



Munich Personal RePEc Archive

## **Markov chains and the problem of keyboard layout for Serbian cyrilic**

Bukvić, Rajko and Vemić, Mirčeta

Geographical Institute “Jovan Cvijić” SASA, Belgrade

2016

Online at <https://mpra.ub.uni-muenchen.de/74862/>  
MPRA Paper No. 74862, posted 02 Nov 2016 11:36 UTC



## МАРКОВЉЕВИ ЛАНЦИ И ПРОБЛЕМ ОДРЕЂИВАЊА РАСПОРЕДА СЛОВА НА ТАСТАТУРИ СРПСКЕ ЋИРИЛИЦЕ

## MARKOV CHAINS AND THE PROBLEM OF KEYBOARD LAYOUT FOR SERBIAN CYRILIC

РАЈКО М. БУКВИЋ

Географски институт „Јован Цвијић” САНУ, Београд, r.bukvic@mail.ru

МИРЧЕТА ВЕМИЋ

Београд, mvemic @eunet.rs

**Резиме:** У раду се разматрају проблеми одређивања оптималног распореда слова на тастатури, који мора бити заснован на карактеристикама и специфичностима сваког конкретног језика. Данашњи стандард за српски језик заснован је на тзв. универзалном распореду и као такав неподесан је. Емпиријском анализом на примеру двају класичних романа српске књижевности, уз примену апаратуре Марковљевих ланаца, анализирани су перформансе постојеће, „српске” тастатуре, и указано је на потребу рада на новом стандарду.

**Кључне речи:** Тастатура, распоред слова, Марковљеви процеси, Марковљеви ланци.

**Abstract:** The paper considers issues related to an optimal keyboard layout which needs to be based on the characteristics and specificities of each particular language. Today's standard for Serbian language is based on the so-called universal layout and, as such, is inadequate. The performances of the today-existing "Serbian" keyboard were analyzed through an empirical analysis of two Serbian classical novels as well as the use of Markov chains apparatus, and a need for further development of a new standard was emphasized.

**Keywords:** Keyboard, letter layout, Markov processes, Markov chains.

### 1. УВОД

Распоред слова на тастатурама раније писаћих машина а данас компјутера проблем је који се у нашим актуелним условима разматра махом на алтернативним форумима или на сајтовима појединих удружења с циљевима очувања ћириличног писма. При томе, он започиње и окончава се на захтевима да се на тастатурама уместо латиничких слова нађу слова српске ћирилице. Сам распоред слова не разматра се, и тако не доводи у питање. Резултат који би се тиме добио (и који је у неким акцијама тих удружења добијен) популарно се назива – „српска тастатура”.

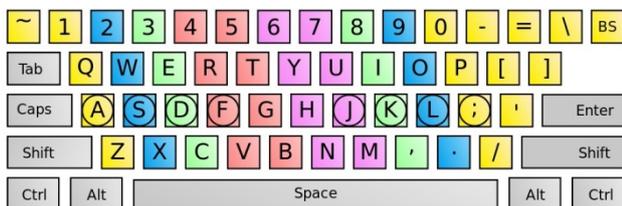
Такав резултат, и уопште такво резонување, нису адекватни. Како се подвлачи у [1], распоред који би се тиме озваничио као „српска тастатура” познати је универзални (QWERTY) распоред (слика 1), тачније његова немачка варијанта (QWERTZ), прилагођена Гајевој („хрватској”) латиници (са словима š, č, đ, ć), од које би се „српска тастатура” разликовала словима љ, њ и ц, смештеним на тастерима q, w и x. Осим визуелно, дакле, та се тастатура суштински не би разликовала од данас коришћених тастатура у енглеској или „хрватској” варијанти.

Пошто је QWERTY распоред (аналогно и QWERTZ) неадекватан не само за српски језик, односно српску ћирилицу, већ и за енглески, за који је и створен, проблем српске тастатуре мора се сагледати у другом светлу, обухватајући већи број (група) питања. Првој од тих група – како је и зашто и код нас прихваћен универзални распоред, при томе касније него у англосаксонским и другим земљама, да ли је и зашто (ни)је разматрана примена ЂЦУКЕН распореда, сачињеног за потребе руског, дакле и осталих словенских језика – ма колико она била важна, ми се овде нећемо бавити. Пажњу ћемо усмерити на други комплекс – на стварање онога што је апострофирано као „српска тастатура”.

