

Цветана Ј. КРСТЕВ
Филолошки факултет
Универзитета у Београду

О ОДАБИРУ ОДРЕДНИЦА ЗА ЕЛЕКТРОНСКИ РЕЧНИК СРПСКОГ ЈЕЗИКА И ЊИХОВОМ ПОВЕЗИВАЊУ*

Потпуно, или што потпуније, препознавање облика речи у писаним текстовима је један од предуслова успешне аутоматске обраде текста. При томе се под препознавањем не подразумева пуко одређивање да ли реч припада неком језику или не, већ се претпоставља да ће за препознате облике речи бити позната и њена својства (канонски облик или лема, граматичка улога, значење, доменско одређење и слично). Електронски речник српског језика је изграђен управо с оваквим циљем. Да би се електронски речник могао дограђивати и допуњавати новим информацијама неопходно је адекватно радно окружење које се заснива на лекикографској бази података. У овом раду представљена је лексикографска база и апликација ЛЕКСИМИРКА и њене могућности на једном примеру обогаћивања садржаја речника и повезивања одредница.

Кључне речи: аутоматска обрада српског језика, електронски речници, корпусна анализа, лексикографска база

О садржају и формату електронског речника српског језика

Рачунарска обрада и разумевање текста на природном језику се, у зависности од постављеног задатка и примењених метода, у већој или мањој мери ослањање на адекватно осмишљене и изграђене лексичке ресурсе. Електронски речници (у даљем тексту е-речници) у смислу (Грос и Перен 1987) су незаменљив лексички ресурс за ове намене. Улога електронских речника није да само препознају да ли речи текста припадају језику (као бинарна функција да/не) већ и да препознатим речима придруже што више корисних информација (врста речи, граматичка својства, домен примене и слично). Да

* Овај рад приказује резултате постигнуте током 2018. и 2019. године у оквиру пројекта Српски језик и његови ресурси (178006) који финансира Министарство просвете, науке и технолошког развоја Републике Србије.

би остварили ову улогу, е-речници треба да буду што обухватнији и богатији чиме се број непрепознатих речи у обрађиваном тексту смањује. Математичким језиком речено, ако је T скуп свих облика речи које се јављају у једном тексту, а R скуп свих облика речи обухваћених е-речником, оптималан однос ова два скупа је $R \cap T = T$, или $T \subseteq R$. Овај циљ је, међутим, тешко достижан, а разлози за то су многобројни што је детаљно образложено у (Витас 2007).

На основу реченог следи да један е-речник треба да садржи све (или већину) облика речи које се могу наћи у тексту, што значи да њихов канонски представник, односно лема, није довољан. Ово је и суштинска разлика између електронских и традиционалних (штампаних) речника који су намењени људима. Приликом одабира одредница за традиционалне речнике и информација које ће им бити придружене претпоставља се да ће просечан корисник, коме је речник намењен, без већих проблема надопунити све оне информације које нису експлицитно наведене или да ће моћи да их пронађе у одговарајућим приручницима, на пример, граматикама.

Ову разлику ћемо илустровати на примеру глагола *певати*. Одредница овог глагола у традиционалном речнику почиње следећим информацијама:

пѐвати , -âм, ијек. пјѐвати , несврш.	(РМС)
пѐвати јек. пјѐвати , -âм несвр.	(ЈРМС)
пѐвати ијек. пјѐвати , -âм (дијал. ијек. пљевати) несврш.	(РСАНУ)

На основу датих информација, као и оних које се подразумевају, читалац лако закључује да је у питању несвршени глагол чији је облик инфинитива *певати* у екавској варијанти, *пјевати* у ијекавској, те да је прво лице једнине презента *певам*. На основу овога ће читалац закључити да се глагол *певати* (као и *пјевати*) мења по V глаголској врсти те му неће бити проблем да изведе остале облике овог глагола.

Опис овог глагола у е-речнику изгледа другачије. Пре свега, глаголи *певати* и *пјевати* се морају одвојено описати, а све остале информације морају бити исцрпно дате и то на такав начин да их рачунарска апликација којој је намењен може недвосмислено протумачити. У табели 1 је приказан опис ова два глагола у е-речнику српског језика. У горњем делу табеле су глаголи *певати* и *пјевати* означени као несвршени (маркер +Impref) који припадају екавском (маркер +Ek), односно ијекавском (маркер +Ijk) изговору.¹ Маркер VI говори да су у питању глаголи, а осим тога их повезује са регуларним изразом, у табели представљеним у облику графа, који омогућава генерисање свих њихових облика. У доњем делу табеле приказани су генерисани облици.² Уз сваки генерисани облик, на пример певала, дата је његова лема, у овом случају облик инфинитива певати, а затим и придружене информације: ради се о глаголу (V), радном глаголском придеву у средњем роду множине (Gpn) или женском роду једнине (Gsf).

