PANEVROPSKI UNIVERZITET
APEIRON
za multidisciplinarne i virtuelne studije
Pan-European University for Multidiscipline & Virtual Studies
Banja Luka

FAKULTET INFORMACIONIH TEHNOLOGIJA
COLLEGE OF INFORMATION TECHNOLOGIES

ITeit 2020

XII međunarodni naučno-stručni skup

Informacione Tehnologije za e-Obrazovanje

ZBORNIK RADOVA
PROCEEDINGS

25-26. 09. 2020.

Banja Luka

POKROVITELJI KONFERENCIJE
AKADEMIJA NAUKA I UMJETNOSTI REPUBLIKE SRPSKE,
MINISTARSTVO PROSVJETE I KULTURE REPUBLIKE SRPSKE,
MINISTARSTVO ZA NAUČNOTEHNOLOŠKI RAZVOJ, VISOKO OBRAZOVANJE I
INFORMACIONO DRUŠTVO REPUBLIKE SRPSKE

PANEVROPSKI APEIRON UNIVERZITET
Banja Luka

XII međunaroni naučno-stručni skup
Informacione tehnologije za e-obrazovanje

ITeO

ZBORNIK RADOVA
PROCEEDINGS

UREDNIK
ZORAN Ž.Avramović

POKROVITELJI KONFERENCIJE:
AKADEMIJA NAUKA I UMJETNOSTI REPUBLIKE SRPSKE,
MINISTARSTVO ZA NAUČNOTEHNOLOŠKI RAZVOJ, VISOKO OBRAZOVANJE I INFORMACIONO DRUŠTVO REPUBLIKE SRPSKE I
MINISTARSTVO PROSVJETE I KULTURE REPUBLIKE SRPSKE

25 – 26. 9. 2020.
Banja Luka

PANEVROPSKI APEIRON UNIVERZITET
Banja Luka

XII međunarodni naučno-stručni skup Informacione tehnologije za e-obrazovanje

ZBORNIK RADOVA

Urednik:
Akademik prof. dr ZORAN Ž. Avramović

Izdavač:
Panevropski univerzitet "APEIRON", Banja Luka, godina 2020.

Odgovorno lice izdavača:
DARKO Uremović

Glavni i odgovorni urednik izdavača:
Prof. dr ALEKSANDRA Vidović

Tehnički urednik:
SRETKO Bojić

Štampa:
CD izdanje

Tiraž:
200 primjeraka

EDICIJA:
Informacione tehnologije - **Information technologies**
Knjiga br. 29

ISBN 978-99976-34-61-0

P O Č A S N I O D B O R :

Akademik prof. dr Rajko Kuzmanović, *predsjednik Akademije nauka i umjetnosti RS (ANURS)*
Mr Srđan Rajčević, *ministar za naučnotehnološki razvoj, visoko obrazovanje i informaciono društvo RS*
Mr Natalija Trivić, *ministar prosvjete i kulture RS*
Prof. dr Zoran Ž. Avramović, *rektor Panevropskog univerziteta APEIRON*
Prof. emeritus dr Dušan Starčević, *redovni član Akademije inženjerskih nauka Srbije*
Doc. dr Siniša Aleksić, *direktor Panevropskog univerziteta APEIRON*
Darko Uremović, *predsjednik Upravnog odbora Panevropskog univerziteta APEIRON*

P R O G R A M S K I O D B O R :