### 2. ДЕФИНИСАЊЕ ПРОБЛЕМА

Одређивање (адекватног) распореда слова има дугу, али и не баш познату историју. Први кораци, стварање универзалног распореда, довољно су осветљени, барем кључни аспект – зашто је изабран тај, по Двораку, „најгори могући” распоред. Много интригантнија су питања како је и зашто тај распоред однео превагу у „слободној тржишној утакмици” (омиљен термин економиста) над алтернативним и супериорним распоредима, утолико пре што су

техничким усавршавањима врло брзо елиминисани главни проблеми (сударање суседних полуца са словима при брзом куцању), који су Шоулза, творца универзалног распореда, практично приморали да створи најгори могући стандард. И зашто каснији покушаји да се тај распоред замени подеснијим, међу којима су најпознатија вишедеценијска настојања Дворака, нису били успешни.



Слика 1: Универзална (QWERTY) тастатура

Управо Двораково искуство, иако с једне стране неуспело, а с друге стране настало и намењено енглеском језику и алфabetу, драгоцено је и као сведочанство, али и зато што је Дворак поставио основне принципе које треба да испуњава добар распоред на тастатури [2]:

- при куцању руке треба да се смењују што чешће,
- за максималну брзину и ефикасност куцања најлакши за укуцавање треба да буду најчешће сусретани симболи (то значи да тастери с тим словима треба да буду у основном реду, где су постављени прсти, и у зони кажипрста и средњих прстију),
- аналогно, ретка слова треба да буду у нижем реду, чије тастере је најтеже притиснути,
- десна рука треба да буде више оптерећена него лева (с обзиром да су људи већином дешњаци), и
- укуцавање диграфа суседним прстима је теже него прстима који су далеко један од другог.



Слика 2: „Српска” тастатура

Из прегледа ових принципа види се да је одређивање оптималног распореда слова сложен проблем, који укључује већи број аспеката и задатака. Међу њима се издвајају како они који се односе на сам текст који треба укуцати – учестаност појављивања појединих слова појединачно и у паровима (тј. прелази из једног у друго слово), тако и они којима се решавају физиолошки и ергономски проблеми – смењивање руку при куцању, оптерећеност појединих прстију, као и десне и леве руке у целини, и др. Од уважавања ових принципа зависи распоређивање слова не само по редовима тастатуре (основни, горњи и доњи), већ и по вертикалним сегментима (различите боје на слици 1), који неједнако оптерећују поједине прсте.

### 3. МЕТОДОЛОШКЕ НАПОМЕНЕ

Учестаност појављивања појединих слова (гласова) прва је карактеристика текста коју треба разматрати. Она се може посматрати као учестаност у оквиру речи, и као општа учестаност. На овом нивоу наших истраживања посматраћемо овај други случај. Пошто су, наравно, текстови неједнаке величине, потребно је учестаност изражавати релативно, тј. у процентима

$$F(Ch_i) = \frac{N_i}{N} \quad (1)$$

Други аспект, с много већим и диференциранијим утицајем односи се на прелазак из једног у друго слово (тј. из једног стања система у друго). Ови прелази изражавају се такође релативно, дакле као емпиријске вероватноће

$$p_{ij} = P(Ch_i \rightarrow Ch_j) \quad (2)$$

које у коначном резултату дају транзитивну матрицу  $P$ . Прелази из једног у друго стање система представљају у овом случају типичну реализацију Марковљевог процеса (видети, на пример, [3]), у коме тај прелаз зависи од стања у коме се систем налази, а не и од претходних стања. Другим речима, појава новог (наредног) слова у тексту зависи само од стања (слова) система у датом моменту, и изражава се одговарајућом транзитивном вероватноћом (2).

