

Душанка Вујовић

ДИГИТАЛНА ИНФРАСТРУКТУРА РЕЧНИКА САВРЕМЕНОГ СРПСКОГ ЈЕЗИКА

Рад је посвећен дигиталној инфраструктури развијеној за потребе израде *Речника савременог српског језика* Матице српске, који има за циљ попис и систематски опис савремене лексике српског језика. За ове потребе развијена је софтверска платформа која подржава структурирани унос и обраду лексикографских података, као и њихово повезивање са електронским корпусом српског језика посебно формираним за ову намену. Посебна пажња посвећена је принципима организације платформе из лексикографског угла, структури лексикографске одреднице у дигиталном окружењу и утицају дигиталних алата на организацију и методологију лексикографског рада, нарочито у погледу избора и употребе корпусних примера.

Кључне речи: Речник савременог српског језика; дигитална лексикографија; електронски корпус; дигитални језички ресурси.

This paper focuses on the digital infrastructure developed for the compilation of the *Dictionary of Contemporary Serbian* (Matica srpska), which aims to record and systematically describe the contemporary vocabulary of the Serbian language. For this purpose, a dedicated software platform has been developed, enabling the structured input and processing of lexicographic data, as well as their integration with an electronic corpus of Serbian specifically compiled for this task. Particular attention is given to the principles underlying the organization of the platform from a lexicographic perspective. The structure of dictionary entries in a digital environment is also examined, as well as the impact of digital tools on lexicographic work, especially regarding the selection and use of corpus examples.

Key words: *Dictionary of Contemporary Serbian*; digital lexicography; electronic corpus; digital language resources.

1. Увод. Лексикографска делатност Матице српске заузима значајно место у српској култури, развоју српске лингвистике, стандардизацији српског језика, као и у систематском бележењу, проучавању и очувању његовог лексичког блага. У оквиру ове традиције настао је и лексикографски пројекат *Речник савременог српског језика* (РССЈ) који се непосредно надовезује на Матичине опште језичке речнике, шестотомни *Речник српскохрватскога књижевног језика* (РМС) и једнотомни *Речник српског језика* (РСЈ). Концепција новог *Речника савременог српског језика* ослања се на традиционална лексикографска решења српске лексикографије примењена у овим речницима, као и на она примењена у *Речнику српскохрватског књижевног и народног језика* (РСАНУ), настојећи да их даље развије и прилагоди условима савремене лексикографске праксе.

Иако су разговори о концепцији Речника, као и одређене организационе припреме, започети много раније, конкретне активности на његовој изради отпочеле су 2020. године. Циљ овог великог лексикографског пројекта јесте сакупљање, систематизација, опис и објављивање око 150.000 речи потврђе-

них у изворима из друге половине XX и првих деценија XXI века (ДРАГИЋЕВИЋ 2026: 5, ВУЈОВИЋ – МИЛОСАВЉЕВИЋ 2023: 135). Његов значај огледа се у томе што ће попунити празнину у опису лексике српског језика из периода од половине XX века до данас, јер су у грађу за речник ушла савремена „дела српских писаца са језичког простора на којем се говори српски језик, дневне новине и часописи, речници, уџбеници за основну и средњу школу, универзитетски приручници, стручна литература, преведена дела најбољих преводилаца“ као и други доступни писани и говорни извори (ДРАГИЋЕВИЋ 2026: 5).

Пошто обим и разноврсност савремене језичке грађе захтевају другачије начине прикупљања, организовања и анализе, него што је то било могуће у традиционалној лексикографској пракси, у изради Речника значајну улогу имају дигитални језички ресурси и алати, који обезбеђују систематско прикупљање, структурирање и претрагу лексичке грађе (GRANGER 2012: 2–3). Зато је у оквиру пројекта прво развијена специјална дигитална инфраструктура¹. Циљ рада је да прикаже њене основне компоненте: софтверску платформу за унос и обраду лексикографских података и електронски корпус српског језика, формиран на основу савремених писаних и говорних извора, као и да покаже на који начин њихова интеграција утиче на организацију лексикографског рада и структуру одреднице.

У савременој лексикографској пракси развијени су различити софтверски системи намењени изради речника. Такви системи усмерени су преваходно на уређивање лексикографских одредница и генерисање речничког текста, док се корпусна претрага по правилу обавља помоћу засебног алата. Међу доступним решењима нисмо пронашли систем који би пружио одговарајућу подршку за сложену прозодијску и морфолошку структуру српског језика и за формат лексикографског чланка какав је успостављен у традиционалним Матичиним речницима РМС и РСЈ. Зато је за потребе израде РСЈ развијено сопствено софтверско решење у виду платформе која обједињује формирање лексикографског чланка и приказ и претрагу електронског корпуса у заједничком радном окружењу, прилагођеном специфичностима српског језика, при чему генерисани речнички текст одговара формату постојећих Матичиних речника (VUJOVIĆ – MILOSAVLJEVIĆ 2025: 426). Развијена софтверска решења веома су сложена и обухватају велики број функционалности, од којих се у раду приказују оне најзначајније за организацију и методологију лексикографског рада. Будући да за српски језик нисмо пронашли у потпуности упоредива решења овог типа², опис дигиталне инфраструктуре у раду заснива се, пре свега, на сопственом искуству у развоју система.

