

УЧЕСТАЛОСТ МАЛИХ И ВЕЛИКИХ СЛОВА, БИГРАМА И ТРИГРАМА У СРПСКОМ ЈЕЗИКУ THE FREQUENCY OF LOWERCASE AND UPPERCASE LETTERS, BIGRAMS, AND TRIGRAMS IN THE SERBIAN LANGUAGE

Вукашин Бабић

РЕЗИМЕ: Ова студија представља свеобухватну анализу учесталости слова, биграма и триграма у српском језику ћириличним писмом. Користећи корпус од приближно 4 милиона карактера из различитих књижевних дела, новина и онлајн енциклопедије, израчунали смо учесталост великих и малих слова, као и биграма и триграма. Наши налази откривају различите обрасце у српском језику, укључујући распрострањеност одређених слова и словних комбинација. Ови резултати су у великој мери усклађени са претходним студијама о српском и хрватском језику, уз извесне варијације због дијалекатских разлика. Ово истраживање даје значајне податке за примену у криптографији, обради природног језика и лингвистичким студијама специфичним за српски језик.

КЉУЧНЕ РЕЧИ: Српски језик, учесталост слова, фреквенција биграма, фреквенција триграма, ћирилично писмо, анализа текста.

ABSTRACT: This study presents a comprehensive analysis of letter, bigram, and trigram frequencies in the Serbian language using the Cyrillic script. Utilising a corpus of approximately 4 million characters from various literary works, newspapers, and an online encyclopedia, we calculated the frequencies of uppercase and lowercase letters, as well as bigrams and trigrams. Our findings reveal distinct patterns in the Serbian language, including the prevalence of certain letters and letter combinations. These results largely align with previous studies on Serbian and Croatian languages, with some variations due to dialectal differences. This research contributes valuable data for applications in cryptography, natural language processing, and linguistic studies specific to the Serbian language.

KEY WORDS: Serbian language, letter frequency, bigram frequency, trigram frequency, Cyrillic script, text analysis.

УВОД

Примарни циљеви ове студије су да пружи свеобухватну анализу учесталости слова, биграма и триграма у српској ћирилици и да створи поуздану референцу за будућа лингвистичка и рачунарска истраживања. Фокусирајући се на ћирилично писмо, желимо да попунимо празнину у постојећој литератури, која је претежно анализирала латинично писмо. Подаци које имамо ограничени су на учесталост слова са неколико веб-сајтова [1] [2], али постоје анализе које су доступне и за друге језике, као што је енглески [3] [4]. Међутим, већина постојећих анализа српског језика фокусира се на латинично писмо, што резултира губитком података за слова 'љ', 'њ' и 'џ', јер су у латиници представљена биграммима 'lj', 'nj' и 'dz'. Сходно томе, недостају нам адекватни подаци за биграме и триграме које они формирају.

Потенцијалне примене овог истраживања су бројне и обухватају више области. У криптографији, ове дистрибуције фреквенција могу помоћи у развоју и анализи алгоритама за шифровање специфичних за српски језик. Разбијање једноставних супституцијских шифара је у великој мери засновано на фреквенцијама слова и биграма, као што су Плејфер [5] и Цезарове шифре [6]. За обраду природног језика, наши налази могу да помогну у креирању прецизнијих језичких модела и побољшају генерисање текста, препознавање говора и системе машинског преводјења за српски језик [7] [8]. У лингвистици ови подаци могу послужити као основа за упоредна проучавања са другим језицима који користе ћирилично писмо [9] и допринети нашем разумевању еволуције и структуре српског језика. Поред тога, ови налази могу имати практичну примену у области као што је дизајн тастатуре [10]. Учесталост биграма и триграма утиче на решења анаграма [11] и на препознавање

речи [12] [13]. Неколико студија [14], [15], [16] сугерисало је да се речи и слова са високом фреквенцијом перципирају брже од оних нискофреквентних.