Добар (идеалан) распоред мора да уважава вредности из транзитивне матрице, у том смислу што ће обезбедити да се највеће вредности из те матрице „прелију” између места (типки) која се могу најлакше реализовати, узимајући у обзир Дворакове принципе, пре свега смењивање руку при куцању, а посебно прелазе из једног у други ред тастатуре. Комплекснија истраживања проблема показале, управо имајући у виду те принципе, да у разматрање треба узети и квадрат транзитивне матрице  $P^2_{(ij)}$ , дакле вероватноће прелаза у два корака. Без тога, чак и уз максимално уважавање принципа груписања најфреквентнијих слова у центру тастатуре и смењивања руку при типкању, може доћи до оптерећивања руку ако се често прелази из једног у други хоризонтални ред.

### 4. РЕЗУЛТАТИ И ДИСКУСИЈА

Емпиријска анализа на роману Сеобе [4], дата у [1], показала је да су далеко најзаступљенија четири вокала: а, е, и, о, што је већ било утврђено у ранијим истраживањима (према [5]). Такав налаз потврђује и анализа урађена за потребе овог рада, на роману На Дрини ћуприја [6] (табела 1). Наравно, свесни смо да су ови примери (узорци) сувише мали за конкретније закључке и да су потребна обимнија емпиријска истраживања, будући да је основни скуп практично неограничен – чини га све што је икада написано на српском језику.

Одређени ниво разлика, не само у разматраним узорцима, условљен је постојањем двају наречја, услед чега ће у једном броју текстова фреквентност

слова (гласова) „и” и „ј” бити већа. Ипак, корелације између добијених резултата, као што се могло претпоставити, врло су високе (табела 2). Практично, то би требало да значи да је основна подела слова по вертикалним сегментима у идеалном распореду у основи опредељена – централна места на тастатури (тастери „F”, „G”, „H”, „J”) требало би да припадну словима „A”, „E”, „I”, „O”. У QWERTY-ју (слика 1) њихова места су далеко од тога, и већ је то довољан разлог да констатујемо да је универзалан распоред (веома) неповољан за српски језик и српску ћирилицу.

**Табела 1:** Фреквенција и ранг слова

	Сеобе		На Дрини ћуприја		Ковачевић	
	F	R	F	R	F	R
А	12,08	1	12,21	1	11,61	1
О	9,78	2	9,91	3	9,59	3
Е	9,53	3	9,24	4	8,77	4
И	9,43	4	9,99	2	9,96	2
Н	4,95	5	5,51	5	5,88	5
У	4,84	6	4,44	7	4,21	9
С	4,69	7	5,01	6	5,19	6
Р	4,29	8	4,35	8	4,97	7
Д	3,93	9	3,97	9	3,95	11
К	3,90	10	3,68	12	3,35	14
Т	3,75	11	3,94	10	4,67	8
В	3,75	12	3,77	11	3,74	12
М	3,49	13	3,40	13	3,64	13
Л	3,14	14	3,07	15	3,05	15
Ј	3,06	15	3,27	14	4,17	10
П	2,92	16	2,64	16	2,90	16
Б	1,79	17	1,53	19	1,48	19
Г	1,72	18	1,91	17	1,69	18
З	1,67	19	1,69	18	1,71	17
Ш	1,55	20	1,24	20	1,15	20
Ч	1,15	21	0,95	21	0,67	23
Ћ	0,91	22	0,68	24	0,67	23
Њ	0,78	23	0,72	23	0,34	27
Х	0,74	24	0,72	22	0,68	22
Ж	0,63	25	0,56	26	0,66	25
Џ	0,58	26	0,62	25	0,72	21
Љ	0,56	27	0,54	27	0,36	26
Ђ	0,30	28	0,26	28	0,12	29
Ф	0,09	29	0,10	29	0,21	28
Џ	0,01	30	0,07	30	0,02	30

Други аспекти разматрања подесности универзалне тастатуре за српску ћирилицу дати су у [1], полазећи од емпиријских анализа на роману Сеобе, представљених транзитивном матрицом и на основу ње сачињеним илустрацијама за „српску” и тастатуру створену за потребе руског (и других словенских језика, односно језика који користе ћирилично писмо). Посебно је потцртана чињеница постојања великог броја нултих вероватноћа (неостварени прелаз) између суседних слова на „српској тастатури”. У неком идеалном распореду то би, наравно, могло бити (бар донекле) поправљено смештањем одговарајућих слова на разне делове тастатуре (принцип смењивања руку при куцању). Колико је то на постојећој тастатури случај остаје за емпиријску проверу у даљем разматрању проблема.

