му свакако само по себи не мора бити мана него, у неку руку, и квалитет више.” (2009: 30).

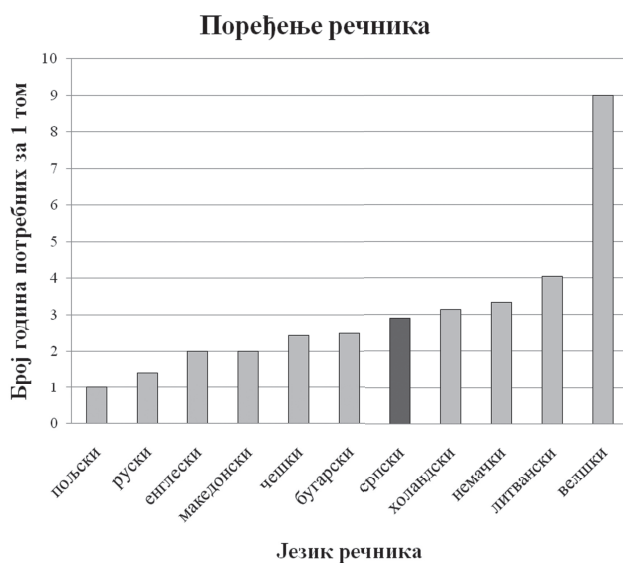
⁵ Тешкоће у типолошком одређењу академијских речника наведене су и у приручнику Ј. Згусте, у коме се као критеријум за класификацију, поред обима и структуре, уводи начин израде ових речника и њихов однос према језичком стандарду у различитим етапама развика (1991: 208).

1.2. *Темпо израде Речника САНУ (у односу на друге академијске речнике у Европи).* Прикупљање грађе за Речник САНУ отпочело је 1850. године, а његов први том објављен је 1959. године. Године 2014. објављен је 19. том, а тренутно су у изради 20, 21. и 22. том у различитим фазама израде. Један том Речника САНУ има 800 страница двостручног текста, у просеку са око 11.500 одредница и излази у просеку на 3 године. У изради сваке одреднице Речника у просеку учествује између 5 и 8 сарадника.

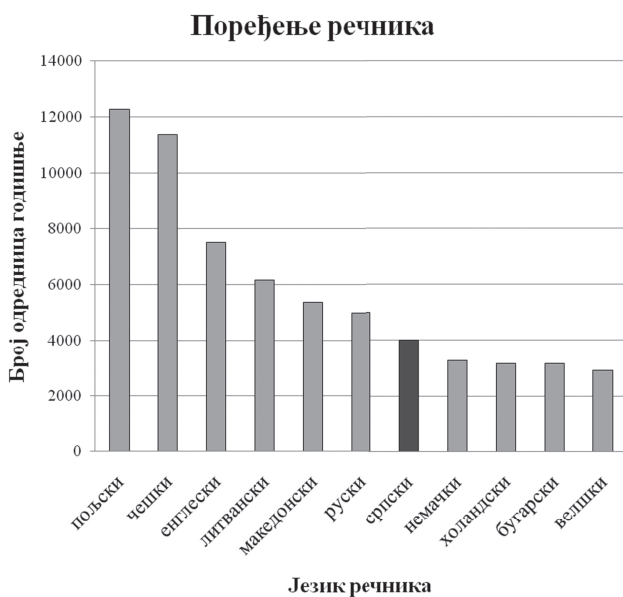
У табели 1 дати су подаци о броју одредница и томова, као и број протеклих година од издавања првог до издавања последњег објављеног тома првог издања тезаурусних (академијских) речника неких европских језика. Подаци о Речнику САНУ истакнути су курсивом, а редослед у табели дат је према азбучном реду назива језика:

Језик	Назив речника	Број одредница (у хиљадама)	Број томова	Број година
бугарски	Речник на българския език	111	14	35
велшки	Geiriadur Prifysgol Cymru	105	4	36
енглески	Oxford English Dictionary	301	20	40
литвански	Didysis lietuvių kalbos žodynas or Akademinis lietuvių kalbos žodynas	500	20	81
македонски	Толковен речник на македонскиот јазик	64.5	6	12
немачки	Deutsches Wörterbuch	350	32	107
пољски	Słownik języka polskiego	135	11	11
руски	Словарь современного русского литературного языка (БАС)	120	17	24
<i>српски (српскохрватски)</i>	<i>Речник српскохрватског књижевног и народног језика САНУ</i>	<i>220</i>	<i>19</i>	<i>55</i>
холандски	Woordenboek der Nederlandse Taal	430	43	135
чешки	Příruční slovník jazyka českého	250	9	22

Табела 1. Преглед броја одредница, томова и година протеклих од издавања првог до издавања последњег публикованог (не обавезно и последњег планираног) тома великих тезаурусних (академијских) речника за европске језике.



Слика 1. Поређење тезаурусних, академијских речника за 11 европских језика према просечном броју година потребних за издавање једног тома.



Слика 2. Поређење тезаурусних, академијских речника за 11 европских језика према просечном броју одредница које се обраде током једне године.

Када се анализирају подаци о темпу израде наведених речника, долази се до закључка да је корелација броја издатих томова и броја година потребних за њихово публикување за велике речнике академијског типа 11 европских језика изразито висока: $R=0,91$, $p<0,0001$.⁶ Такође, корелација броја одредница и броја година потребних за њихово публикување износи $R=0,80$ $p<0,01$. То значи да се Речник САНУ израђује приближно истим темпом којим се израђују и други тезаурусни речници академијског типа у Европи.⁷ Графички приказ поређења просечног броја година потребних за издавање једног тома различитих речника дат је на слици 1, док је просечни број одредница израђених током једне године у различитим речницима приказан на слици 2.

1.3. *Комплексноћ фазне израде Речника САНУ*. Израда Речника САНУ представља један од најзначајнијих и најобимнијих пројеката Института за српски језик САНУ. У изради овог речника тренутно учествује 40 сарадника, а од 1893. године⁸ до данас у његовој припреми и изради суделовало је преко 120 лексикографа.⁹

Речник САНУ осмишљен је тако да представља целокупну лексику савременог српског (српскохрватског) језика у периоду од два века њеног лексичког, семантичког, творбеног, стилистичког и др. развитка.¹⁰ Ово

⁶ *Корелација* је поступак којим се утврђује да ли између варирања две варијабле постоји веза. Уколико две варијабле варирају на сродан начин, тј. ако су промене вредности једне варијабле праћене променама вредности друге, онда тврдимо да те варијабле корелирају, тј. да су повезане. Корелација се изражава коефицијентом корелације (R) који може имати вредности од -1 до 1 . *Сигнафичност*, тј. вредност статистика „ p ”, указује на то колика је могућност да се првобитни резултат добије у поновљеном истраживању, са другим подацима.