2. УЛОГА ДИГИТАЛНИХ РЕСУРСА У САВРЕМЕНОЈ ЛЕКСИКОГРАФИЈИ. Лексикографија као дисциплина која се бави прикупљањем, описом и систематизацијом лексичког фонда једног језика, традиционално је била везана за штампане медије. Речници су настајали као обимна штампана дела, која су представљала

¹ Под дигиталном инфраструктуром у раду подразумевамо целовит систем софтверских алата, база података и електронских ресурса који чине технолошку основу за лексикографски рад.

² У Матици српској, за пројекте из домена историјске лексикографије, развијен је софтвер специјализован за обраду старих текстова.

результат дуготрајне и систематске лексикографске обраде. Развој дигиталних технологија и интернета последњих деценија довео је до значајних промена у савременој лексикографији, што се огледа у новим моделима израде и дистрибуције речника (FUERTES-OLIVERA 2018: 1–2). У многим језицима дошло је до постепеног преласка са традиционалних штампаних речника на електронске и онлајн лексикографске ресурсе, који отварају нове начине претраге, повезивање са језичким корпусима и динамично ажурирање лексичких података. Иако су први електронски речници често представљали само дигиталну верзију штампаних издања, савремени лексикографски ресурси све више користе могућности електронских медија и доносе другачији методолошки приступ у организацији и представљању лексичке грађе (GRANGER 2012: 2; DE SCHRYVER 2003: 146). Истовремено, језички корпуси, као електронске збирке аутентичних језичких података, играју све значајнију улогу у савременој лексикографији и представљају један од основних извора за лексикографску анализу (GRANGER 2012: 3). За разлику од неких других европских језика, српска лексикографија још увек се, у великој мери, ослања на традиционални модел штампаних речника. Ипак, у савременим лексикографским пројектима, као што је Матичин пројекат израде РССЈ, све је присутнија употреба дигиталних алата, електронских база података и језичких корпуса, што постепено мења начин на који се лексичка грађа прикупља, обрађује и систематизује.

2.1. ДИГИТАЛНА ОСНОВА ПРОЈЕКТА РССЈ. Поједини аспекти дигиталне инфраструктуре за израду *Речника савременог српског језика* већ су разматрани у ранијим радовима, у којима су представљени програм за унос лексикографских јединица и електронски корпус *Срјко* (Вуловић – Милосављевић 2023; ВУЈОВИЋ – МИЛОСАВЉЕВИЋ 2025). У овом раду даје се општи преглед дигиталног окружења у коме се одвија пројекат израде Речника, чији је циљ систематски опис лексике савременог српског језика, са посебним освртом на језичку употребу у новијем периоду. На тај начин овај Речник треба да допуни и савремени постојеће лексикографске ресурсе, узимајући у обзир промене које су се у језику догодиле у последњим деценијама, као и нове методолошке приступе у савременој лексикографији.

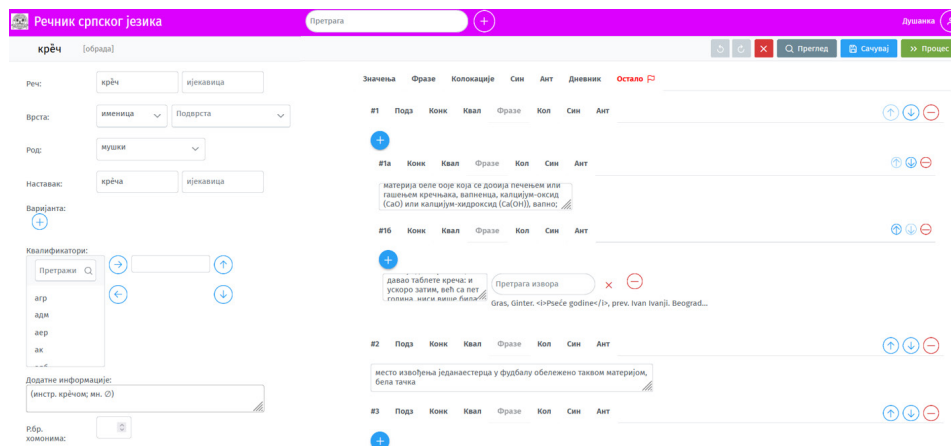
За разлику од ранијих речничких подухвата, који су у великој мери почивали на традиционалним системима прикупљања и обраде грађе, израда Речника савременог српског језика од почетка се заснива на дигиталној инфраструктури и корпусној грађи. Управо тај технолошки оквир значајно утиче на организацију лексикографског рада, начин обраде језичке грађе и структуру самог речничког чланка.

