

Маријан Јелић

О НЕКИМ ПРОБЛЕМИМА УТВРЂИВАЊА ОДРЕДНИЧКЕ РЕЧИ У ФРЕКВЕНЦИЈСКИМ РЕЧНИЦИМА

У раду су анализирани неки проблеми који се јављају приликом утврђивања одредничке речи у фреквенцијским речницима. Циљ рада је да се на одабраним примерима фреквентних речи српског језика (суплетивни облици, варијанте, вишеверсне речи, вишечлани географски називи) укаже на неуједначености које постоје у одређивању одредничке речи у различитим фреквенцијским речницима. Истраживање показује да постоји потреба, поготово за најфреквентније речи, да се одредничка реч утврђује на исти начин, како би добијени резултати били тачни и упоредиви.

Кључне речи: фреквенцијски речник, пребројавање, одредничка реч, варијанте, вишеверсне речи.

This paper analyzes how lemmata are determined in frequency dictionaries, and for this purpose we chose those words and word groups that are more frequent and for which there are different approaches to processing (*sebe, se, bitiljesam*, variants, multi-class words, two-part names), which caused significant differences in the obtained results. The research shows uneven treatment of determining lemmata in frequency dictionaries, which questions the accuracy of the results obtained in this way, which makes those dictionaries or results more or less unsuitable for comparisons.

Keywords: frequency dictionary, counting, lemmata, variants, multi-category words.

1. Увод. У претходних педесетак година у српској лингвистици објављено је неколико десетина фреквенцијских речника (многи од њих биће поменути у наставку рада). Код нас се углавном израђују фреквенцијски речници књижевних дела, док у светској лингвистици постоји велики број различитих врста фреквенцијских речника (националног језика, говорног језика, регионални, дијалекатски, историјски, речници најфреквентније лексике за учење језика итд.). С обзиром на то да је софтвер помоћу којег се овакви речници израђују све квалитетнији, може се очекивати већа продукција ових речника. Поступак израде фреквенцијског речника прилично је једноставан: након сегментације текста (корпуса) на појединачне речи, која се обавља помоћу софтвера, неопходно је извршити лематизацију ради утврђивања основних облика речи, при чему се овај корак често изводи ручно због потенцијалних грешака. Након овог процеса следи бројање речи и навођење њиховог укупног броја тако што се код променљивих врста речи узимају у обзир сви њихови облици. Одредничке речи наводе се у основном облику уз давање податка о броју појављивања (фреквенцији) у тексту (корпусу). У изради фреквенцијских речника, дакле, главни задатак је пребројавање речи, одн. утврђивање њихове фреквенције. Одредничка реч утврђује се на исти начин као и у описним речницима или би бар тако требало да буде.

1.1. У досадашњој домаћој теоријској лексикографији проблем утврђивања одредничке речи, одн. листе одредничких речи (енгл. headword list) у фреквенцијском речнику није посебно разматран, осим кратких напомена о томе како ће поједине речи или групе речи бити третиране (примери ће бити наведени у наредним одељцима), иако се у пракси јављају специфични проблеми који су само делимично слични проблемима утврђивања одреднице у дескриптивној лексикографији (в. Николић 2014; ТАФРА 1994; 2016). О томе како обрадити поједине речи или групе речи (изговорне и правописне варијанте, хомоними, вишечлани називи / лексеми, компаративи и суперлативи, скраћенице итд.) у европској лингвистици писали су о овој теми у појединим радовима или у уводима фреквенцијских речника истраживачи (да поменемо само неке): Michael McCarthy, Douglas Biber, Adam Kilgarriff, František Čermák, Ольга Ляшевская и други. У вези са нашом темом важна је напомена коју даје Adam Kilgarriff: „Building a headword list for a new dictionary (or revising one for an existing title) has never been an exact science, and little has been written about it” (KILGARRIFF 2013: 4).

Према страниј литератури (пре свега европској), проблеми о којима говоримо решавају се на различит начин у зависности од више фактора, а најважнији су: врста фреквенцијског речника, односно његова сврха и намена, величина корпуса и могућности софтвера – избор одговарајућег решења нужно се прилагођава могућностима софтвера (за израду фреквенцијских речника).

