

Саша П. МАРЈАНОВИЋ*
Универзитет у Београду
Филолошки факултет

Оригинални научни рад
Примљен: 28. 2. 2022.
Прихваћен: 10. 5. 2022.

КОРПУСНА АНАЛИЗА ПОРЕДБЕНЕ ФРАЗЕМЕ КАО БЕЛА ЛАЛА**

У раду се представљају резултати корпусне анализе поредбене фраземе *као бела лала* у српском језику. Формалне одлике ове фраземе (њена структура, устаљеност и парадигматичност) и значењске одлике (експресивност, идиоматичност и конотативност) испитане су на грађи ексцерпираној с мрежног претраживача *Google*. У раду се такође износе закључци о домену употребе ове фраземе, и осветљава се и њена мотивацијска база. Циљ је рада да се покаже да се одлике не само наведене, већ и осталих фразема, на одговарајући начин могу испитати уз помоћ електронских корпуса састављених од интернетске грађе.

Кључне речи: поредбена фразама, *као бела лала*, фразеологија, корпус, речник, српски језик.

1. Уводна разматрања

1.1 **Уводна разматрања.** Повећи број скорашњих фразеолошких истраживања у Србији и окружењу (нпр. Видовић Болт 2011; Пејановић 2015; Вуловић 2015; Штрбац 2018) спроводи се на грађи ексцерпираној из постојећих једнојезичних речника, међу којима предњаче МФР, РМС и, упадљиво ређе, РСАНУ. Међу овим трима речницима, најобухватнији је лексикографски извор за српску фразеологију – РСАНУ; он је заснован на импозантној грађи од око 5.000 претежно књижевних дела и периодике, којој је придружено око 300 збирки речи из народних говора (уп. Ристић/Ивановић 2011: 531; Стијовић 2017: 202; Ивановић и др. 2016: 143). Уколико би се грађа за РСАНУ дигитализовала, за шта, како сазнајемо из управо цитираних рефе-

* sasa.marjanovic@fil.bg.ac.rs

** Рад је настао у оквиру пројекта *Језици и културе у времену и простору* (бр. 178002), који је до 2020. године финансирао Министарство просвете, науке и технолошког развоја Републике Србије.

ренци, постоје намере, претпостављамо – ако се рачуна пропорционално – да би она садржала до 300 милиона појавница. Међутим, иако ова грађа наизглед сачињава огромну количину података за поуздан лексикографски опис, она, како то показују фреквенцијска испитивања фразеолошког фонда, нису довољна.

Наиме, Р. Мун (1999: 267) је, пошто је испитала учесталост фразема у референтном електронском корпусу енглескога језика од 211 милиона појавница, у којем је удео писаног језика велик, изнела податке да се од 4.000 фразема, за које просечан изворни говорник може проценити да су учестале у језику и које су укључене у један од фразеолошких речника енглеског језика, за само 750 фразема, а то је нешто мање од 20% грађе, проналази потврда једном на два милиона појавница; за 750 фразема налази се 3–5 потврда на десет милиона појавница, за додатних 1.500 јављају се до три потврде, док за преостали део грађе постоји мање од једне потврде на истом броју појавница (уп. Мун 1998: 59–60). Р. Мун је своје податке поредила и с налазима других истраживача (Мун 1998: 64–68), и показала да су они у потпуној корелацији. Ова ауторка истакла је такође да удео фразема варира у зависности од жанра текстова који су укључени у корпус, али је утврдила да је и у тада постојећим корпусима говорног енглеског језика удео фразеологије мали (Мун 1998: 68–74). Слично њој, И. Пенадес Мартинез (2012: 58–59) показала је да ни у транскрибованом корпусу од 27 сати сниманог разговора на шпанском језику број фразема није велик: на 322.000 појавница ова је ауторка поткрепила тек 515 глаголских фразема.

У класичним електронским корпусима (као што су, иначе, корпуси на којима су радиле Р. Мун и И. Пенадес Мартинез), како то и Ж.-П. Колсон (2003: 47) потврђује, многе се глаголске фраземе францускога и низоземског језика које је овај аутор испитивао, али и већине осталих европских језика, јављају с учесталошћу мањом од једне потврде на милион корпусних појавница. Из тога Колсон (2003: 59) закључује да би се електронски корпуси морали састојати од милијарди појавница како би садржали довољан број фраземских потврда, те тако били репрезентативни за одговарајућа истраживања. С тим у вези, истичемо и то да је Р. Мун (2007: 1050–1051) таблично показала број потврда за четири фраземе енглеског језика на корпусном узорку од 1, 20, 50, 100 и 450 милиона појавница и закључила да је корпус од 50 милиона минимум за фразеолошка истраживања, а да тек корпуси од 100 милиона појавница дају одговарајућу слику. То значи да МФР и РМС, који се обично користе у домаћим фразеолошким истраживањима и чији корпус чини око 800 дела, тј. око 50 милиона појавница, нису засновани на довољно репрезентативној грађи. Наш закључак подупиरे и недостатак примера уз већи број фразема у овим речницима. Наспрам њих, рекло би се да је грађа РСАНУ довољна. Но, будући да се РСАНУ израђује исписивањем извора на листиће (Ивановић и др. 2016: 143), где, сасвим сигурно, исписивачи бележе само потврде за лексеме и остале језичке јединице које им се учине „интересантним” (Павковић 2014: 57), а да се не служе електронским конкорданцијама које би им приказале *све* потврде у изворима, износимо сумњу да

обрађивачи и редактори не располажу целокупном грађом ексцерпираним из извора од горе претпостављених 300 милиона појавница, а то значи да остварени лексикографски опис може бити искривљен или пак само непотпун. Уосталом, С. Ристић и Н. Ивановић (2011), те Т. Самарџић (2011) показали су на конкретним примерима како постојећа обрада није потпуна и како се употребом електронских корпуса лексикографски опис може унапредити.