3. ДИГИТАЛНА ИНФРАСТРУКТУРА ЗА ИЗРАДУ *РЕЧНИКА САВРЕМЕНОГ СРПСКОГ ЈЕЗИКА*. Дигитална инфраструктура развијена за потребе израде *Речника савременог српског језика* заснива се на софтверским платформама за унос и обраду лексикографских података и за приказ и анализу електронског корпуса српског језика као основног извора језичке грађе. Читав систем реализован је као веб-апликација која подржава истовремени рад више корисника и централизовано управљање лексикографским подацима. Апликацији се приступа путем стандардних веб-прегледача (као што су *Mozilla Firefox*, *Google Chrome*,

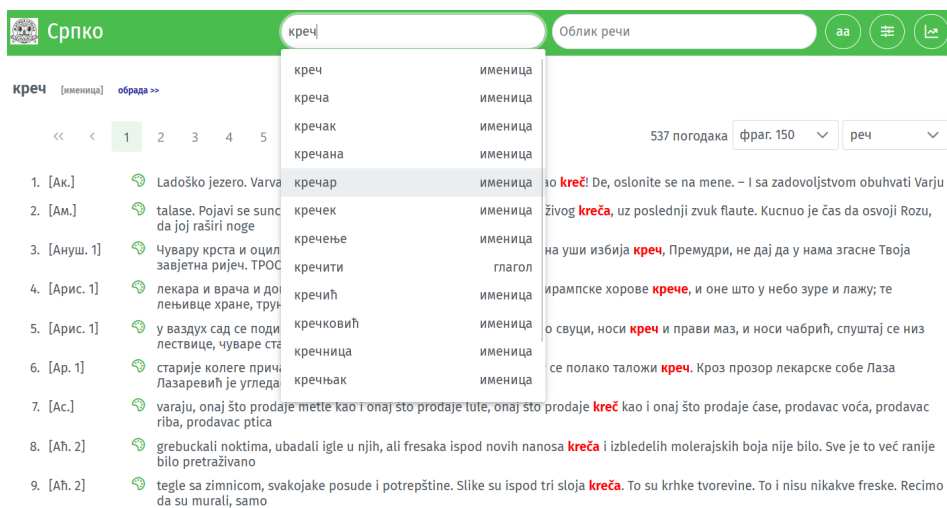
Safari, Opera, Microsoft Edge), без потребе за додатном инсталацијом софтвера на радне станице. Истовремено, корисничка права дефинисана су одвојено за корпус и за платформу за израду речника тако да се свака од платформи може користити и независно једна од друге, у ширем истраживачком контексту (VUJOVIĆ – MILOSAVLJEVIĆ 2025).

Основни принцип функционисања овог система јесте структурирање лексикографских података у оквиру базе података, што чини основу за њихову систематску организацију, претрагу и даљу обраду. За разлику од традиционалног приступа, у коме је речнички чланак представљен као линеарни текст, у дигиталном окружењу он се организује као скуп међусобно повезаних података који се уносе у посебна поља и могу се различито комбиновати у зависности од начина приказа (De SCHRYVER 2003: 156). Повезивањем овако организованих лексикографских података са корпусним примерима, успоставља се непосредна веза између описа лексичке јединице и њене употребе у аутентичном језичком контексту, што представља један од основних захтева савремене корпусно засноване лексикографије (GRANGER 2012: 3). Лексеме се тако истовремено сагледавају као структурирани скуп података и као елемент живе језичке употребе у свим контекстима у којима се налазе у корпусу.

3.1. ОРГАНИЗАЦИЈА ДИГИТАЛНОГ СИСТЕМА ЗА ИЗРАДУ РЕЧНИКА САВРЕМЕНОГ СРПСКОГ ЈЕЗИКА. За потребе израде *Речника савременог српског језика* развијен је интегрисани дигитални систем који подржава унос, обраду и систематизацију лексикографских података, као и рад са језичком грађом на којој се заснива њихова анализа. Основу система чине два функционално повезана, али концепцијски различита подсистема: софтверска платформа за израду речника (Сл. 1), намењена уређивању лексикографског чланка, и електронски корпус српског језика, који служи као извор за претрагу и избор примера употребе лексеме у процесу њене лексикографске обраде (Сл. 2 и 3). Ови подсистеми међусобно су повезани тако да корисник може директно пренети корпусни пример у структуру лексикографског чланка, што значајно олакшава избор и обраду лексичке грађе.