1.2. Приликом одређивања одредничке речи у описној лексикографији јављају се многобројни проблеми, а међу крупнијима су: једночлане и двочлане одредничке речи, суплетивни облици, конверзија, разграничење синонимије и хомонимије, разграничење конверзије и полисемије, односно могућности да неке речи могу припадати двама или трима врстама речи (вишеврсне речи), недоследна правописна норма. Према мишљењу Бранке Тафре „Izvori većine problema pri uspostavi natuknica leže u nepotpunim gramatičkim opisima i u mnoštvu neusuglašenih leksikoloških pristupa leksičkoj jedinici” (ТАФРА 2016: 517). Лексиколошка сазнања и лексикографска пракса неретко се разилазе, па је тако, на пример, у лексиколошким радовима реч *свакако* и речца (према Ристић 1993), док је у РМС и РСЈ само прилог.

Поменути проблеми постоје и приликом одређивања одредничке речи у фреквенцијским речницима. Може се рећи да је непостојање јединствене методологије за утврђивање одредничке речи у описној лексикографији највећи проблем (или бар један од већих проблема) за састављача фреквенцијског речника, будући да састављач фреквенцијског речника приликом утврђивања одредничке речи полази од постојећих описних речника. Тако се, на пример, у делу *Фреквенцијски речници школске лектиуре* наводи да је „утврђивање одредница вршено” према „моделу датом у РМС” (ИВАНОВИЋ 2018: 5).

1.3. При изради фреквенцијског речника јављају се и додатне недоумице које су узроковане, пре свега, специфичним циљевима који се постављају приликом његове израде. Тако, на пример, у описним речницима (често) постоји велики број упућивања (на пример, у РСЈ лексема *казивалац* је упућена скраћеницом *в.* на лексеми *казивач*), док таквих упућивања нема у фреквен-

цијским речницима. У дескриптивним речницима утврђивање укупног броја речи или њихово пребројавање не поставља се као сврха, па се питање раздвајања или обједињавања сродних одредница не проблематизује. У фреквенцијским речницима, међутим, наведена питања имају суштинску важност за правилну обраду и презентовање фреквенције.

1.4. Краћи, енклитички облик заменице *себе* обрађен је у РМС и РСЈ као посебна одредница, и притом *се* може бити и речца. Како у таквим ситуацијама да поступа састављач фреквенцијског речника? Неке речи у описним речницима су вишеврсне (могу припадати двома или трима врстама речи). Тако, на пример, реч *kag(a)* може бити прилог и везник (према РМС), односно прилог, везник и речца (према РСАНУ и РСЈ). Да ли је у фреквенцијском речнику обрадити као јединствену реч или као две, односно три речи? Како такве речи пребројавати по врстама: бројати је и међу прилоге и међу везнике (и међу речце) или утврдити у колико примера је прилог, а у колико је везник (и у колико речца)? Потреба да се свака реч изброји, да се утврди тачан број речи (свих речи, појединачних речи, свих речи одређене врсте) – каква постоји приликом израде фреквенцијских речника, или у анализама које најчешће иду уз фреквенцијске речнике, изискује одређени приступ који није нужан приликом израде описних речника.

1.5. Циљ рада је да се размотре проблеми утврђивања одредничке речи у фреквенцијским речницима на одабраним примерима фреквентних речи српског језика (*себе, се*) као и на примерима типске лексике: суплетивни облици, варијантна лексика, вишечлани географски називи, вишеврсне речи, код којих су уочене недоследности у досадашњем представљању у различитим фреквенцијским речницима, као и да се укаже на потребу, поготово за најфреквентније речи, да се одредничка реч утврђује на исти начин, како би добијени резултати били тачни и упоредиви. У наредним одељцима биће посебно анализирано за сваку од наведених група речи како је утврђивана одредничка реч поређењем поступака који су примењени у постојећим фреквенцијским речницима.

2. СЕБЕ, СЕ. Заменица *себе* је лична заменица за свако лице. Јавља се и у краћем, енклитичком облику *се* (*мислићи на се, чељљачи се*). Међутим, реч *се* може да буде и речца (*свађачи се*). Будући да се у готово свим, доступним нам, фреквенцијским речницима ова заменица (и речца) налази међу 10 најфреквентнијих речи у српском језику, од прворазредног је значаја да њена обрада у фреквенцијским речницима буде уједначена.