Предност ове анализе је у томе што испитује ћирилично писмо, дајући податке за свако слово, и велико и мало, као и за биграме и триграме. Све досадашње анализе су се ослањале на донекле ограничен корпус текста, наш корпус је приближно три пута дужи од до сад највећег који је анализиран [2] [17].

ИЗБОР УЗОРКА ТЕКСТА

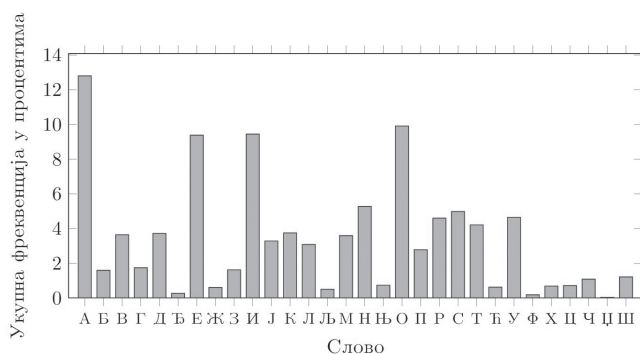
Да бисте анализирали језик, прво је потребно одабрати довољно дугачак и релевантан текст на одабраном језику. Пошто тренутно не постоји релевантан корпус ћириличног српског текста, морали смо да саставимо сопствени. Текст мора да садржи добар спој различитих књижевних жанрова: белетристика, новински чланци, научни радови, и правни документи. Наравно, уколико би се користио неки други текст као основа, резултат би био нешто другачији. Одабрали смо корпус текста од приближно 4 милиона карактера из различитих књижевних дела, новина и онлајн енциклопедије. Корпус обухвата позната српска дела као нпр. *Сеобе* Милоша Црњанског, *Моји изуми* од Николе Тесле, *Кроз васиону и векове* Милутина Миланковића, *Хајдук Станко* Јанка Веселиновића, између осталих. Укључена је и поезија, модерна књижевност Милорада Павића, као и књижевност страних писаца на српском језику. Коришћени су и новински чланци из *Политике*, *Новости* и других листова из ширег временског периода. У појединим деловима текста уочене су и исправљене правописне грешке ради прецизности анализе. Сви сабрани текстови су на еквивалентном изговору, који је данас доминантан у српском језику.

Одабир књижевних дела и других текстова за овај корпус био је заснован на њиховој разноликости и репрезентативности у оквиру српске књижевности и културе. Изабрана дела обухватају класична дела српске књижевности која су кључна за разумевање развоја језика и књижевних токова у Србији. Дела као што су *Моји изуми* Николе Тесле и *Кроз васиону и векове* Милутина Миланковића не само да представљају важне текстове у домену науке и технике, већ и пружају увид у стручни језик и терминологију. Новински чланци су коришћени због њиховог дугог историјског континуитета и утицаја на јавни живот у Србији, те осликавају еволуцију српског језика у медијима. Укључивање поезије и модерне књижевности доприноси анализи књижевног језика у различитим стиливима и формама, док књижевност страних писаца на српском језику даје ширину у приступу преводу и адаптацији језика. Циљ је био да се обухвати што шири спектар језичких форми и стилова, како би корпус био довољно разнолик и погодан за анализу језичких образаца и фреквенције у ћириличном српском тексту.

РЕЗУЛТАТИ

Резултати добијени за учесталост великих и малих слова приказани су у табели 1. Колоне табеле, с лева на десно, приказују слово, учесталост великог слова, учесталост малог слова, укупну учесталост, ранг великог слова, ранг малог слова и укупан ранг. Ови резултати се незнатно разликују од оних приказаних у [1] и [2]. Упоредјујући, уочавамо високу конзистентност, посебно за најчешћа и најређа слова. Самогласници 'А', 'О', 'Е' и 'И' су најфреквентнији у све три анализе, при чему је 'А' на првом месту са око 12%. Међу сугласницима, 'Н', 'Р', 'С' и 'Т' су најзаступљенији. Специфична српска слова попут 'Ђ', 'Ж', 'Љ', 'Њ', 'Ћ', 'Џ' и 'Ш' показују сличне фреквенције у анализама где су присутна. Најређе слово у свим сетовима је 'Џ' са око 0,04%, док је 'Ф' такође међу најређим. Мање разлике у фреквенцијама могу се приписати различитим корпусима или методама анализе. Упркос малим варијацијама, општи поредак фреквенција остаје доследан, потврђујући карактеристичне образце српског језика, укључујући високу заступљеност самогласника и специфичних ћириличних слова.

