

**ПРИНЦИПИ ОБРАДЕ И НАЧЕЛА ПРИРЕЂИВАЊА ИМЕНИЦА
У ПРОЦЕСУ ЛЕМАТИЗАЦИЈЕ КОРПУСА *НОВИНА СЕРБСКИХ* ОД 1813. ДО 1868. ГОДИНЕ¹**

Проф. др Зорица Несторовић
стручни консултант – редактор у процесу лематизације
(2012–2013); стручни консултант (2015)

Ас. мср. Катарина Беговић
стручни консултант – редактор у процесу лематизације (2015)

Шта представља одредницу у лематизацији именица у корпусу 'Новина сербских' (1813–1868)?

Одредница у оквиру процеса лематизације именица у корпусу *Новина сербских* (1813–1868) дефинисана је у складу са резултатима истраживања дисциплина науке о језику као и појединачних проучавања језика српских писаца 18. и 19. века. У том смислу успостављени су посебни параметри одређивања одредница на основу којих се одредницом сматра именица која припада књижевном језику или дијалекту (у случајевима када се према својим фонетским и морфолошким особинама разликује од облика те именице у књижевном језику (на пример: одредница у књижевном језику: **кромпирь** (*кромпира, кромпире, кромпири, кромпиры*) / одредница у дијалекту: **крумпирь** (*крумпире, крумпире, крумпирь*)). Такође, услед присуства елемената српскословенског, рускословенског, књижевног руског и народног језичког слоја у оквирима језика *Новина сербских* (1813–1868), статус одреднице приликом лематизације именица овог корпуса додељен је и свим именицама које овим слојевима припадају. На пример, као појединачне одреднице издвојене су именице **благодѣяніе** (*благодѣянія, благодѣяніама, благодѣянія, благодѣяніе, благодѣяніемъ, благодѣяніе, благодѣяніемъ, благодѣянія, благодѣяніама, благодѣянію, благодѣяніамъ, благодѣяніе, балгодѣянія, балгодѣяніама, балгодѣяніе, благодѣянію*), **благодѣй** (*благодѣя, благодѣю, благодѣю*) и **благодѣяње** (*благодѣяня, благодѣянья*), а превод за све три одреднице је исти и гласи *милост*. Статус одреднице су, аналогно томе, добиле и све именице које припадају страним језицима, али се у корпусу *Новина сербских* (1813–1868) јављају транскрибоване.

Имајући на уму лингвистичке критеријуме постављене при дефинисању појма одреднице у вези са славеносрпским језиком, сваки облик именице који се услед ортографских, а не фонетских или пак морфолошких разлога, пише другачије у односу

¹ **НАПОМЕНА:** Наведене принципе и начела приређивања именица у процесу лематизације корпуса *Новина сербских* (1813–1868) поставила је у првој фази пројекта проф. др Зорица Несторовић, тадашњи стручни консултант – редактор у процесу лематизације (2012–2013), а садашњи стручни консултант (2015), и они су се односили на све врсте именица осим на ономастичке појмове, јер они тада нису били обрађивани изузев у такозваној типологизацији која ће бити објашњена у упутству за кориснике. У наставку рада на пројекту (2015; стручни консултант – редактор у процесу лематизације ас. мср. Катарина Беговић), они су и даље примењивани. Будући да се у овом тренутку (мај 2015) лематизује преостали део корпуса именица у *Новинама сербским* (1813–1868), као и да је неопходно довршити започету ревизију обрађеног материјала, која је престала услед прекидања рада на пројекту 2013. године, као и урадити завршну ревизију целокупног лематизованог материјала, коректуру (одређивање шта је штампарска, а шта дактилографска грешка), индекс имена, топонима и скраћеница, претпостављамо да ће се установљеним принципима обраде и начелима приређивања придружити још неки елементи у зависности од корпуса који тек треба да буде лематизован, а које ће профилисати садашњи редактор у процесу лематизације ас. мср. Катарина Беговић. У овом тексту биће представљени сви они поступци лематизације и начела приређивања који су важили за радну верзију обраде корпуса (периоди 2012–2013, 2015), а који ће у крајњој верзији његове обраде у појединим тачкама бити допуњени.