2.1. Милка Ивић у раду *Један проблем словенске синтагматичке осветљен трансформационим методом (граматичка улога морфеме се у српскохрватском језику)* говори о рефлексивној морфемци *се* (а за сам термин каже да је „услован, оправдан искључиво утврђеном лингвистичком традицијом” (Ивић 1961–192: 138)) и посматра је из синтаксичке перспективе. Иако се морфема *се* назива рефлексивном (у домаћим и словенским граматицама), она, закључује М. Ивић, „само у једном врло ограниченом делу своје употребе стоји у односу „комбинаторичне варијанте” према рефлексивној заменици

себе. У осталом (и највећем делу својих граматичких служби ова морфема не означава рефлексивност” (Ивић 1961–192: 138).

2.2. У новијим граматикама се реч *се* углавном не помиње као речца, са ретким изузецима – у *Граматици српског језика* Душке Кликовац: „повратна речца *се*, која се налази у саставу неправих повратних глагола” (Кликовац 2003: 115) и *Савременом српскохрватском језику I* Михаила Стевановића: „У науци о језику речцом се назива и *се* у саставу повратних глагола када се ова реч значењем више не подудара с акценатским обликом акузатива заменице *себе*” (Стевановић 1986: 384). С друге стране, на пример, у *Нормативној граматици српског језика* реч *се* је наведена само међу заменицама, као ненаглашени облик заменице *себе* у акузативу, који се употребљава „за грађење повратних и безличних глагола”, а заменица *себе* „може бити значењски врло блиска морфема *се* у саставу повратних глагола” (Пипер и др. 2022: 103). У делу који се тиче повратних глагола наведено је да су то они глаголи „у чијем је сложеном граматичком облику увек повратна енклитичка заменичка морфема *се*” (Пипер и др.: 197). Дакле, према граматичкој литератури није сасвим јасно да ли је *се* само заменица или је и речца.

2.3. Заменица *себе* и њена краћа варијанта *се* у РМС и РСЈ обрађене су на сличан начин – засебно, као две одреднице. Код речи *се* у оба речника је под 1. наведено да је то краћи (РМС), одн. енклитички (РСЈ) облик акузатива повратне заменице „себе”. Под тачком 2. *се* је обрађено као речца, али је у подтачки 2.а. у РМС наведено: „*йоврайна речца у сасйаву йравих йоврайних йлајола у значењу акузатива йоврайне заменице себе, 'се'*”, а под 2.б.: „*йоврайна речца у сасйаву узајамно-йоврайних и рефлексивних йлајола уойщие'*” (РМС), док је у РСЈ значење речи *се* у повратним глаголима обједињено у једној подтачки (2.а.): „*йоврайна речца у сасйаву йоврайних йлајола'*” (РСЈ). Дакле, према РМС код правих повратних глагола имамо повратну речцу „у значењу акузатива повратне заменице *себе*, док је према РМС то само речца.

Видимо да реч *се* у нашим описним речницима може бити и заменица и речца, али и то да она није на исти начин обрађена у поменути речницима. Ако је *се* и заменица, а знамо да јесте, онда је прва дилема да ли ту реч треба одвајати (као посебну одредницу) од *себе*? Ако би *се* било обрађено заједно са *себе*, да ли би *се* које је речца било посебна одредничка реч? Можемо додати још једну недоумицу: ако је *се* део повратних глагола, има ли оправдања то *се* у фреквенцијским речницима бројати као заменицу или речцу?

2.4. У свим доступним нам фреквенцијским речницима у српској лингвистици *себе*, *се* је обрађено као једна одредница и то углавном само као заменица (негде је указано на то да може бити и речца). Тако Смиљка Васић, говорећи о најфреквентнијим речима у речницима које је израдила, наводи: „Заменица *себе*, се јавља се у сва четири речника и има у садашњем истраживању 4. и 5. ранг, а у ранијим 3. и 4. ранг” (Васић 1998: 27). И у *Дечијем фреквенцијском речнику* лексема *себе*, *се* је у 10 најфреквентнијих речи корпуса (Лукић 1983). У дисертацији *Политички дискурс у йаролама, слојанима и йрафийима йројшесйа йројшив НАТО бомбардовања СР Јујославије 1999*. Тамара Груден је израдила и одговарајући фреквенцијски речник за исти

корпус и у том фреквенцијском речнику лексема *себе, се* налази се међу 10 најфреквентнијих (Груден 2016).

