

Александар Костић

https://doi.org/10.18485/mks_dh_skn.2021.1.ch2

Универзитет у Београду

Филозофски факултет

Лабораторија за експерименталну психологију

akostic@f.bg.ac.rs

Зорица Витић

Универзитет у Београду

Филолошки факултет

zvitic@fil.bg.ac.rs

ДИСТРИБУЦИЈА ВЕРОВАТНОЋА ПАДЕЖНИХ ОБЛИКА ИМЕНИЦА У ДЕЛИМА Св. САВЕ И САВРЕМЕНОМ СРПСКОМ ЈЕЗИКУ

Сажетак: Анализе дистрибуције вероватноћа падежних облика именица у делима Св. Саве и савременом језику урађене су на основу дигиталне верзије *Корпуса српског језика* и *Фреквенцијског речника савременог српског језика* (Котић, Ђ. 1999) са намером да се утврди у којој мери се по овом критеријуму језик којим је писао Св. Сава разликује од савременог српског језика. Налази ових анализа указују на изразите дијахроне промене када је реч о дистрибуцији вероватноћа падежних облика именица. Истовремено, указују и на то да, упркос наведеним променама, систем одржава константном удаљеност од максималне ентропије. Другим речима, све промене вероватноћа падежних облика су дозвољене, под условом да ентропија падежног система буде на преферентној удаљености од максималне ентропије.

Кључне речи: Корпус српског језика, списи св. Саве, српски језик, падежни облици, дистрибуција вероватноћа, ентропија

Средином педесетих година прошлог века проф. Ђорђе Костић је осмислио *Корпус српског језика* и руководио његовом израдом до почетка шездесетих година, када је овај пројекат прекинут. Реч је о дијахроном корпусу од 11 милиона речи од којих је свака реч ручно граматички обрађена и који захвата српски језик од XII века до савременог језика. Један од сегмената Корпуса чине српскословенски текстови настали од XII до XVIII века. Овај сегмент садржи око 500.000 речи и подељен је на два подузорка: житија српских владара и архиепископа – најзаступље-

нији и незаобилазни жанр српске средњовековне књижевности који садржи око 250.000 речи и Старе српске повеље и писма у издању Љубомира Стојановића. Део подузорка који се односи на житија српских владара и архиепископа је лематизован и граматички обрађен, док је део који захвата Старе српске повеље и писма граматички обрађен, а његова лематизација је у току.¹

За дела из подузорка житија српских владара и архиепископа направљени су фреквенцијски речници који су у издању Службеног гласника објављени у осам томова са пратећом софтверском апликацијом. У првих шест томова фреквенцијских речника обрађена житија илустрјују књижевни језик од половине XIII до половине XIV века (оба Доментијанова житија, Теодосијево *Житије светог Саве* и обимни зборник *Житија краљева и архиепископа српских* архиепископа Данила II и његових настављача).

Седми и осми том (Костић, Ђ. 2012) обухватају примере српскословенског језика од првих сачуваних споменика (крај XII века) до почетка XVIII столећа. Ова грађа је и жанровски најразноврснија: укључена су житија (*Житије св. Симеона* од светог Саве и Стефана Првовенчаног (Ђоровић 1936: 15–74); Цамблаково *Житије св. Стефана Дечанског* (Шафарик 1859: 35–94); *Житије цара Уроша* од Патријарха Пајсија (Руварац 1867: 209–239); анонимно *Житије Јоакима Сарандапорског*²), служба и канон (*Савина Служба св. Симеону, Заједнички канон св. Симеону и св. Сави четвртога гласа* Теодосија Хиландараца³), најстарија повеља и типик (*Хиландарска повеља* и *Карејски типик* светог Саве), као и одломци из позних *Хроника* грофа Ђорђа Бранковића.⁴

Поред великог временског распона који покрива и разноврсних текстова бројних средњовековних писаца од којих је сачињен, овај део фреквенцијских речника издваја се и по томе што су у њему сабрани списи светог Саве, као јединствен и заокружен узорак језика с почетка XIII века. Према издању Владимира Ђоровића (*Списи св. Саве*, Београд – Ср. Карловци 1928) у корпус су најпре унесена два црквено-правна списка: оснивачка *Хиландарска повеља* Стефана Немање из 1198. године, најстарији сачувани правни акт на српскословенском, као и *Карејски типик* (1198/1199. година),

1 О специфичностима анотације српскословенског у оквиру Електронског корпуса српског језика в. Витић, Костић 2015: 315–323.

2 Светитељ који се подвизавао у близини Криве Паланке (Осоговска планина, место Сарандапор). Време његовог живота није тачно утврђено (највероватније XI–XII), али спада у круг раних, општештованих јужнословенских светитеља. Изд. Новаковић 1867: 242–264.

3 Изд. Радојичић 1955/56: 137–151.

4 *Хронике* грофа Ђорђа Бранковића (1645–1711), до сада издате само у одломцима (Новаковић 1872: 135–162), представљају прекретницу у развоју српске историографије, „знаменити прелаз од средњовековног схватања српске историје модерном“ (Радојичић 1926: 6).

правила за монашки живот у малој испосници коју је свети Сава основао у административном средишту Свете Горе, Кареји (Ђоровић 1928: 1–13).

Укључена су и два Савина права литерарна састава – најстарије српскословенско житије и служба. *Житије светог Симеона* (1208), ктиторско житије оснивача Студенице, замишљено као уводни део *Студеничког типика* (Ђоровић 1928: 151–175), одликује се специфичном композицијом и једно је од најлепших и најскладнијих житија у читавој старој књижевности, док *Служба светом Симеону* (Ђоровић 1928: 176–186), прва у низу служби којима је успостављано штовање светитеља из светородне лозе Немањића, представља саме почетке српске поезије. Рад на *Служби* Сава је започео у Хиландару 1200. године, а допунио ју је и довршио после повратка у Србију, око 1207. године.