на основни облик те именице није лематизован као посебна одредница. Другим речима, ортографија по себи не узима се као дистинктивно обележје. Дobar пример за то представља недоследно писање танког и дебелог јера на крају именице. Чак у делима једног истог писца насталим на славеносрпском језику честа је појава наизменичне употребе оба наведена ортографска решења. Тако, на пример, лексема из корпуса *Новина сербских* чији је превод *безбедност* појављује се у облицима номинатива и са дебелим и танким јером на крају: **безбѣдностъ** / **безбѣдность**. Будући да су у питању само различити ортографски облици једне исте речи, у корпусу постоји само једна одредница, а не две. Облици деклинације ове именице (*безбедности*, *безбедность*, *безбѣдности*, *безбѣдностъ*, *безбѣдность*, *безбѣдностю*, *безбѣдностію*, *безбѣдностју*, *базбѣдностъ*, *безбѣдностъ*) подведени су под једну одредницу чији је коначан облик изабран према учесталости ортографског решења у до сада обрађеном корпусу.

Овом примеру додајемо још један који је веома репрезентативан за корпус *Новина сербских* (1813–1868) јер илуструје потребу да се направи разлика између оних именица које долазе из неког другог језика од различитог ортографског бележења неке именице већ одомаћене у нашем лексичком фонду или оне која припада основној лексици српског језика. Лексема чији је превод *анђео* појављује се у корпусу *Новина сербских* у следећим облицима номинатива: **ангелъ** / **ангѣео** / **анђео**. Именица **ангелъ** издвојена је у посебну одредницу зато што припада другом језичком наслеђу и по томе испуњава услов да буде посебна одредница, тако да су и сви облици њене деклинације овој одредници придружени. Сви облици деклинације именица **ангѣео** и **анђео** подведени под једну одредницу чији ће ортографски облик бити изабран на крају процеса лематизације корпуса у целини, и то према учесталости ортографског решења у облицима, јер су у питању само различите ортографске варијанте једне исте речи.

Као што се у литератури о овој теми често истиче на основу бројних примера, у позајмљеницама се јављају удвојене графеме услед оригиналне графичке језика из којег је реч позајмљена, а честа је и њихова појава у облицима речи које припадају српском језику услед, на пример, означавања дугог слога, испадања гласа х, или на морфемским спојевима... За све примере таквих облика именица не постоји посебна одредница (будући да она не испуњава већ дефинисане услове у вези са тим шта може бити дефинисано као одредница), већ се на основу тренутне учесталости прави једна одредница за коју се везују све њене варијанте настале искључиво из ортографских разлога. Тако су, на пример, за одредницу **баллъ** (превод: *бал*) везани следећи облици: *баллу*, *баллъ*, *балове*, *баалу*, *баалъ*, *баловима*, *баломъ*, *баллови*, *бала*, *балу*, *балъ*, *балова*, *баловы*.

У случајевима када је облик именице представљао евидентну и недвосмислену словну грешку (штампарску или дактилографску), лематизатори су на основу семантике погрешно написаног облика и ширег контекста јављања у реченици или делу текста, утврђивали да ли је у питању грешка или није. Када су на основу тих семантичких параметара утврдили да се јавља грешка (штампарска, то јест она настала у штампи новина, или дактилографска, то јест она настала приликом прекуцавања текста за потребе лематизације), они су тај облик придруживали одредници надређеној за тачан облик именице. У следећем примеру облици који су дати црвеном бојом представљају пример евидентних и недвосмислених словних грешака: *благодѣянія*, *благодѣяніяма*, *благодѣянія*, *благодѣяніе*, *благодѣяніемъ*, *благодѣяніе*, *благодѣяніемъ*, *благодѣянія*, *благодѣяніяма*, *благодѣянію*, *благодѣяніямъ*, *благодѣяніе*, *балгодѣянія*, *балгодѣяніяма*, *балгодѣяніе*, *благодѣянію* и сви они су везани за одредницу **благодѣяніе** (превод: *милост*).