Међутим, краћи облик *се* најчешће се јавља као део повратног глагола, дакле, као речца (изузев код правих повратних глагола). Ако би се *се* које је заменица одвојило од *се* које је део повратних глагола (изузев код правих повратних глагола) и оног *се* које служи за грађење пасивних и безличних реченица, заменица *себе, се* засигурно не би била међу најфреквентнијим речима у српском језику.

2.5. Изузетак у поступању са речју *се* у изради фреквенцијског речника налазимо у *Речнику њриповедака Лазе Лазаревића* ауторке Браниславе Јелић (2008) и у *Речнику њоезије Милана Ракића* аутора Васе Павковића (1984) (ови су речници, чини се, рађени по истим принципима). У овим речницима *се* које је део повратних глагола није убрајано и код заменице *себе, се*, па отуда ова заменица у овим речницима не спада ни међу десет најфреквентнијих заменица. И не само то, неубрајањем *се* међу заменице или речце значајно се умањује укупан број текућих речи¹. У ова два речника повратни глаголи су наведени засебно у односу на неповратне (такав поступак, колико знамо, примењен је само у тим фреквенцијским речницима), те је услед такве обраде глагола укупан број глагола знатно повећан.

2.6. Бавећи се лексичком фреквенцијом Вукових приповедака, утврдили смо да се у Вуковим приповеткама реч *се* јавља 2020 пута, а као заменица јавља се само 170 пута. Она се најчешће јавља као део повратних глагола (око 1900 пута). Ако би се свако *се* бројало као заменица, онда би *себе, се* заузимало четврто место по фреквентности у Вуковим приповеткама. Међутим, ако бисмо у *себе, се* убројали само оно *се* које јесте заменица, онда би *себе, се* заузимало 83. место по фреквентности у Вуковим приповеткама. С друге стране, број текућих речи био би мањи за око 1.900 речи. Видимо да начин обраде ове лексеме битно утиче на резултате које добијамо (да ли ће реч бити међу најфреквентнијима, колики ће бити укупан број текућих речи).

2.7. Из жеље да се што више посла у изради фреквенцијског речника пренесе на рачунар тежи се што једноставнијим решењима. Можда је то разлог што је *се* обрађено заједно са *себе* и углавном само као заменица, јер би додатан рад био, који не може да изврши рачунар, да се издвоји *се* које није заменица већ је речца. На такав закључак упућује и поступак примењен у књизи *Фреквенцијски речници школске лекције*, чији је саставни део *Фреквенцијски речник њрозе Боре Сџанковића*, у којем „повратни глаголи нису посебно обележени” јер је то „био прилично обиман посао, који је захтевао да се сваки облик глагола посебно провери” (Ивановић 2018: 6). Чак ни код глагола који су само повратни (нпр. *вриољити се, досећати се, смејати се*) није наведено *се* (у поменутом *Речнику* је наведено: *вриољити, досећати, смејати*). Међутим, у првој књизи (Ивановић 2016), у којој су дати фреквенцијски речници књижевних дела *Каг су цветале њишке* (Драгослава Миха-

¹ Текуће речи подразумевају укупан број речи у једном узорку. Користи се и термин токени.

иловића), *Вечийи младожења* (Јакова Игњатовића) и *Писмо Хараламију* (Доситеја Обрадовића) – повратни глаголи су посебно обележени.

2.8. У сваком случају, ако се израђује фреквенцијски речник на мањем корпусу, могуће је ручном обрадом утврдити тачно колико пута *се* је заменица а колико пута речца, и то навести (нпр. *се* зам. 35, речца 355); притом, требало би додатно размотрити да ли *се* које је део повратних глагола треба бројати код одредничке речи *се(бе)*. Међутим, ако се ради на великом корпусу (са преко милион речи), није могућа ручна обрада, а софтвер, колико знамо, не може да разликује *се* које је заменица од оног које је речца и тада се може дати само укупан број појављивања речи *се* и нагласити да је та реч и речца (а не само заменица).