⁷ Додатна статистичка анализа (регресиона анализа, чији се коефицијент добија квадрирањем вредности корелационог коефицијента) указује да на темпо објављивања утичу и други фактори осим броја одредница, односно да број одредница објашњава свега 64% варијансе брзине објављивања речника. Остали фактори (36%) могу бити: број сарадника у колективу, ефикасност обуке нових сарадника, правилна расподела посла, експлицираност правила за израду речника, структура лексикографског чланка итд., што би могло да буде предмет посебног истраживања. Све наведене факторе (не само дигитализацију грађе) требало би узимати у обзир када се говори о убрзању израде речника.

⁸ Године 1893, 5. априла, Стојан Новаковић је изнео *Предлог Српској краљевској академији да се ошћочне кућњење грађе за академијски Српски Речник, и да се за њај ѡсоа усћанови у Академији Лексикографски одсек*. Стога се ова година обично узима за годину у којој започиње израда Речника САНУ.

⁹ Податак је добијен анализом списка извора за Речник САНУ, који је објављен у његовом последњем, 19. тому (РСАНУ XIX: VIII–CXIII).

¹⁰ Овим својим карактеристикама Речник САНУ показује актуелност, која се огледа и у чињеници да се његова горња хронолошка граница стално помера према савременом стању српског језика.

представљање лексике одвија се кроз лексикографски опис *књижевног* и *народног* језика, као и кроз одређење њиховог међусобног односа у речничком чланку. Речнички опис језика врши се у складу са утврђеним правилима анализе, дефинисања и опримеравања речи у речнику, која чине део традиције савремене српске лексикографије.

Израда Речника САНУ може се описати кроз следеће основне кораке, које наводи и Л. Згуста (1991: 212): (1) *сакуиљање грађе*, (2) *одабир и обликовање фонда одредница*, (3) *конституисање лексикографског чланка* и (4) *уређење одредница према одабраном критеријуму*.

(1) При *сакуиљању грађе* пошло се од књижевнојезичке праксе Доситеја Обрадовића за којом је следила свеобухватна језичка реформа (правописна, граматичка, лексичка) Вука Стеф. Карацића. Другим речима, доња граница сакупљања грађе подразумева период у историји развоја српског језика који је поставио темељ савременој књижевнојезичкој норми. Поред историјских принципа, лексичка грађа за Речник САНУ ексцерпирана је и према жанровским, културним, стилистичким и др. принципима, те се тако за изворе књижевног језика узимају писци који показују оригиналност, самостално владање језиком и поседују језичку и стилску компетенцију за унапређење језичке заједнице (Ивановић 2015: 33–34).¹¹ Са друге стране, када су у питању народни говори, ови критеријуми подразумевају избор речи и њихових значења, диференцијалних у односу на књижевни језик.

(2) За овим кораком следе *одабир и обликовање одредница*. Како истиче М. Николић (2014), одабир и обликовање одредница подразумевају стално промишљање принципа скупљања и селекције језичке грађе кроз њено представљање и обраду у оквирима претходно задатих граматичких и семантичких категорија. Саставни део поманутог процеса чине бројна отворена питања у обради, попут нпр. питања навођења синтагматских одредница, проблеми хомонимије, конверзије и лексикализације, питања (не)одређеног вида придева, лексикализације граматичких облика лексема (нпр. у примерима типа *класало жићо*, *надошла река* и сл.), проблеми оправданости реконструкције инфинитива и сл. као и многи други који се тичу одабира конкретних одредница и њихових значења која ће ући у речник (Николић 2014: 62–67).

¹¹ Важно је нагласити да примери из дела ових писаца нису ексцерпирани насумично, већ према одређеним принципима који су имали за циљ да укажу на језичке специфичности њиховог израза. Ови принципи обухватали су синтаксичке, семантичке и стилске критеријуме (према: Белић 1926: 116).

(3) Правилима је прописана и већина лексикографских поступака при *конструисању лексикографског чланка*, као и типова информација који се дају уз сваку обрађену лексему у Речнику, као нпр.: степеновање одредница, акцентовање, одређење правописних, екавских/ијекавских и др. варијанти, граматичка квалификација, етимологија, упућивачка квалификација, стилска и термилошка квалификација, начин дефинисања значења, израза и фигуративних значења, унакрсно упућивање значења, опримеравање, начин идентификације коришћених извора и др. (в. Упутства РСАНУ 1959 [2007]).

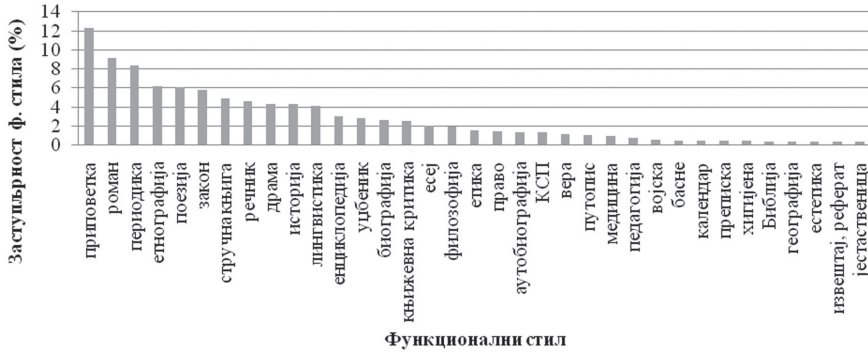
(4) Једном обрађене одреднице *уређују се* према унапред утврђеном критеријуму (у случају Речника САНУ азбучним редом, а у оквиру хомографских одредница редоследом према типовима и месту акцената), уз системско уједначавање, унакрсно упућивање и испоређивање обрађене лексике у вишим фазама речничке израде.

