

Белешка о дигитализацији речника

Душко М. Витас, Цветана Ј. Крстев, Ранка М. Станковић



Дигитални репозиторијум Рударско-геолошког факултета Универзитета у Београду

[ДР РГФ]

Белешка о дигитализацији речника | Душко М. Витас, Цветана Ј. Крстев, Ранка М. Станковић | Српски језик и његови ресурси | 2019 | |

10.18485/msc.2019.48.3.ch3

<http://dr.rgf.bg.ac.rs/s/repo/item/0004944>

Дигитални репозиторијум Рударско-геолошког факултета Универзитета у Београду омогућава приступ издањима Факултета и радовима запослених доступним у слободном приступу. - Претрага репозиторијума доступна је на www.dr.rgf.bg.ac.rs

The Digital repository of The University of Belgrade Faculty of Mining and Geology archives faculty publications available in open access, as well as the employees' publications. - The Repository is available at: www.dr.rgf.bg.ac.rs

Душко М. ВИТАС, Цветана Ј. КРСТЕВ, Ранка М. СТАНКОВИЋ
Филолошки факултет
Универзитета у Београду

БЕЛЕШКА О ДИГИТАЛИЗАЦИЈИ РЕЧНИКА

У раду ће се анализирати ограничења која проистичу из линеарног процеса традиционалне израде речника на примеру Речника САНУ. Начин да се превазиђу ова ограничења се састоји у формирању електронске лексикографске базе која не представља само пуку дигиталну транскрипцију папирног издања речника. Посебно се указује на чињеницу да текст речника може представљати корпус и приказују се одабрани примери анализе таквог корпуса формираног из текстове 1. и 19. тома Речника САНУ.

Кључне речи: лексикографија, рачунарска лексикографија, информатика, информациони систем

1. Увод

Информатичарско интересовање за речнике има најмање два извора. С једне стране, процес производње речника (и енциклопедија) може да послужи као прототипски пример информационог система на коме се може илустровати конверзија традиционално започетог пројекта у информатизовано окружење, али и методе изградње информационог система започетог *ab ovo*. С друге стране, различите информатичке апликације, почев од најнаивнијих игара речима, па до комплексних система за рачунарски посредовану комуникацију са човеком, садрже у позадини речнике у специфичним форматима. Отуда употребљивост једног језика у информатизованом друштву зависи непосредно од доступности таквих речника у дигиталном облику. Могућност губитка употребне моћи једног језика у информатизованом друштву не погађа непосредно међуљудску комуникацију, али онемогућава говорнике једног језика да користе технолошке могућности које имају говорници информатички боље опремљених језика. Назначимо да ово питање није питање локализације (или прилагођавања) софтверских пакета, већ питање суштинске интеграције српског језика у различите апликације да се не би морале користити само посредством других језика. У извесном смислу, ова ситуација

наликује на однос какав постоји између језика који имају писмо и језика који су остали без њега.

Квалитетан речнички ресурс за српски језик би могао бити Речник САНУ (РСАНУ 1959–2019, даље само Речник), али, као што је показано у (Витас и др., 2015), овај се пројекат одвија сувише споро да би могао да одговори на растуће информатичке захтеве. При томе, како је на истом месту назначено, питање евентуалне модернизације рада на овом пројекту се тек однедавно артикулише на начин који би омогућио пуну интеграцију српског језика у амбијент какав нуде модерне информатичке технологије.

Шта више, и други домаћи лексикографски пројекти се одвијају на сличан начин. Циљ овог рада је да укаже на могућа ограничења у информатизацији лексикографског посла у домаћој средини, као и да прикаже један прилаз у решавању овог проблема.

2. Линеарност процеса израде речника

Рад на Речнику САНУ, како је то уосталом дефинисано Законом (Закон, 2005), јесте процес који се састоји у објављивању томова речника. Из овако одређеног задатка следи да је резултат пројекта израде Речника његова штампана верзија у одређеном броју томова. Одавде даље следе два важна својства процеса његове израде:

- процес је строго линеаран у времену и
- резултат рада се повремено објављује у облику штампаног, дакле папирног, тома овог Речника.