3. Бити, ЈЕСАМ (СУПЛЕТИВНИ ОБЛИЦИ). Глагол *биџи* у презенту поред регуларног облика (*будем*) има и суплетивну форму *јесам*. У РСАНУ *јесам* је само упућено на *биџи*. У РМС је дата информација да се ради о 1. лицу јединине презента помоћног глагола *биџи*, док су у РСЈ наведени и остали облици овог глагола у презенту, као и енклитичке варијанте, те је наведено да је то суплетивни презент глагола *биџи*. Питање је да ли је оваква обрада глагола (*биџи* и *јесам*) јасна препорука састављачу фреквенцијског речника шта треба да ради и како да поступи приликом обраде најфреквентнијег глагола у српском језику.

3.1. Вера Лукић у студији *Дечија лексика речи биџи и јесам* наводи засебно, као две речи (Лукић 1982). Исто решење налазимо и код Здравка Вукчевића у *Фреквенцијским речницима Горској вијенца и Луче Микрокозма* (Вукчевић 2005), као и у *Фреквенцијском речнику Пријоведака Лазе Лазаревића* Браниславе Јелић (Јелић 2008), те у фреквенцијским речницима које је приредила Зорица Ивановић (Ивановић 2016; 2018).

Исти поступак примењује Смиљка Васић за именице које немају правилне облике множине, те наводи да су дате „као посебне одреднице (на пример: **човек** – **људи**)” (Васић 1998: 48). Међутим, за глагол *биџи* и његов суплетивни облик *јесам* поступак је супротан – „дати су као једна одредница са скупном фреквенцијом” (Васић 1998: 50). И у *Фреквенцијском речнику четири јеванђелиста* (у *Полазним основама новије српске прозе*, књ. 1) глагол *биџи* и његов суплетивни облик у презенту *јесам* обрађени су као једна реч (наведено је само *биџи*) (Васић 1996).

3.2. Мишљења смо да суплетивне облике треба придружити основном облику. Управо тако су поступили и аутори речника *A Frequency Dictionary of Czech*: „Suppletive and other irregular forms are subsumed under the base form, i.e. the headword **člověk** “man” also includes its plural form **lidé** “people”; the headword **dobrý** “good” also includes its comparative form **lepší** “better” etc.” (Čermák – Křen 2011: 5)

4. ВАРИЈАНТЕ. Бавећи се утврђивањем тачног броја речи у Вуковом Српском рјечнику, Славко Вукомановић (1999) указује на чињеницу да у том речнику постоји велики број варијаната и да је укупан број речи у њему

знатно мањи него што је Вук навео (47.427 речи). Вукомановић се пита да ли су те варијанте (нпр. *фала*, *вала* и *хвала*) „... засебне лексеме или су то само лексикографске одреднице исте речи?” и даје одговор: „Шта су онда све оне Вукове многобројне одреднице које се у различитим гласовним варијантама својим везују за једну примарну реч тако што означавају исти појам и што су с том речи у најбитнијим акустичким елементима истоветне, подударне. Из свега што је напред речено мислим да је јасно да су то само гласовне варијанте једне лексеме, исте речи” (Вукомановић 1999: 117). У РСАНУ варијанте неке речи често су наведене код стандардне речи (нпр. многобројни дијалектизми – нпр. *ђевојка* под *девојка/дјевојка*) или су наведене засебно с упућивањем на (стандардну) реч (нпр. *Белоџрађанин* је дијалектизам уз који се даје упућивање на стандарду варијанту *Беоџрађанин*). Како поступати са оваквим и другим варијантама при изради фреквенцијских речника – да ли те речи наводити у фреквенцијским речницима одвојено или заједно? Поступак упућивања једне речи на другу из описних речника и уопште појаву варијаната у фреквенцијским речницима могуће је разрешити или у правцу формирања засебних одредница или у правцу формирања једне заједничке одреднице.