Откривено је укупно 3.905.445 слова, од којих 103.293 мала и 3.802.152 велика слова. Поред тога, било је 3.050.516 биграма, са 675 јединствених комбинација, и 2.267.549 триграма, са 6.810 различитих образаца.



Слика 1: Хистограм укупних фреквенција слова у процентима.

Учесталости биграма су приказане у табели 2. У табели су редови означени првим словима биграма, а колоне другим словима биграма. Табела 3 представља најчешће биграме и триграме заједно са њиховим бројем појављивања. Десет најчешћих биграма су <ЈЕ>, <НА>, <КО>, <ДА>, <РА>, <НИ>, <СТ>, <КА>, <ТА> и <НЕ>, што је блиско усклађено са налазима у [17]. Најчешћи триграми су у складу са налазима у [17] и [18].

Вреди напоменути да не постоји јака корелација између релативних фреквенција малих и великих слова. Најчешћа мала слова су традиционално самогласници, искључујући <У>, док су два најчешћа велика слова <П> и <С>. Поред тога, примећујемо да су међу најчешћим биграммима ретки они који почињу самогласником.

Ови резултати се углавном поклапају са анализом хрватског језика, осим што учесталост слова 'Г', 'Ј' и 'Е', као и биграма 'ИЈ' и <ЈЕ> и триграма 'ИЈЕ', већи је у хрватском због анализе ијекавског изговора [18] (од одговарајућих ћириличних еквивалената: 'И', 'Ј', 'Е', 'ИЈ', 'ЈЕ', 'ИЈЕ'). Српски ијекавски изговор имао би готово идентичне резултате као и хрватски језик.

Истраживање учесталости слова и n-грама спроведено је на различитим језицима, пружајући увид у обрасце који могу имати ширу лингвистичку релевантност. Иако је овај рад фокусиран на српски језик, студије спроведене на другим писмима и језицима нуде вредне методологије и налазе који могу унапредити разумевање анализе учесталости кроз различите језике.

На пример, студије као што је рад Бризберта и Њуа [14] о учесталости слова и норми речи у енглеском значајно су допринеле психолингвистици, идентификујући образце који утичу на обраду језика у различитим контекстима. Иако су били фокусирани на енглески језик, статистички модели које су користили за анализу учесталости слова у великим корпусима могу се прилагодити сличним истраживањима у српском језику. Поред тога, примена анализе учесталости биграма и триграма у језицима са различитим морфолошким структурама, као што су руски или грчки, може помоћи у тумачењу образаца учесталости у српском.

Штавише, студије учесталости у другим словенским и језицима заснованим на ћирилици истакле су како фонолошке и ортографске структуре утичу на расподелу слова и n-грама. Анализа [19] показује блиску везу између учесталости слова и синтаксичких улога, пружајући користан оквир за даља истраживања у српском језику. Укључивањем оваквих студија, овај рад позиционира анализу учесталости српског језика у шири лингвистички оквир, истичући међујезичке сличности и разлике које доприносе дубљем разумевању јединствених карактеристика српског језика.

Због специфичности српског језика, где једном гласу одговара једно слово, овај рад пружа вредне увиде који могу послужити као основа за фонетске анализе. У српском, свако слово представља одређени фонем, што олакшава и унапређује процес анализе фонетских структура. Ова једноставна корелација између графемских и фонемских јединица омогућава детаљније и прецизније проучавање звука у језику, у поређењу с другим језицима где може доћи до

сложенијих односа између слова и гласова. На пример, у многим западним језицима, једно слово може представљати више различитих звукова. Ова сложеност може отежати фонетске анализе и захтевати додатне слојеве разматрања у анализи [20]. У том смислу, резултати овог истраживања могу бити од посебне користи лингвистима и истраживачима, пошто пружају чврсту основу за проучавање фонетике српског језика и његових особености, те истовремено истичу разлике у анализи језика.