Треба истаћи и то да се многе заједничке именице у корпусу *Новина сербских* (1813–1868) бележе великим словом иако нису део било каквог назива, под утицајем стране ортографије. У тим случајевима облици заједничких именица написани великим словом придруживани су њима одговарајућој одредници са малим почетним словом (**администраторъ, адмиралъ, запалење, звезда, зверница, зверъ** итд.).

На самом крају, наглашавамо да ће коначан ортографски облик одреднице бити утврђен по завршеној лематизацији целог корпуса, будући да ће само тада статистички валидним методом бити могуће утврдити који је најфреквентнији ортографски облик који ће уједно представљати и ортографски лик саме одреднице.

Неколико напомена у вези са принципима лематизације

1. Уколико у корпусу именица *Новина сербских* (1813–1868) не постоји ниједан пример номинатива једнине оне именице за коју се прави одредница, лематизатор сам одређује облик номинатива једнине за ту именицу. Том приликом дужан је да се руководи свим граматичким принципима и достигнућима српске науке о језику која се бавила овом тематиком. Један од тих принципа је да су се *танко јер* (ѣ) и *дебело јер* (ѣ) писали на крају оне речи која се завршавала сугласником. Такође се употребом *танког* и *дебелог јер* ортографски остваривала разлика између именица мушког и женског рода тако што се *дебело јер* (ѣ) писало на крају именица мушког рода, а *танко јер* (ѣ) на крају именица женског рода. Такође, истраживачи бележе да се *танко јер* (ѣ) писало и на крају речи које су се завршавале палаталним сугласником (*помоћь*) мада је и то недоследно јер је у пракси било облика и са *дебелим јер* (ѣ) на том месту.

2. Имајући на уму нестандардизованост славеносрпског језика, као и у науци о језику вишеструко потврђен став о условима које реч треба да испуни да би се могла прихватити аутономним обликом именице, напомињемо да је увек поштован оригинални текст *Новина сербских* (1813–1868), односно да су се одреднице образовале на основу примера који су потврђени у њиховом корпусу. Као пример навешћемо нека од решења примењена приликом лематизације овог корпуса *Новина сербских*.

Као што је из литературе познато, графема **щ** се у текстовима написаним славеносрпским језиком јавља како у транскрибованом тако и у нетранскрибованом облику. У првом наведеном случају она се некад транскрибује као **шт**, а некад као **шч**. Будући да ти процеси нису стандардизовани, а да се у корпусу *Новина сербских* (1813–1868), на пример, номинатив именице *башта* појављује у облицима **баща** / **башта** / **башча**, сваки од наведених облика има статус посебне одреднице јер, уколико у корпусу постоје облици **башта** и **башча**, а нема стандарда према којем би требало урадити транскрипцију облика **баща** (што постојање облика **башта/башча** додатно потврђује), нема правила према којем бисмо облике деklinације именице **баща** везали за одредницу **башта** односно **башча**. Но, будући да се у преводу све три одреднице везују за савремени облик БАШТА, корисник ће претраживањем корпуса моћи да види све примере за све три одреднице **баща** / **башта** / **башча** а научни истраживачи филолошких дисциплина имаће увид у разноврсност именичке грађе. Исто је урађено и у примерима **баштица, башчица, бащован, башчованъ, баштованъ**.

Будући да код различитих аутора постоје различити начини обележавања **вокалног р** у овом периоду (1813–1868) – **р, р', ер, њр, рѣ, рѣ, њр** – у овој фази рада на лематизацији корпуса именица *Новина сербских* (1813–1868) облици су раздвајани према следећем принципу: а) облици са **р** – посебна одредница; б) облици са **ер** –

посебна одредница; в) облици са **рѣ** / **рѣ** / **р'** – посебна одредница и г) облици са **ѣр** / **ѣр** – посебна одредница.