Ако се ради о (гласовно сличним) варијантама (нпр. *човек/чоек*, *девојка/ђевојка*, *вече/вечер*) које имају исто значење и употребу, чини се да има више смисла у фреквенцијским речницима обрадити их заједно, као једну реч, па их тако и бројати (као што сугерише и Вукомановић у поменутом раду). Међутим, у постојећим фреквенцијским речницима такве речи углавном су обрађене засебно, као посебне одреднице. У таквом приступу најдаље се отишло у *Фреквенцијским речницима Горској вијенца и Луче Микрокозма* (Вукчевић 2005) у којим су посебно наведене чак и екавске и јекавске варијанте (нпр. *време* и *вријеме*, *колевка* и *колијевка*, *млеко* и *млијeko* итд.). Ипак, неке варијанте су наведене заједно (*каг(a)*, *џог(e)*, *говијек(a)*, *докле(н)*, *јер(e)*, *још(џи, џе)* итд. Иако су варијанте са покретним вокалима у фреквенцијским речницима углавном обрађене заједно, ни то није правило у свим фреквенцијским речницима. Тако су у *Фреквенцијском речнику њрозе Боре Сиџанковића* (у *Фреквенцијским речницима школске лектиуре*, друга књига) ове варијанте наведене засебно: „Тамо где је било гласовних промена, као што је покретни вокал, навођена су оба примера као посебне јединице (*каг/кага*)” (Ивановић 2018: 6). Ако је корпус мањи, варијанте се могу уз додатну мању ручну обраду објединити; међутим, ако је корпус већи, варијанте могу бити бројне и веома разноврсне и тада обједињавање зависи од софтверских могућности.

5. ВИШЕВРСНЕ РЕЧИ. Познато је да у српском језику има и речи које у одређеном контексту припадају једној врсти речи, а у другом некој другој врсти речи. Дакле, ради се о једној речи која може припадати већем броју врста речи. У описним речницима оне су најчешће обрађене тако да су најпре дата значења за једну врсту а потом за другу. Али, ни то није увек случај, па тако има и лексема код којих је само на почетку дата напомена да реч припада двама врстама, али даље код навођења значења нема напомене о којој врсти се ради (тако је, на пример, обрађена заменица *сав* у РМС и РСЈ). На који начин треба обрадити (и бројати) вишеврсне речи у фреквенцијским речницима?

Ако за такву реч само наведемо колико је пута употребљена, неће се знати колико пута је једна врста, а колико пута нека друга врста (приликом утврђивања броја речи по врстама). Или би требало пребројати и навести колико је пута употребљена као једна, а колико као друга врста речи (нпр. *како* прил. и везн. 43/76)?

Само је Здравко Вукчевић, колико нам је познато, све вишеврсне речи раздвојио и навео их као посебне одредничке речи. Тако се у *Фреквенцијским речницима Горској вијенца и Луче Микрокозма* као одредничка реч *када* јавља два пута – једном као прилог, а други пут као везник, дакле као хомоними (што је одступање од онога што налазимо у описним речницима). Избројати колико је нека лексема употребљена као једна, а колико као друга (и евентуално трећа) врста речи могуће је, колико знамо, само ручно, а ово је опет изводљиво само у раду са мањим корпусима.

6. Вишечлани (ГЕОГРАФСКИ) НАЗИВИ. Чак и када је нека одредничка реч у описним речницима утврђена на исти начин то не мора да буде случај и у фреквенцијским речницима за исту ту реч. Овде, пре свега, мислимо на вишечлане одредничке речи. Иако постоји више различитих врста вишечланих назива (или вишечланих лексичких јединица), ми ћемо се у овом раду бавити само вишечланим географским називима.