Слика 1: Страница за унос и обраду лексикографског чланка



Слика 2: Страница за претрагу у електронском корпусу



Слика 3: Страница за селекцију и унос корпусних примера у поља значења лексикографске одреднице

Софтвер за рад са корпусом и речником обухвата више типова података који су у систему организовани према својој намени. Структурирани подаци (са унапред познатом структуром) као што су речнички чланци, лексеме са свим својим граматичким облицима, библиографски описи изворе, статистике за извештаје, подаци о одлукама у вези са избором речи које ће ући у Речник, и подаци о корисницима и правима приступа, систематски се чувају у релационој бази података (*MariaDB*) која обезбеђује њихово ефикасно управљање и контролу. Оваква организација података омогућава јасно дефинисање улога обрађивача, редактора и уредника у процесу израде речника, контролу измена и доследност у обради лексикографских јединица.

Неструктурирани подаци, као што су документи који чине корпус и генерисани Речник у PDF и Word формату, чувају се непосредно у фајл-систему сервера на коме се извршава програм (VUJOVIĆ – MILOSAVLJEVIĆ 2025: 437).

Ефикасна претрага текстуалних података и брзо проналажење тражених речи у свим контекстима остварени су помоћу посебног сервиса развијеног специјално за ову намену. Овако устројен систем подржава ефикасно управљање обимном грађом и доприноси систематичнијој и уједначенијој лексикографској обради, јер се појединачни елементи чланка могу лакше контролисати и међусобно упоређивати.

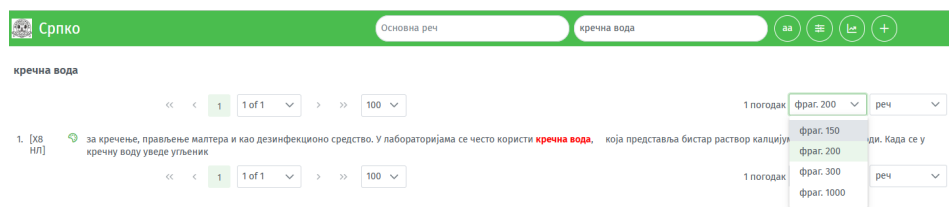
3.1.1. ПЛАТФОРМА ЗА ИЗРАДУ РЕЧНИКА. Начин организације платформе за израду речника може се сагледати на слици 1. На њој је приказано радно окружење у коме се врши унос и преглед свих елемената лексикографског чланка у оквиру јединственог система. Софтверска платформа за израду речника заснива се на принципу структурираног уноса података при чему се поједини елементи лексикографског чланка уносе у засебна поља базе података. Овакво решење доприноси прецизнијем организовању лексикографских информација, њиховом лакшем претраживању, као и аутоматском генерисању различитих приказа речничког текста. Појединачна поља за унос леме, њених варијаната, основних и додатних граматичких информација и квалификатора употребне вредности, јасно су разграничена и доступна за уређивање на левој половини екрана. На десној половини, преко менија се отварају поља за унос значења, подзначења и примери употребе у конкретном контексту, са пољима за конкретне библиографске изворе из корпуса. Поља за унос значења и подзначења организована су хијерархијски, при чему се свако од њих може самостално додавати, брисати, уређивати и премештати у оквиру структуре чланка, што чини лексикографски опис детаљним и флексибилним. Мени садржи и поља за унос квалификатора везаних за појединачна значења, за унос колокација и фразеологизама, синонима и антонима. У систем је интегрисан и дневник рада у коме је могуће пратити који лексикограф ради на којој одредници и у којој је фази обрада одређене лексеме. Картица „остало“ отвара поље за унос питања и коментара у вези са обрадом речи. То поље је обележено црвеном заставицом ради лакше уочљивости. Ту се налази још и дугме за снимање измена, као и картица „процес“ која служи да се одредница пошаље редактору или уреднику или, ако је потребно, врати обрађивачу на дораду. Ова платформа представља централно радно окружење у коме се лексикографски подаци систематизују, проверавају и обликују у коначну форму, будући да се у њој врше унос, исправке, допуне и уједначавање појединих елемената структуре у складу са унапред дефинисаном организацијом чланка и форматом предвиђеним за штампану верзију речника.