Ову оцену подупиру и наши корпусни налази од пре скоро деценију (в. Марјановић 2014), где смо на примеру поредбене фраземе *пасти као зрела крушка* показали да постојећи речници неадекватно обрађују ту фразу. Наш рад у томе смислу није једини: Ј. Паризоска и И. Филиповић Петровић (2016) поредиле су обраду 83 фраземе у Бенешихеву речнику с потврдама у двама хрватским референтним електронским корпусима од, редом, 216,8 милиона и 1,9 милијарди појавница, те нашле да обрада фразема у Бенешихевом речнику не одражава садашње стање језика, да се у питање доводе како устаљеност употребе и облика фразема у речнику, тако и њихова варијантност (уп. и Филиповић Петровић 2018: 96–138). Хрватске ауторке (Паризоска/Филиповић Петровић 2016: 155) истакле су да би у савременој лексикографији требало представљати употребни модел језика, до којег се најлакше долази претраживањем великих, али балансираних, корпуса, тј. корпуса који би, поред књижевних дела, новина и часописа, садржали грађу с интернетских форума и блогова. Има домаћих аутора који се с овим приступом не би сложили (нпр. М. Радовић-Тешић, 2009: 251), стога што некњижевним текстовима поричу значај, али приступ хрватских ауторки слаже с другим, на светском нивоу, утицајним истраживачима: на пример, Ж.-П. Колсон (2007: 1072) позива се на горепоменуто фреквенцијска истраживања Р. Мун и износи податак да се за неке фраземе, као што је *spill the beans* (енг. 'изланути се') (за коју је у највећим класичним електронским корпусима било 0,56 потврда на милион појавница), у томе тренутку на интернету преко мрежног претраживача *Google* могло пронаћи чак 34.000 потврда. Слично поређење аутор је спровео и за француску фразу *empêcheur de tourner en rond* (фр. 'онај ко квари расположење') и њене варијанте (Колсон 2007: 1074): у новинским корпусима од свега неколико десетина милиона појавница нашло се тек неколико потврда, док је с интернета врло лако екстраховано 500 потврда, које су омогућиле да се стекне увид у заступљеност фраземских варијанти. Аутор на основу тога закључује да интернет пружа у кратком временском периоду најпотпунији корпус фразеолошке грађе (уп. и Мун 2008; Шатара 2008). То несумњиво показује да су корпуси засновани на интернетској грађи, која је диверсификована, које је лако доступна и која се данас свакодневно може осавремењивати, од непроцењиве важности у различитим лингвистичким – па тако и фразеолошким – истраживањима. Приступ назван *Web as Corpus* (енг. 'интернет као корпус'), у којем се интернетски садржај и интернетски претраживачи користе као извор корпусне грађе, у последње су се време уобичајили у корпусној лингвистици (уп. Гато 2014). На закључак да су интернетски корпуси, који се одликују нецензурисаним и размерно спонтаним језиком, добар извор за фразеолошка истраживања наводе и подаци Р. Мун (2007: 1051–1052), која обавештава о томе да су у таблоидној периодици у поређењу с осталим новинским текстовима фраземе далеко учесталије.

1.2. **О овоме раду.** Циљ је овога рада да се на примеру српске поредбене фраземе *као бела лала* покаже да интернетска грађа може послужити и у истраживању фразеолошке грађе *савременога* српског језика и да због релативно великог броја потврда фразеологу може дати поуздане резултате, те да у фразеолошким истраживањима не би требало полазити од речника заснованих на данас застарелом и недовољно обимном корпусу. Рад се, стога, састоји од три одељка: после ових уводних разматрања о значају интернетских корпуса у фразеолошким истраживањима (§1), у наредном одељку најпре говоримо о поставци сопственог истраживања, како бисмо потом изнели резултате корпусне анализе поредбене фраземе *као бела лала* на интернетској грађи (§2); коначно, у трећем одељку резимирамо закључке анализе и отварамо перспективу за могуће примене добијених резултата (§3).

2. Корпусна анализа

2.1. **Израда електронског корпуса.** Како бисмо испитали формалне и значењске одлике поредбене фраземе *као бела лала* и показали може ли анализа електронског корпуса састављеног од интернетског садржаја дати довољан број потврда и поуздане податке о фразеолошкој грађи српскога језика, претражили смо предметну фразему на мрежном претраживачу *Google* (<https://www.google.com>). Најпре смо претраге спровели на наведеном облику фраземе, а потом на могућим варијантним ликовима њених саставница: у претрагу је, стога, укључен како ијекавски лик придевске саставнице (*бијела*), тако и варијантни гласовни и правописни ликови везничке саставнице (*ко* и *к'о*). Даље, све варијантне ликове фраземе претражили смо и у множинском облику (*као беле лале* итд.). При претраживању нисмо ограничавали временски распон ни интернетски домен, нити дефинисани језик интернетских страница.