Грађи је придодато и писмо које је свети Сава послао студеничком игуману Спиридону са свог другог путовања у Свету земљу (1234/35). Посланица је састављена на српскословенском језику, али у њеном другом делу приметне су и црте народног језика (Ђоровић 1928:187–189).

Иако постоји штампана конкорданција списа светога Саве (Трифунковић, Јовановић, Јухас: 1980), у којој су, додуше, по узору на библијске конкорданције, изостављене заменице, предлози, везници, речце и помоћни глаголи, електронска верзија српскословенских лематизованих и граматички обрађених текстова, поред увида у изворни текст и фреквенцијски речник, пружа знатно веће могућности за претраживање по различитим критеријумима, самим тим и нови и другачији увид у језичко-стилска, или чисто граматичка својства Савиних дела и дела осталих српских хагиографа.

Поред сегмента *Корпуса српског језика* који захвата српскословенске текстове, посебан део чини узорак савременог српског језика који садржи око пет милиона речи и који је подељен на пет подзорака (дневна штампа, поетски текстови, литерарна проза, политички и научни текстови). Крајем деведесетих година прошлог века целокупан материјал *Корпуса* је дигитализован, а 1999. године из узорка дневне штампе и поетских текстова сачињен је седмотомни фреквенцијски речник савременог српског језика који садржи око 60.000 одредница и око 250.000 граматичких облика одредница.⁵ На основу граматички обрађених узо-

5 Последњих година интензивно се ради на лематизацији узорка литерарне прозе (узорак од око 1000.000 речи). Када лематизација буде завршена, овај узорак ће бити прикључен узорцима дневне штампе и поетских текстова и на основу ова три узорка биће направљен нови фреквенцијски речник.

рака дневне штампе и поетских текстова који садрже око 2.000.000 речи, проф. Костић је сачинио детаљну евиденцију вероватноћа граматичких облика у српском језику (Костић, Ђ., 1965а).⁶

Имајући у виду чињеницу да постоји детаљна и поуздана евиденција вероватноћа граматичких облика у делима светог Саве и савременом српском језику, могуће је направити поређење њихових вероватноћа и тиме добити прелимирани увид у промене до којих је дошло у последњих осам векова развоја српског језика. Имајући у виду обим овог рада, ограничићемо се на поређење вероватноћа падежних облика именица. Неизбежно је, међутим, да ово поређење изведемо на узроцима упадљиво различитих величина, што би могло да изазове сумњу у његову поузданост. Док се узорак савременог језика састоји од око два милиона речи, дела светог Саве садрже 9.332 речи, од чега 2.370 именица. Ранија истраживања, међутим, показују да је величина овог узорка (2.370 именица) довољна да се добију поуздане вероватноће падежних облика именица (Kostić, Ilić i Milin, 2008).⁷

Закључци овог поређења морају да се узму с опрезом, уз претпоставку да писани језик светог Саве није битно различит од језика аутора осталих житија која чине узорак српкословенског језика. Неопходан је и додатни опрез приликом извођења закључака јер није извесно у којој мери је језик светог Саве, као ни језик осталих аутора житија, репрезентативан када је реч о српкословенском језику. Поузданију слику о вероватноћама појединих речи и граматичких облика у српкословенском језику добићемо када у оквиру *Корпуса српског језика* буде завршен пројекат израде целовитог фреквенцијског речника српкословенских текстова који би укључио и Старе српске повеље и писма.

Вероватноће падежних облика именица и њиховог граматичког рода и броја у делима светог Саве и савременом српском језику упоређићемо кроз следеће анализе:

а) Поређење вероватноћа граматичког броја, без обзира на падеж и граматички род.

6 Граматичка обрада текстова савременог српског језика изведена је на основу система анотације који је разликовао око 2.500 граматичких облика. Евиденцију вероватноћа граматичких облика у савременом српском језику проф. Костић је објавио 1965. године у студији под насловом *Структура употребне вредности граматичких облика у српскохрватском језику* коју је издао Институт за експерименталну фонетику и патологију говора.

7 Налази истраживања у коме су испитани критеријуми оптималне величине језичког узорка из кога се добијају поуздане вероватноће граматичких облика указују на то да се узорак од око 2000 именица може сматрати поузданим јер дистрибуција вероватноћа њихових падежних облика високо корелира са дистрибуцијом на узорку од 2 милиона речи ($r=0.99$) /Kostić, Ilić i Milin, 2008/.

б) Поређење вероватноћа граматичког рода, без обзира на падеж и граматички број;

в) Поређење падежних облика јединине и множине, без обзира на граматички род;

г) Поређење вероватноћа падежних облика сва три рода, без обзира на граматички број;

д) Поређење вероватноће падежних облика без обзира на граматички род и број.