**Табела 2:** Коefицијенти корелације фреквенција и рангова слова из табеле 1

	Фреквенције		Рангови	
	2	3	2	3
1	0,997703	0,989227	0,992881	0,970319
2		0,994738		0,982822

Оно што је, међутим, већ и на основу анализе дате у наведеном раду показано, а потврђује се и на роману На Дрини ћуприја, јесу велике транзитивне вероватноће на различитим редовима тастатуре, а пре свега код слова из основног реда: с, д, ф, г, х, ј, к, л (с изузетком слова а). У целом том низу вероватноћа, заправо, налазе се такорећи саме нуле, с изузетком прелаз „са” (вероватноћа 0,149). Према Андрићевом роману, одговарајуће прелазне вероватноће основног реда дате су у табели 3. Слика је врло илустративна.

**Табела 3:** Суседне вероватноће прелаз у основном реду „српске тастатуре”

7,2	21,2	0,2	0,0	159,2	9,8	111,6	71,0	
а	54,1	139,2	с	0,0	13,3	д	0,0	
	1,2		0,0	1,0	0,0	0,0	26,0	
337,4	28,6	24,5	118,7	0,0	0,7	60,2	0,6	45,9
0,0	ф	0,0	0,0	г	0,0	0,0	х	0,0
2,0	0,0	14,3	1,9	0,0	20,3	0,0	22,1	17,1
0,2	110,6	112,5	72,0	77,0	372,1	316,0	158,4	0,1
0,1	ј	9,0	0,1	к	19,2	1,9	л	0,2
18,6	4,8		0,7					
58,3	0,0	0,0	0,0	0,0	0,0	0,0	3,1	
5,2	ч	0,0	0,0	ћ	0,0	0,0	ж	

Још један моменат, који у раду [1] нисмо разматрали, заслужује пажњу. Реч је о оптерећености руку и прстију, као важан критеријум који опредељује (не)адекватност распореда слова. Табела 4 даје одговарајућу анализу на текстовима двају класичних романа српске књижевности, коју су начинили аутори овог рада, и анализу коју је навео Ковачевић у [5]. Прихваћено је његово методолошко решење (накнадно додавање тачке и запете), због чега су збирови већи од 1.

**Табела 4:** Оптерећеност руку и прстију

Рука	Прст	Сеобе	На Дрини ћуприја	Ковачевић
Лева	Мали	12,64	12,75	11,97
	Домали	5,48	5,80	5,55
	Средњи	14,04	13,82	13,44
	Кажипрст	15,39	15,60	16,76
Свега		47,55	47,98	47,72
Десна	Кажипрст	18,74	19,03	20,29
	Средњи	16,29	15,11	14,90
	Домали	13,74	13,85	13,54
	Мали	7,45	6,35	6,17
Свега		56,22	54,34	54,90

Као што се види, разлике посматраних романа и онога што даје аутор [5] не би се могле сматрати баш занемарљивим, при чему је нарочито значајно мање оптерећење кажипрста на обема рукама у односу на

Ковачевића. Али, важније је нешто друго. Ковачевић, наине, сматра да ови резултати (последња колона у табели 4) потврђују адекватност универзалне (тј. „српске“ тастатуре), имајући у виду већу оптерећеност десне од леве руке, као и опадајућу оптерећеност прстију обе руке од кажипрста до (до)малог прста. Његово резоновање и закључивање, међутим, прилично је чудно – домали прст десне руке двоструко је оптерећенији од малог, и готово исто оптерећен као и средњи прст (у анализираним романима та разлика је нешто мало већа). И док за први од наведених факата истиче да је мали прст ипак више оптерећен, имајући у виду да покрива више слова и да тиме прелази већи пут (што је начелно тачно, али је питање да ли је самерљиво), други факт аутор просто прећуткује. Као и чињеницу да је на левој руци мали прст готово подједнако оптерећен као и средњи прст. Пошто њему припада слово „а“, јасни су и узрок и решење наведеног проблема – као најфреквентније то слово мора се преместити у центар тастатуре.