ЗАКЉУЧАК

Ова анализа учесталости слова, биграма и триграма у српској ћирилици нуди драгоцене увиде у статистичка својства језика. На основу корпуса од око 4 милиона карактера, наша студија открива различите фреквенције тернс између самогласника и сугласника, идентификује уобичајене комбинације слова и пружа дубину нашем разумевању српске морфологије. Налази потврђују претходне студије док се баве празнинама у постојећој литератури, посебно у вези са ћириличним писмом.

Ови резултати имају значајне импликације на криптографију, обраду природног језика и лингвистику, на пример: могу да информишу развој прилагођених алгоритама за шифровање [5] [6], побољшају језичке моделе за задатке као што су генерисање текста и машинско превођење [7] [8] и служе као ресурсе за упоредна проучавања са другим словенским језицима [9].

Међутим, постоје ограничења у саставу корпуса, који се првенствено састоји од књижевних дела и новинских чланака. Будућа истраживања би могла имати користи од проширења корпуса како би укључили шири спектар типова текста и испитивањем дистрибуције фреквенција у различитим жанровима или временским периодима.

У закључку, ова студија пружа солидну основу за будућа истраживања српске лингвистике и рачунарске обраде језика, нудећи како теоријски увид, тако и практичне примене. Редовно ажурирање оваквих анализа биће од кључног значаја за одржавање њихове релевантности како језик настави да се развија.

РЕФЕРЕНЦЕ

- [1] "Frequency of Letters and Words in the Serbian Language." 2005, accessed: 2024-08-09. <https://www.elitecurity.org/t72747-Ucestalost-slova-reci-srpskom-jeziku>.
- [2] STT Media, "Character Frequency: Serbian.", 2024, accessed: 2024-08-10. <https://www.sttmedia.com/characterfrequency-serbian>.
- [3] Jones, Michael N, and Douglas JK Mewhort. 2004. "Case-Sensitive Letter and Bigram Frequency Counts from Large-Scale English Corpora." *Behavior Research Methods, Instruments, & Computers* 36 (3): 388–96. <https://doi.org/10.3758/bf03195586>.
- [4] Mayzner, Mark S, and Margaret E Tresselt. 1962. "The Ranking of Letter Pairs and Single Letters to Match Digram and Single-Letter Frequency Counts." *Journal of Verbal Learning and Verbal Behavior* 1 (3): 203–7. [https://doi.org/10.1016/s0022-5371\(62\)80032-2](https://doi.org/10.1016/s0022-5371(62)80032-2).
- [5] Sharma, Nikhil, Himmat Meghwal, Munish Mehta, and Tarun Kumar. 2018. "A Review on Playfair Substitution Cipher and