У оквиру система имплементирано је аутоматско генерисање прелоњеног текста речника у PDF и Word форматима са свим елементима макроструктуре. Тиме се обезбеђује да се у сваком тренутку добије увид у текућу, форматирану верзију речника, са бројем страница и заглављима за сваку, што значајно олакшава контролу и планирање будућег штампаног издања. Функционалност читавог система проверена је током израде *Нацрта за Огледну свеску Речника савременој српској језика* који је завршен у фебруару 2026. У њему је приказано око сто лексикографских одредница, а састављен је уз примену дигиталне платформе. Тако су у пракси тестиране све кључне

функције програма, од уноса и обраде података, преко претраге корпуса и интегрисања примера употребе у лексикографски чланак, до генерисања коначног текста речника у форми припреме за штампу. Иако су, у овој фази пројекта, активности усмерене на стварање основе за израду Речника у штампаном облику, начин на који је систем конципиран допушта да се лексикографски подаци, по потреби, прилагоде и електронском облику представљања.

3.1.2. ПЛАТФОРМА ЗА РАД СА КОРПУСОМ. Поред платформе намењене изради речничког чланка, важан део дигиталног система чини и платформа за рад са корпусом, чија је организација приказана на сликама 2 и 3. Овај део система фокусиран је на рад са језичком грађом, односно на њено претраживање, анализу и селекцију речи које ће ући у речник, као и примера употребе који ће се даље користити у лексикографској обради (Сл. 3). Платформа има два поља за претрагу (Сл. 3 и 4), прво је намењено претрази по основном облику речи, а друго претрази по било ком облику речи, по фрази, изразу или делу текста.

Приликом претраге корисник сам одређује ширину контекста у којем се приказују резултати и то могу бити фрагменти од 150, 200, 300 или 1000 речи, или читава реченица у којој се налази тражени пример (Сл. 4).



Слика 4: Поља за претрагу и подешавање ширине контекста

Резултати претраге корпуса приказују се у облику конкорданци са назначеним изворима (Сл. 2), како је и уобичајено у електронским корпусима. Маркирањем одређеног контекста може се извршити његово убацивање у поље за пример употребе изабраног значења или подзначења неке речи (Сл. 3) чиме се аутоматски пример појављује и у платформи за израду Речника. Заједно са примером употребе преноси се и информација о извору у ком се налази.

Поткорпус	Број извора	Број речи	Удео
Књижевни	774	47.237.445	27,02%
Новински	163.451	95.922.558	54,87%
Разговорни	5.690	26.443.513	15,13%
Научни	10.070	4.163.683	2,38%
Административни	153	1.065.547	0,61%
УКУПНО	180.138	174.832.746	100,00%

Слика 5: Преглед статистике по поткорпусима

Грађа корпуса груписана је по поткорпусима (књижевни, новински, разговорни, научни и административни), што служи као основ за анализу

употребе лексике у различитим функционалним и стилским доменима. Укупан број речи у корпусу, за сада је 174.832.746. На слици бр. 5 види се тренутни преглед статистике по поткорпусима у коју су укључени број извора, број речи и процентуални удео сваког од поткорпуса у укупној грађи. Заступљеност појединих поткорпуса одражава, с једне стране, свесну одлуку Уредништва о избору извора, а с друге, објективну доступност грађе у дигиталном облику. Избор књижевних дела, превода и стручне литературе сачинила је Комисија Матице српске, док је избор електронских и осталих извора био у надлежности лексикографског тима.

Прикупљање грађе врши се комбиновањем више приступа: преузимањем постојећих дигиталних текстова, дигитализацијом штампаних извора и транскрипцијом аудио-визуелног материјала. Разговорни део корпуса, у највећој мери формиран је аутоматском транскрипцијом садржаја са интернет платформи (емисије и поткасти на Јутјубу), уз накнадну проверу и делимичну ручну обраду. Важно је нагласити да корпус није коначно затворен систем. Његово формирање је још увек у току, пре свега због ограничене доступности појединих текстова у дигиталном облику, нарочито у домену књижевности. Велики број књижевних извора мора ручно да се дигитализује, што је веома захтеван и дуготрајан процес. У том смислу, постојећи статистички подаци представљају тренутно стање, које се континуирано мења у складу са даљим богаћењем корпуса.

Осим претраге речи, у платформу за корпус интегрисане су и напредне функције филтрирања резултата и генерисања извештаја (статичких и динамичких), на основу различитих параметара, као што су учесталост појављивања, припадност поткорпусу или статус обраде лексеме. Ови алати имају важну улогу у процесу селекције лексичке грађе. Захваљујући њима, платформа за рад са корпусом превазилази функцију класичног претраживача текстова и постаје аналитички инструмент који омогућава систематско праћење употребе речи и доношење методолошки утемељених и аргументованих одлука о избору речи које треба да уђу у Речник.

Овако устројена дигитална инфраструктура значајно мења организацију лексикографског рада. Она уводи истовремени рад већег броја лексикографа у заједничком систему, систематско праћење обраде лексичких јединица и њихову редактуру, као и генерисање различитих извештаја о унетим подацима.