¹ У е-речнику српског језика акценти нису забележени, јер је речник намењен обради писаних текстова који их такође не бележе.

² Читљивости ради, графом је представљено генерисање само неких облика глагола; ти облици су и представљени у доњем делу табеле.

Опис глагола *певати* и *пјевати* дат у горњем делу табеле служи за систематизовање информација у речнику и његово одржавање, док се облик дат у доњем делу табеле ефективно користи за рачунарску обраду текста. Више о структури е-речника и информацијама у њему може се наћи у (Крстев 2008). Овде приказан формат е-речника није једини који се користи. У аутоматској обради се често користи и такозвани MULTTEXT формат (Крстев и др. 2004). Без обзира на формат е-речника све њих одликује исцрпност и експлицитност бележења информација.

Леме (речник Delas)	
певати, V1+Imperf+Ek пјевати, V1+Imperf+Ijk	
Облици (речник Delaf)	
певати, певати. V:W; певам, певати. V:Pxs; певају, певати. V:Pzp; певаше, певати. V:Yys:Azp; певај, певати. V:Yys; певајте, певати. V:Yyp; певајући, певати. V:S; певао, певати. V:Gsm; певала, певати. V:Gpn:Gsf; певаћемо, певати. V:Fxp; певаћете, певати. V:Fyp; певан, певати. V:Tms; певано, певати. V:Tns	
пјевати, пјевати. V:W; пјевам, пјевати. V:Pxs; пјевају, пјевати. V:Pzp; пјеваше, пјевати. V:Yys:Azp; пјевај, пјевати. V:Yys; пјевајте, пјевати. V:Yyp; пјевајући, пјевати. V:S; пјевао, пјевати. V:Gsm; пјевала, пјевати. V:Gpn:Gsf; пјеваћемо, пјевати. V:Fxp; пјеваћете, пјевати. V:Fyp; пјеван, пјевати. V:Tms; пјевано, пјевати. V:Tns	

Табела 1. Глаголи *певати* и *пјевати* у е-речнику српског језика.

О одабиру одредница за електронски речник српског језика

Одреднице глагола се у традиционалним, штампаним речницима наводе, по правилу, у облику инфинитива, док се други глаголски облици не појављују као засебне одреднице, осим изузетно, на пример,

ће, енкл. 3. л. јд. и мн. през. од *хтети* (PMC)

ће, енклитички облик 3. л. јд. и мн. през. од *хтети* (JPMC)

У речничком чланку глагола се по правилу наводи облик или наставак првог лица јединине презента, док се неки други облици, али никада сви, наводе изузетно у опису промене глагола. На пример,

вѣсти , вѣзѣм несврш. (р. прид. вѣзао, вѣзла, вѣзло; трп. прид. вѣзен)	(РМС)
вѣсти (вѣсти), вѣзѣм несвр. (трп. прид. вѣзен)	(ЈРМС)
вѣсти и вѣсти , вѣзѣм несвр. (имп. вѣзи; импф. вѣзијах и вѣзѣх; аор. вѣзе; р. пр. вѣзао, вѣзла, вѣзло; пр. пр. вѣзѣв(ши); трп. вѣзен (непр. вѣжен))	(РСАНУ)

С друге стране, трпни придеви, као и радни глаголски придеви и прилози времена садашњег употребљавају се некада и као прави придеви. У граматичици (Стевановић 1966) наводи се да се радни глаголски придев употребљава и у служби именских речи (т. 672), да се трпни придев као и радни употребљава придевски (т. 676), а да облик глаголског прилога садашњег, осим прилошке, има и придевску службу (т. 680). Такође, у граматичици (Станојчић/Поповић 2010) је речено да: „Прави описни придеви настају и попридевљавањем четири нелична глаголска облика – трпног и радног глаголског придева и садашњег и прошлог глаголског прилога, дакле, остваривањем трансформационог принципа” (т. 328). У граматичици (Пипер/Клајн 2014) наводи се да трпни глаголски придев, осим што служи за грађење партиципског пасива, може да се, у принципу, употреби и атрибутивно (т. 319), те да се изванредно, мањи број радних придева од непрелазних глагола употребљава и атрибутивно, аналогно правим придевима (т. 315). У истој граматичици се наводи да се „облик са истом основном као глаголски прилог несвршеног вида понекад употребљава у придевској функцији придева са значењем сталне особине” (т. 328).

Четири нелична глаголска облика, трпни и радни глаголски придев и садашњи и прошли глаголски прилог, наводе се у речницима као засебне одреднице у појединачним случајевима. Тако се у предговору (РМС 1967-1976: књ. I стр. 12) наводи: „Глаголски придеви који су постали прави описни придеви, као *зрео*, *врео*, *утучен*, *отворен*, *виђен* уносили су се као посебне речи”, а у предговору (РСАНУ: књ. I стр. XXXIII): „Глаголски придеви (по пореклу) дају се као засебне одреднице када су постали прави придеви (напр. *врућ*, *зрео*, *држећ*, *васпитан*, *питом*, *ваљан*, итд.)”.