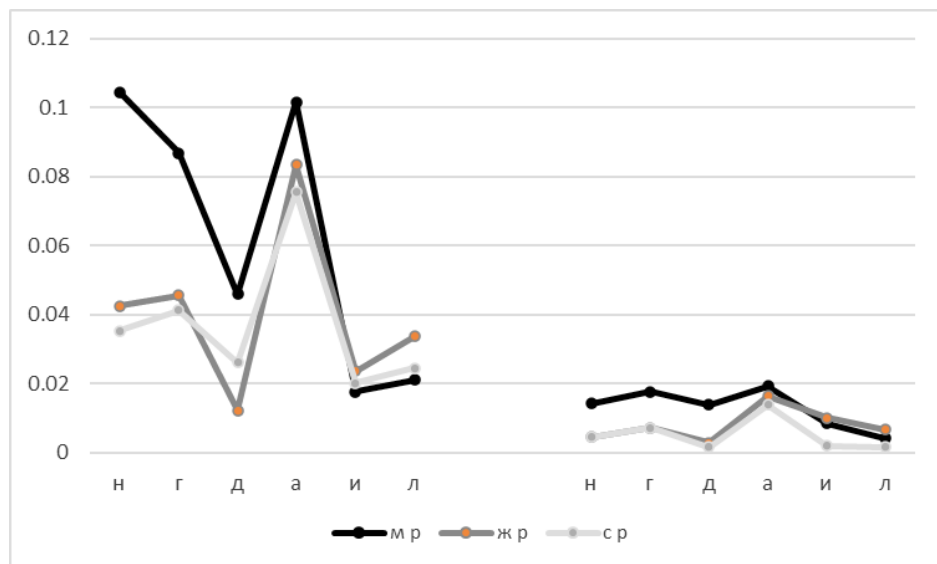
Вероватноће јављања падежних облика именица сва три рода у јединини и множини у делима светог Саве и савременом српском језику дате су у Табели 1.⁸

Табела 1. Пропорције јављања падежних облика именица мушког, женског и средњег рода у јединини и множини у делима светог Саве и у савременом српском језику

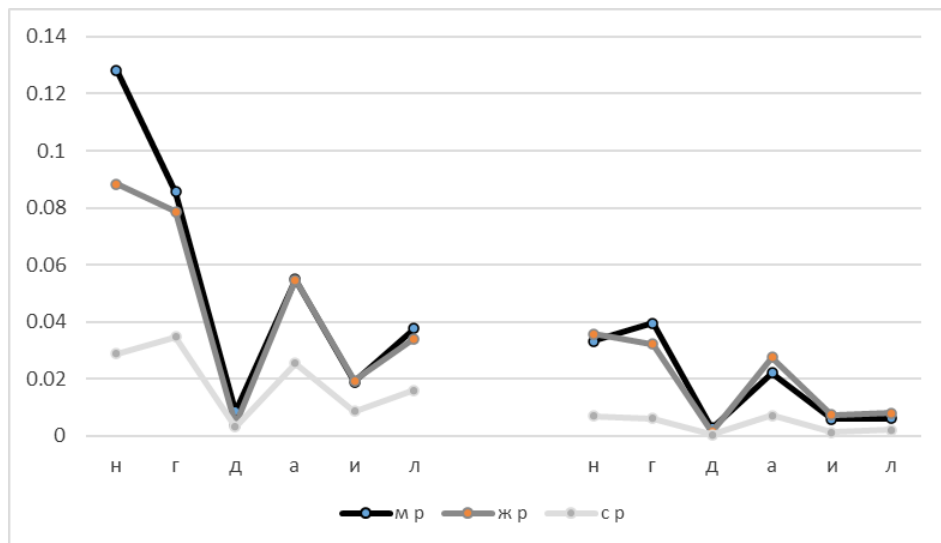
Ј	св. САВА			С. Ј.		
	мр	жр	ср	мр	жр	ср
н	0,104641	0,042616	0,035443	0,1283	0,08841	0,02881
г	0,08692	0,04557	0,04135	0,0856	0,07875	0,03465
д	0,045992	0,012236	0,02616	0,00867	0,00377	0,00312
а	0,101688	0,083544	0,075527	0,05486	0,0548	0,02551
и	0,017722	0,023629	0,020253	0,01895	0,01939	0,00862
л	0,021097	0,033755	0,024473	0,03768	0,03419	0,01606
М						
н	0,014346	0,004641	0,004641	0,03326	0,03577	0,00694
г	0,017722	0,007173	0,007173	0,03961	0,0322	0,00614
д	0,013924	0,002954	0,001688	0,0028	0,00157	0,00039
а	0,019409	0,016456	0,013924	0,02211	0,0275	0,00728
и	0,008439	0,010127	0,00211	0,00597	0,00734	0,00134
л	0,004219	0,006751	0,001688	0,00611	0,00799	0,00212

8 Из Табеле 1 видимо да је изостављен вокатив. У граматичку обраду узорака савременог српског језика, па самим и у евиденцију вероватноћа падежних облика, овај падеж није укључен, иако је у неким случајевима маркиран јединственим флективним суфиксом. Његова фреквенција је прикључена фреквенцији номинатива, а разлог оваквом решењу лежи у чињеници да у случају вокатива постоји конфундирање падежа и синтаксичке функције/значања (1 функција/значање). Овакво решење, међутим, није примењено и на српскословенским текстовима у којима су дате вероватноће вокатива.

На Графиконима 1 и 2 приказан је однос пропорција падежних облика именица сва три рода и оба граматичка броја.



Графикон 1. Пропорције падежних облика именица мушког, женског и средњег рода у једнини и множини у делима светог Саве.