Из изложеног се види да се улога лексикографа у процесу израде Речника САНУ састоји у трансформацији грађе организоване према нормативним, лингвистичким и културним принципима у структурирани низ одредница на основу правила која се примењују и коригују кроз фазну израду речничког текста. Коначан производ изложених корака и фаза обраде представља опис лексема као јединица у укупности својих облика и значења, као и опис њиховог односа према стандарднојезичкој норми. Лексичке јединице описују се као део апстрактног језичког система (према терминологији де Сосира: *langue*), али и као део говорне праксе (*parole*), што се у првом случају остварује системским повезивањем одредница на парадигматској равни, а у другом формулисањем правила спојивости на синтагматској равни и дефинисањем инваријантних лексичких реализација на основу употребе у контексту (Ristić, Ivanović, *in press*).

1.4. *Састав грађе за израду Речника САНУ*

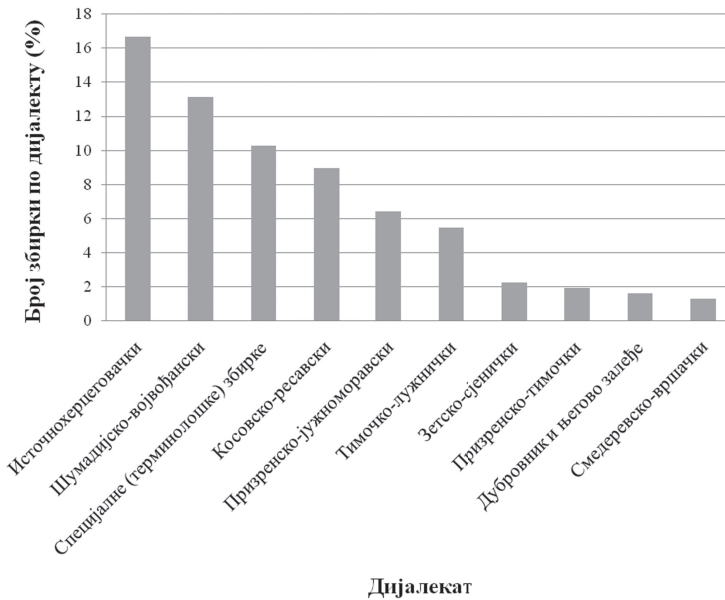
1.4.1. *Састав грађе ексерциране из писаног језика*. Уколико се анализира састав грађе сачињене на основу писаног језика према функционалностилским карактеристикама, запажа се да велику већину чине прозна дела (пре свега приповетке и романи), потом периодика (сачињена од дневних и недељних листова), етнографски списи, поезија, законски списи, речници, приручници, драмска и историјска дела итд. (слика 3).

Писани извори



Слика 3. Састав корпуса писаног језика који се користи при изради Речника САНУ. Таг „КСП” означава дела која се у грађи квалификују као „култура, социологија, политика” (према: Белић 1926).

Збирке речи по дијалектима



Слика 4. Преглед заступљености дијалекатских ареала у збиркама речи које представљају грађу из српског народног језика. Таг „Дубровник и његово залеђе” обухвата збирке које припадају штокавском ијекавском наречју са специфичним одликама говора дубровачког подручја (према: Тривунац 2002: 9).

1.4.2. *Састав грађе из народног језика*. Уколико се анализира састав грађе из народног језика у Речнику САНУ, уочава се да број збирки речи варира у односу на тип дијалекта. Највећи број збирки речи припада источнохерцеговачким говорима, потом шумадијско-војвођанским, призренско-тимочким итд. (в. слику 4).¹²

Изложено указује на то да грађа за Речник САНУ није скупљана насумично, већ у сагласности са моделом књижевног језика у његовој теоријској основи. Овај модел први пут је експлицитно дефинисан у спису *Српски речник књижевног и народног језика* А. Белића из 1926. године (Белић 1926: 30–112; Ивановић 2015), мада је имплицитно присутан у српској академијској лексикографији од оснивања Речника САНУ. Модел савременог језика у основи Речника САНУ сачињавају *савремени књижевни језик* (језик као објекат културе српског народа) и *савремени народни језик* (језик као систем појмова и односа који карактеришу народни живот и обичаје у различитим дијалекатским ареалима). Гледано са аспекта савремености, у питању је *живи језик*, језик који његови савременици виде и чују и којим владају као језичке личности (исп. Ристић 2012: 11). Док изворе *књижевног језика* у Речнику САНУ чине књиге и часописи штампани на штокавском говорном подручју од почетка 19. до краја 20. века,¹³ изворе *народног језика* чине збирке речи, које се од половине 19. до краја 20. века прикупљају по разним дијалекатским ареалима целокупног штокавског подручја. Збирке речи у грађи за Речник САНУ потичу из различитих периода употребе српског језика (најстарија збирка потиче из 1848. године), те представљају велику културну, лингвистичку и уопште етнографску вредност.¹⁴

¹² Додуше, уколико упоредимо број дијалекатских зона и број лексема који је у њима прикупљен, примећујемо да нису све збирке речи сакупљане према истом принципу. Тако на пример у оквиру зетско-сјеничког ареала постоји свега 7 збирки речи, али је број одредница који је у њима прикупљен несразмерно већи у односу на све остале збирке речи по другим дијалектима. Корелација збирки речи и логаритмованог броја одредница које су у њима прикупљене износи $R=0,73$, $p<0,05$, а тачка која на графикону означава зетско-сјенички дијалекат у великој мери одскаче са корелационе линије.

¹³ О хронолошком прегледу књижевних извора за Речник САНУ у контексту динамике језичке стандардизације и промене доминантних центара у развоју српског језика у 19. и 20. веку в. у Ристић и др. 2012: 84.

¹⁴ Тако нпр. збирка Андрије Јовићевића (учитеља из Црне Горе), настала у последњим десенијама 19. века, у 18.599 забележених речи (са примерима употребе) фиксира различите аспекте дијалекатске употребе српског језика у ријечком крају Црне Горе (ослањајући се нарочито на фонетске, морфолошке и творбене особине дијалектизама). Са друге стране, збирка од 6.684 речи, сликара Новака Радонића из

1.4.3. *Подела речничке грађе на њишове*. Грађу за израду Речника САНУ чине три основна дела: лексикографски листићи, библиотека Института за српски језик САНУ и објављени томови Речника САНУ.