У табели 2 су дати примери обраде придева насталих попридевљавањем из неличних глаголских облика. За придеве настале од трпних радних придева, ако су обрађени, у речницима РМС и ЈРМС се наводи да представљају облик трпног придева одговарајућег глагола док се у речнику РСАНУ то некада наводи (видети примере за *исеван* и *основан*). За придеве настале од радних глаголских придева се у речнику РМС не наводи веза са одговарајућим глаголом, док се у ЈРМС та веза некада експлицитно наводи, а у неким случајевима је то и једини опис одговарајућег облика (видети пример *отупео*). Слично је и са речником РСАНУ (видети примере *доснео* и *отупео*). За придеве настале од прилога времена садашњег се ни у једном речнику не наводи веза са одговарајућим глаголом, а такав је случај и са малобројним примерима придева насталих од прилога времена прошлог – на пример, у речнику ЈРМС

се наводи само један такав пример *бивши*.³ Међу узорком примера из РСАНУ јавља се само придев *дошавши* (и *новодошавши*) настао од прилога времена прошлог, док се у СрпКор јављају осим њих још и *почивши* (и *блаженопочивши* и *благопочивши*). У РСАНУ *дошав(ши)* је засебна одредница (која само упућује на *доћи*), као и *новодошавши*.

испеван	РМС	испеван , -а, -о, ијек. испјеван 1. <i>трп. прид. од испевати 2. који је од много певања изгубио глас</i>
	ЈРМС	∅
	РСАНУ	испеван ијек. испјеван , -а, -о који је изгубио свежину, истрошен, промукао од певања
основан	РМС	основан , -а, -о 1. <i>трп. прид. од основати 2. заснован на разлозима, поткрепљен разлозима, доказима, оправдан</i>
	ЈРМС	основан , -а, -о 1. <i>трп. прид. од основати 2. на разлозима, на аргументима, на доказима утемељен, разложан, оправдан; ~ сумња, ~ захтев</i>
	РСАНУ	основан , -а, -о 1. <i>трп. прид. од основати (се) 2. заснован на разлозима, утемељен, оправдан 3.</i> <i>покр. који је почео да се пара, из ког испадају конци</i>
распеван	РМС	распеван , -а, -о, ијек. распјеван 1. <i>трп. прид. од распевати (се) 2. а. весео, ведар; испуњен песмом б. који се богато, звучно прелива, сличан певању (о гласу) в. пун песничког надахнућа, поезије.</i>
	ЈРМС	распеван јек. распјеван , -а, -о 1. <i>трп. прид. од распевати (се) 2. весео, ведар; испуњен песмом 3. који се звучно прелива, мелодичан.</i>
презрео	РМС	презрео , -ела, ело (ек. и ијек.; ретко ијек. презрио) 1. <i>превише зрео 2. а</i> <i>фиг. који је прешао године живота уобичајене за какво стање, занимање и сл., престар (о људима) б. који је прерано постао зрео, који показује својства зрелости пре уобичајеног времена.</i>
	ЈРМС	презрео , -ела, о (ретко јек. презрио, -рела) 1. <i>превише зрео; уп. презрети: ~ јабука 2. а</i> <i>фиг. који је прешао године живота уобичајене за какво стање, занимање и сл., престар: ~ девојка б. који је прерано постао зрео; ~ дечак.</i>

³ Прилог времена прошлог глагола *бити* се данас ретко користи. Ипак, међу примерима из РСАНУ могу се наћи и ови примери: „Постао подофициром, па, бивши рањен ... врати се кући с нешто пензије.” (Симо Матавуљ *Биљешке једног писца*, Београд, 1939) (уз одредницу *пензија*) и „Бивши у Карловци горњи изиђем на гроб нашега песника будилника, Лукијана Мушицког.” (у листу Јавор 1874) (уз одредницу *будилник*).