Резултате сваке претраге аутоматски смо екстраховали с мрежног претраживача. При екстракцији резултата за сваки је резултат уз контекстуализовану појавницу фраземе аутоматски додат њен претражени лик, година кад је појавница објављена на интернету, као и хипервеза ка страници на којој је резултат пронађен, ради евентуалних додатних провера контекста. Резултати код којих је било поклапања аутоматски су уклоњени. Резултате смо похранили у табличну базу података. Базу смо потом вишеструко аотирали: за сваку смо појавницу фраземе додали и ознаку које је структуре, уписали смо лематски облик њеног колоката (а у случају глаголских колоката, које смо, иначе, и очекивали, и ознаку глаголског облика, лица и броја у којем се колокат појављује), субјекат колоката и његову архисему, као и значење фраземе у контексту. Аотирана база чини електронски корпус за наше истраживање.

2.2. **Резултати корпусне анализе.** Актуелност и значај анализе описаног корпуса – састављеног од интернетске грађе екстраховане путем мрежног претраживача *Google* – подупире чињеница да се за поредбену фразему

као бела лала у класичном електронском корпусу *СрпКорп2013*¹ од 122 милиона појавница налази свега 7, а у интернетскоме електронском корпусу *srWaC*² од 550 милиона појавница само 39 потврда. Наш електронски корпус обухвата чак 774 потврде предметне фраземе. Потврде су објављене на интернету у временском распону од 20 година, тј. од 2001. до 15. септембра 2021. године. Дистрибуција потврда по годинама није равномерно распоређена: за прву деценију (2001–2011) прикупљене су 102, а за другу (2012–2021) 542 потврде, од чега број потврда за период 2019–2021. чини 32,1% грађе. За 16,8% потврда није било могуће утврдити годину. Овај размерно велик број потврда, нарочито за скорашњи период, омогућује нам да спроведемо квантификацију аотираних података на *савременој* грађи и изведемо закључке о формалним и значењским одликама предметне фраземе у *савременоме* српском језику. Другим речима, корпусном анализом испитујемо, с једне стране, устаљеност и састав фраземе (§2.2.1), њену структуру (§2.2.2) и парадигматичност (§2.2.3), те, с друге стране, њену експресивност (§2.2.4), идиоматичност (§2.2.5), и конотативност (§2.2.6). Пошто на експресивност фраземе утиче и њена мотивацијска база, током корпусне анализе се осврћемо и на ово питање (§2.2.4).

2.2.1. *Устаљеност*. Квантификовањем варијантних ликова фраземе долазимо до два налаза у вези с њеним лексичким саставом. *Прво*, да везничка саставница варира у великом броју: лик *као* присутан је у свега 41% случајева, док су ликови *ко* и *к'о* чешћи (редом 57,5%, и 1,5%). Ово је варирање очекивано, јер је већина потврда део инстантне и неприпремљене комуникације корисника интернета у виду статуса и коментара с друштвених мрежа, реакција на новинске чланке, у виду постова учесника форума, објава с блогова и сличних интернетских садржаја, као и у виду реакција других корисника на те објаве. Таква комуникација, мада у писаном облику, прилично је блиска неформалном говорном језику, у којем се скупина *-ао* по правилу сажима у *-о*. *Друго*, дошли смо до неочекиваног налаза да се придевска саставница у свега 9 појавница јавља у ијекавском лику (1,2% корпусне грађе). Извор тих појавница представљају црногорски сајтови с доменом *.me* или *.net*. Мали број ових појавница наводи нас на закључак да се фразема тек периферно користи ван екавског ареала. Овакву дијатопијску обележеност нисмо, на пример, нашли за фразему *пасти као зрела крушка* (в. Марјановић 2014: 379–380). Но, у 107 потврда с црногорских сајтова (14% корпусне грађе), где се у левом и/или десном контексту, унутар исте или ван те реченице, јасно види доследна употреба лексема с ијекавским рефлексом јата — примећује се екавски лик придевске саставнице (в. примере у табелици 1 и 2). С обзиром на то да је број интернетских страница с црногорским доменом неупоредиво мањи него број страница с доменима регистрованим у Србији, налазимо да је

¹ Доступно на: <http://www.korpus.matf.bg.ac.rs/korpus/korpus2013adv.php> (последњи приступ: 10. 9. 2021).

² Доступно на: <https://www.clarin.si/noske/all.cgi/first?corpname=srwas> (последњи приступ: 10. 9. 2021).

овај постотак статистички значајан и закључујемо да се предметна фразема у Црној Гори усталила у екавском облику.

2.2.2. *Структура*. Квантификовањем различитих колоката фраземе долази се до њене структуре и категоријалног значења. Фразема у само 14 појавница (1,8% грађе) колоцира с придевима: *миран* (9 пута), *леп* (2), *опуштен*, *фин*, *покоран* (1); док су у осталим случајевима, као што смо претпоставили (уп. 2.1), колокати глаголи. Од 98,2% колоката, глагол *бити* је најучесталији, али се јавља у само 53 појавнице. То значи да је у 6,8% грађе фразема придевскога категоријалног значења, док је у осталим случајевима прилошка (в. примере у табелици 1). Придевско категоријално значење заступљено је у 19 случајева у потврдама с подручја Црне Горе, док су остале потврде објавили екавци, тачније говорници с подручја Србије. С обзиром на то што је интернетских страница с црногорским доменима мање, закључујемо да је придевска употреба, статистички гледано, нешто активнија у Црној Гори него у екавском ареалу. Кад је у питању прилошко категоријално значење, примећује се и то да је дисперзија колоката велика: од укупно 241 глагола с којима фразема колоцира у корпусној грађи, само 32 имају појавну учесталост 5 и већу од тога броја. Ови глаголи улазе у састав 54% корпусних потврда:

ићи (48)	вратити (15)	плаћати (9)	падати (6)
п(ј)евати (41)	спавати (14)	трчати (9)	стајати (6)
радити (27)	гласати (13)	устајати (9)	возити (6)
играти (24)	потписати (12)	доћи (8)	ћутати (6)
лет(ј)ети (24)	изаћи (12)	признати (8)	прихватити (5)
слушати (17)	јести (14)	проп(ј)евати (7)	примити (5)
хт(ј)ети (16)	отићи (11)	процв(ј)етати (7)	вакцинисати се (5)
платити (15)	носити (10)	с(ј)ед(ј)ети (7)	заспати (5)

Велика дисперзија колоката поткрепљује чињеницу да у структуру фраземе не улази глагол, већ да је он њен колокат. Кад се попис учесталих глаголских колоката ближе погледа, запажа се да међу њима преовлађују глаголи с обележјем динамичности, а чини се да су глаголи стања у целомитом попису искључени, али би то требало даље испитати. Затим, запажа се да су у приказаном попису глаголи несвршеног бројнији од глагола свршеног вида (редом 19 и 13). При томе, глагол *хтети/хтјети* у корпусним се потврдама јавља као колокат фраземе у склопу футурског облика кад говорник уверава да ће се радња извршити пошто је саговорник претходно одрекао ту могућност.

2.2.3. *Парадигматичност*. Квантификовањем појавног облика фраземе и њеног колоката изводи се закључак о степену зависности њеног облика од контекста, тј. уклапања њенога лематског облика у контекст. Једински облик фраземе (*као бела лала* итд.) потврђен је у 398 (51,4% грађе), а множински облик (*као беле лале*) у 376 случајева (48,6%), што значи да је предметна фразема, по терминологији Ж. Финк-Арсовски (2002: 21–25), делимичне парадигматичности. Да би се одредио степен репродуктивности фраземе у контексту (уп. Марјановић 2017: 95–140), потребно је видети да ли парадиг-

матичност фраземе иде у пару с парадигматичношћу колоката. Анализа корпуса показује да се уз множински облик фраземе у свим случајевима користи колокат у множини. Кад би се тај колокат пребацио у једнину, и фразема би се очекивано употребила у једнини. То је тип репродуктивности који смо 2017. означили као 2:1. Корпусна грађа, међутим, показује да се уз једински облик фраземе не користи увек колокат у једнини. Наиме, 16 појавница (2% грађе) карактерише се типом репродуктивности 2:0. То значи да глагол има потпуну парадигматичност, а фразема нулту, тј. да се њен облик у том контексту не мења зависно од множинског облика глагола. С обзиром на мали постотак овог типа репродуктивности, може се закључити да је главна тенденција у корпусу репродуктивност типа 2:1. Кад се ближе погледа облик глагола, који се одликује високом степеном парадигматичности, констатујемо да се глаголски колокат јавља у следећим облицима:

футур (199) футуroid (51)	презент (157) да + презент (146)	потенцијал (130) перфекат (75) аорист (16)
------------------------------	-------------------------------------	--

Квантификација аотираних података показује да се колокат најчешће налази у футуру, било да је стандардни или нестандардни футур, тзв. футуroid (32,3%), у презенту (20,3%), као допуна у презенту с везником *да* (18,9%), и то најчешће у безличним конструкцијама (*има да* + презент), те потенцијалу (16,8%), док се аорист и перфекат јављају у мањем броју случајева. Кад је реч о лицу, најзаступљеније је, сасвим очекивано, треће лице (66,8%), следе друго (25,2%) и прво (8%). Кад се укрсте подаци за лице и облик, добије се следећи налаз: глагол у 2. лицу налази се најчешће у футуру (47%) и у презенту као допуна безличне конструкције (21%), док су остали облици упадљиво ређи; у 1. лицу подједнако се јављају сви облици, док је у 3. лицу најучесталији футур (26%), којем следе презент (22%), потенцијал (19,5%) и презент као допуна безличне конструкције (18%). Ови резултати могу бити смерница за припрему примера у речнику.

2.2.4. *Експресивност*. Све поредбене фраземе, па тако и предметна, поседују изразиту експресивност (уп. Марјановић 2017: 154–155). На експресивност фразема утиче њихова мотивисаност. Мотивацијска база предметне фраземе није, међутим, прозирна за просечног говорника српскога језика, па је њена експресивност, чини нам се, израженија. На експресивност, иначе, како показују све корпусне потврде, указује њена употреба у крајње неформалној комуникацији. Може се рећи да је из синхронијског аспекта реч о, како то И. Видовић-Болт (2011: 47–48) назива, нелогичном или апсурдном мотиву. Непрозрачношћу подупире чињеница да је од 32 студената друге године француског језика, књижевности и културе на Филолошком факултету у Београду, којима је српски језик матерњи, у електронској анкети коју смо спровели у мају 2021. године, на питање зашто се каже *као бела лала* 94% одговорило да не зна и покушало у другоме делу одговора на наиван начин ремотивисати базу фраземе. У тим одговорима испитаника листом је поминуто фитоним.