3.2. СТРУКТУРА ЛЕКСИКОГРАФСКЕ ОДРЕДНИЦЕ У ДИГИТАЛНОМ ОКРУЖЕЊУ. Увођење дигиталног окружења у израду речника не утиче само на технички аспект лексикографског рада, већ доводи и до промена у начину на који се лексикографска одредница концептуализује и обликује. За разлику од традиционалног модела, у коме је одредница схватана као јединствен текстуални низ, у дигиталном систему она се успоставља као резултат поступног повезивања различитих типова података који се уносе и обрађују током рада. Такав приступ омогућава већу прецизност у разграничењу значења, доследнију употребу квалификатора и систематичнији избор примера, будући да се сваки елемент одреднице формира посебно и у директној вези са језичком грађом.

Илустративан пример овакве организације види се на сликама 6 и 7. На њима се налази страница за обраду лексеме *дигитализовати*. Видимо да

су лема и њена граматичка обележја (Сл. 6), значења и примери употребе (Сл. 7) унети у засебна поља. Значења су организована хијерархијски, са могућношћу даљег гранања на подзначења, док се уз свако значење додају одговарајући примери из корпуса са назначеним изворима.

The screenshot shows the 'Речник српског језика' (Serbian Language Dictionary) interface. The main section is titled 'дигитализовати' (digitize) with a sub-label '[обрада]' (processing). It contains several input fields for entering a lemma and its grammatical features:

- Реч:** (Word) with a text input field containing 'дигитализовати' and a checkbox for '(се)' (se).
- Врста:** (Type) with a dropdown menu set to 'глагол' (verb).
- Наставак:** (Suffix) with a text input field containing '-ујем' and a checkbox for 'ијекавица' (Ijekavian).
- Презент:** (Present) with a text input field containing 'екавица' and a checkbox for 'ијекавица' (Ijekavian).
- Варијанта:** (Variant) with a blue plus icon and a dropdown menu.
- Глаголски вид:** (Verb form) with a dropdown menu set to 'свршен и несвршен' (perfective and imperfective).
- Глаголски род:** (Verb gender) with a dropdown menu set to 'Глаголски род' (Verb gender) and a checkbox for 'прикажи' (show).
- Квалификатори:** (Qualifiers) with a checkbox for 'има посебна значења за - се' (has special meanings for - se).

At the bottom, there is a search bar with the label 'Квалификатори:' and a 'Претражи' (Search) button.

Слика 6: Изглед поља за унос леме, њених граматичких обележја и квалификатора

The screenshot shows the 'Речник српског језика' (Serbian Language Dictionary) interface for the word 'дигитализовати' (digitize). The main section is titled 'дигитализовати' with a sub-label 'глагол' (verb). It contains several tabs for different views: 'Значења' (Meanings), 'Значења (се)' (Meanings (se)), 'Фразе' (Phrases), 'Колокације' (Collocations), 'Син' (Syntax), 'Ант' (Antonyms), 'Дневник' (Journal), and 'Остало' (Other). The 'Значења' tab is selected, showing a list of meanings and their usage examples:

- #1** Подз: Конк: Квал: Фразе: Кол: Син: Ант:
 - пребацити, пребацивати (податке, слике, звуке и сл.) у дигитални (машински читљив) запис, тзв. бинарни код.
 - План је да се дигитализује целокупни архив јавног сервиса ... У: <Nova ekonomija>/i>, 2021.
 - снимали и дигитализовали грађу. <Kosovo и Метохија: јуче, данас, сутра>/i>
 - САНУ дигитализује своје збирке у којима се <Nova ekonomija>/i>, 2021.
 - овима ... да ти дигитализујем и он је заиста све то мени <Agelast>/i>, 2022. (YouTube)

Слика 7: Изглед поља за унос значења речи, примера употребе и извора

Оваква организација олакшава уједначавање лексикографских решења унутар читавог речника, јер се појединачне одлуке могу контролисати и упоређивати у оквиру јединственог система. Овај помак може се довести у везу са савременим лексикографским приступима који указују на то да се структура речника и његових јединица мора преиспитати у складу са функцијама које речник треба да оствари. То подразумева и редефинисање неких од традиционалних структурних решења као што су линеарна организација чланка или алфабетни распоред одредница (уп. Gouws 2009: 4; Tarp 2012: 113).