отупео	РМС	отупео , -ела, -ело ијек. отупио, -пјела, -пјео а. <i>туп, неосетљив, огугало</i> б. <i>који је изгубио оштрину, интензитет</i>
	ЈРМС	отупео , -ела, -о јек. отупио , -пјела, -о <i>рад. прид. од отупети у атриб. функцији</i>
	РСАНУ	отупео , -ела, -ело ијек отупио, -пјела, -пјео 1. а. <i>који не реагује на подражаје, туп</i> б. <i>који је изгубио оштрину, интензитет</i> в. <i>у коме се не виде осећања, реакције на подражаје, безизражајан</i> . 2. <i>у прид. служби: р. пр. од отупети (1а).</i>
доспео	РМС	∅
	ЈРМС	доспео , -ела, -о јек. доспио , -пјела, -пјело а. <i>који се потпуно развио, зрео; стасао: ~ воће, ~ деца</i> . б. <i>фин. који је доспео, приспео за плаћање, исплату: ~ меница, доспели дугови.</i>
	РСАНУ	доспео , -ела, -о ијек. доспио , -пјела, -пјело а. <i>који се потпуно развио, који је постао погодан за одређену намену; стасао, зрео: доспели усев, доспело воће, доспела деца</i> . б. <i>фин. који је приспео за плаћање, исплату.</i>
претећи	РМС	претећи -а, -е ијек. пријетећи , <i>који је пун претње; ~ говор, ~ став</i>
	ЈРМС	претећи јек. пријетећи , -а, -о <i>који је пун претње; ~ писмо, ~ гест</i>
важећи	РМС	∅
	ЈРМС	важећи , -а, -е <i>који још важи, који је на снази, пуноважан: ~ пропис, ~ исправа, ~ гест</i>
	РСАНУ	важећи , -а, -е <i>који важи, који вреди, који је на снази; исп. постојећи, пуноважан.</i>

Табела 2. Примери одредница из речника РМС, ЈРМС и РСАНУ које се одnose на придеве настале од трпних и радних глаголских придева и прилога времена садашњег.⁴

Ова анализа начина описа придева насталих попридевљавањем неличних глаголских облика у речницима српског језика, илустрована табелом 2, мора се узети условно. Тек када текстови ових речника буду доступни у адекватном дигиталном облику моћи ће да се уради темељна анализа.

⁴Симбол ∅ значи да не постоји као засебна одредница; примери из РСАНУ нису дати за оне облике који нису ни могли бити укључени до двадесетог тома.

Као што је речено у претходном одељку, организација и садржај електронских речника разликује се од традиционалних. Тако у е-речнику сви облици глагола треба да буду засебне одреднице којима се представник – лема, то јест глагол у инфинитиву – придружује као једна од информација. Поставља се, међутим, питање, шта треба да буде тај представник за рецимо *распеван*, а шта за *опеван*? Ако су у питању облици трпног глаголског придева, онда је јасно да то треба да буду облици инфинитива *распевати* и *опевати*:

распеван, распевати. V: Tms

опеван, опевати. V: Tms⁵

Питање је шта треба да буде тај представник у случајевима када се ови облици користе као прави придеви, имајући у виду да је *распеван* засебна одредница у традиционалним речницима, док *опеван* није. На пример:

...prevagu je davao tanano vedroj muzici Mocarta i kristalno jasnoj muzici Johana Sebastijana Baha, a ne Vagnerovoj monumentalno **raspevanoj**. (Ото Бихаљи Мерин, *Градитељи модерне мисли*, СрпКор⁶)

Sve što je 'mrtvo' uzdrhtalo je . Ne samo **opevane** zvezde, mesec, šume, cveće, nego i pikavac u pepeljari... (Ото Бихаљи Мерин, *Градитељи модерне мисли*, СрпКор)

Слична питања се намећу и у случајевима радних глаголски придева и прилога времена садашњег када се користе као прави придеви:

Добро **ускисао** и **порастао** хлеб је шупљикав, растрешен, бухав ... лакше сокове прима, па се брже и свари. (др М. Батут-Јовановић, *Сељачки хлеб и задружне пећи*, 1905, пример из САНУ уз *бухав*)

Postadosmo **putujuć**i pisari i uzemoso seliti svoja pera i črnilnice preko voda i granica carevina (Милорад Павић, *Хазарски речник*, СрпКор)

Одговори на ова питања се не могу наћи у традиционалним речницима због начина на који се у њима обрађују примери попридевљених глаголских придева и прилога – они се јављају или уз засебну придевску одредницу ако је она у речнику дата, или, у неким случајевима уз глаголску одредницу. На пример, у речнику РСМ срећемо следећи случај:

зачађати, -ам, сврш. = зачађити *зачађавити*. – Око **зачађала** стакла ... облијеће жућкаста лептирица. *Андр. И. Фиг.* Можда има још која искра у ... зачађалу његову срцу. *Коз. Ј.*

На узорку примера из речника САНУ који обухвата књиге 1, 2, 18, 19 и 20⁷ и садржи близу 2 милиона речи проналази се 50 примера употребе при-

⁵ Код Tms у е-речнику српског језика говори да се ради о облику трпног глаголског придева у мушком роду једине.

⁶ Корпус савременог српског језика СрпКор2013 (<http://korpus.matf.bg.ac.rs>) (Витас и Крстев 2012), (Утвић 2014)

⁷ Оваква анализа је омогућена отпочетом дигитализацијом РСАНУ и рашчлањавањем његовог садржаја (Стиловић и др. 2017), (Стиловић и Станковић 2018).