С друге стране, два су испитаника у одговору истакла да је *бела лала* друго име за невесту, уз накнадну ограду да су то научили од наставника у гимназији. Према делу је логично да, како су та два испитаника образложила, мотивацијску базу чини покоран однос новопридошле жене према старијим члановима својте, асоцијативна испитивања наивне слике невесте и њеног односа према члановима породице у коју је ушла, у српском језику односно култури не потврђују те одговоре (уп. Стефановић 2010а; Стефановић 2010б). У изналажењу могуће мотивацијске базе ми смо кренули од РСАНУ. У њему је (РСАНУ, том 11, стр. 202) дат стални епитет *бела лала*, уз који стоји значењска парафраза „млађа женска особа у кући”. Ако се још пажљиво прочита цитат из РСАНУ („Овако новодоведене Младе предавају имена челядма у оној кући, гди се доведу и прочима сродницима ... г) Омлађини женској ... бела лала”), закључује се да су у питању тзв. невестинска имена (уп. Сикимић 1998), тј. имена које је невеста имала обичај надевати новим укућанима. У овоме случају, реч је о именовању детета женског пола. С овим сазнањем може се претпоставити да је значење фраземе засновано на, према терминологији Д. Стаменовића (2011: 198), релативно објективном мотиву: деца су у традиционалној породици у подређеном односу према старијим члановима домаћинства, па су у обавези да без поговора извршавају све њихове наредбе, а, додатно, девојчице у подређенијем односу према дечацима (уп. Требјешанин 1993: 304–311).³

2.2.5. *Идиоматичност*. С обзиром на управо објашњену мотивацијску базу, може се претпоставити да се значење поредбене фраземе сужава на значење доминантних обележја сложене лексеме *бела лала*, а то би биле 'послушност', 'беспоговорност' и 'очекиваност'. Но, будући да је та лексема за просечног говорника српскога језика непозната, а фразема немотивисана, можемо закључити да је реч о фраземи глобалног значења коју одликује изразита идиоматичност (уп. Марјановић 2017: 142–151). Квантификавањем значењских обележја која парафразирају глобално значење, и која су ручно унета у базу, показујемо тенденције развоја фраземе у денотативном значењу. У табелици 1) дајемо прво парафразу придевскога категоријалног значења, затим његову бројчану заступљеност (прво апсолутну, па постотну) и на крају потврде које илуструју то значење. Потврде у оквиру одговарајућег значења уједно илуструју и устаљеност фраземе, њену структуру, експресивност и конотативна обележја. Фразема је у потврдама истакнута масним слогом, а колокати (те, евентуално, синтаксичка структура у којој се налази) подвучени:

³ Уп. фр. *obéir à quelqu'un comme un enfant* 'послушати кога као дете' (Марјановић 2017: 326, 646).

Таблица 1. Придевско значење

парафраза	ПОТВРДА	БРОЈ
'крајње послушан'	<p>Не дирај ми Ниницу. Биће ко бела лала кад га дова- тим. <ana.rs></p> <p>Да су ты децу лешили на време, били би они ко беле лале. <vreme.com></p> <p>Можда сад и нисте за чиповање, ал' ако вас човјек добро препане, бићете као беле лале, заборавићете и права, и приватност, и све. <forum.cdm.me></p> <p>Док је литијаша, ово ће да се дешава, али ћемо их смирити сад за изборе, биће ко бијеле лале. <m.cdm.me></p>	53 (7%)

На исти начин поступамо у табlici 2) за прилошко категоријално значење:

Таблица 2. Прилошка значења

парафраза	ПОТВРДА	БРОЈ
(1) 'у крајњој мери'	<p>Био си ти миран као бела лала и прихватио све што је наређено, а сад кад су мјере укинуте раскокодакао си се. <m.cdm.me></p> <p>Сви који потурају приче како треба да будемо мирни као беле лале су за мене [...] ботови, или им је добро неко испрао мозак. <juznatribina.net></p>	14 (2%)
(2) 'у крајњој мери'	<p>Ја бих стимулисала (ћубривом, наравно) и оставила напољу. Мени су две процветале, а две још нису (ал' цветаће ко беле лале). <svetbiljaka.com></p> <p>Ти си од уласка Марка [...] процветала као бела лала! Ти си била весела сваки дан, ти си летела по имању! <pink.rs></p>	16 (2%)
(3) 'онако како је нај-боље и очекивано'	<p>Када крене да багује, уместо слушки учитава УСБ, ја му убацим софтвер изнова и ради као бела лала... <mobilnisvet.com></p> <p>Одскора возим ц макс, [...] ишло се на дуже релације и утисци су добри, све ландара, без обзира на нове гуме и центрање трапова, тресе на 120 да би на 140 ишао ко бела лала... <mojaraza.rs></p> <p>Ако хоћеш максимално да уштедиш – неки људи су стављали озлоглашене Оптисан рефракторе за guiding и они су радили као беле лале. <forum.astronomija.org.rs></p>	18 (2%)