- Frequency Analysis Attack on Playfair." In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, 1–9. IEEE. <https://doi.org/10.1109/icoei.2018.8553837>.
- [6] Shrivastava, Gaurav, Ravindra Sharma, and Manorama Chouhan. 2013. "Using Letters Frequency Analysis in Caesar Cipher with Double Columnar Transposition Technique." *International Journal of Engineering Sciences & Research Technology* 2 (6): 1475–78.
- [7] Khristoforov, Stanislav, Vladimir Bochkarev, and Anna Shevlyakova. 2020. "Recognition of Parts of Speech Using the Vector of Bigram Frequencies." In *Analysis of Images, Social Networks and Texts: 8th International Conference, AIST 2019, Kazan, Russia, July 17–19, 2019, Revised Selected Papers* 8, 132–42. Springer. https://doi.org/10.1007/978-3-030-39575-9_13.
- [8] Ney, Hermann, and Ute Essen. 1991. "On Smoothing Techniques for Bigram-Based Natural Language Modelling." In *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 825–28. IEEE Computer Society. <https://doi.org/10.1109/ICASSP.1991.150464>.
- [9] Grigas, Gintautas, and Anita Juškevičienė. 2018. "Letter Frequency Analysis of Languages Using Latin Alphabet." *International Linguistics Research* 1 (1): 18–31. <https://doi.org/10.30560/ilr.v1n1p18>.
- [10] Zhai, Shumin, Michael Hunter, and Barton A Smith. 2000. "The Metropolis Keyboard—an Exploration of Quantitative Techniques for Virtual Keyboard Design." In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, 119–28. <https://doi.org/10.1145/354401.354424>.
- [11] Dominowski, Roger L, and Carl P Duncan. 1964. "Anagram Solving as a Function of Bigram Frequency." *Journal of Verbal Learning and Verbal Behavior* 3 (4): 321–25. [https://doi.org/10.1016/s0022-5371\(64\)80073-6](https://doi.org/10.1016/s0022-5371(64)80073-6).
- [12] Broadbent, DE, and Margaret Gregory. 1968. "Visual Perception of Words Differing in Letter Digram Frequency." *Journal of Memory and Language* 7 (2): 569. [https://doi.org/10.1016/s0022-5371\(68\)80052-0](https://doi.org/10.1016/s0022-5371(68)80052-0).
- [13] Rice, Glenn A, and David Owen Robinson. 1975. "The Role of Bigram Frequency in the Perception of Words and Nonwords." *Memory & Cognition* 3 (5): 513–18. <https://doi.org/10.3758/bf03197523>.
- [14] Brysbaert, Marc, and Boris New. 2009. "Moving Beyond Kučera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English." *Behavior Research Methods* 41 (4): 977–90. <https://doi.org/10.3758/bm.41.4.977>.
- [15] Appelman, Ira B, and Mark S Mayzner. 1981. "The Letter-Frequency Effect and the Generality of Familiarity Effects on Perception." *Perception & Psychophysics* 30: 436–46. <https://doi.org/10.3758/bf03204839>.
- [16] Bryden, MP. 1968. "Symmetry of Letters as a Factor in Tachistoscopic Recognition." *The American Journal of Psychology* 81 (4): 513–24. <https://doi.org/10.2307/1421055>.
- [17] STT Media, "Syllable Frequency: Serbian.", 2024, accessed: 2024-08-09. <https://www.sttmedia.com/syllablefrequency-serbian>.
- [18] Puač, Sonja. 2002. "Congruences and Cryptography - Diploma Thesis.", accessed: 2024-08-09. <https://www.mathos.unios.hr/~mdjmic/uploads/diplomski/PUA02.pdf>.
- [19] Khokhlova, M., & Benko, V. (2020). Size of corpora and collocations: the case of Russian. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, 8(2), 58–77.
- [20] Hayden, R. E. (1950). The relative frequency of phonemes in General-American English. *Word*, 6(3), 217–223. Taylor & Francis. <https://doi.org/10.1080/00437956.1950.11659381>



Вукашин Бабић, студент, Електротехнички факултет, Универзитет у Београду
Контакт: vukasin.babic06@gmail.com
Области интересовања: обрада природних језика, дигитална обрада сигнала