Посебно место у оваквом приступу има хијерархијска организација значења, која омогућава да се значењска структура лексеме постепено разрађује и прецизније диференцира. Основна значења могу се даље гранати у подзначења, при чему се сваки ниво описује у непосредној вези са конкретним примерима употребе. Управо у том сегменту долази до изражаја улога корпуса као примарног извора података, будући да је избор примера заснован на систематичној анализи аутентичне језичке грађе. Корпусни примери служе не само за илустрацију значења, већ и његово прецизније разграничење, јер указују на различите контексте употребе, колокацијске обрасце и стилску вредност лексеме. Стога се структура одреднице не формира унапред као затворен систем, већ се постепено обликује у процесу рада, у интеракцији између лексикографске анализе и корпусне грађе. Илустративан пример таквог процеса пружа лексема *йахуљица*, која је у корпусу потврђена у низу секундарних значења, од назива за прехранбени производ од житарица,

преко назива за најмлађи узрасни разред у скијашком такмичењу и замрзнути ембрион, до експресивног означавања припадника миленијумске генерације. Управо корпус омогућава да се уоче све ове употребе и да се на основу броја и распоређености потврда процени да ли заслужују статус засебног значења у Речнику или представљају казационалну појаву. Тако је нпр. лексема *йахуљица* у корпусу потврђена са око 700 примера, од чега се око 170 односи на значење 'прехрамбени производ од житарица', које у ранијим Матичиним речницима није забележено. Фреквенција од 170 појављивања поуздан је показатељ да значење 'прехрамбени производ од житарица' треба да се узме у обзир приликом обраде речи *йахуљица*. Као што наглашавају Аткинс и Рандел (2008: 48), само употреба која је истовремено учестала и распрострањена у различитим типовима текста може поуздано бити сматрана делом стабилног лексичког језгра, а не пролазном или казационалном појавом. Тако корпус не служи само као извор примера, већ непосредно учествује у обликовању значењске структуре одреднице.

У том контексту, велики значај има начин на који се корпусна грађа укључује у процес израде лексикографске одреднице. У оквиру развијеног система предвиђено је да се примери употребе не бирају издвојено од процеса лексикографске обраде, већ да се селекција врши унутар посебног окружења за рад са корпусом, у коме су резултати претраге приказани у облику конкорданци. Лексикограф може на основу таквог приказа да анализира различите контексте употребе, упоређи примере и одабере оне који најпрецизније илуструју одређено значење или подзначење (уп. PRINSLOO 2012: 184). Одабрани примери затим се непосредно повезују са одговарајућим сегментима лексикографског чланка и преносе у платформу за његову израду, где се даље уређују и прилагођавају финалном облику (Сл. 3). На тај начин успоставља се континуитет између анализе језичке грађе и формулисања лексикографског описа, чиме се знатно умањује могућност произвољног избора примера и обезбеђује већи ослонац одреднице на стварну језичку употребу. Овакво решење може се посматрати као значајан искорак у односу на традиционалну лексикографску праксу јер се процес ексерпције примера и лексикографска обрада лексеме одвијају у јединственом радном окружењу.

4. ЗАКЉУЧНА РАЗМАТРАЊА. Развијена дигитална инфраструктура за израду Речника савременог српског језика показује да увођење дигиталних језичких ресурса и алата не представља само техничко унапређење лексикографског рада, већ доводи и до промена у његовој организацији и методологији. Интеграција електронског корпуса са софтверском платформом за унос и обраду лексикографских података омогућава систематско прикупљање и анализу језичке грађе, као и њено непосредно укључивање у структуру лексикографског описа, нарочито кроз избор и повезивање корпусних примера са одређеним значењима у лексикографском чланку.

У таквом дигиталном окружењу лексикографска одредница више се не обликује као јединствен линеарни текст, већ као структура међусобно повезаних података који се постепено формирају у процесу рада. Овакав начин организације доприноси већој прецизности у опису значења, уједначенијој

формулацији појединих елемената одреднице и већој контроли над током лексикографске обраде.

Истовремено, упркос значајној улози дигиталних алата, кључна улога у обликовању лексикографског чланка остаје у домену лексикографа. Формирање и разграничење значења, утврђивање њихових међусобних односа, избор релевантних примера, као и издвајање фразеолошких и других типичних употреба, заснивају се на стручној процени лексикографа и на његовој интерпретацији језичке грађе. Дигитално окружење у том процесу не може га заменити, већ му пружа непосредан и систематичан увид у обимну корпусну грађу, различите извештаје засноване на њој, већу прегледност података и флексибилније обликовање појединих елемената одреднице. На тај начин лексикограф је у позицији да на основу ширег и поузданијег увида у језичку употребу доноси утемељене и доследне одлуке у процесу обраде лексеме.