<p>(4) 'лако, без икаквих потешкоћа'</p>	<p>Ја једва на очи гледам. Одавно нисам осетила овакав умор, баш ми је недостајало <u>да заспим ко бела лала</u>. <ucibibapl.wordpress.com> Синоћ сам одгледао две епизоде у цугу (3 и 4) и после <u>спавао ко бела лала</u>. <parapsihopatologija.com> Једва чека да поврати стару линију, па из породиштва трчи у теретану... Дижки своје дете, сестро, усисавај, пери, па <u>ћеш да вратиш</u> линију <u>као бела лала!</u> <najboljamatanasvetu.com> Ако је шраф пластичан, а кућиште где се шрафи метално, отопи га са лемилицу и ушрафи други. Довољно да га само пробјеш, после <u>ће изађе ко бела лала</u>. <zafc.forumsr.com> Нема ту неке технике, узми кило кључ и лагано натежи на 10 кг, <u>одвиће се ко беле лале</u>. <zafc.forumsr.com> Када сам један пешкир натопио нитроразређивачем и истраљао све то, <u>сишле су</u> [мрље] <u>као беле лале</u>. Додуше и лак с паркета... <receptomania.com></p>	<p>52 (7%)</p>
<p>(5) 'беспоговорно, неминовно, онако како се мора'</p>	<p>Он је подсетио да је „друга страна” пре десетак дана рекла да „ни за живу главу” не иде на сутрашњи дијалог, а било је јасно да <u>ће ићи „као бела лала”</u>. <radiokontaktplus.org> Нема другог начина осим да га пустиш да огладни, а онда му даш да једе искључиво оно што треба да једе. <u>Јешће све ко бела лала</u>. <b92.net> Кога протера, <u>има да пије</u> антинал <u>као бела лала</u> и да пита има ли још. Јадна мајка ономе ко се буде нећкао, тога плажа до краја одмора неће видети. <putovanja.info> Тај ваш је веома харизматичан. Прво неће ништа да потписује, али после <u>ко бела лала</u> све <u>потпише</u>. <rtcg.me> <u>Има да гласају</u> за војску <u>ко бела лала</u>, тих 9 само служи да подиже руку, нема право да размишља. <b92.net> Нећемо да <u>платимо</u> унапред. – <u>Оћете, оћете, ко бела лала</u>, само полако. <b92.net> <u>Вратиће се</u> у парламент <u>као беле лале</u> и то брзо. Стиже директива из иностранства, па морају сад наћи оправдање. <rtcg.me> Затварао би их и слатко се на њима иживљавао. Спроводио бих европске вредности на сваком кораку. <u>Певали би</u> сви химну Европе <u>као беле лале!</u> <forum.krstarica.com> Хитно и код нас да се уведе овај систем ислеђивања. <u>Певали би ко беле лале!</u> <telegraf.rs> Шта је било, мора да се плаћа? Е море, <u>ће да плаћате ко беле лале</u>, још и те како. Оћете за цабе да се грете, неће да може!? <kurir.rs> <u>Има да шетате ко беле лале</u>, ко вас пита ел оћете. <twitter.com></p>	<p>621 (80%)</p>

Квантификовањем потврда према значењским парафразама долази се до податка да је најзаступљеније и статистички најзначајније прилошко значење под бројем (5) – оно се јавља у чак 80% потврда. Њему се може додати и придевско значење (7%) те прилошко значење под (1) (2%), с којима дели напред поменута сва три денотативна обележја. Разлика је између придевскога и првога прилошког значења у томе што се код придевскога основа поређења изоставља, а код прилошкога не. Даље, квантификовањем података о архисеми лексема на месту субјекта глаголског колоката, запажа се да се на месту субјекта листом користе синтагме с обележјем 'живо' и 'људско'. Овај налаз потврђује досадашња знања о антропоцентричности фразеолошког фонда (уп. Марјановић 2017: 252–254; в. и Вуловић 2015: 137). Штавише, прегледом анотираних потврда закључује се и то да се у наведеним значењима јављају како потврде с екавског ареала, тако и потврде с црногорских сајтова. Удео црногорских потврда знатнији је код придевског значења (36%), док се код двају прилошких значења јавља у око 10% свих потврда, што сразмерно одговара укупном уделу црногорских потврда (уп. 2.2.1).

Удео црногорских потврда занемарив је, пак, код прилошких значења под бројем (3) и (4) — укупно 4 потврде (тј. 6% потврда у оквиру обају значења укупно), па се може претпоставити да су ова два значења у претежној употреби у Србији. Иначе, потврде за значења под бројем (3) разликују се од осталих по томе што је код њих субјекат глаголског колоката искључиво синтагма с обележјем 'неживо'. С друге стране, код потврда за значење (4) јављају се како субјекатске синтагме с обележјем 'неживо', тако и оне с обележјем 'живо', односно 'људско'. Субјекатске синтагме с обележјем 'неживо' налазе се код ова два значења у укупно 23 потврде, што чини свега 3% целог корпуса. Коначно, истичемо да фразема у потврдама ових двају значења, којих је укупно 9% и којима се квалификује извршавање радње по уобичајеном правилу, у значењској структури носи, од претходна три обележја, једино обележје 'очекиваност', с тиме што се у значењу (4) активира и обележје 'лакоћа'.

Код прилошког значења под бројем (2), које се у 2% потврда остварује уз глаголске колокате као *цв(ј)етати*, *проц(ј)ветати*, *белети се/бијељети се* и сл. примећује се да није реч о мотивацијској бази изнетој у одељку 2.2.4. У овим је потврдама *бела лала* заправо фитоним у основном – неидиоматском – значењу ('украсна биљка из рода *Tulipa*'), чији је цвет беле боје па сликовито поређење служи да експресивизује основно или пренесено значење глагола који се у основном значењу односи на цвет. Ове потврде, стога, издавајмо од осталих, и поређење у њима сматрамо хомонимним у односу на фразему у осталим потврдама. Могуће је да ће се ово значење у будућности развити због непрозирности мотивацијске базе (уп. 2.2.4).