дева насталог од прилога времена садашњег.⁸ Међу овим примерима наилазимо на следеће случајеве:

1. Пример употребе придева насталог од прилога времена садашњег је дат уз глагол:

Чујеш и како глас бруји ... и како расте и пада као брз талас **бујајућих** вода. (Ivan Kozarac, Proza, 1947) – уз одредницу *бујати*.

2. Придева настао од прилога времена садашњег је засебна одредница уз коју је дат и одговарајући пример:

Очаравајућа способност вечнога живота и не изгледа више баш тако неприступачна. (Борислав Пекић, *Златно руно IV*, 1980) – уз одредницу *очаравајући*.

3. Пример употребе придева насталог од прилога времена садашњег је дат уз неку другу одредницу:

У овом се каже да **лутајуће** мине могу бити или вежбовне мине ... или мине положене пре новембра 1918 године. (Mornarički glasnik, god. I, Beograd 1933) – уз одредницу *вежбовни*.

Сличне примере срећемо у РСАНУ и у случају придева насталих од радних глаголских придева (261 случај): примери се јављају уз глагол (**васкрслој Србији уз васкрнути**), придев је засебна одредница (**оћелавелим плишем уз оћелавео**), или се примери јављају уз неке друге одреднице (**застарелим обичајем уз повлачити**), док су у случају придева насталих од трпних глаголских придева (96 случајева) придеви или засебне одреднице (**виђенијим кућама уз виђен**) или се јављају уз неке друге одреднице (**одабраније писце уз по**).

Из овога проистиче да би се питање попридевљених неличних глаголских облика могло у е-речнику решавати на више начина. Једна могућност је да се они игноришу, али већ и површна анализа њиховог појављивања у текстовима на савременом српском језику (видети податке из табеле 3) показује да би то довело у аутоматској анализи текстова до великог броја непознатих речи.

текст	трпни	радни	прилог в.с.
Д. Албахари <i>Мамац</i>	141	6	0
В. Арсенијевић <i>У потпалуљу</i>	161	13	12
Д. Киш <i>Пешчаник</i>	295	29	4
Г. Олујић <i>Излет у небо</i>	172	24	4
А. Тишма <i>Употреба човека</i>	701	66	9

⁸ Овде смо се ограничили само на оне придеве настале од прилога времена садашњег који су већ засебне одреднице у е-речнику (видети образложење даље у тексту) и на оне њихове облике који се не поклапају с обликом прилога.

С. Ваљаревић <i>Комо</i>	58	1	1
Д. Великић <i>Римски прозор</i>	358	19	10

Табела 3. Коришћење придјева насталих од неличних глаголских облика у одабраним романима српских писаца⁹

Друга могућност је да се они сврстају уз глагол од кога су настали, како што се то некада чини за неке од ових придјева у неким речницима. Илустрирамо овакво решење примерима:

- Пружили су му руку и повукли га уз **одроњену** страну брега,...
- Његов лелек пред сумњом, **ђикнулом** из повређеног мозга...
- Њено посртање без Верине **подупируће** руке,... (А. Тишма *Уномреба човека*)

Уколико би се прихватило овакво решење облици из претходних примера **одроњену**, **ђикнулом** и **подупируће** би се прикључили осталим облицима глагола *одронити*, *ђикнути* и *подупирати*, а одговарајући улази у е-речнику били би:

одроњену, одронити. V:aefs4g

ђикнулом, џикнути. V:aefs6g

подупируће, подупирати. V:aefs2g¹⁰

Последица усвајања оваквог решења била би да би при напредном консултовању корпуса, на пример СрпКор-а, упит [lemma="rodupirati"] вратио, између осталих и следеће редове конкорданци:

Rinsvind je ramenom **podupirao** uzdrhtalu policu, a kolenima vraćao na mesto Čarli koji se **podupire** štapom da mu uši ne bi utrnule od silnog stajanja ...dlan ruke položio na sedište do prolaza, kao da se tako najbolje **podupirem**. Tankim pritkama, kojima se **podupiru** pasulj i grašak, preti velika opasnost, Njeno posrtanje bez Verine **podupiruće** ruke, sred vike čiji smisao ne razume, Ona je obično ležala nalakćena, **podupirući** obraz dlanom, i gledala skoro ne

Иако овакви резултати могу да буду корисни, многи корисници СрпКор-а би вероватно желели да могу да претражују одвојено по глаголским а одвојено по придјевским облицима. На неадекватност оваквог решења указују и придјеви и именице изведени из попридевљених глаголских облика. Неки од њих су:

- ...пловила је уз саму обалу једна величанствена трбушина у коју би могли стати сви наши данашњи **загорели** и **незагорели** паприкаши. (Г. Олујић *Излет у небо*)
- У **полуупржен** лук ставити кашику брашна. (Пата *Кувар*)

⁹ Видети фусноту 8.