Први њиш извора чине лексикографски листићи који садрже речи са примерима употребе (контекстуалним реализацијама) и другим подацима (о аутору, делу и страни, дијалекатском ареалу из кога је реч потекла и сл.). Листићи имају различиту гранулацију, квалитет, дебљину, позадину, боју и величину, а процењује се да њихов укупни број износи око 6.000.000. Распоређени су по секцијама, од којих свака секција обухвата одређен, приближно једнак број узбучених речи. Број до сада обрађених секција износи 642, док број необрађених тренутно износи нешто више од 500. У просеку, једна секција грађе заузима око 24 странице у објављеном речнику, те се на основу овога може проценити да ће Речник САНУ имати још нешто више од 15 томова (након до сада објављених 19).

Други њиш извора чине публикације из којих је ексцерпирана грађа на листићима (књиге, часописи, енциклопедије, стручна дела и др.). Ове публикације штампане су на штокавском говорном подручју од почетка 19. до краја 20. века, и броје преко 3.000.000 страница текста. У изворе грађе убраја се и дијалекатски материјал (збирке речи и речници) сакупљан и израђиван од половине 19. до краја 20. века, као и скуп посебних збирки и речника (ономастичких и термилолошких назива, страних речи, жаргонских речи, скраћеница и др.).

Трећи њиш основних извора за Речник САНУ чине његови објављени томови (19 томова, 14.961 страна, од *A* до *йејогласник*). Поред тога што представља извор информација о значењима речи српског језика, као и о култури и обичајима српског народа, овај речник у дигиталном облику представљао би полазну основу за различите врсте напредних претрага у системским истраживањима и метаанализи лексикографског текста из наведених области (као и из граматике, семантике, синтаксе, фразеологије, психоллингвистике, етнолингвистике, теоллингвистике, лексикологије, теоријске лексикографије и др.).¹⁵ Таква

Бачке, указује на лексичко-фолклорне елементе српског буњевачког говора друге половине 19. века. Међу збирке речи од велике културне и научне вредности сврстава се и збирка Владана Арсенијевића из 1900. године. Ова збирка броји преко 250.000 лексичких уноса организованих према тематским целинама (делови тела, народна ношња, обичаји и др.).

¹⁵ Значај метаанализе за системска истраживања у практичној и теоријској лексикографији истицан је и у досадашњим радовима из ових области, а као пример најновијих резултата в. радове у колективној монографији *Савремена српска лексикографија у теорији и пракси* (Драгићевић 2014).

истраживања унапредила би и рад на што потпунијем скупу правила и упутстава за обраду различитих лексикографских типова.

2. Одабир начина дигитализације грађе за Речник САНУ

При одабиру начина дигитализације грађе за Речник САНУ треба имати у виду да су очекивања која, лексикографи имају, врло широка и, у исто време, врло специфична. Према њиховим захтевима, успешност дигитализације грађе за Речник САНУ мерљива је њеном искористљивошћу при изради речничког текста, при чему се мора узети у обзир што више аспеката употребе те грађе у процесу састављања речничког (лексикографског) чланка.

У могуће начине дигитализације речничке грађе спадају скенирање (са или без оптичког препознавања карактера) и дактилографско прекуцавање. Док прекуцавање обезбеђује потпуну претраживост грађе, скенирани материјал се може подвргнути процесу оптичког препознавања карактера (Optical character recognition или скраћено OCR) или посебном типу аотације, како би се омогућила барем делимична претраживост материјала у бази.¹⁶

2.1. *Оптимальан начин дигитализације лексикографских листића.* Лексикографски листићи представљају примарну грађу за израду Речника САНУ, па је при изради речника неопходна идентичност њихове дигиталне и физичке верзије. Потпуна претраживост материјала обезбедила би се дактилографским прекуцавањем (оптичко препознавање карактера у случају десетина различитих рукописа није изводљиво), а потом би требало омогућити даљу аотацију и напредну претрагу, како у домену систематског приписивања обавезних података примерима (стандардна скраћеница извора и година издања), тако и у домену морфолошког, синтаксичког и др. аотирања.

¹⁶ Тачност препознатог материјала никада не достиже ниво који се постиже дактилографским прекуцавањем. Процент варира зависно од степена очуваности књиге, типа графије, квалитета папира и сл., те се у случају лексикографских потреба при изради Речника САНУ оптичко препознавање карактера може користити искључиво као помоћна техника или се мора комбиновати са коректуром према штампаном издању публикација. Наравно, треба утврдити који од начина ((1) дактилографско прекуцавање, (2) аотација фотографија/скенова према пагинацији оригиналног издања из којег је вршена ексцерпција или (3) оптичко препознавање и коректура) даје најбоље резултате у погледу квалитета (тј. тачности и претраживости) материјала, временских и финансијских улагања, па према томе изабрати оптимални начин дигитализације за сваки од типова грађе.

Графичке карактеристике листића (начин исписивања, тип рукописа, боја мастила и сл.) у неким случајевима представљају значајан податак при обради (тип извора, период ексцерпције и сл.). Стога би уз прекуцан листић било потребно обезбедити и скенирану верзију.¹⁷ Скениране листиће требало би анотирати основним системом ознака (одредница, циљна реч, извор, пагинација) и повезати их са прекуцаном верзијом.

Када говоримо о **архивирању** овако вредне грађе, општеприхваћени начин архивирања представља микрофилмовање материјала, те би листиће у идеалном случају требало и микрофилмовати.

Дакле, лексикографски листићи, као основни део грађе за израду Речника САНУ у идеалном случају дигитализовали би се на три начина, како би се задовољиле три различите потребе: (1) унапређење и повећање доступности грађе за лексикографски и истраживачки рад, (2) стварање драгоцене базе која би омогућила напредне претраге, чиме би се стекли многи нови увиди и (3) презервација и архивирање културног блага.