Закон не одређује технологију којом ће појединачни томови бити произведени, нити судбину узгредних производа лексикографског рада. Наведена својства процеса израде речника, с обзиром на временски распон између првог и последњег објављеног тома, носе извештан број неминовних последица на садржај појединачних томова.

Линеарност процеса израде Речника је одређена обликовањем резултата рада на пројекту кроз штампане томове који следе азбучни редослед. Неке од последица овако организованог процеса израде су:

- обрада одреднице на одређено слово, нпр. *p*, не може бити ни сачињена, ни објављена пре него што се заврши обрада свих одредница које јој претходе (и то у облику штампаних томова);
- линеарно одвијање процеса израде речника не може да опише динамичке процесе у језику као што су нове одреднице или нова значења одредница које се налазе у већ објављеним томовима;
- допуњавање грађе током времена, нужно мења дистрибуцију коришћених извора што потенцијално води ка неуједначеностима у дистрибуцији извора и природи изабраних примера између почетних и завршних томова;

- како је полазна концепција речника утврђена на почетку израде Речника, објављени томови следе полазне принципе у обликовању речничког чланка, па се новине у поимању својстава српске лексике и њених међуодноса углавном не могу узимати у обзир
- како је речник првенствено намењен истраживачкој и високо стручној публици, други типови корисника и њихових потреба нису уграђени у текст речника, а друкчије потребе и друкчији погледи на грађу не могу бити приказани у линеарном процесу израде;
- и као најзначајније, кретање кроз речник је могуће једино преко одреднице: потрага за информацијом која није одредница није могуће друкчије до линеарном претрагом.

Последица линеарног процеса израде Речника је његово поступно формирање у облику штампаних томова. Али током рада на Речнику, технологија припреме штампаног издања је значајно промењена. Коначни текст првих томова је обликован у класичној штампарији¹, док су рукопис последњих објављених томова припремили сами лексикографи у *MS-Word*-у (уз накнадно формирање текста). Ова технолошка иновација је само привидна: рачунар се у поступку припреме рукописа за штампу користи као писаћа машина. Ово значи да је електронска верзија рукописа само „рукопис” који је припремљен графички у формату Речника, али не и сам Речник. Другим речима, текст појединачних томова у дигиталном облику није исто што и њихов папирни еквивалент. Шта више, овакав облик дигиталног речника не подлеже аутоматској контроли садржаја (почев од употребљених карактерских скупова па до експлицитних ознака у микроструктури речничког чланка).

Промена у технологији припреме рукописа није утицала значајније на утврђени процес израде Речника: резултат рада је и даље објављивање новог тома. Статус електронске верзије припремљеног текста, будући да је реч о узгредно производу у припреми коначне верзије тома, није дефинисан: његова даља употреба² није до скоро била предмет посебне бриге. Непосредна последица оваквог прилаза је да електронска верзија текста, припремљена за штампу, није предмет адекватног архивирања, није организована на начин који омогућава његову поновну употребу, а није ни заштићена од могућих злоупотреба. Најзад, штампана верзија је ограничена физичким димензијама појединачних томова чиме су постављена ограничења на обим речничког чланка, избор, број и дужину примера итд.

¹ Тек у 3. тому постоје подаци о штампарији.

² Овде под употребом се подразумева архивирање, модификовање и дистрибуција дигиталне верзије одштампаног тома.

3. Дигитални облици српских речника

За разлику од одштапаног речника на папиру, његов привремени живот у дигиталном облику омогућио је развој различитих платформи које омогућавају претрагу штампаних томова преко веба. Овде се разликује неколико начина употребе таквих ресурса. Најједноставнији пример су скениране верзије папирних речника са препознатим карактерима, обично у формату pdf. Међу оваквим речницима се може пронаћи већина домаћих речника укључујући и речнике издате у Институту за српски језик САНУ³ и Матици српској⁴. Овакав облик дигитализованих речника, поред своје несумњиве корисности (доступан је где год је доступан и веб), ипак је само дигитална транскрипција папирне верзије.