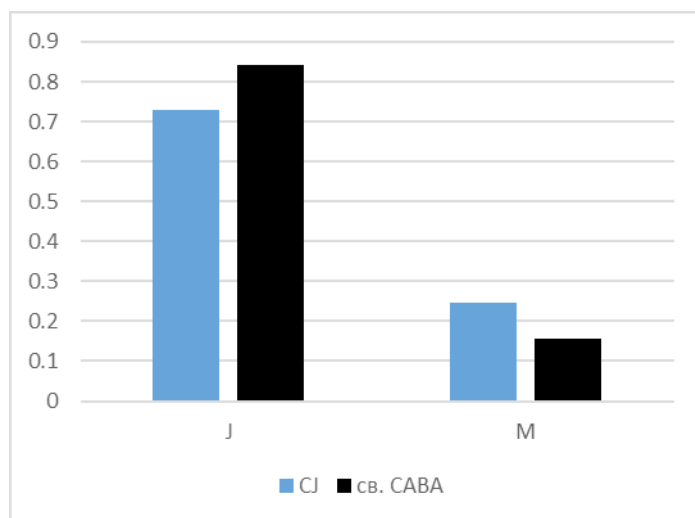


Графикон 2. Вероватноће падежних облика именица мушког, женског и средњег рода у једнини и множини у савременом српском језику

Прегледом Графикана 1 и 2 уочавамо значајне разлике у дистрибуцији вероватноћа падежних облика именица у два анализирана узорка. С обзиром на то да је на графиконима приказан велики број тачака и да су они због тога донекле тешки за интерпретацију, у даљем тексту ћемо приказати анализе редоследом који је дат на претходним странама.⁹

а) Разлике у вероватноћама именица у једнини и множини

Однос вероватноћа именица у једнини и множини у делима светог Саве и савременом српском језику дат је на Графику 3.



Графикон 3. Пропорције именица у једнини и множини у делима светог Саве и савременом српском језику

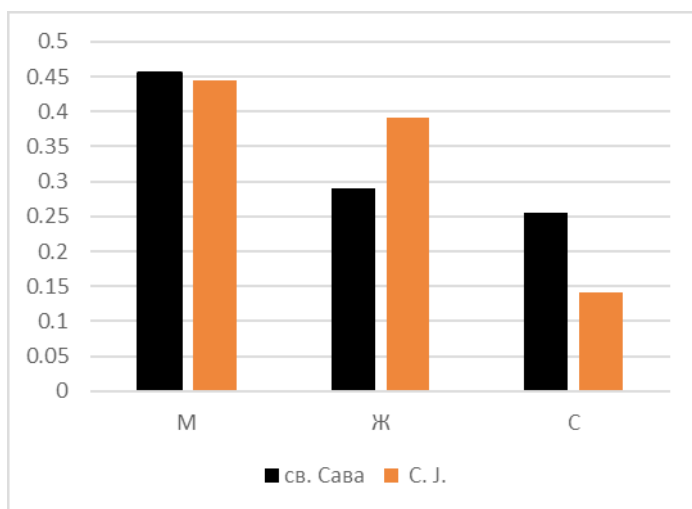
Прегледом Графикана 3 констатујемо да су именице у једнини у делима светог Саве учесталије (0.84) у односу на њихову учесталост у савременом српском језику (0.73). Чини се ваљаном претпоставка да разлоге овој разлици треба тражити или у идиосинкразији стила светог Саве или у специфичностима жанра. Мање је вероватна претпоставка да су, дијахроно гледано, именице у једнини из неког разлога временом постале учесталије.¹⁰

9 Иако би вероватоће падежних облика требало приказати кроз номиналну скалу, из чега следи коришћење хистограма, ради јасноће приказа користићемо линијске графиконе. Изузетак су прикази вероватноћа граматичког броја и граматичког рода код којих немамо велики број тачака, тако да приказивање хистограмом даје задовољавајућу јасноћу.

10 Да ли је реч о идиосинкразији стила могло би се утврдити на основу поређења списка светог Саве и осталих аутора житија. Уколико би се показало да су пропорције гра-

б) Разлике у вероватноћама именица мушког, женског и средњег рода

На Графикону 4 приказане су пропорције именица мушког, женског и средњег рода у два узорка. У делима светог Саве доминирају именице мушког рода, док су вероватноће именица женског и средњег рода изразито мање и приближно једнаке (М – 0,46; Ж – 0,29; С – 0,25). Оваква дистрибуција није уочена у савременом српском језику где су именице мушког и женског рода приближно једнаке вероватноће, док је вероватноћа именица средњег рода упадљиво мања (М – 0,44; Ж – 0,39; С – 0,14).



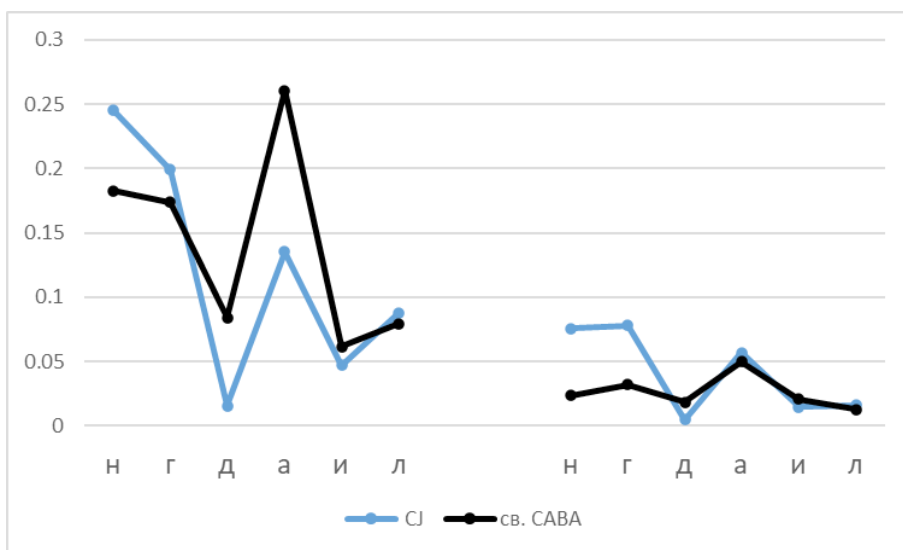
Графикон 4. Пропорције именица мушког, женског и средње рода у делима светог Саве и савременом српском језику