2.2. Оптималан начин дигитализације библиотечких извора. При дигитализацији библиотечких извора за израду Речника САНУ требало би имати у виду њихову структуру, функцију, вредност, физичко стање и формат. Од обавезних речника и приручника, који се користе као допуна грађе при изради лексикографског текста, требало би сачинити јединствену базу података, којом би се омогућила лака и брза претрага током израде текста речника. Ту спадају енциклопедије, речници, дијалекатски речници и др. При изради ове базе требало би тежити што већој тачности дигитализованог материјала, односно што мањем проценту текстуалних грешака.¹⁸ Остатак библиотечких извора могао би се скенирати, потом и анотирати према оригиналној пагинацији издања из којих је ексцерпција вршена, јер искуство показује да оптичко препознавање не доноси потпуну идентичност са штампаним

¹⁷ Овде се пре свега мисли на случајеве када на листићима не постоји податак о извору, већ се извор идентификује помоћу графичких карактеристика, као што су изглед налепљеног исечка из речника, тип рукописа, боја мастила (нпр. збирка речи Владана Арсенијевића), испис писаћом машином и сл. Такође, слике листића биле би корисне и у случајевима када је рукопис нечитак готово свима, осим увежбаном оку искусних лексикографа.

¹⁸ Један од препоручених начина представља дактилографско прекуцавање, а други скенирање које би се обавезно комбиновало са оптичким препознавањем и коректуром текста. Треба утврдити који од ова два начина даје тачнији резултат и већу финансијску искористљивост. Тренутно се чини да је то први начин, барем када су у питању обавезни приручници који се користе као допуна грађе Речника САНУ.

материјалом. На тај начин, лексикографима би се олакшала и убрзала провера конкорданција на листићима.

2.3. *Оптимальан начин дигитализације објављених томова Речника САНУ*. Од објављених томова Речника САНУ у дигиталном облику очекује се да омогуће различите врсте напредних претрага, које обезбеђују увиде у методолошке поступке и решења која су корисна за даљи рад на Речнику и лексикографску обуку. Стога је, уз постизање идентичности дигиталног и штампаног материјала, неопходно експлицитно формализовати речнички текст према типовима информација у речничком чланку (нпр. информације које су у штампаној верзији речника означене графички: заглавље одреднице, тип и место акцента, морфолошке информације, квалификатори, издвајање значења, али и информације које нису ни на који начин експлицитне: типови дефиниције итд.). Ова формализација постиже се аотирањем сваког од типова информација посебним тагом (нпр. помоћу XML¹⁹ система ознака са значењем: „одредница”, „акцент”, „квалификатор”, „пример”, „извор” и сл.), као и утврђивањем њихове хијерархије и односа припадности унутар лексикографског чланка.

3. Дигитализација грађе за Речник САНУ у светлу савремених приступа

3.1. У првом делу рада дат је опис функције и намене Речника САНУ и његове улоге у српској науци и култури. Поред састава језичког корпуса на коме је овај речник заснован, као и израде речничког текста по фазама, указано је и на адекватне начине дигитализације грађе за овај речник у складу са потребама његове даље израде. Имајући ово у виду, у делу који следи рећи ћемо нешто више о могућој примени тако припремљене грађе у састављању и издавању Речника САНУ, али и о другим предностима коришћења дигитализоване грађе, нпр. у сфери презентације материјалне културе и сл.

У планирању процеса дигитализације речничке грађе пошло се од идеје о унапређењу постојећег лексикографског рада, који се спроводи од стране обученог и специјализованог колектива, уз помоћ рачунарских технологија. То значи да се интегрисање рачунарских технологија у сам процес израде Речника САНУ посматра првенствено као подр-

¹⁹ XML је стандардни скуп правила за дефинисање формата података у електронској форми. Следећи правила XML стандарда, корисници дефинишу сопствене (XML) формате података, које могу користити за њихово складиштење, обраду и размену (Wikipedia).

шка методологији његове традиционалне израде и садашње форме (уп. Ристић и др. 2012: 96–98). Као што је већ наглашено у првом делу рада, такво становиште је произашло из улоге Речника САНУ (као академијског речника и капиталног културно-научног пројекта) у процесима описивања лексичког система српског језика и изграђивања његове норме. Такође, произашло је и из уверења да лексикографско искуство, у свом садашњем виду, представља систем правила која је могуће (и неопходно) експлицирати у форми детаљних упутстава за рад на речнику. Уз ово, свакако треба имати у виду и друге стручне потребе на пројекту, као нпр.: стабилизацију динамике израде појединачних томова овог речника, боље повезивање чланова ауторског тима ангажованих на пословима његовог састављања, техничку редакцију, рационално коришћење стручних кадрова, тј. правилну расподелу посла и сл.

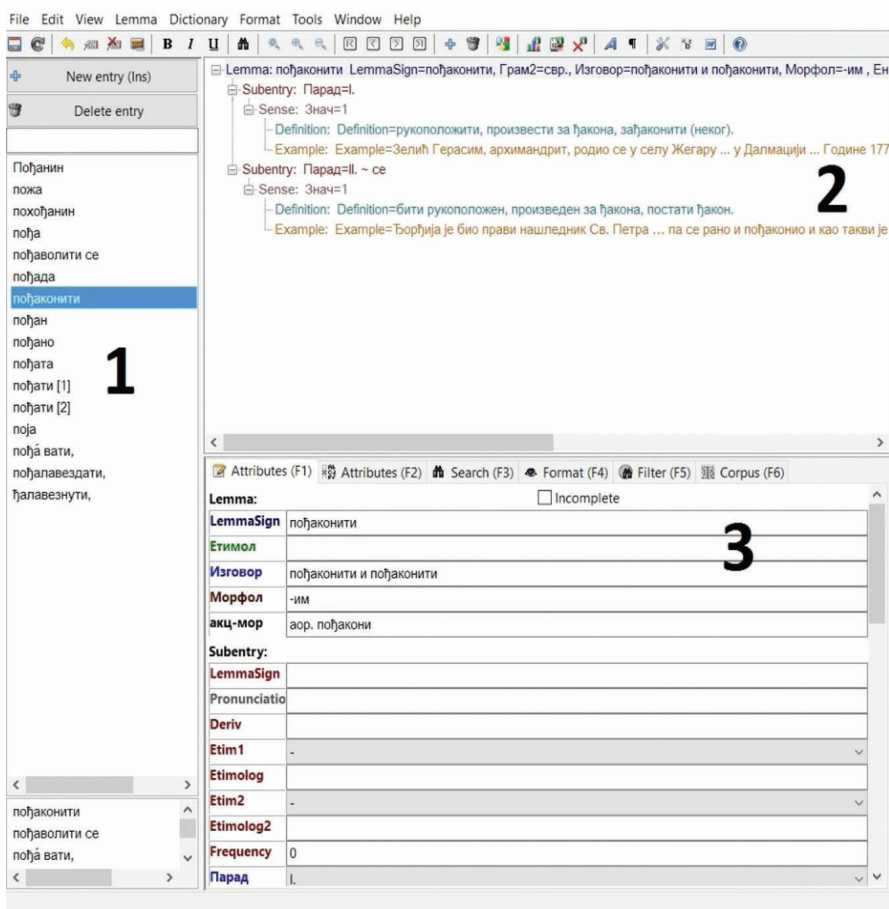
У даљем излагању биће представљен нацрт рачунарског окружења које би интегрисало дигитализовану грађу (в. т. 1.4.3 и 2) и ставило је на располагање лексикографу у оквиру јединственог *система за израду речника*.²⁰ Тај систем обухвата две целине: (1) део за унос речничког текста и (2) део за различите врсте претрага дигиталних колекција листића, библиотечких извора и објављених томова Речника САНУ. Уз ово, у засебном делу рада представимо и могућности даље експлоатације дигитализоване речничке грађе у сфери језичке културе.