¹⁰ ф – женски род, s – једнина, 2, 4, 6 – генитив, акузатив, инструментал, а – позитив

- ...на другој су се продавале **малоупотребљаване** кочије, увезене из Париза 1814. године; (Н.В. Гогољ *Нос* 1995)
- ... da je njegov goli bes usmeren samo protiv njene brižljive, iako nespretno primenjene **negovanosti**. (Е. Јелинек *Пијанисткиња*)
- ...sa superiornom dosadom ispod konvencionalne i korektne **kvazizainteresovanosti**,... (О. Давичо *Песма*)

Уколико би се одреднице е-речника повезивале деривационим релацијама, то значи да би у случају оваквог решења придев *незагорео* био директно повезан са глаголом директно повезан са глаголом *загорети*, а именица *негованост* са глаголом *неговати*.

Због свих ових разлога је за е-речник српског језика одабрано треће решење у коме су сви прави придеви који су настали попридевљавањем неличних глаголских облика засебне одреднице у речнику лема Delas. Тако облици **одроњену**, **ђикнулом** и **подупируће** из раније датих примера имају као нормирани облик *одроњен*, *ђикнуо* и *подупирући* – придев у номинативу једнине мушког рода – а одговарајући улази у е-речнику су:

одроњену, одроњен. A+trpni:aefs4g

ђикнулом, ђикнуо. A+radni:aefs6g

подупируће, подупирући. A+prilog:aefs2g

Из ових примера се види да сви овако уведени придеви имају ознаку која указује да су настали од одређеног неличног глаголског облика, али не и ког глагола.

У е-речнику српског који у овом тренутку има 195.000 одредница (које генеришу скоро 2,5 милиона облика речи на начин приказан у табели 1) налазе се, као засебне одреднице, велики број придева насталих попридевљавањем неличних глаголских облика, а њихов број и дистрибуција дати су у табели 4. Треба имати у виду да су придеви настали од трпних придева обрађивани систематски, када и одговарајући глаголи, док су остали унети у е-речник када су уочени у неком обрађиваном тексту.

	трпни	радни	прилог в.с
укупно	11.352	321	205
одрични	594	16	2
с префиксом <i>ново-</i>	36	2	0
с префиксом <i>полу-</i>	34	6	1

Табела 4. Број придева насталих попридевљавањем неличних глаголских облика и придева насталих из њих префиксацијом (*не-*, *ново-*, *полу-*) у е-речнику српског.

3. О повезивању одредница у е-речнику српског језика

Као што је речено у претходном одељку, придеви настали попридевљавањем неличних глаголских облика имају у е-речнику ознаке која указују од кога глаголског облика су настали – трпног или радног глаголског придева или прилога времена садашњег – али нема назнаке глагола чији су то облици. Оваква информација може бити значајна за кориснике корпуса које би, на пример, могло интересовати који су то све глаголи од чијих су прилога времена садашњег настали придеви који су у употреби. Ова информација је важна и за аутоматску обраду; илустрације ради, ако бисмо сегмент „Њено посртање без Верине **подупируће** руке...” желели да трансформишемо у „Њено посртање без Верине руке која **подупири...**” потребно би било да постоји веза између придева *подупирући* који је настао попридевљавањем прилога времена садашњег и одговарајућег глагола *подупирати*. Одговарајуће леме и њихови облици би сада у е-речнику били представљени на следећи начин:

Delas	подупирати, V17+superlema=подупирати подупирући, A3+prilog+superlema=подупирати
Delaf	подупири, подупирати.V+superlema=подупирати:Pzs подупируће, подупирући.A+prilog+superlema=подупирати:acfs2g

У оваквом решењу се као посебна ознака неке одреднице у речнику лема (*Delas*) појављује +superlema, која као вредност има одредницу из истог речника с којом дата одредница успоставља везу. Ове ознаке се на начин који приказан у одељку 1 (табела 1) прослеђују и одредницама у речнику облика (*Delaf*). Уколико се даље прошири овакав начин допуњавања е-речника у њему би се појавиле и следеће одреднице:

Delas	одобравати, V+suplema=одобравати одобравање, N+глаголска+suplema=одобравати неодобравање, N+глаголска+одрич+suplema=одобравати одобраван, A+трпни+одрич+suplema=одобравати неодобраван, A+трпни+suplema=одобравати одобравајући, A+прилог+suplema=одобравати неодобравајући, A+прилог+одрич+suplema=одобравати
--------------	--

Оваква допуна е-речника омогућава, између осталог, корисницима који га користе за претрагу корпуса (директно, преко програмског система Unitex¹¹, или индиректно кроз тагирање и лематизацију Корпуса савременог српског језика (Утвић 2011)) да постављају упите било преко основних лема или преко њихових „суперлема”. Тако би, на пример, коршћењем система Unitex корисник за претраживање корпуса преко основних лема могао користити упите <одобравати.V>, <одобравање.N>, <одобравајући.A>, <одобраван.A>, <неодобравајући.A>, док би упит <suplema=одобравати> могао користити уколико жели да добије све облике глагола *одобравати* као и облике