в) Разлике у вероватноћама падежних облика именица без обзира на граматички род

На Графикону 5 приказан је однос вероватноћа падежних облика именица јединине и множине у делима светог Саве и савременом српском језику, без обзира на њихов граматички род. Уочавамо неколико изразитих разлика у дистрибуцији вероватноћа између два узорка. Номинатив

матичког броја приближно једнаке, одбацила би се претпоставка о специфичностима стила светог Саве и прихватила почетна претпоставка о специфичностима жанра. Требало би, међутим, извести додатно поређење текстова житија са текстовима српских средњовековних повеља и писама. Тек уколико би се и у овом узорку добио сличан налаз, могла би се детаљније размотрити могућност систематске дијахроне промене односа вероватноћа јединине и множине именица у два узорка.

и генитив једнине и множине фреквентнији су у савременом српском језику, док су датив и акузатив изразито фреквентни у делима светог Саве, али само у једнини. Посебно изненађује релативно висока учесталост датива једнине у текстовима светог Саве (0,084) ако се има у виду да је овај падежни облик у једнини најнефреквентнији у савременом српском језику (0,015).¹¹ Неочекиван налаз у списима светог Саве добијен је и за акузатив једнине чија је учесталост већа од номинатива једнине, који је у савременом српском језику стабилно најучесталији падежни облик.



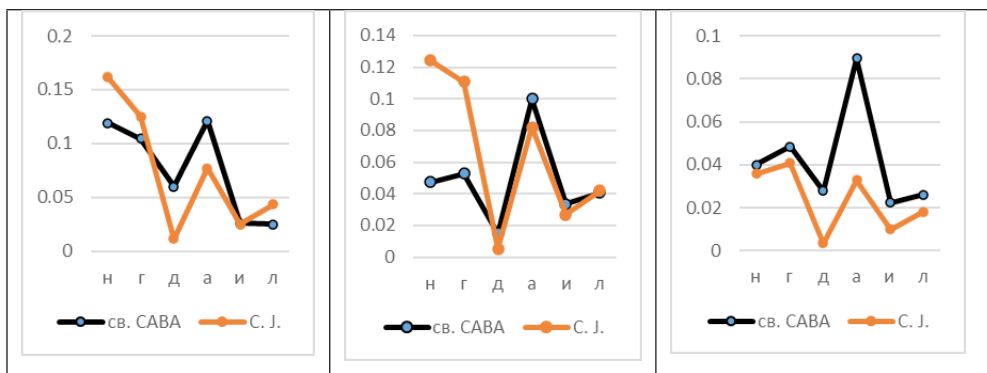
Графикон 5. Пропорције падежних облика именица једнине и множине, без обзира на граматички род

Упркос наведеним разликама, између два узорка добијена је статистички значајна корелација ($r=0.77$, $p<0.01$).

г) Поређење вероватноћа падежних облика именица сва три рода, без обзира на граматички број

Пропорције падежних облика именица мушког, женског и средњег рода у два узорка, без обзира на граматички број, приказане су на Графикону 6.

¹¹ Овако висока пропорција датива у делима светог Саве може се објаснити чињеницом да је у питању посебна старословенска, па и српкословенска конструкција апсолутног датива, који се користи уместо многих зависних реченица.

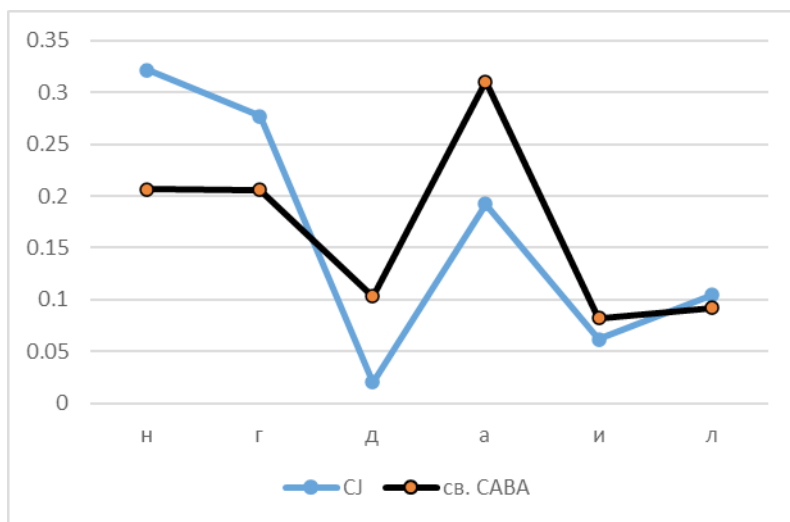


Графикон 6. Пропорције падежних облика именица мушког (леви графикон), женског (средњи графикон) и средњег рода (десни графикон) у делима светог Саве и савременом српском језику, без обзира на грама-тички број

Прегледом Графикона 6 код именица женског и средњег рода кон-статујемо изразите разлике у дистрибуцији вероватноћа падежних облика у два узорка. Именице женског рода у номинативу и генитиву су мање учестале у списима светог Саве, док су именице у дативу и акузативу у истом узорку уочљиво учесталије у средњем роду. Већа учесталост датива и акузатива уочена је и у мушком роду. У оквиру сваког од три грама-тичка рода није достигнута статистички значајна корелација између два узорка.