3.2. *Нацрт система за израду речника*. Посматрано из угла лексикографа, средишњи део система за израду Речника САНУ чинио би део за унос речничког текста. Оптимално решење за овај део корисничког окружења представља лексикографски софтвер, наменски конструисан или прилагођен изради Речника САНУ. Главна предност овог софтвера лежи у могућности структурирања речничког чланка, експлицитним дефинисањем свих типова информација из одреднице, као и одређивањем хијерархијских односа међу њима, што одговара потребама за формализацијом речничког текста које су описане у т. 2.3.²¹ На овај начин добија се речнички текст описан према типовима информација, а истовремено и хијерархијски уређен. На слици 5 виде се три различита поља за унос података при обради речничког чланка.

²⁰ Термин *систем за израду речника* у значењу комплексног софтвера организованог око програма за унос и измену одредница користе Аткинсова и Рандел (Atkins and Rundell 2008, у: Abel 2012: 85). Уз овај термин, у науци се за исти појам користе још и термини *систем за њирењу речника за издавање*, *софтвер за састављање речника*, *софтвер за њродукуцију речника* и др. (Abel 2012: 85).

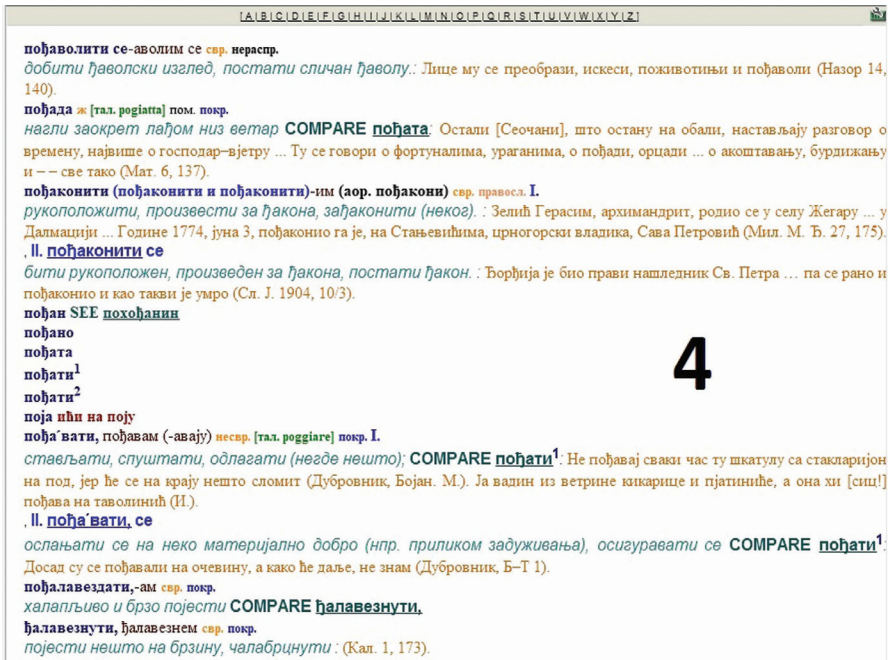
²¹ Преглед специјализованих програма за израду речника, као и опис њихових могућности и функција в. у Abel 2012.

У поље 1 уносе се одреднице које ће бити обрађене. У пољу 3 врши се унос података према елементима речничког чланка (овде је приказан само један од постојећих прозора поља 3), а преглед структуре чланка, укључујући и тагове налази се у пољу 2. Коначно, на слици 6 приказано је поље 4, где се види финални изглед речничког чланка.



Слика 5. Приказ лексикографског софтвера TschwaneLex (D. Joffe and G.M. de Schryver 2002, <http://tshwanedje.com/tshwanelex/>) према пољима за унос и форматирање текста: 1) списак одредница; 2) хијерархија речничког чланка од одреднице према примеру; 3) поље за унос различитих типова информација (акценат, морфологија, квалификатори, дефиниције, примери и сл.).

Из приказаног се види да лексикографски софтвер конструише хијерархијску структуру речничког текста, која, опет, одређује садржај одреднице. То значи да програм овог типа захтева од лексикографа



Слика 6. Приказ лексикографског софтвера TschwaneLex: 4) речнички текст, форматиран према подацима из поља 2) и 3).

да сваки елемент у грађи (нпр. глагол *појаконити* (се) у примеру: *Ђорђија ... се рано ... појаконио*) одреди прво по значењу: [„бити рукоположен, произведен за ђакона”], а затим и по односима припадности тог значења другим елементима унутар одреднице, од најужег према најопштијем: [значење повратног облика глагола] →припада [пољу повратни облици глагола], →припада [одредници *појаконити*]). Тиме се омогућује дефинисање и испоређивање елемената на истом нивоу хијерархије у речничком чланку, што значајно повећава прегледност обраде и помаже лексикографу да не пропусти унос неког од великог броја података које је потребно унети у текст обраде. Преглед текста одреднице *појаконити*, кодиран помоћу XML система ознака (в. т. 2.3) види се на слици 7.