## 5. ЗАКЉУЧАК

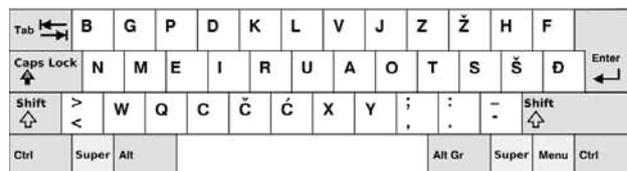
Бивша Југославија прихватила је у фебруару 1976, као чланица ИСО (ISO, International Organization for Standardization), стандард тастатура за све електричне и механичке машине за словеначко, српскохрватско, македонско и албанско језичко подручје за ћирилицу и латиницу. Тај стандард заснован је на универзалном распореду, заправо на његовој QWERTZ (немачкој) варијанти, у којој су замењена места словима Z и Y, и извршене одређене корекције (углавном на типкама специфичних карактера) које се односе на слова Ч, Ћ, Ш, Ђ и Њ, Љ и Џ.



Слика 3: Последњи модел АЕГ писаће машине из 1910. године

Историја прихватања универзалног стандарда, тј. QWERTZ варијанте у Србији (и Југославији) није до сада проучена. Ковачевић у [5] наводи да су прве (писаће) машине у Југославију дошле из Немачке (с већ извршеном корекцијом Z и Y) и да је Југославија, уз још неке земље, ту измену прихватила. Ова констатација, међутим, лако се демантује – на слици 3 приказана је тастатура последњег модела АЕГ писаће машине из 1910. године, на којој се ћирилични распоред слова значајно разликује од прихваћеног QWERTZ стандарда.

Питање како је и зашто прихваћен постојећи стандард, очигледно остаје отворено. С друге стране, отворено је и друго питање: зашто се и даље држимо тог стандарда? Утолико пре што је у међувремену нестало тзв. српскохрватско језичко заједништво, које је, ма колико лабаво, могло бити некакав алиби за чување заједничког стандарда и у овој области.



Слика 4: Преправљена тастатура Ерцеговића

У [5] се констатује да су данас (тј. пре тридесетак и нешто година) идеје о националним тастатурама напуштене и да се поставља захтев да се тастатуре потпуно изједначе (!), као и да се та акција спроводи у оквиру ИСО. Понашање наших надлежних органа и тела током свих ових година било је, доиста, свесно или несвесно на линији таквих идеја и захтева. Ипак, то је врло чудно, не само када је реч о идеји националне тастатуре, већ и када је реч о једноставној ефикасности. Универзална тастатура, наине, показала се као неподесна и за енглеско говорно подручје, за које је створена, а утолико више се као таква мора показати када је реч о другим језицима и писмима.

Најзад, кад је реч о националним особеностима (у било ком домену), неприхватљивост насиља које провејава из истакнутог става и захтева само се појачава. Њихово одбацивање реактивираће данас већ заборављене покушаје да се изгради национална тастатура (видети на пример Ерцеговићев, слика 4), односно морало би да доведе до нових идеја и нових предлога. Све њих, наравно, потребно је темељно анализирати и тестирати, између осталог и уз примену апаратуре Марковљевих ланаца, као што је то демонстрирано у раду [1] и у овом раду.

## ЛИТЕРАТУРА

- [1] Буквић, Рајко М. QWERTY, ЙЦУКЕН и српска ћирилица: треба ли нам нови стандард?, *Зборник Матице српске за друштвене науке*, 67(1) 2016 1–17.
- [2] Power Typing – online free typing tutor, <http://www.powertyping.com>
- [3] Таха, Хэмди. *Введение в исследование операций*, Издательский дом Вильямс, Москва, 2001, 912 с.
- [4] Црњански. Милош. *Сеобе 1*, [www.tamoiovde.files.wordpress.com](http://www.tamoiovde.files.wordpress.com)
- [5] Ковачевић, Миљко. *Савремена типографија 1.*, Београд: Образовни графички центар „Милан Ракић“, 1981, 229 стр.
- [6] Андрић, Иво. *На Дрини ћуприја*, [www.bookzz.org](http://www.bookzz.org)