д) Поређење вероватноћа падежних облика без обзира на грама-тички род и број.

Пропорције падежних облика именица изведене из два узорка, без обзира на грама-тички род и број, приказане су на Графикону 7.



Графикон 7. Пропорције падежних облика именица, без обзира на граматички род и број

Ово поређење може се сматрати резимеом претходних налаза из кога се добија јаснија слика односа вероватноћа падежних облика у два узорка. Номинатив и генитив су уочљиво фреквентнији у савременом српском језику у односу на учесталост ових падежа у списима светог Саве. Са друге стране, датив и акузатив су фреквентнији у списима светог Саве. Ни овог пута није добијена статистички значајна корелација између два узорка.

Поређењем поменутих разлика вероватноћа у два узорка уочавамо одређену правилност - ове разлике такођећи идентичне али у супротном смеру: збирна разлика у пропорцијама номинатива и генитива између два узорка износи 0.18, док збирна разлика пропорција датива и акузатива износи 0.20. Овакав налаз сугерише међусобне компензације, о чему ће бити више речи у завршном делу ове студије.

Из претходних анализа можемо да закључимо да између списка светог Саве и савременог српског језика постоје изразите разлике у дистрибуцији вероватноћа падежних облика именице. Поставља се питање да ли су уочене разлике несистематске или постоје ограничења које их у већој или мањој мери детерминишу. Може се поставити и питање да ли

упркос уоченим разликама ипак постоји нека карактеристика дистрибуције вероватноћа падежних облика која је у распону од осам векова остала непромењена? Одговор на ово питање потражићемо у теорији информација, области која је наоко удаљена од лингвистичких разматрања али је применљива на језичке феномене.

Један од основних појмова теорије информације је *ентропија*. У теорији информације она може да се схвати као мера уређености неког система или као просечна количина информације коју емитује неки систем (Једначина 1).

$$H(X) = -p_{(x)} \log_2 p_{(x)} \quad (1)$$

У Једначини 1 p је пропорција једног елемента система, а вредност ентропије се добија сабирањем производа пропорције елемената и количине информације коју садржи сваки елемент система ($\log_2 p$). Нумеричка вредност ентропије зависи од хомогености дистрибуције вероватноћа елемената система. Уколико су вероватноће свих елемената система једнаке, систем се налази у стању *максималне ентропије*. Свако повећање разлике у вероватноћама елемената смањује ентропију уколико је њихов број константан. Важно је напоменути да вредност ентропије може да варира и у зависности од броја елемената система. Уколико је број елемената већи, у начелу ће и вредност ентропије бити већа, што спречава поуздано поређење ентропије два система са различитим бројем елемената. Начин да се упореди ентропија два система, без обзира на број њихових елемената, је примена мере познате под називом *Шенонова једнакост* (Shannon's equitability) која је изражена као количник добијене (H) и максималне ентропије (H_{max}) /Једначина 2/.

$$E_H = H / H_{max} \quad (2)$$

Шенонова једнакост је мера удаљености система од максималне ентропије и њена вредност се креће од 0 до 1. Системи који се одликују већом хомогеношћу дистрибуције вероватноћа, тј. код којих су разлике у вероватноћама елемената мање, имаће већу вредност E_H , тј. биће ближи максималној ентропији.

Имајући у виду импликације Шенонове једнакости, можемо поставити питање да ли између списка светог Саве и савременог српског језика постоје разлике у удаљености од максималне ентропије, тј. да ли су дијахроне промене у дистрибуцији вероватноћа падежних облика име-

нице у распону од осам векова довеле и до промене удаљености од максималне ентропије. Претходно приказана анализа у којој је за два узорка разматран однос вероватноћа падежних облика без обзира на граматички род и број може да се схвати и као поређење два система са шест елемената за чију дистрибуцију вероватноћа може да се израчуна ентропија. Максимална ентропија за поменуте системе изражена је као логаритам са основом 2 броја елемената (6).

Анализа удаљености од максималне ентропије (E_H) дистрибуције падежних облика *без обзира на граматички род и број* у два узорка показује да у протеклих осам векова није дошло до значајнијих промена. Вредност Шенонове једнакости за списе светог Саве износи 0.85, док је за савремени српски језик њена вредност 0.82. Анализа у којој је испитана удаљеност од максималне ентропије за дистрибуције падежних облика једнине и множине, *без обзира на граматички род* (12 елемената), дала је сличне резултате: за списе светог Саве вредност Шенонове једнакости износи 0.80, док је за савремени српски језик њена вредност 0.83. Најзад, у анализи у којој је испитана удаљеност од максималне ентропије за падеже именица сва три рода и два броја (36 елемената) за списе светог Саве вредност Шенонове једнакости износи 0.87, а за савремени српски језик 0.85. Другим речима, вредност Шенонове једнакости у распону од осам векова варира + - 1,5%.

У разматрању добијених налаза почињемо од разлога због којих је дошло до у разлика у дистрибуцији вероватноћа падежних облика именица у два узорка. Да бисмо добили јаснији увид у ове разлике постављамо питање *пореkla вероватноће* самог падежног облика, без обзира на време у коме је настао узорак текста из кога је његова вероватноћа изведена. Другим речима, постављамо питање зашто су неки падежни облици учесталији, а неки мање учестали. Разлог овим разликама треба тражити у чињеници да је учесталост (фреквенција) падежног облика *сума учесталости његових синтаксичких функција/значења* које се реализују у реченичном контексту. Падежни облици у српском језику у збиру садрже око 190 различитих синтаксичких функција/значења које се драстично разликују по својој учесталости (Костић, Ђ., 1965б).¹²

12 Нагласимо да учесталост падежног облика није директно повезана са бројем синтаксичких функција/значења, тј. не следи да ће падежни облици са већим број функција/значења бити фреквентнији Датив и локатив, на пример, имају приближно једнак број функ-

Посматрано дијахроно, разлике у дистрибуцији вероватноћа падежних облика именице могу да потичу из два извора: а) промене вероватноћа појединих функција/значења у оквиру датог падежног облика, што доводи до повећања или смањења фреквенције тог облика и/или б) премештања реализације одређене функције/значења у други падежни облик чиме се учесталог тог облика повећава, а учесталост облика који је изгубио дату функцију/значење смањује.