Нешто сложенији прилаз скенираној верзији речника српског подразумева растављање текста речника на речничке чланке и њихово обликовање у језицима за обележавање, у дијалектима XML-а. Један такав речник се може наћи на веб-страни предузећа Србсофт⁵ која објављује *Речник српског језика*, „урађен је на основу 18 томова Речника Српске академије наука и уметности и 3 последња тома Речника Матице српске”. За разлику од прве групе дигитализованих речника, овај систем омогућава претраживање текста ових речника кроз одговарајућу сумеђу. Занимљиво је да се у решењу Завода за унапређење васпитања и образовања ова дигитализована компилација два угледна речника назива „рукопис додатног наставног средства Програмски пакет РАС”⁶ што потврђује став из претходне тачке према коме је Речник САНУ само онај који је штампан (у више томова). Друкчије обликовање истог садржаја, дакле, није више Речник САНУ чак и када се предочи у форми „рукописа”.

Нешто друкчији прилаз дигитализованим речницима нуди платформа Расковник⁷ Института за српски језик и Центра за дигиталне хуманистичке науке. Ова платформа приказује више дигитализованих штампаних речника. Речнички чланци су описани у складу са ТЕI-препорукама, а сами текстови похрањени у XML-базу података што омогућава претраге које превазилазе оквире штампаних верзија као што су опсцене речи или речи грчког порекла у Вуковом *Српском рјечнику* (Тасовац, 2017).

Ови различито организовани дигитализовани речници су довршена дела која се од своје папирне верзије разликују према начину претраге. Одавде следи да они могу бити од помоћи лексикографу у консултовању извора, али садрже и већину недостатака које одређује линеарни процес стварања речни-

³ *Речник САНУ* се налази на <https://books.google.rs/> полазећи од папирне верзије са Универзитета у Индијани.

⁴ *Једнотомник*, као и други мањи речници се налазе, нпр, на <https://repo.etfovac.com/srpski-jezik/> (посећено 01.07.2019.)

⁵ <http://www.srpskijezik.com/Home/Store>

⁶ <https://spaces.hightail.com/space/YOq5sKGvpc/files/fi-0ed700f5-8d2d-430f-8573-df2a226f94-bf/fv-8299eb9a-c879-4dd1-b9f2-44b8822dc1c8/1.%20Zavod%20za%20unapre%3-Fivanje%20vaspitanja%20i%20obrazovanja.pdf>

⁷ <http://raskovnik.org/reznici> (посећено 01.07.2019.)

ка. Основни недостатак, према нашем мишљењу, је методолошка неуједначеност у избору и опису одредница, као и у структурирању самог речничког чланка.

За разлику од речника наведених у т. 3, пројекат Викиречника⁸ илуструје један друкчији прилаз изградњи речника у дигиталном окружењу. Упркос високој неуједначености чланака, овај пројекат показује да је могуће не само превазићи линеарни ток израде речника него, што је још значајније, уградити у речник информације које се не могу представити у штампаним речницима, као што је изговор или повезивање са енциклопедијском информацијом.

4. Превазилажење линеарности

Превазилажење недостатака које долазе од линеарности процеса израде речника се постиже организовањем радног процеса око лексикографске базе података. Овакве базе података, које су постале уобичајени део савременог организовања лексикографског процеса⁹, су централни део ове производне активности (Еткинс и др. 2008).

База се састоји из записа који на формализован начин прецизно описују микроструктуру речничког чланка укључујући и различите врсте упућивања на друге чланке или њихове делове. У случају миграције традиционално започетог лексикографског пројекта ка лексикографској бази, опис микроструктуре садржи опис усвојене структуре чланка, али потенцијално и различите надградње које нису биле предвиђене или нису могуће у продукцији папирне верзије речника. У том светлу, лексикографска база је средство за акумулацију лексикографског (и лингвистичког) знања о језику која, као динамички објекат, обезбеђује ажурност кроз интеракцију са корпусом, хомогенизацију употребљених ознака унутар базе, али и различите погледе на садржај базе: речник постаје *вишеслојна* конструкција из које је могуће извозити различито обликоване речнике различитог садржаја (Пустојевски и др. 2017).