3. Период од 1813. до 1868. обележен је укрштањем и смењивањем различитих тенденција у развоју лексике *Новина сербских*. Лематизација именица које припадају њиховом корпусу у наведеном периоду потврдила је постојање значајног броја **хомонима**, **хомографа** и **хомоформа**. Стога су у овој фази лематизације у вези са овим питањем постављени следећи принципи:

Дистинктивно обележје одреднице је њено значење, што значи да без обзира на чињеницу што се именице пишу исто, често исто и изговарају, мора се имати на уму да ли оне имају различита значења потврђена контекстом. У том случају свака именица постаје одредница уколико има аутономно значење. У првој фази лематизације се у преводу наводе сва значења која реч тако написана може имати јер се, на пример, приликом претраживања појма ДЕО могу појавити и облици одредница **чест** и **част** које имају и то значење, а не само значење *част*. У таквим случајевима корисник се упозорава напоменом: * *Под овом одредницом јављају се и други облици речи;*

Будући да је услед постојања хомоформа могуће да се приликом претраживања добију и облици речи које нису именице, то се такође назначавала звездицом и објашњава у пратећој напомени која гласи: * *Под овом одредницом јављају се и други облици речи.*

Питање превода одредница

Одреднице се преводе на савремени српски језик. Превод се састоји од једне одговарајуће лексеме која представља најбољи превод одреднице на савремени српски језик. Међутим, то у великом броју случајева није могуће остварити, па се у преводу даје опис значења одреднице сведен на минималан број речи довољан да кориснику на адекватан начин укаже на њено право значење. Превод одреднице на савремени српски језик усаглашен је са важећим правописним правилима.

Будући да се у значајном броју примера приликом одређивања значења одреднице јављају бројни синоними, сви они се бележе да би корисник могао да претражује базу одредница имајући на уму и оне појмове који су синонимни термину којим лематизатор преводи одредницу. Лексеме које означавају слична значења одреднице наводе се у преводу једна иза друге и раздвајају интерпункцијским знаком (;).

Познато је да је у наведеном периоду развоја српског књижевног језика (1813–1868) у лексичкој грађи изузетно велики број позајмљеница (турцизама, германизама, галицизама, англицизама...), од којих су неке и данас у употреби у савременом српском језику. Стога превод у таквим случајевима најчешће укључује и тај облик као основну именицу превода, као и ону именицу која у савременом српском језику постоји да означи исти појам. На пример, одредница **анџар** преводи се као *нож*, али и као *ханџар*; **маневарѣ** као *маневар*, али и као *војна вежба*... Циљ оваквог превођења је да се кориснику пружи што је могуће више начина претраживања, али и информација о значењу речи.

Штампарске и дактилографске грешке

У свим примерима увек се поштује оригинални текст. Уколико се уоче грешке у писању речи, увидом у контекст односно упоређивањем са оригиналом, утврђује се да ли је у питању штампарска грешка (она која постоји у оригиналу) или дактилографска грешка (што значи да је реч у оригиналу тачно одштампана, али је приликом прекуцавања дактилограф начинио коректорску грешку). У случају штампарске грешке, облик речи се не исправља већ се у напомени назначавача да је у питању штампарска грешка. Такође, облици који су погрешно одштампани у оригиналу везују се за ону одредницу која је одговарајућа оном њиховом облику који је без грешке. Уколико је у питању дактилографска грешка, то се у овој фази лематизације бележи да би се у њеној завршној фази и у коначној коректури те грешке настале приликом прекуцавања текста исправиле сравњивањем са оригиналом. И у овом случају, као и у претходном, облици који су погрешно прекуцани од стране дактилографа везују се за ону одредницу која је одговарајућа облику без грешке.