2.2.6. *Конотативност*. Прегледом корпусних потврда могуће је утврдити и која се конотативна обележја фраземе *као бела лала* активирају у контексту. То се формално може испитати анализом значењске прозодије контекста. У највећем броју потврда у контексту се активирају негативна конотативна обележја предметне поредбене фраземе (89%), док се у мањем броју јављају

позитивна (11%). Негативна се обележја, пре свега, активирају код корпусних потврда које имају придевско значење које смо у одељку 2.2.5. дали у табlici 1), као и код потврда у којима фразема има прилошка значења која смо у табlici 2) означили бројевима (1) и (5). Те потврде засићене су лексемама које указују на негативну значењску прозодију због тога што међу њима доминирају директивност и комисивност, тј. беспоговорна се послушност намеће противно вољи лица означеног субјекатском синтагмом. Овако изражена негативна конотативност у складу је с општим фразеолошким тенденцијама (уп. Марјановић 2017: 160–162). С друге стране, у потврдама код којих фразема остварује денотативна значења обележена бројем (2), (3) и (4) активирају се позитивна конотативна обележја, јер се радње којима се фразема прилошки квалификује оцењују као прихватљиве и пожељне.

3. Закључна разматрања

3.1. **Налази.** У овоме смо раду на примеру анализе једне фраземе српскога језика показали да електронски корпус састављен од интернетске грађе екстраховане путем мрежног претраживача *Google* представља тренутно најобухватнији, најактуелнији и најдоступнији извор грађе за фразеолошка истраживања српскога језика. Корпус је дао 774 потврде поредбене фраземе *као бела лала* у контексту и омогућио је да се испитају све формалне и значењске одлике фраземе. Корпусна анализа дала је *седам* главних резултата. *Прво*, да се фразема у екавском лику употребљава у Србији и, што није било очекивано, у Црној Гори, а да је употреба саставнице с ијекавским рефлексом јата статистички безначајна. *Друго*, да се фразема у највећем броју случајева прилагођава контексту (*као бела лала : као беле лале*) у зависности од броја у којем је колокат. *Треће*, да се као колокат сасвим ретко јављају придеви, у маломе броју случајева глагол *бити*, а у свим осталим случајевима низ глагола с обележјем динамичности. *Четврто*, да је колокат у преко 50% потврда у футуру и у склопу конструкције *има да* + презент, те да је друго лице врло често. *Пето*, да се фразема користи у неформалној комуникацији ради повећања експресивности исказа. *Шесто*, да се у највећем броју случајева фразема користи уз субјекат с обележјима 'живо' и 'људско', те да су ту доминанта денотативна значењска обележја 'послушност' и 'беспоговорност'; да се фразема у тим потврдама може парафразирати дефиницијом 'беспоговорно, неминовно, онако како се мора'; да се у малом броју случајева фразема користи уз субјект с обележјем 'неживо' и да има другачије значење, којим доминирају једино обележја 'очекиваност' и 'лакоћа', а да се у веома малом броју случајева јавља хомонимна фразема која се заснива на фитонимском мотиву. И коначно, *седмо*, да придевско и најзаступљеније прилошко значење носе негативна конотативна обележја, а да прилошко значење у којем се фразема користи уз субјекат за неживо имају позитивну конотативност.

3.2. **Перспектива.** Корпусне резултате до којих смо дошли требало би сад упоредити с постојећим лексикографским описом у једнојезичним и двојезичним речницима српскога језика и указати на њихова поклапања и размимоилажења, те на тај начин дати оцену како корпусне анализе, тако и лексикографског описа. То, међутим, изискује додатни простор, те ова тема остаје за наредно истраживање.

Лексикографски извори

- МФР:** J. Matešić (ur.), *Frazeološki rječnik hrvatskoga ili srpskog jezika*, Zagreb: Školska knjiga, 1982.
- РМС:** М. Стевановић (ур.), *Речник српскохрватскога књижевног језика*, I–VI, Нови Сад (– Загреб): Матица српска (– Матица хрватска), 1967–1976.
- РСАНУ:** *Речник српскохрватског књижевног и народног језика*, I–XXI, Београд: Институт за српски језик САНУ, 1959–2019.

ЛИТЕРАТУРА

- Видовић Болт 2011:** I. Vidović Bolt, *Životinjski svijet u hrvatskoj i poljskoj frazeologiji I*, Zagreb: Hrvatska sveučilišna naklada.
- Вуловић 2015:** Н. Вуловић, *Српска фразеологија и религија. Лингвокултуролошка истраживања*, Београд: Институт за српски језик САНУ.
- Гато 2014:** М. Gatto, *Web as Corpus. Theory and Practice*, Bloomsbury Publishing.
- Ивановић и др. 2016:** Н. Ивановић, М. Јакић, С. Ристић, Грађа речника САНУ – потребе и могућности дигитализације у светлу савремених приступа, у: С. Ристић и др. (ур.), *Лексикологија и лексикографија у светлу савремених приступа*, Београд: Институт за српски језик САНУ, 132–154.
- Колсон 2003:** J.-P. Colson, Corpus linguistics and phraseological statistics: a few hypotheses and examples, in: Н. Burger et al. (eds.), *Flut von Texten – Vielfalt des Kulturen*. Baltmannsweiler: Schneider Verlag Hohengehren, 47–59.
- Колсон 2007:** J.-P. Colson, The World Wide Web as a corpus for set phrases, in: Н. Burger et al. (eds.), *Phraseologie / Phraseology, Ein internationales Handbuch des zeitgenössischen Forschung*. Berlin–New York: De Gruyter, 1071–1077.
- Марјановић 2014:** С. Марјановић, Лексикографска обрада фраземе (*насти*) као зрела крушка и њени кореспонденти у францускоме језику, *Научни састанак слависта у Вукове дане*, 43/1, 371–383.
- Марјановић 2017:** С. Марјановић, *Поредбене фраземе с компонентом сомте/као у француском и српском језику* (необјављена докторска дисертација), Београд: Филолошки факултет.
- Мун 1998:** R. Moon, *Fixed Expressions and Idioms in English. A Corpus-Based Approach*, Oxford: Clarendon Press.