Уколико промену учесталости падежних облика посматрамо кроз динамику промена вероватноћа и/или премештање реализације одређених функција/значења из једног падежног облика у други, указују нам се наговештаји принципа који обезбеђују стабилност падежног система, а који нису лингвистички већ су когнитивно мотивисани. Увидом у синтаксичке функције/значења уочавамо да не постоји значењска инваријанта која би била специфична одлика датог падежа.¹³ Такође, у свим падежима доминирају једна или две функције, док велики број функција/значења има изразито ниску учесталост. Уколико падежни облик садржи велики број функција/значења ниске учесталости, и уколико се овај број не компензује високом учесталашћу једне или две функције/значења, падежни облик ће бити нискофреквентан и одликоваће се великом количином информације (бита). Сваки даљи прираштај у броју нискофреквентних функција/значења и/или смањење фреквенције најфреквентнијих функција/значења повећава количину информације датог падежног облика, из чега следи повећање његове когнитивне комплексности, а самим тим и когнитивног оптерећења приликом његове обраде. Достизање критичне тачке когнитивне комплексности, изражене кроз количину информације датог падежног облика, мотивише динамику промене вероватноћа падежа тако што се критични ниво комплексности редукује премештањем одређеног броја функција/значења у други падежни облик чија комплексност још није достигла критичну вредност информационог оптерећења. Из овога следи да је одсуство значењске инваријанте синтаксичких функција/значења у оквиру датог падежног облика мотивисано оптимизацијом његове когнитивне комплексности, а не лингвистичким разлозима.

Објашњење разлика у вероватноћама падежних облика, дато у претходним параграфима, може се прихватити као полазна основа за

ција/значења (датив – 22, локатив – 21) али је локатив пет пута учесталији од датива (датив – 0.02, локатив – 0.1).

13 Тако је, на пример, за акузатив доминантна функција објекта али акузатив се јавља и у значењу времена, места итд.

објашњење дијакроних промена уочених на два анализирана узорка. Можемо претпоставити да је до ових промена у распону од осам векова дошло варирањем вероватноћа појединих функција/значања односно премештањем њихове реализације из једног падежа у други у тренутку када је количина информације коју је носио падеж у коме је она до тада била реализована достигла критичан ниво когнитивне комплексности. Ово је, међутим, само претпоставка коју би требало емпиријски проверити. За савремени српски језик постоји евиденција вероватноћа функција/значања падежних облика (Костић, Ћ., 1965),¹⁴ али не и за српскословенске текстове. Имајући то у виду, објашњење добијених разлика у два временска узорка предложено у овој студији може да добије емпиријску проверу тек када се утврди вероватноћа функција/значања падежних облика и у узорку српскословенских текстова.

Поред налаза који говоре о разликама у дистрибуцији вероватноћа падежних облика именица између два узорка, у овој студији испитана је и удаљеност њихове ентропије од максималне ентропије. Добијени налази указују на то да се у распону од осам векова по овом критеријуму нису догодиле значајне промене. Штавише, можемо да констатујемо да се удаљеност од максималне ентропије креће у оквиру минималне маргине варирања.

Импликације овог налаза наговештавају могући индекс стабилног стања система (у овом случају падежног система именица). Као што је констатовано у претходном параграфу, у распону од осам векова удаљеност ентропије падежних облика именице од максималне ентропије остала је непромењена, упркос понекад драматичним променама вероватноћа појединих падежа. Ова чињеница указује на то да је при оптимизацији когнитивног оптерећења појединачних падежа (тј. количине информације коју садржи падежни облик) ирелевантан начин на који ће когнитивни систем ту оптимизацију да реализује. Отуда и одсуство значењске инваријанте када је реч о синтаксичким функцијама/значањима које садржи дати падеж јер се, посматрано из перспективе развоја језика, оптимизација когнитивног оптерећења одвијала по информационам, а не значењским критеријумима.

У циљу одржања стабилног стања падежног система когнитивни систем се суочава са два ограничења:

а) Количина информације коју носи појединачни падежни облик унутар падежног система не сме да пређе критичну границу когнитивне комплексности.

14 Изутетак је датив за који немамо евиденцију вероватноћа функција/значања.

б) Дистрибуција вероватноћа падежних облика (тј. њена ентропија) мора да буде у оквирима оптималне удаљености од максималне ентропије.

Дијахроно посматрано, било која синтаксичка функција/значење, у тренутку достизања критичног нивоа когнитивне комплексности падежног облика у коме је до тада реализована, може да буде премештена у било који други падежни облик, при чему ова промена није несистематска. Она мора да буде реализована у оквиру решења које ће одржати дистрибуцију вероватноћа облика (тј. ентропију) у оквирима преферентне удаљености од максималне ентропије. Све док је ентропија падежног система изведена из дистрибуције вероватноћа падежних облика у овим оквирима, падежни систем је у стабилном стању.