Досадашњи рад на Речнику савременог српског језика био је усмерен на постављање поузданих теоријских, методолошких и техничких основа за његову израду, а као конкретан резултат тог рада настао је Нацрт за Огледну свеску. Израда Нацрта показала је да су у оквиру развијеног система успешно проверене кључне функције уноса, обраде и обликовања лексикографских података у речнички чланак, чиме је потврђена практична применљивост система у лексикографском раду. У том смислу, развијена дигитална инфраструктура може се посматрати као поуздан оквир за даљи рад на Речнику и истовремено као подстицај за постепено унапређивање српске лексикографске праксе у условима дигиталног окружења, не као замена за лексикографски рад, већ као његово методолошко унапређење. Рад показује да интеграција корпусних ресурса и лексикографске обраде у јединственом дигиталном окружењу представља ефикасан модел за организацију савременог лексикографског рада.

ЦИТИРАНА ЛИТЕРАТУРА

- Вуловић, Душанка, Бранко Милосављевић. О програму за унос и обраду речи новог вишетомника Матице српске из лексикографског угла. *Актуелна истраживања лексикологије и лексикографије српског језика*. Андрићград: Андрићев институт. 2023, 133–149.
- Драгићевећ, Рајна. Предговор за Нацрт Огледне свеске Речника савременог српског језика (необјављени текст). 2026.
- РМС: *Речник српскохрватског књижевног језика*, I–VI. Нови Сад: Матица српска (и Матица хрватска I–III), 1967–1976.
- РСАНУ: *Речник српскохрватског књижевног и народног језика*, 1–. Београд: САНУ – Институт за српски језик САНУ, 1959–.
- РСЈ: *Речник српског језика*. Нови Сад: Матица српска, 2007.

*

- ATKINS, B. T. Sue, Michael RUNDELL. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press, 2008.
- DE SCHRYVER, Gilles-Maurice. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16/2 (2003): 143–199.
- FUERTES-OLIVERA, Pedro A. Lexicography in the Internet era. Pedro A. Fuertes-Olivera (ed.). *The Routledge handbook of lexicography*. London: Routledge, 2018, 1–14.

- GOUWS, Rufus H. Sinuous Lemma Files in Printed Dictionaries: Access and Lexicographic Functions. Sven Tarp (ed.). *Lexicography in the 21st Century*. Amsterdam – Philadelphia: John Benjamins Publishing Company, 2009, 3–21.
- GRANGER, Sylviane. Introduction: Electronic Lexicography – From Challenge to Opportunity. Sylviane Granger, Magali Paquot (eds.). *Electronic Lexicography*. Oxford: Oxford University Press, 2012, 1–11.
- PRINSLOO, Danie J. The Role of Corpora in Future Dictionaries. Sandro Nielsen, Sven Tarp (eds.). *Lexicography in the 21st Century*. Amsterdam – Philadelphia: John Benjamins Publishing Company, 2012, 181–206.
- TARP, Sven. Theoretical challenges in the transition from lexicographical p-works to e-tools. Sylviane Granger, Magali Paquot (eds.). *Electronic Lexicography*. Oxford: Oxford University Press, 2012, 107–118.
- VUJOVIĆ, Dušanka, Branko MILOSAVLJEVIĆ. Korpus „Srpko” kao tehnološka osnova za izradu Rečnika savremenog srpskog jezika Matice srpske. Jasmina Moskovljević Popović, Ranka Stanković (eds.). *South Slavic Languages in the Digital Environment (JuDig)*. Beograd: Filološki fakultet, 2025, 425–440.

Dušanka Vujović

DIGITAL INFRASTRUCTURE OF THE DICTIONARY OF CONTEMPORARY SERBIAN

S u m m a r y

This paper presents the digital infrastructure developed for the compilation of the *Dictionary of Contemporary Serbian Language* (Matica srpska), aimed at the systematic description of the contemporary lexicon of standard Serbian. The system is based on the integration of a purpose-built software platform for lexicographic data processing and a specially designed electronic corpus serving as the primary source of linguistic evidence. The paper outlines the organization of the platform, focusing on structured data input and the representation of dictionary entries as interconnected data sets rather than linear textual units. Particular attention is given to the role of the corpus in the lexicographic process, especially in the selection, analysis, and direct incorporation of authentic usage examples into dictionary entries. The digital environment supports a more systematic and transparent lexicographic workflow, contributing to greater precision in meaning differentiation and consistency in description. At the same time, the central role remains with the lexicographer, whose expert judgment governs the identification, formulation, and organization of meanings. The system has been tested through a draft sample booklet (*Nacrt za Oglednu svesku*), comprising approximately one hundred entries, confirming its practical applicability and establishing a reliable foundation for further work on the dictionary.

Универзитет у Новом Саду
Филозофски факултет
Одсек за српски језик и лингвистику
Др Зорана Ђинђића 2, 21000 Нови Сад, Србија
dusanka.vujovic@ff.uns.ac.rs

(Примљено: 31. марта 2026;
прихваћено: 19. маја 2026)