У овако конципираном процесу, штампани томови речника се јављају као повремени извод одговарајућег дела базе. У случају када је речник доступан преко вебa, а његов комплетан садржај видљив, штампани том има првенствено архивски значај. Увођење лексикографске базе у процес израде речника подразумева, поред задатака у пројектовању базе (Павловић-Лажетић 1996), одређивање за начин организовања корпуса и његове експлоатације, као и одређивање система за писање речника (*DWS – dictionary writing system*) који мора бити тесно повезан са структуром базе и корпуса. Одлуке о ове две компоненте су дакле најуже повезане са концепцијом базе и морају омогућавати приступ пројектованим пољима која чине речнички чланак.

⁸ <https://sr.wiktionary.org/>

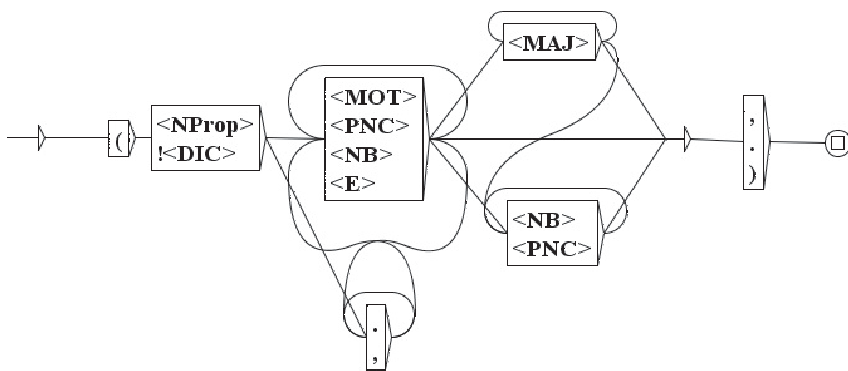
⁹ Погледати сумарни преглед о системима за писање речника и системима за управљање корпусом на http://www.elixicography.eu/wp-content/uploads/2015/04/ENeL_WG3_Vienna_DWS_CQS_final_web.pdf

Елементи овако конципираног рада на изради речника САНУ су дати детаљније у (Станковић и др. 2018а), (Станковић и др. 2018б) (СТИЈОВИЋ и др. 2017), (СТИЈОВИЋ и др. 2018).

5. Речник као корпус

Целина речника је својеврстан корпус који омогућава увид у различита својства речника невидљива листањем његове папирне верзије. Упоредили смо неструктуриране дигиталне верзије 1. и 19. тома¹⁰ са циљем да испитамо могућност аутоматске анотације сировог текста, али и да сагледамо својства језика који чини овај текст. Први том садржи око 550.000 облика речи на 690 страна штампаног текста, а 19. том око 672.000 облика речи на 800 страна.

Како текст Речника, као што је раније речено, није структурирани документ, неопходно је раздвојити поједине делове речничког чланка у засебне јединице као што су одредница, граматичка дефиниција, поједина значења, пример, итд. Део примера је и упућивање на извор у коме је цитирано значење потврђено. Издвајање једног одређеног броја дела цитираних извора се може извршити графом на слици 1.



Слика 1. Граф који екстрахује део цитираних извора у РСАНУ

Примери овако екстрахованих извора су дати у табели 1.

х послова (Гар. И. 2, 16). На глас о рус
ну мараму (Босна, Мургић Ф., ЗНЖ 8, 97).
м чесмама (Мус. 1908, 36/3). абдестана ж
абдицира (Павлин. 3, 22). абдов, абдова
х, штраух (Фрушка Гора, Шкар. 2, 235). а
ве у абез (Батут 1, 152). абемарија ж по
авемарија (Дубровник, Зоре 2). абен1, -а
кабаницу (Поп. Д. Ј. 1, 148). абен2, -а
бено воће (Ниш, Стој. Д.). фиг. рђав, бе