- Мун 2007:** R. Moon, Corpus linguistic approaches with English corpora. *in:* Burger, Harald (ed.). *Phraseologie / Phraseology, Ein internationales Handbuch des zeitgenössischen Forschung*, Berlin–New York: De Gruyter, 1045–1059.
- Мун 2008:** R. Moon, Conventionalized *as*-similes in English. A problem case, *International Journal of Corpus Linguistics*, 13/1, 3–37.
- Павковић 2014:** В. Павковић, Грађа за Речник САНУ, у: Рајна Драгићевић (ур.), *Савремена српска лексикографија*, Београд: Филолошки факултет, 55–59.
- Паризоска/Филиповић Петровић 2016:** J. Parizoska, I. Filipović Petrović, Uporabni model jezika u leksikografskoj obradi frazeološke varijantnosti u hrvatskome. *u:* S. L. Udier i K. Cergol Kovačević (ur.), *Metodologija i primjena lingvističkih istraživanja*, Zagreb: Srednja Europa, 147–158.
- Пејановић 2015:** А. Пејановић, *Синтакса, семантика и лингвокултурологија компаративних фразеологизама*, Бања Лука: Паневропски универзитет Апеирон.
- Пенадес Мартинез 2012:** I. Penadés Martínez, *Gramática y semántica de las locuciones*, Alcalá de Henares: Universidad de Alcalá.
- Радовић Тешић 2009:** М. Радовић-Тешић, *С речима и речником*, Београд: Учитељски факултет.
- Ристић/Ивановић 2011:** С. Ристић, Н. Ивановић, Предлог за модернизацију рада на Речнику САНУ, у: Срето Танасић (ур.), *Граматика и лексика у словенским језицима*, Нови Сад – Београд: Матица српска – Институт за српски језик САНУ, 529–553.
- Самарцић 2011:** Т. Самарцић, Електронски корпуси као извор нове грађе за лексикографски опис српског језика, *Научни састанак слависта у Вукове дане*, 40/1, 413–427.
- Сикимић 1998:** Б. Сикимић, Невестинска имена: Од хипокористика до пејоратива, *Српски језик*, 3/1–2, 27–55.
- Стаменковић 2010:** Д. Стаменковић, Придевске поредбе с називима животиња у енглеском и српском језику, *Зборник Матице српске за филологију и лингвистику*, ЛП/2, 169–189.
- Стефановић 2010а:** М. Стефанович, „Чужие” среди „своих” и „свои”: О структуре концептов *зять*, *невестка* и *мама* на материале асоциативних словарей сербског и руског језика, *в:* И. А. Стернин (ред.), *Коммуникативные исследования*, Воронеж: Издательство „Истоки”, 147–156.
- Стефановић 2010б:** М. Стефановић, Језички концепт „туђих” међу „својима” у српској и руској породици, *Славистика*, 14, Београд, 102–115.
- СТИЈОВИЋ 2017:** Р. Стијовић, Грађа за Речник САНУ – благо које треба сачувати (о дигитализацији листића), *Наш језик*, 38/3–4, 201–207.
- Требјешанин 1991:** Ж. Требјешанин, *Представа о детету у српској култури*, Београд: Српска књижевна задруга.
- Филиповић Петровић 2018:** I. Filipović Petrović, *Kad se sretnu leksikografija i frazeologija. O statusu frazema u rječniku*, Zagreb: Srednja Europa.

- Финк Арсовски 2002:** Ž. Fink-Arsovski, *Poredbena frazeologija: pogled izvana i iznutra*, Zagreb: Filozofski fakultet.
- Шатара 2008:** С. М. Хатара, Expression idiomatiques en interface bilingue électronique, in: С. Mellado Blanco (ed.), *Colocaciones y fraseología en los diccionarios*, Frankfurt am Main: Peter Lang, 321–332.
- Штрбац 2018:** Г. Штрбац, *Фразеологија у човеку и човек у фразеологији*, Нови Сад: Филозофски факултет.

Saša Marjanović

CORPUS ANALYSIS OF THE SIMILE *KAO BELA LALA*

Summary

The paper presents the results of the corpus analysis of formal and semantic properties of the simile *kao bela lala* in Serbian. The formal properties of this simile (its composition and structure, fixedness and paradigmaticity) as well as its semantic properties (idiomaticity, expressiveness and connotative meaning) were examined on the representative material of the formal and informal written language extracted from a contemporary internet electronic corpus based on the Google search engine results. Conclusions on the sphere of use of this simile are also provided on the basis of the corpus analysis. In addition, the paper sheds light on its motivation basis, as well as the awareness of the average speakers of the Serbian language of the motive behind the simile. The goal of the paper is to show that the properties of the analyzed, but also of other similes and phrasemes in general, can indeed be properly examined with the help of internet-based electronic corpora.

Key words: simile, *kao bela lala*, phraseology, corpus, dictionary, Serbian.