¹¹ Unitex/Gramlab, софтверски пакет отвореног кода за обраду корпуса (<https://unitexgramlab.org/>)

именица и придева изведених из њега. Тако би, примера ради, примена упита <suplema=одобравати> у програмском систему Unitex на узорак примера из РСАНУ дала 32 поготка, а међу њима су:

Кнез Милош волео је **неодобравање** него пецкаву шалу. (Мил. М. Ђ. 2, РЈА)
 Овим се речима **одобрава** пиће. (Ник. В. 2, 418).
 ...народа, претставници Северног Вијетнама **одобравају**. (Пол. 1957, 15850/6).
 ...свадљивом, парничарском, али ... **одобравајућем** расположењу. (Пек. Б. 11, 139).
 Знао је да се на тај начин, **одобравајући** јој и повлађујући, дете квари.
 Нису **одобравали** не само брак него ни плотску везу међу човек...
 ...ми је највише годило да поберем његово **одобравање**. (Десн. 1, 166).
 ...се служим ни при грађењу буџета, ни при **одобравању**. (Н. Скуп. 1, 642).
 ...и усмене лицитације ... **одобравање** територијални надлежни инспектор (3б. зак.

Увођење додатних информација у е-речник значајно отежава његов развој и одржавање. На пример, ако бисмо одричне облике придева и именица желели да повежемо с њиховим афирмативним облицима, а не са глаголом од кога су настали, њихове „суперлеме” би биле друкчије, што значи и да би била потребна и другачија верзија речника, на пример:

Delas одобравање, N+глаголска+suplema=одобравање
 неодобравање, N+глаголска+одрич+suplema=одобравање
 одобраван, A+трпни+одрич+suplema=одобраван
 неодобраван, A+трпни+suplema=одобраван
 одобравајући, A+прилог+suplema=одобравајући
 неодобравајући, A+прилог+одрич+suplema=одобравајући

Ефикасно одржавање е-речника који садрже богате информације, као и могућност производње разних изведених облика – како у погледу формата тако и садржаја – могуће је једино ако се е-речник конципира као база података. Модели лексикографских база података могу бити разноврсни, и неки од њих су погоднији за развој традиционалних речника намењених човеку, док су други више прилагођени е-речницима који су намењени информатичким применама. С обзиром да је наш е-речник ове друге врсте за његов даљи развој је одабран модел заснован на *lemon* (Lexicon Model for Ontologies) моделу (Макреј и др. 2012) који је до сада успешно коришћен за развој лексичких ресурса за аутоматску обраду језика. Развијена је онлајн апликација за развој и одржавање е-речника, названа Лексимирика, која користи централизовану лексичку базу података (Станковић и др. 2018). Осим што омогућава истовремени рад више корисника, овако конципирана база података омогућава проверу конзистентности података, повезивање одредница међусобно, као и њихово повезивање са спољашњим ресурсима – семантичким мрежама, корпусима и другим речницима.

Увођење „суперлеме” за повезивање придева насталих из неличних глаголских облика са одговарајућим глаголом је само једна од веза која је успостављена у е-речнику српског језика кроз базу и апликацију Лексимирика. Примери других врста веза су:

- Варијантни облици:

кафа, N+var=кава
кава, N+var=кафа
моделирати, V+var=моделисати+var=моделовати
моделисати, V+var=моделирати+var=моделовати
моделовати, V+var=моделирати+var=моделисати

- Облици различитог изговора:

девојка, N+Ijk=дјевојка
дјевојка, N+Ek=девојка

- Деривациони облици:

акцијаш, N+fem=акцијашаца
акцијашаца, N+mas=акцијаш
акцијашев, A+присвојни+noun=акцијаш
акцијашаца, A+присвојни+noun=акцијашаца

Све успостављене везе су у оквиру Лексимирке детаљно описане, као и механизми за њихово аутоматско успостављање на основу информација већ садржаних у речнику – на тај начин успостављено је више од 89.000 веза између одредница у е-речнику. Лексимирка такође омогућава извоз е-речника у различитим форматима уз одабир информација које су прилагођене потребама корисника и њихових апликација.

4. Закључак

У овом раду смо представили неке карактеристике е-речника српског језика и дилеме које се јављају при одабиру одредница и представљања информација у њему, посебно у односу на њихово решавање у традиционалним речницима. Ове дилеме су илустроване на примерима придева насталих попридевљавањем трпног и радног глаголског придева и прилога времена садашњег и прошлог. Показано је да је једна од важних додатних информација у е-речнику повезивање одредница, али да се оно не може ефикасно спровести уколико се е-речник не конципира као лексикографска база. Лексимирка укратко представљена у овом раду представља једну такву базу података и апликацију која омогућава успостављање веза између одредница, њихово одржавање, као и експортовање резултата у формату прилагођеном кориснику.