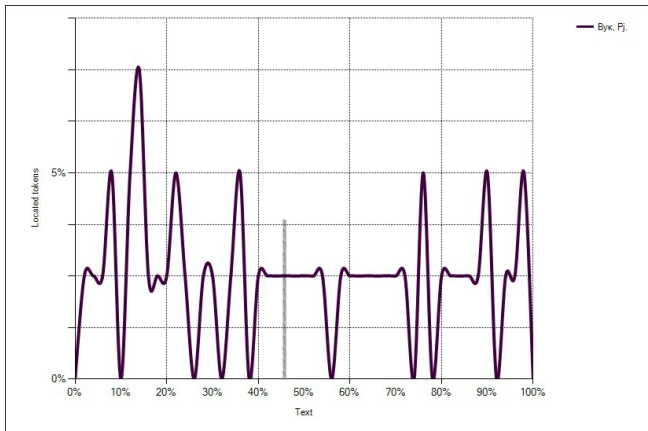
¹⁰Први том обухвата одреднице од А до Богољуб, а 19. том од оцат до петогласник.

онајпречи (Новић 1, 116). абено прил. на
рау живи (Новић 3, 32). Кад [ко] живи л
бритвица (Новић 6, 132). Ђелавој глави
а абењача (Мил. М. Ђ. 9, 226). абер и аб
ништавило (Цар. БК 1901, 1599). абердак,
ердар (1) (Вук, Рј.; Р–К 2). абердан, -а

Табела 1. *Извори екстраховани графом са слике 1.*

Други графови препознају друге типове цитирања, а сви заједно омогућавају да се аутоматски издвоје поља у којима је назначен извор. Најчешће цитирани извор у анализираним томовима је Вуков *Рјечник*, а дистрибуција цитирања је приказана на слици 2, где вертикална сива линија обележава границу између текста 1. и 19. тома. Крива на слици 2. приказује како се дистрибуира ознака за Рјечник у линеарном тексту¹¹.

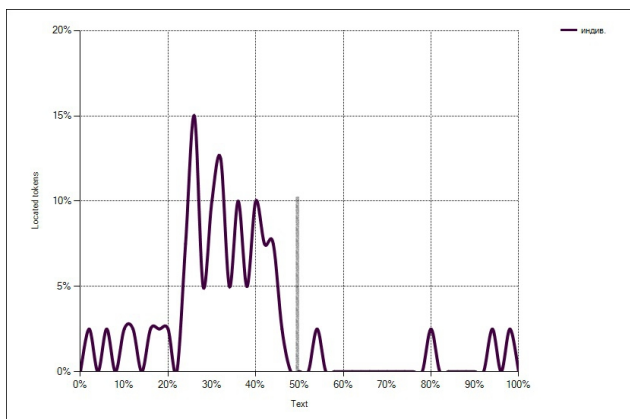
Разлика у дистрибуцији овог извора између 1. и 19. тома је, по свој прилици, последица проширивања полазног корпуса новим изворима који су „разредили” појављивања примера из *Рјечника*, што је очекивани резултат у линеарном процесу израде Речника.



Слика 2. *Дистрибуција извора Вук, Рј. у 1. и 19. тому РСАНУ*

Као илустрација промене у атрибуцији обележја се може посматрати расподела квалификатора *индив* (= индивидуално, само код једног писца) који реализује 90% својих појављивања у 1. тому као што илуструје слика 3.

¹¹ Графици на сликама 2, 3, 4 и 5. имају следеће значење: дуж *x*-осе су приказана места на којима се појављује одређени феномен у целини корпуса, а на *y*-оси се читава колико је процената од укупног броја појављивања анализираниог феномена реализовано у посматраној тачки на *x*-оси.



Слика 3. Дистрибуција квалификатора индив.