Полазећи од тога да се реч као основна јединица у фреквенцијском речнику може дефинисати као садржај између две белине, у фреквенцијским речницима готово да и нема двочланих одредничких речи. Тако, на пример, Владимир Вукомановић Растегорац, пишући о фреквенцији најчешћих имена заступљених у предзмајевској поезији за децу, наводи да „делови географских назива који се састоје из две речи се такође узимају као посебне одреднице”, као, на пример, Нови и Сад (2017: 71). Готово сви састављачи фреквенцијских речника поступају на исти начин када су у питању географски називи. Само у раду Милице Динић Маринковић наилазимо на другачији приступ: „Као резултат овог приступа у фреквенцијском речнику ЗУНС5 се као посебне јединице налазе и вишечлане лексичке јединице – термилошке синтагме и 'оними'” (2016: 255). Мирослав Николић, говорећи о проблемима утврђивања одредничке речи у лексикографији, за дефиницију да је реч 'одсечак исказа ограничен у писму белинама' каже да је практична, али да је лингвистички ирелевантна (2014: 62). Приступ у обради двочланих географских појмова који је доминантан погрешан је јер представља одступање од начина обраде (утврђивање одредничке речи) поменуте лексике у нашој описној лексикографији – двочлани географски називи обрађују се као јединствена одредничка реч, на шта упућује и М. Николић: „Неспорна су, наравно, властита имена (топоними, хидроними итд.) типа *Владичин Хан*, *Нови Сад*, *Херцеџ Нови* и сл., која се као таква наводе у Речнику САНУ” (2014: 62–63). Са развојем софтвера у 21. веку у европској корпусној лингвистици све више се тежи ка решењима која ће регистровати и различите врсте вишечланих назива (енгл. multiwords, multiword expressions, phrasal verbs, idioms).

7. Закључак. Претходна анализа показује да има доста неуједначеног поступања када је у питању процес утврђивања одредничке речи у фреквен-

цијским речницима, што доводи у питање и тачност тако добијених резултата и што чини те речнике, односно добијене резултате мање или више непогодним за поређења. Постоји потреба да се, поготово за најфреквентније речи, одредничка реч утврди на исти начин – најпре за речи које спадају у најфреквентније, као што су *себе*, *се* и *бићу/јесам*, као и за неке групе фреквентнијих речи: суплетивне облике (уопште), вишечлане (географске) називе, вишеверсне речи, варијанте. Тако, на пример, по нашем мишљењу, инфинитив *бићи* и презент *јесам* не би требало одвајати и наводити као засебне одреднице јер се ради о истом глаголу; исто тако треба поступати и са другим суплетивним облицима (нпр. *човек* и *људи*), вишечланим (географским) називима (нпр. *Нови Сад*), варијантама (речима које су гласовно сличне и имају потпуно исто значење и употребу, нпр. *девојка*, *дјевојка* и *ђевојка*, *каг* и *кага*). Списак дилема који је у раду изнесен није коначан. Важно је истаћи и да, с једне стране, реализација предложених решења у великој мери зависи од величине корпуса и могућности софтвера, и, с друге стране, у случају мањег корпуса спремности аутора фреквенцијског речника да део посла обави ручно. Реално је очекивати да ће у скорој будућности софтвер за израду фреквенцијских речника (попут платформе Sketch Engine) бити још квалитетнији и омогућити да учешће човека буде сведено на минимум, односно да ће се већина послова које данас још увек морамо радити ручно обављати помоћу рачунара.

ЦИТИРАНА ЛИТЕРАТУРА

- Васић, Смиљка. *Полазне основе новије српске прозе. Књ. 1, Речник четири јеванђелиста*. Београд: Институт за педагошка истраживања, 1996.
- Васић, Смиљка. *Полазне основе новије српске прозе. Књ. 2, Хазарски речник Милорада Павића: фреквенцијски речник* (1. изд.). Београд: Завод за уџбенике и наставна средства – Институт за педагошка истраживања, 1998.
- Вукомановић, Славко. Речи и њихове одреднице у Вуковом другом српском рјечнику. *Српски језик* 4/1–2 (1999): 203–225.
- Вукомановић РАСТЕГОРАЦ, Владимир. Лексика као огледало времена у контексту предзмајевске поезије за децу. *Детињство* XLIII/1 (2017): 69–79.
- Вукчевић, Здравко. *Фреквенцијски речници Горској вијенца и Луче Микрокозма*. Подгорица: ЦАНУ, 2005.
- Груден, Тамара. *Политички дискурс у паролама, слоанима и графичким процесама процеса НАТО бомбардовања СР Југославије 1999*. Докторска дисертација, Универзитет у Београду, 2016.
- Динић Маринковић, Милица. О најфреквентнијим лексичким јединицама у уџбенику историје за пети разред. Зорка Кашић (ур.). *О језику и језицима*. Нови Сад – Београд: ДПЈС, Филозофски факултет Универзитета у Новом Саду – Филолошки факултет Универзитета у Београду, 2016, 253–270.
- Ивановић, Зорица (прир.). *Фреквенцијски речници школске лектире*, Књ. 1. Параћин: Библиотека „Др Вићентије Ракић”, 2016.
- Ивановић, Зорица (прир.). *Фреквенцијски речници школске лектире*, Књ. 2. Параћин: Библиотека „Др Вићентије Ракић”, 2018.
- Ивић, Милка. Један проблем словенске синтагматике осветљен трансформационим методом (граматичка улога морфеме се у српскохрватском језику). *Јужнословенски филолоџ* 25 (1961–1962): 137–151.
- Јелић, Бранислава. *Речник ириводеака Лазе Лазаревића*. Нови Сад: Матица српска, 2008.