ДОДАТАК

Табела 1: Учесталост и ранг великих и малих слова

Слово	Фреквенција – велико слово	Фреквенција – мало слово	Фреквенција – укупно	Ранг – велико слово	Ранг – мало слово	Ранг – укупно
А	7228	468515	475743	3	1	1
Б	5478	56635	62113	8	19	19
В	4858	137156	142014	11	12	12
Г	2257	65746	68003	17	17	17
Д	3930	141308	145238	12	10	11
Ђ	654	9410	10064	26	28	28
Е	1945	364497	366442	18	3	4
Ж	531	23043	23574	27	26	26
З	3493	59783	63276	13	18	18
И	5043	363739	368782	9	4	3
Ј	3136	125052	128188	15	14	14
К	6760	139606	146366	5	11	10
Л	1535	118630	120165	20	15	15
Љ	206	19471	19677	29	27	27
М	5041	135061	140102	10	13	13
Н	6261	199494	205755	6	5	5
Њ	1015	27386	28401	23	22	22
О	6258	380655	386913	7	2	2
П	9059	99524	108583	1	16	15
Р	2717	176759	179476	16	8	8
С	8925	185439	194364	2	6	6
Т	7183	157423	164606	4	9	9
Ћ	97	23989	24086	30	25	25
У	3440	177754	181194	14	7	7
Ф	784	6410	7194	24	29	29
Х	1718	24985	26703	19	24	24
Ц	737	27111	27848	25	23	23
Ч	1269	40810	42079	22	21	11
Џ	227	1159	1386	28	30	30
Ш	1508	45602	47110	21	20	20

Табела 2: Учесталост биграма

	А	Б	В	Г	Д	Ђ	Е	Ж	З	И	Ј	К	Л	Љ	М	Н	Њ	О	П	Р	С	Т	Ћ	У	Ф	Х	Ц	Ч	Ш	
А	12	2539	21695	4121	20434	1775	297	4081	10003	649	17320	20311	24200	2062	20993	26065	7157	21801	4810	21429	12242	18397	1435	2682	585	2692	4284	5787	172	6819
Б	8084	7	13	106	7687				120	22084	330	1	2882	731	14	642	32	7630		7242	303			3991			2		17	1
В	31334	70	891	1	29604	1	25281	106	231	3142	2213	1	4047	31	26505	1	7721	1161	17		4452					592	120		731	
Г	13822		141	18	1237	3	2313		115	2455	2	1	7422	76	73	942	135	17659		6417	39	29		3685	5	17	1		15	
Д	45835	694	2958	728			16149	5	82	15300	327	10	1021	254	945	8265	937	16406	5	7507	1569			6848	1	6	9	6	105	
Ђ	1876						3946			624			3			1	335				15			3189						
Е	399	4389	7744	9170	19957	3083	3	1872	4815	226	1772	14856	13794	2952	14392	26352	3035	5674	4543	12888	11144	17062	7710	272	463	1283	1603	5132	300	5434
Ж	3549	339	35	1	455	75	8338			6593	191	1	2	170	36	1372	431	106		74	2			1315				1	28	
З	21094	2118	3431	2033	2137	1	4597	2	3	6702	172	1	1494	21	1712	6414	4	2123		1146	1	1		1946			1	1	1	
И	673	851	10848	3112	5663	620	73	1201	12640	17	22877	13141	19441	1058	30712	18660	1367	13558	1837	6615	14815	12006	2537	83	890	13893	9312	8820	211	9689
Ј	17900	295	627	90	431		61648	6	278	10727	25	791	225	14	363	1867	41	5330	291	37	2178	349		12577	13	19	169	186	5	110
К	37770		2289	17	4	1	7903			11907	13		3123	320	258	836	637	49505		8489	1259	1880	101	11212	6	7	248	107	1	175
Л	32709	301	49	48	230		21591	1	22	35382		661	11		262	1891		18621	76	2	387	457		6075	74	81	185	82	1	17
Љ	5260	13			161		4873			3087		145				571	9	335			381	1		4412			25	3	25	
М	30678	853	18		55		19576	3	3	11323	31	99	1322	2170	3	2460	206	17329	694	2239	530	149	22	8073	52	121	147	62	5	49
Н	55843	31	28	1024	3091	298	35837	165	175	40979	16	2091	47	40	11	7		31176	140	218	3740	2798		9953	58	34	2246	509	108	125
Њ	5329			1	1		13578			5572		3				8		829			254			2567			28	6		
О	175	11004	32054	21409	24799	1658	91	4572	4372	305	29770	8385	11564	2725	23623	18816	1207	69	6620	17206	23672	9266	2442	421	1783	699	1173	5851	19	6577
П	14008						6796			6098	13	186	3444	517	2	317	30	35352	4	31650	1640	531	20	6194	15	25	140	87	2	806
Р	43016	1875	3724	1305	1393	834	33621	1505	1231	27704	227	1975	1559	276	1099	3965	370	22349	1296	4469	2431	92	11765	150	711	1537	1224	46	1559	
С	24873	132	17154	9			32535			7893	572	16650	10286		4713	4628		3618	6449	4389	1	38727		16889	257	517	448	9	2	
Т	36759	3	6180	21	70		17524		12	30148	470	2651	1217	369	332	3283	462	29236	909	16105	1024	3		8481	15	90	70	30	15	
Ћ	3783		4			38	7363			6583		105			4	554		277			12			2985			2			
У	180	3024	2548	5145	7477	691	67	3190	3235	108	3485	5074	2649	374	4015	5563	330	1600	4224	5235	5916	8241	4883	15	28	1129	934	3843	29	4396
Ф	1061	4		6	8		1222			2095	47	2	99	3	1	8		807		599	152	47		811				10	2	
Х	2643	5	1180		1		685			843	6	1	444		61	419	5	3823	1	962	1	900	24	1296		6	12	2	1	
Ц	6077	11	266	2	1		5453			9706	3	261	23	7	21	186	33	474	2	1821				2033						
Ч	6686		196		1		10321			10080	123	3871	250	3	39	3752	86	1267	3	128		47		3765	1		27			
Ш	296	66					180			433							109				15			164						
Щ	5399		473				8739			4779	34	2631	1431	880	56	956	1238	447	109	10		13445	859	1259			66	205		