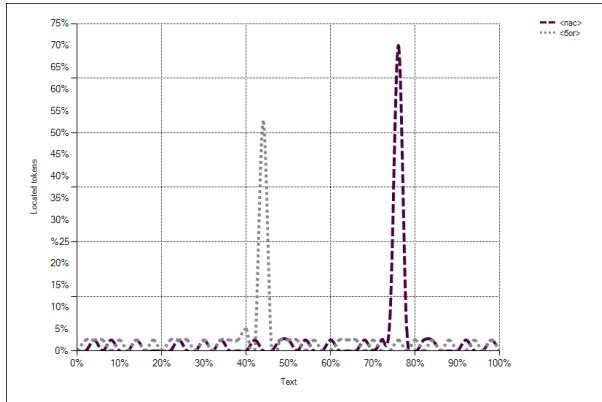
Погледајмо даље дистрибуцију неких именица у оба анализирана тома. Подаци о фреквенцији су дати у табели 2. Именица <бог> у различитом облицима је описана чланком у 1. тому, а <нас> у 19. Дистрибуција обе именице (у било ком облику) је приказана на слици 4. Овај график указује да се обе именице интензивно користе у зони око речничког чланка који их описује, али се користе и у другим деловима текста, првенствено у примерима¹². Из угла корисника речника може бити значајно и корисно имати увид и у друге примере у којима се користи одређена реч. Ова врста веза, коју није могуће утврдити у штампаном издању, могуће је успоставити кроз лексикографску базу (као део који интерагује са корпусима). Из угла лексикографа, овакви примери показују да је изванредан број примера у изворима садржан у објављеном делу речника и да могу представљати додатну грађу пиликом обраде одређене одреднице.

	Том 1	Том 19	Збирно
<бог.N> ¹³	874	336	1210
<нас.N>	121	716	837

Табела 2. Високофреквентне именице у 1. и 19. тому РСАНУ

¹² Према слици 4, одредница <бог> остварује половину својих појављивања у зони чланка који је описује, а другу половину у осталим деловима корпуса.

¹³ Ознака .N обезбеђује да се у претраживању корпуса препознају само именички облици леме наведене међу угластим заградама < и >.



Слика 4. Дистрибуција одредница <бог> и <пас>

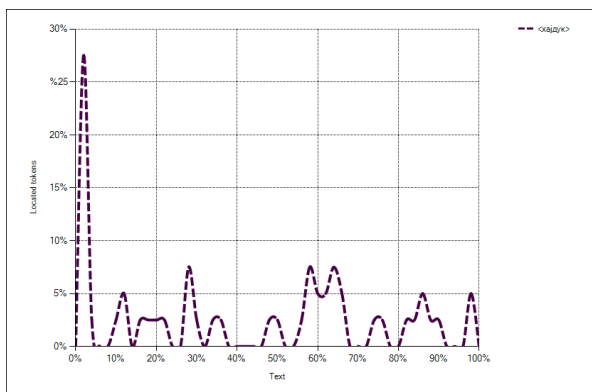
Грамматички опис уз одредницу *бог* је дат (ако се изоставе акценти и дужине) са:

бог, бога м (мн. богови, заст. бози, ген. богова, заст. бога). У тексту речника су потврђени и други облици флективне парадигме ове именице као што показује следећи списак (у колони обележеној са *f* су фреквенције облика, редом, у корпусу речника и у чланку одреднице *бог*):

Облик	<i>f</i>	Облик	<i>f</i>
(Б б)о ^г ¹⁴	465 (133)	<i>богови</i>	20 (3)
(Б б)ога	350 (115)	<i>богова</i>	18 (5)
(Б б)огу	151 (47)	<i>боговима</i>	9 (0)
(Б б)оже	127 (49)	<i>бози</i>	3 (2)
(Б б)огом	64 (31)	<i>богове</i>	2 (0)
		<i>бозима</i>	1 (1)

Погледајмо још пример именице *хајдук* за коју ће речнички чланак бити формиран у последњим томовима речника. Ипак, под одредницом *ајдук* наилазимо на упућивање са скраћеницом в.: *ајдук, -ука м (вок. ајдуче) (...) в. хајдук*. Поред овог упућивања, именица *хајдук* се у различитим облицима јавља у целом корпусу 1. и 19. тома преко 80 пута на начин приказан дијаграмом на слици 5, а преко четвртине својих појављивања има уз одредницу <*ајдук*>.

¹⁴ (Б|б) је регуларни израз који значи Б или б, овде Бог или бог.



Слика 5. Дистрибуција облика именице хајдук

Како су у оквиру чланка *ајдук* описана значења која се односе и на одредницу *хајдук*, простом трансформацијом могуће је аутоматски унапред формирати део чланка за ову одредницу.