- Кликовац, Душка. *Грамаџика српској језика*. Београд: Српска школска књига, 2003.
- Лукић, Вера. *Дечја лексика*. Београд: Институт за педагошка истраживања – Просвета, 1982.
- Лукић, Вера. *Дечји фреквенцијски речник*. Београд: Институт за педагошка истраживања – Просвета, 1983.
- Николић, Мирослав. Утврђивање одреднице у српској лексикографији. *Савремена српска лексикографија у теорији и пракси*: (колективна монографија). Београд: Филолошки факултет, 2014, 61–68.
- Павковић, Васа. *Речник њезије Милана Ракића*. Нови Сад: Матица српска, 1984.
- ПИПЕР, Предраг, Иван Клајн, Рајна Драгићевић. *Нормативна грамаџика српској језика* (4. измењено и допуњено изд.). Нови Сад: Матица српска, 2022.
- Ристић, Стана. Партикуле и њихови функционални еквиваленти (лексичко-семантичке у функционалне карактеристике). *Јужнословенски филолоџ* XLIX (1993): 75–93.
- РМС: *Речник српскохрватској књижевној језика*, I–VI. Нови Сад: Матица српска, 1967–1976.
- РСАНУ: *Речник српскохрватској књижевној и народној језика*, 1–. Београд: Српска академија наука и уметности – Институт за српски језик, 1959–.
- РСЈ: *Речник српској језика* (ред. М. Николић). Нови Сад: Матица српска, 2011.

*

- ČERMÁK, František, Michal KRÉN. *A Frequency Dictionary of Czech: Core Vocabulary for Learners*. New York: Routledge, 2011.
- KILGARRIFF, Adam. Using Corpora as Data Sources for Dictionaries. Howard Jackson (ed.). *The Bloomsbury Companion to Lexicography*. London: Bloomsbury Publishing, 2013, 77–96.
- ТАФРА, Бранка. Грамаџички критерији за гјечничку натукницу. *Расправе ЗНЈ* 20 (1994): 300–312.
- ТАФРА, Бранка. Двојбе при одређивању натукнице. Стана Ристић, Иван Лазич Коњик, Ненад Ивановић (ур.). *Лексиколоџија и лексикографија у свејилу савремених њрисијуа*. Београд: Институт за српски језик САНУ, 2016, 515–532.

Marijan Jelić

ON SOME PROBLEMS OF DETERMINING LEMMATA IN FREQUENCY DICTIONARIES

Summary

This paper analyzes how lemmata are determined in frequency dictionaries, and for this purpose we chose those words and word groups that are more frequent and for which there are different approaches to processing (*sebe, se, biti/jesam*, variants, multi-class words, two-part names), which caused significant differences in the obtained results. The research shows uneven treatment of determining lemmata in frequency dictionaries, which challenges the accuracy of the results obtained in this way, which makes those dictionaries or results more or less unsuitable for comparisons. There is a need to approach determining lemmata in a uniform way, especially for the most frequent words. It would be beneficial if solutions were found for as many existing ambiguities as possible when determining lemmata in frequency dictionaries.

It is possible that some better solutions would involve greater human involvement, at least at this moment. On the other hand, it is realistic to expect that in the near future, by using even better software for creating frequency dictionaries, this disadvantage will be removed or at least reduced, and that this part of the work will also be performed using a computer, and not done manually.

Универзитет у Новом Саду
Педагошки факултет
Подгоричка 4, 25000 Сомбор, Србија
jeliciso@uns.ac.rs

(Примљено: 27. марта 2024;
прихваћено: 20. маја 2024)