Ономастика

Приликом лематизације ономастичких појмова у првој фази рада вођено је рачуна о следећим принципима обраде:

а) У првој фази рада се као посебне одреднице лематизују сви појединачни ортографски начини писања неког антропонима (лична имена, презимена, надимци), топонима (имена места), етника (имена народа, племена, становника неког места), ктетика (присвојни придеви који су изведени из имена места), егзонима (називи за стране топониме који се разликују од изворног облика тог топонима) и ендонима (страни топоними у изворном облику, али су у корпусу *Новина сербских* дати у транскрибованом облику). Ово начело се у првој фази разликује од начела за остале врсте именица и важи само за ономастику. Уведено је да би се уочиле све варијанте бележења ономастичког појма и да би се на крају у завршној фази обраде након истраживања сваког појединачног примера како у ужем, етимолошком/дијалектолошком контексту, тако и ширем, културноисторијском контексту, одредило значење. На основу тога, у завршној фази лематизације, биће урађени ИНДЕКСИ ИМЕНА И ТОПОНИМА.

б) Превод ономастичког појма увек се даје с обзиром на стање у савременом језику. На пример, у корпусу *Новина сербских* (1813–1868), име великог старогрчког песника појављује се у облицима одредница **Омир** и **Хомер**. То значи да постоје две одреднице (**Омир**, **Хомер**) и да су за њих везани сви њима одговарајући облици. Међутим, њихов превод је један и он гласи *Хомер*. То значи да ће корисник, када у претраживач укуца антропоним ХОМЕР, моћи да претражује сва она места где се појављују сви облици одредница **Омир** и **Хомер**. Исто важи и за топониме оличене у одредницама **Виenna** и **Беч**, које се лематизују као посебне одреднице, а у преводу везују за топоним *Беч*.

У овој фази рада приликом превођења антропонима поштује се оригинални текст. То конкретно значи да превод одреднице поштује облик одреднице у славеносрпском језику у случајевима као што су **Јoан** чији ће превод такође бити *Јoан*. Адекватно томе **Јoхан** се преводи као *Јохан*, **Јован** као *Јован*, а **Јокан** као *Јокан*. Тек у завршној фази лематизације биће приступљено критичком приређивању ових ономастичких појмова односно писања, на пример, графема **ј**, **в** или **х** тамо где се у савременом српском језику пишу и то према начелима критичког издања у угластим

заградама: *Jo[v]ан, Jo[x]ан, Ми[x]ајло, Живо[j]ин*. Тада ће се приступити и повезивању више различитих варијанти једног антропонима са једном одредницом у преводу. У завршној фази превод ће у сваком смислу поштовати стандарде савременог српског језика.

У овој фази рада презимена која се у оригиналу славеносрпске одреднице завршавају на **ч**, и у преводу се завршавају на **ч** (примери: Исаилович, Маркович, Обрадович...). Такође, уколико се у оквиру одреднице ономастичког појма појављује цртица, она се за сада преноси и у превод одреднице.

Облици присвојних придева који су саставни део ономастичког појма такође се лематизују. Будући да је веома честа појава ономастичких појмова који су свом саставу имају присвојни придев, приликом лематизовања врши се типологизација у којој се ова појава бележи избором међу понуђеним опцијама. Ову техничку могућност типологизације везане за обраду ономастике у првој фази лематизације осмислили су проф. др Александар Костић и програмер Марко Јевтић и она је описана у упутству за кориснике. У упутству за кориснике наглашено је да у претраживач увек треба укуцати одређени појам па ће тако у случају када, на пример, корисник претражује појам АЗОВСКО МОРЕ, бити потребно да укуца цео појам. Када се ради о примерима имена владара у оквирима корпуса *Новина сербских*, међу антропонимима се јављају и облици попут *Фридрих Први*, али и *Фридрих I*. У тим случајевима превод гласи управо као и одредница, па се кориснику сугерише да претраживање може обавити и према том параметру.

Скраћенице

У корпусу *Новина сербских* изузетно су учестале различите врсте скраћеница, међу којима су и оне везане за означавање година, месеца, дана, титула, различитих облика етикеције, војничких звања... Скраћенице се у првој фази лематизације обележавају знаком (+). У завршној фази лематизације биће урађени индекс скраћеница и њихово разрешење.

Београд, 20. мај 2015.