Извори и литература

- Витић, Зорица и Александар Костић. „Специфични захтеви анотације српскословенског у оквиру Електронског корпуса српског језика.“ *Дигиталне библиотеке и дигитални архиви*. А. Вранеш, Љ. Марковић (ур.). Београд: Филолошки факултет, 2015. 315–323. Штампано.
- Kostić, Aleksandar, Svetlana Ilić, Petar Milin. „Апроксимација вероватноћа и оптимална величина јежићког узорка“. *Психологија*, 41, 1, 36-51.
- Костић, Александар и Зорица Витић. „Софтверска апликација за претрагу и анализу српских средњовековних текстова. *Дигиталне библиотеке и дигитални архиви*. А. Вранеш, Љ. Марковић (ур.). Београд: Филолошки факултет, 2015. 303–313. Штампано.
- Костић, Ђорђе. *Структура употребне вредности граматичких облика у српском језику*, Институт за експерименталну фонетику и патологију говора, Београд, 1965а.
- Костић, Ђорђе. *Синтаксичке функције и значења падежа у српскохрватском језику*, Институт за експерименталну фонетику и патологију говора, Београд, 1965б.
- Костић, Ђорђе. *Квантитативни опис структуре српског језика*, књ. 1-3, *Српски језик од XII до XVIII века, Доментијан. Житије св. Симеона, Житије св. Саве*, I–III, Институт за експерименталну фонетику и патологију говора, Лабораторија за експерименталну психологију, Филозофски факултет Универзитета у Београду, ЈП Службени гласник, 2009. Штампано.
- Костић, Ђорђе. *Квантитативни опис структуре српског језика*, књ. 4, *Српски језик од XII до XVIII века, Теодосије Хиландарац, Житије св. Саве*, Институт за експерименталну фонетику и патологију говора, Лабораторија за експерименталну психологију, Филозофски факултет Универзитета у Београду, ЈП Службени гласник, 2010. Штампано.
- Костић, Ђорђе. *Квантитативни опис структуре српског језика*, књ. 5-6, *Српски језик од XII до XVIII века, Архиепископ Данило II, Животи краљева и архиепи-*

- скопа српских*, I-II, Институт за експерименталну фонетику и патологију говора, Лабораторија за експерименталну психологију, Филозофски факултет Универзитета у Београду, ЈП Службени гласник, 2010. Штампано.
- Костић, Ђорђе. *Квантитативни опис структуре српског језика*, књ. 7–8, *Српски језик од XII до XVIII века, Житија, канон, хронике*, I-II, Београд: Институт за експерименталну фонетику и патологију говора, Лабораторија за експерименталну психологију, Филозофски факултет Универзитета у Београду, ЈП Службени гласник, 2012. Штампано.
- Костић, Ђорђе. *Фреквенцијски речник савременог српског језика*, I-VII, Институт за експерименталну фонетику и патологију говора, Лабораторија за експерименталну психологију, Филозофски факултет Универзитета у Београду, 1999.
- Новаковић, Стојан. „Прилози к историји српске књижевности“. *Гласник Српског ученог друштва*. 5 (1867) 242–264. Штампано.
- Новаковић, Стојан. „Нешто из Хроника деспота Ђорђа Бранковића“. *Глас Српског ученог друштва*. 33 (1872) 135–162. Штампано.
- Радојичић, Ђорђе Сп. „Теодосијев канон општи Симеону Немањи и Сави“. *Јужнословенски филолог*. 21, књ. 1–4 (1955–56) 137–151. Штампано.
- Радојичић, Никола. „О Хроникама грофа Ђорђа Бранковића“. *Прилози за књижевност, језик, историју и фолклор*. 6/1 (1926) 1–45. Штампано.
- Руvaraц, Иларион. „Житије цара Уроша од патријарха Пајсија“. *Гласник Српског ученог друштва*. 5 (1867), 209–239. Штампано.
- Трифуновић, Ђорђе, Томислав Јовановић и Љиљана Јухас. *Азбучни показатељ речи у списима светога Саве*. Београд: Народна библиотека Србије, 1980. Штампано.
- Ђоровић, Владимир. *Списи св. Саве*. Београд – Ср. Карловци: СКА, 1928. Штампано.
- Ђоровић, Владимир. „Житије Симеона Немање од Стефана Првовенчаног“. *Светосавски зборник*. 2 (1936) 15–74. Штампано.
- Шафарик, Јанко. „Живот краља Стефана Дечанског“. *Гласник Друштва српске словесности*. 11 (1859) 35–94. Штампано.

Aleksandar Kostić

Laboratory for Experimental Psychology
Faculty of Philosophy, University of Belgrade
zvitic@fil.bg.ac.rs

Zorica Vitić

Faculty of Philology, University of Belgrade
akostic@f.bg.ac.rs

**PROBABILITY DISTRIBUTION OF NOUN CASES
IN WRITINGS OF ST. SAVA AND THE CONTEMPORARY SERBIAN
LANGUAGE**

Summary: The aim of this study was to get an insight into differences in probabilities of noun cases in writings of St. Sava and contemporary Serbian language where those two samples span eight centuries of Serbian written language. The analyses were performed on digital version of *Corpus of Serbian Language* and *Frequency Dictionary of Contemporary Serbian Language* (Kostić, Đ., 1999). The outcome of these analyses indicate conspicuous diachronic differences in probability distributions of noun cases. In addition, it was demonstrated that in spite of the observed differences the distance from the maximum entropy of these probability distributions remains constant. In conclusion, all diachronic changes in probability distributions of noun cases are permissible as long as they remain at preferred distance from the maximum entropy.

Keywords: writings of St. Sava, Serbian language, inflected cases, probability distribution, entropy.