Табела 3: Најчешћи биграми и триграми.

Ранг	Биграм	Фреквенција биграма	Триграм	Фреквенција триграма
1	ЈЕ	61648	СТА	12579
2	НА	55843	КОЈ	12524
3	КО	49505	ИМА	12254
4	ДА	45835	ПРЕ	10807
5	РА	43016	ИЈЕ	10468
6	НИ	40979	ОСТ	9613
7	СТ	38727	АЛИ	9373
8	КА	37770	АКО	9045
9	ТА	36759	САМ	8956
10	НЕ	35837	ПРИ	7983
11	ЛИ	35382	КАО	7694
12	ПО	35352	ОЈИ	7679
13	РЕ	33621	ОВА	7378
14	ЛА	32709	ОВО	7298
15	СЕ	32535	ВОЈ	7170
16	ОВ	32054	НИЈ	7017
17	ПР	31650	ПРО	6991
18	ВА	31334	БИЛ	6959
19	НО	31176	СТИ	6832
20	ИМ	30712	СВЕ	6592
21	МА	30678	ИЛИ	6455
22	ТИ	30148	ГОВ	6395
23	ОЈ	29770	ШТО	6356
24	ВЕ	29604	АТИ	6097
25	ТО	29236	ЈЕД	6014
26	РИ	27704	ИЈА	5951
27	ВО	26505	КАД	5939
28	ЕН	26352	ЕНИ	5901
29	АН	26065	ИЛА	5891
30	ВИ	25281	ОВИ	5810
31	СА	24873	ПОС	5749
32	ОД	24799	СВО	5688
33	АЛ	24200	РЕД	5635
34	ОС	23672	ТРА	5585
35	ОМ	23623	ОВЕ	5536
36	ИЈ	22877	ЛИК	5476
37	РО	22349	НОГ	5358
38	БИ	22084	АЛА	5345
39	АО	21801	ОЈЕ	5315
40	АВ	21695	СКО	5137
41	ЛЕ	21591	ЕЛИ	5098
42	АР	21429	ЕГО	5093
43	ОГ	21409	СТО	5073
44	ЗА	21094	ЕДА	5052
45	АМ	20993	АВА	5047
46	АД	20434	АДА	4875
47	АК	20311	АЈУ	4869
48	ЕД	19957	СКИ	4798
49	МЕ	19576	ИЛО	4781
50	ИЛ	19441	ЗНА	4717