Важна одлика лексикографских софтвера јесте могућност подешавања њихових елемената према специфичним потребама и методологији различитих лексикографских школа и ауторских колектива речника, као и према стручно-техничким смерницама. Интегрисањем поменутих смерница у саму структуру лексикографског софтвера добио би

```

</Lemma>
- <Lemma id="61" LemmaSign="пођаконити" Гран2="142" Морфол="-им " акц-мор="аор. пођакони" ЕнцПлг="102" Изговор="пођаконити и пођаконити">
- <Subentry id="71" Парад="205" Етим2="0" Етим1="0">
- <Sense id="63" прагма="0" Знач="1">
- <Definition id="64" Definition="рукоположити, произвести за ђакона, зађаконити (некор)."/>
- <Example id="65" Извор="0" Source="(Мил. М. Ђ. 27, 175). " Example="Зелић Герасим, архимандрит, родно се у селу Жегару ... у Далмацији ... Године 1774, јуна 3, пођаконио га је, на Стањевићима, црногорски владика, Сава Петровић"/>
- </Sense>
- </Subentry>
- <Subentry id="72" Парад="206" Етим2="0" Етим1="0">
- <Sense id="68" прагма="0" Знач="1">
- <Definition id="69" Definition="бити рукоположен, произведен за ђакона, постати ђакон. "/>
- <Example id="70" Извор="0" Source="(Сл. Ј. 1904, 10/3). " Example="Ђорђија је био прави наследник Св. Петра ... па се рано и пођаконио и као такви је умро"/>

```

Слика 7. Одредница *пођаконити* у Речнику САНУ, у форми XML скрипта.

се програм који одражава основне принципе израде Речника САНУ у главним сегментима продукције његовог текста.

3.3. Други важан део система за израду речника чиниле би могућности напредног претраживања дигиталне грађе. Овај део система може се описати кроз три засебне и међусобно повезане целине: 1) база лексикографских листића као колекција примера са допунским информацијама (година и место издања, жанр/дијалекатски ареал, коментар лексикографа и слика скенираног листића уз сваки пример); затим, 2) база библиотечких извора, укључујући и друге речнике и приручнике, који се консултују при изради Речника САНУ²² и 3) база објављених томова Речника САНУ. Таква организација система за израду речника омогућила би да се лексеми, при обради, приступа на три начина: (1) као циљној речи у примеру који је претходно одабран од стране ексцерптора; (2) као циљној речи у корпусу конкорданција из библиотечког материјала; и (3) као одредници која је регистрована и обрађена у разним општим и специјалним речницима. Приказ изложене организације система за израду Речника САНУ в. на слици 8.

3.4. *Дигитализација грађе и унапређење језичке културе*. Поред научних и стручних резултата, дигитализација језичких ресурса за Речник САНУ обухватала би и резултате из области језичке и речничке културе. Ово би подразумевало боље упознавање корисника речника са првим типом грађе – лексикографским листићима – који, поред своје лингвистичке вредности, имају и статус објеката од културно-историјског значаја.

²² Као што смо већ раније напоменули у т. 2.2, приручнике који се консултују при изради Речника САНУ требало би интегрисати у јединствену базу података (оделиту од осталих библиотечких извора, али у каснијој фази са њима интегрисану) која би била организована према одредничким речима и структурирана према информацијама које се могу наћи у овим речницима. Према архитектури и функцији, ова база би требало да буде веома слична бази заснованој на објављеним томовима Речника САНУ. Стога, на слици 8, ове две базе постављамо у заједничко поље, одељено испрекиданим линијама.



Слика 8. Организација система за израду Речника САНУ.

Скенирани листићи могли би бити обједињени у оквиру платформе која би чинила основ за израду виртуелног музеја културе српске речи у току два века постојања савременог вуковског језика. Музејски експонати могли би бити и сами листићи, али и појмови који су на њима описани. Музеј би омогућио да посетилац у претрази комбинује различите врсте (мета)података са листића (нпр. „покрајинска реч” + „тематска група X” + „у збиркама речи”), а као резултат би се добијале информације о пореклу и заступљености овог типа лексике по различитим ареалима штокавског говорног подручја. Тако би нпр. претрага заступљености тематске групе лексема *разбој* и *његови делови* у листићима из збирки речи у грађи за Речник САНУ могла дати преглед покрајинских назива за *разбој* и *његове делове* (нпр. *горњачка*, *долњачка*, *ђуџијагач*, *ирит*, *кљачњак*, *клада* и др.) у различитим ареалима штокавског говорног подручја (Банат, Црна Гора; Левач, Пирот; Дубица и др.). Резултати претраге могли би се допуњавати остензивним „дефиницијама” односно сликовним приказом ентитета и(ли) приказима самих листића, што би обогатило искуство истраживања језичке грађе и културног блага које је кроз њу приказано. Изложене могућности презентације и организације речничке грађе у складу су са савременим тенденцијама ка повезивању лексикографских и музеолошких знања, а у циљу повећања искористљивости културних и педагошких потенцијала великих лексикографских пројеката (према: Колгушкина, у штампи).

4. Закључак

На основу анализа које су у раду изнете дошло се до следећих закључака:

1) Грађа за Речник САНУ прикупљена је наменски и према одређеним класификационим принципима и правилима. Ови принципи узимају у обзир функционалностилски, нормативни и граматички статус речи у лексикону српског језика, као и регионалну заступљеност речи и њихових значења у дијалекатским ареалима штокавског говорног подручја.

2) Темпо израде Речника САНУ веома је сличан темпу израде речника овог типа у европским језицима који су израђени без информатичке подршке.

3) Темпо израде речника не зависи само од броја одредница (64%), него и од других фактора које би требало утврдити у будућим истраживањима.

4) При избору начина дигитализације грађе треба узети у обзир физичке карактеристике и стање саме грађе, очување њене целовитости у даљој изради Речника САНУ, реалне могућности конкретног типа дигитализације, квалитет дигитализованог материјала, а затим и финансијске и временске факторе у овом процесу. Одређивање приоритета корака у дигитализацији требало би да се одвија у складу са њиховом функцијом, као и са њиховим доприносом квалитету и брзини израде преосталих томова Речника САНУ.

5) Најважније добробити од планираног и представљеног процеса јесу: очување културног блага, повећање доступности грађе, побољшање квалитета израде Речника САНУ. Поред тога, постоје и други важни резултати и могућности које би дигитализација пружила, као што су брзина израде, ефикасније коришћење обрађеног материјала и његова презентација ширем кругу корисника.

*

На основу свега што је изложено у раду, истичемо да би модернизацију процеса израде Речника САНУ у ери дигиталних технологија требало започети што пре, водећи рачуна о свим наведеним потребама, а највише о стручним, научним и културним (тим редоследом).