У овом раду су за потребе илустровања употребе придева насталих из неличних глаголских облика коришћени Корпус савременог српског језика, други литерарни текстови као и један узорак примера из Речника српскохрватског књижевног и народног језика. Експлоатација примера из РСАНУ је указала на значај досадашњих резултата у дигитализацији овог речника као и неопходност наставка започетих активности, јер овај речник осим своје основне намене представља и неисцрпан извор података за научна истраживања.

ЛИТЕРАТУРА

- Вигас 2007:** Душко Вигас, „О проблему не(пре)познате речи у обради текстова на српском језику”, *Зборник Матице српске за филологију и лингвистику*, L, 111–120.
- Вигас/Крстев 2012:** Vitas, Duško, and Cvetana Krstev, ”Processing of Corpora of Serbian Using Electronic Dictionaries”, *Prace Filologiczne*, 63, 279–292.
- Грос/Перен 1987:** Gross, Maurice, and Dominique Perrin, *Electronic dictionaries and automata in computational linguistics, LNCS 377*, Berlin: Springer.
- Крстев и др. 2004:** Krstev, Cvetana, Dusko Vitas, and Tomaž Erjavec. ”Morpho-Syntactic Descriptions in MULTEXT-East-the Case of Serbian.” *Informatica (Slovenia)*, 28.4, 431–436.
- Крстев 2008:** Cvetana Krstev, *Processing of Serbian – Automata, Texts and Electronic dictionaries*. Belgrade: Faculty of Philology, University of Belgrade.
- Макреј и др. 2012:** McCrae, J. Et al. ”Interchanging lexical resources on the semantic web”. *Language Resources and Evaluation*, 46(4), 701–719.
- Станковић и др. 2018:** Stanković, Ranka, Cvetana Krstev, Biljana Lazić and Mihailo Škorić, ”Electronic Dictionaries – from File System to lemon Based Lexical Database”, *Proc. of the 11th LREC – W23 6th Workshop on Linked Data in Linguistics : Towards Linguistic Data Science (LDL–2018)*, Miyazaki, Japan, May 7–12, 2018, 2018, eds. John P. McCrae et al., ISBN 979-10-95546-19-1.
- СТИЈОВИЋ И ДР. 2017:** Рада Стијовић, Олга Сабо и Ранка Станковић, „Речник САНУ као база терминолошких речника (на примеру речника кулинарства)”, у: *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. doi. org/10.18485/msc.2018.47.1.ch40
- УТВИЋ 2014:** Miloš Utvić, *Izgradnja referentnog korpusa savremenog srpskog jezika* (необјављена докторска дисертација, Филолошки факултет Универзитета у Београду).
- УТВИЋ 2011:** Милош Утвић, „Анотација Корпуса савременог српског језика”. *ИНФОтека* 12, 39–51.

Речници

- РМС 2007:** *Речник српскога језика*, Нови Сад: Матица српска.
- РМС 1967–1976:** *Речник српскохрватскога књижевног језика*, књ. 1–6, Матица српска [– Матица хрватска], Нови Сад [– Загреб].
- РСАНУ 1959–2019:** *Речник српскохрватског књижевног и народног језика САНУ*, књ. I–XX, Београд: Институт за српски језик САНУ.

Граматице

Станојчић/Поповић 2010: Живојин Станојчић и Љубомир Поповић, *Грамматика српског језика за гимназије и средње школе*, Завод за уџбенике, Београд.

Стевановић 1966: Михаило Стевановић, *Грамматика српскохрватског језика за више разреде гимназије*, пето издање, Цетиње: Обод.

Пипер/Клајн 2014: Предраг Пипер и Иван Клајн, *Нормативна граматика српског језика*, Нови Сад: Матица српска.

Cvetana J. Krstev

ABOUT THE SELECTION OF ENTRIES FOR THE ELECTRONIC DICTIONARY OF SERBIAN AND THEIR LINKING

Summary

Full recognition of words in written texts is one of the prerequisites for the successful natural language processing. By the recognition of words, we do not consider merely the decision whether a word belongs to a language processed or not. We presume that for a recognized word its properties would be known as well (grammatical role, lemma, meaning, domain of usage, etc.) The electronic dictionary of Serbian was compiled to fulfil these requirements. In order to develop and maintain the electronic dictionary efficiently and enable incorporation of rich information in it the adequate environment is crucial that is based on a lexicographic database. In this paper a lexicographic database and an application Leximir is presented and some of its functionalities are illustrated on an example of the enrichment of dictionary content by linking its entries.