Конечно размотримо још скраћивање примера унутар одредничког чланка. Као пример погледајмо пример уз придев *аветињски* ~ *Забележио [је] тај аветињски покрет живе статуе* (Андрић 5, 232). Корисник речника, који би пожелео да види шири контекст овог примера како би га боље разумео, би морао да потражи издање *Травничке хронике*, наведено у изворима, па да на страни 232 задовољи своју радозналост. Али ако претпоставимо да се у корпусу налази електронски текст овог романа, са обележеном пагинацијом, онда је проширивање примера на шири контекст сасвим једноставна операција. Један пример тако проширеног цитата би могао бити:

И конзул и његов тумач изишли су као из гробнице. Давна је био блед као покојник, са капљама хладног зноја на челу. Давил је ћутао све до куће. Али је, међу оних неколико најтежих страхаота које је у току година доживео у Травнику, забележио тај аветињски покрет живе статуе.

Дигитализовани извори у облику који омогућава претраживање могу допринети опису српске лексике и на други начин. Наиме, како је ексерпција из извора за Речник била парцијална, могуће је да потврде речи или њихових облика изостану из грађе (Витас 2007). Како бисмо испитали да ли су сви необични примери из извора представљени у Речнику, састављен је следећи експеримент, полазећи од електронске верзије романа Јакова Игњатовића *Милан Нарањкић*¹⁵ који се налази у списку извора за Речник (као Игњ. Ј. 9). Овај текст је у формираном корпусу Речника цитиран за тридесетак примера. Занимало нас је да ли се могу аутоматски издвојити они облици речи у овом роману за које не постоји адекватна одредница или опис у Речнику. Ограничили смо се на речи који почињу словом *а* и које би могле бити описане у 1.

¹⁵ Овај роман се наводи у изворима са ознаком (Игњ. Ј. 9) која се односи на издање СКЗ из 1900.

тому Речника. Поређењем два корпуса наилазимо на више примера таквих облика, а они су приказани у додатку 1. Циљ овог експеримента није био да тумачимо разлоге због којих су ове одреднице и облици изостављени, већ да прикажемо могућу интеракцију дигиталне библиотеке, уређене према критеријумима описаним у (Витас, 2016), и потенцијалне лексикографске базе. Наиме, овај експеримент показује да ће се нове одреднице наћи не само у новој грађи, већ и у делима која су била парцијално ексцерпирана.

6. Закључак

У раду је указано на проблеме које носи процес линеарне израде Речника као и на могућности превазилажења тих проблема организацијом речничке грађе у облику лексикографске базе података. Посебно су назначене могућности које доноси формирање објављених томова Речника као корпуса и његове интеракције са дигиталним библиотекама. Ипак, како је процес израде речника, у основи, процес изградње једног информационог система, изван је наших могућности да предвидимо критичне тачке у информатизацији овог пројекта. Оваква анализа би подразумевала израду неопходних информатичких докумената за реализацију таквог пројекта.

Додатак 1. Примери из романа *Милан Наранџић (1. део)* Јаше Игњатовића који нису илустровани у Речнику.

Нумерација страна је према првом издању романа чији се дигитални облик налази на <http://digital.bms.rs/ebiblioteka/publications/view/1383> док је дигитализована верзија употребљена за сравњивање са Речником према издању овог дела Матице Српске из 1949. У овом издању је наведено „Објашњење мање познатих речи и израза” које даје значење ниже наведених речи:

абонират као облик трпног придева – »Ви, фрајлице, често у театар идете?« / »*Абонирата* сам.« (стр. 208)

адресират као облик трпног придева – Један дан дође у руке Бранку једно писмо, мушком руком писано и *адресирато* на Паулину. (стр. 211)

академичан – Но једног је једним ударцем тако набоксирао, да је почео крв пљувати. Ту же га декану, држи се *академичан* концес. Једва га ослободе, издржавши ариште осам дана. (стр. 74)

акомпанирати – Сад даду около њега, над главом, испод ногу и код бока шест свећа запалити, па онда запојаше сви у хору Хинеско Јектеније, а Цигани су морали *акомпанирати*. (стр. 59)

алагер – Играо сам добро билијара, па сам сваки дан на *алагеру* што добијао. (стр. 117).