ЛИТЕРАТУРА

Белић, А. 1926. *Српски речник књижевнога и народнога језика*. Рукопис. Архив САНУ.

- Витас, Д., Ц. Крстев 2015. Нацрт за информатизовани речник српског језика. *Научни сасијанак славистија у Вукове дане 44/3*. МСЦ: Београд, 105–116.
- Драгићевић, Р. (ур.) 2014. *Савремена српска лексикографија у теорији и пракси* (колективна монографија). Београд: Филолошки факултет.
- Згуста, Ј. 1991. *Приручник лексикографије*. Сарајево: Свјетлост.
- Ивановић, Н. 2015. Лексикографска мисао Александра Белића на почетку двадесетог века. *Међународни научни сасијанак славистија у Вукове дане 44/1*. Београд: МСЦ, 27–37.
- Колгушкина, Н. В. (у штампи). Средства музейной педагогики в изучении курса отечественной лексикографии в вузе. У: *Современные проблемы лексикографии*. Санкт Петербург: Российская академия наук.
- Николић, М. 2014. Утврђивање одреднице у српској лексикографији. У: *Савремена српска лексикографија у теорији и пракси*, Р. Драгићевић (ур.). Београд: Филолошки факултет, 61–68.
- Речник САНУ: *Речник српскохрватског књижевног и народног језика САНУ*, 1–19. 1959–2014. Београд: Српска академија наука и уметности, Институт за српски језик САНУ.
- Ристић, С., Н. Ивановић 2011. Предлог за модернизацију рада на Речнику САНУ. У: *Грамаишка и лексика у словенским језицима*, зборник радова. Нови Сад – Београд, Матица српска – Институт за српски језик САНУ, 529–553.
- Ристић, С., М. Јакић, А. Марковић, Т. Самарцић, Н. Ивановић 2012. Значај дигитализације језичких ресурса Речника САНУ за развој лексикографије и очување културне баштине. *Дигитализација културне и научне баштине, универзитетски рејоцијоријуми и учење на даљину*, књ. III: *Дигитални извори у друштвено–хуманистичким исцртаживањима*, Београд: Филолошки факултет, 79–108.
- Ристић, С. 2012. Израда нормативног речника савременог српског језика. *Научни сасијанак славистија у Вукове дане 41/1*. Београд: МСЦ, 5–17.
- Сабо, О., Д. Витас 1988. Могућности осавремењивања израде речника на примеру Речника српскохрватског књижевног и народног језика САНУ. У: *IV међународни научни скуп Рачунарска обрада језичких података*. Порто-рој: Институт Јожеф Стефан, 375–384.
- Тешић, М. Р. 2009. *С речима и речником*. Београд: Учительски факултет.
- Тривунац, Р. 2002. Уз рјечник дубровачког говора. У: *Рјечник дубровачког говора*. Бојанић, М., Р. Тривунац (саст.). *СДЗб XLIX*. Београд, 9–18.
- Ујучиштва за израду Речника САНУ*. Београд: Институт за српски језик САНУ, 1959 [2007].

*

- Abel, A. 2012. Dictionary writing systems and beyond. In: *Electronic Lexicography*. Granger, S., M. Paquot (eds.). Oxford: Oxford University Press, 83–106.
- Meijs, W. 1996. Linguistic Corpora and Lexicography. *Annual Review of Applied Linguistics*, vol. XVI. Cambridge: Cambridge University Press, 99–114.
- Meyer, Ch., I. Gurevych 2012. Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography. In: *Electronic*

- Lexicography*. Granger, S., M. Paquot (eds.). Oxford: Oxford University Press, 259–291.
- Ristić, S., N. Ivanović (in press). Contemporary Serbian descriptive lexicography. In: *International Handbook of Lexis and Lexicography*. London: Springer.
- Schryver, G. M. 2003. Lexicographers' Dreams in the Electronic–Dictionary Age. *International Journal of Lexicography*, vol. XVI/2. Oxford: Oxford University Press, 143–199.
- Weiner, E. 2009. The Electronic OED: The Computerization of a Historical Dictionary. In: *The Oxford History of English Lexicography*, vol. I. Cowie, A. P. (ed.). Oxford: Oxford University Press, 378–409.

Nenad B. Ivanović, Milena M. Jakić, Stana S. Ristić

THE REQUIREMENTS AND PERSPECTIVES OF DIGITIZING
LANGUAGE MATERIAL OF THE *DICTIONARY OF THE SERBIAN
ACADEMY* IN THE LIGHT OF MODERN APPROACHES

S u m m a r y

In this paper we describe the significance of the DSASA in Serbian and European lexicography. We have reached the following conclusions:

1) The language corpus for the DSASA was collated according to specific set of principles.

2) The dynamics of the production of the DSASA resembles the dynamics of the production of other dictionaries of similar scope and correlates the number of its entries.

3) Digitizing procedures need to take into account the physical state of the material, its significance and function in the compilation of the dictionary, as well as the other possibilities that open up.

4) The most important benefits from the digitization are: the preservation of cultural heritage, greater availability of the material, increased speed of production, more efficient use of the collated material and its presentation aimed at a wider circle of users.

CIP – Каталогизација у публикацији
Народна библиотека Србије, Београд

811.16'373(082)

811.16'374(082)

ЛЕКСИКОЛОГИЈА и лексикографија у светлу савремених приступа : зборник научних радова / уредништво Стана Ристић, Ивана Лазић Коњик, Ненад Ивановић. – Београд : Институт за српски језик САНУ, 2016 (Београд : Чигоја штампа). – 577 стр. ; 24 cm

На спор. насл. стр.: *Lexicology and Lexicography in the Light of Contemporary Approaches*. – Радови на више језика. – Тираж 300. – Напомене и библиографске референце уз текст. – Библиографија уз већину радова.

ISBN 978-86-82873-56-3

1. Ристић, Стана, 1950– [приређивач, сакупљач]
 2. Лазић Коњик, Ивана, 1978– [приређивач, сакупљач]
 3. Ивановић, Ненад, 1979– [приређивач, сакупљач]
- а) Словенски језици – Лексикологија – Зборници
 - б) Словенски језици – Лексикографија – Зборници

COBISS.SR-ID 228321036