алатура – Ако је високоучена и високовоспитована, нек донесе много новача, тојест, кажем за мене. Напоследку: не дати да се први дан женидбе *алатура* интабулира. У овом случају најбоље је кад муж нема шта да се на њега интабулира. Јер, ако се интабулира, муж ће морати жени сваког њеног имендана лепо гратулирати, да му се у чему год воља испуни, и онда жена носи капут, а муж визитлу. (стр. 133).

анаван (у оригиналу *ан-аван*) – Онда га скину, па удри с њим у францес. Не зна како ће; уфате га двојица испод руке, па вуци *анаван*, гурај га, па ето шасекоте, па после с њим чардаш, па мазур, па тако даље, чак до зоре. (стр. 60)

ЛИТЕРАТУРА

- Витас 2007:** Душко Витас, „О проблему не(пре)познате речи у обради текстова на српском језику”, *Зборник Матице српске за филологију и лингвистику*, L, 111–120.
- Витас/Крстев 2015:** Душко Витас и Цветана Крстев, „Нацрт за информатизовани речник српског језика”, *Научни састанак слависта у Вукове дане*, 44/3, 105–116
- Витас 2016:** Душко Витас, „Инфраструктура за изучавање и обраду српског језика”, у: *Зборник Института за српски језик САНУ III* (уредник Срето Танасић), Београд: Институт за српски језик САНУ, 115–129.
- Еткинс и др. 2008:** Atkins, B. T. Sue; Michael Rundell. *The Oxford Guide to Practical Lexicography*. Oxford University Press.
- Закон 2005:** *Закон о Речнику Српске академије наука и уметности*, „Службени гласник Републике Србије”, број 110/05.
- Павловић-Лажетић 1996:** Gordana Pavlović-Lažetić, *Osnove relacionih baza podataka*, Београд: Matematički fakultet.
- Пустојевски и др. 2019:** Pustejovsky, James; Olga Batiukova. *The Lexicon*. Cambridge University Press.
- Станковић и др. 2018а:** Stanković, Ranka; Cvetana Krstev, Biljana Lazić, Mihailo Škorić. Electronic Dictionaries – from File System to lemon Based Lexical Database, John P. McCrae et al. (eds.) *Proc. of the 11th LREC – W23 6th Workshop on Linked Data in Linguistics : Towards Linguistic Data Science (LDL 2018)*, Miyazaki, Japan, ISBN 979-10-95546-19-1
- Станковић и др. 2018б:** Stanković, Ranka; Rada Stijović, Duško Vitas, Cvetana Krstev, Olga Sabo. The Dictionary of the Serbian Academy: from the Text to the Lexical Database, *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, pp. 941–949.
- СТИЈОВИЋ/САБО/СТАНКОВИЋ 2017:** Рада Стијовић, Олга Сабо, Ранка Станковић, „Речник САНУ као база термилошких речника (на примеру речника кулинарства)”, *Словенска терминологија данас*, Београд: Српска академија наука и уметности, 109–123.
- СТИЈОВИЋ/СТАНКОВИЋ 2018:** Рада Стијовић и Ранка Станковић, „Дигитално издање Речника САНУ: формални опис микроструктуре Речника САНУ”, *Научни састанак слависта у Вукове дане*, 47/1, 427–440.
- РСАНУ 1959–2019:** *Речник српскохрватског књижевног и народног језика САНУ*, књ. I–XX, Београд: Институт за српски језик САНУ.
- Тасовац 2017:** Караџић, Вук Стефановић, *Српски рјечник*: дигитално издање (ур. Тома Тасовац) Београд: Институт за српски језик и Центар за дигиталне хуманистичке науке, 2017. ISBN: 978-86-82873-57-0.

Duško M. Vitas, Cvetana Krstev, Ranka M. Stanković

A NOTE ON A DICTIONARY DIGITIZATION

Summary

The paper analyses the limitations of the linear process of the traditional dictionary production and illustrates them on the example of the SASA Dictionary. These limitations can be overcome by the establishment of a lexicographic database that represents more than the digital transcription of a paper dictionary edition. It is additionally stressed that a dictionary text represents itself a valuable corpus for various research purposes which is illustrated by a few examples of the analysis of such corpus compiled from the first and 19th volume of the SASA Dictionary.