Prof. dr Zoran Ž. Avramović, *Akademik Ruske akademije transportnih nauka, Akademik Ruske akademije prirodnih nauka, Akademik Ruske akademije elektrotehničkih nauka, redovni član Inženjerske akademije Srbije*
Prof. emeritus dr Dušan Starčević, *redovni član Akademije inženjerskih nauka Srbije, potpredsjednik*
Prof. dr Branko Latinović, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Prof. dr Gordana Radić, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Prof. dr Leonid Avramovič Baranov, *Russian University of Transport (MIIT - RUT), Moskva, Rusija*
Prof. dr Wang Bo, *Ningbo University of Technology, China*
Prof. dr Hristo Hristov, *University of Transport "T.Kableshkov", Bulgaria*
Prof. dr Sanja Bauk, *Durban University of Technology, South Africa*
Prof. dr Dragica Radosav, *Tečnički fakultet, Zrenjanin, Srbija*
Prof. dr Yuri M. Inkov, *Russian University of Transport (MIIT - RUT), Russia*
Prof. dr Efim N. Rozenberg, *Research Institute in Railway Transport, Russia*
Prof. dr Emil Jovanov, *University of Alabama in Huntsville, USA*
Prof. dr Vojislav Mišić, *Ryerson University, Toronto, Canada*
Prof. dr Nedim Smailović, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Prof. dr Goran Đukanović, *Panevropski univerzitet APEIRON Banja Luka, BiH*

O R G A N I Z A C I O N I O D B O R :

Prof. dr Branko Latinović, *Panevropski univerzitet APEIRON Banja Luka, BiH, predsjednik*
Mr Dražen Marinković, *Panevropski univerzitet APEIRON Banja Luka, BiH, sekretar Konferencije*
Sretko Bojić, *Panevropski univerzitet APEIRON Banja Luka, BiH, tehnički urednik*
Mr Dalibor Drljača, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Mr Igor Grujić, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Marijana Petković, *Panevropski univerzitet APEIRON Banja Luka, BiH, PR konferencije*
Vladimir Domazet, *Panevropski univerzitet APEIRON Banja Luka, BiH, tehnička podrška*
Radovan Vučenović, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Alen Tatarević, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Stana Mišić, *Panevropski univerzitet APEIRON Banja Luka, BiH, logistika*

R E C E Z E N T S K I O D B O R :

Prof. dr Željko Stanković, *Panevropski univerzitet APEIRON Banja Luka, BiH, predsjednik*
Prof. dr Milan Marković, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Docent dr Tijana Talić, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Docent dr Siniša Tomić, *Panevropski univerzitet APEIRON Banja Luka, BiH*
Docent dr Saša Salapura, *Panevropski univerzitet APEIRON Banja Luka, BiH*

XII međunarodni naučno-stručni skup
Informacione Tehnologije za elektronsko Obrazovanje
ITeO 2020
Banja Luka, 25 - 26. 09. 2020. godine

PANEVROPSKI UNIVERZITET APEIRON
za multidisciplinarne i virtuelne studije
Pan-European University for Multidiscipline & Virtual Studies
Banja Luka

PRIMJENA INFORMACIONO - KOMUNIKACIONIH TEHNOLOGIJA U PROUČAVANJU JEZIKA

- uvodno predavanje na XII međunarodnom naučno-stručnom skupu
Informacione tehnologije za e-obrazovanje - ITeO -

Nedim Smailović, Zoran Ž. Avramović
Panevropski univerzitet APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina
nedim.i.smailovic@apeiron-edu.eu, zoran.z.avramovic@apeiron-edu.eu

Apstrakt: *U ovom radu prikazane su neke mogućnosti savremene računarske lingvistike. Analiza je usmjerena u četiri pravca. Prvi dio se odnosi na savremeni jezik kolumnista četiri balkanska elektronska medija. Oni pišu o različitim temama, različitim jezičkim stilovima ali analiza jasno ukazuje na velike sličnosti u nekim segmentima. U drugom dijelu statistički su analizirana tri poznata književna djela balkanskih pisaca. Djela su različita po sadržaju (psihološki i filozofski roman, komedija sa autobiografskim sadržajem, roman iz doba realizma), a vremenski raspon između objavljivanja ovih djela je blizu osamdeset godina. Predmet analize trećeg dijela ovog rada je poznati roman Žila Verna „20.000 milja pod morem" u prijevodu na četiri jezika: njemački, francuski, engleski i hrvatski jezik. Cilj analize je da se ukaže na međusobne sličnosti i razlike koje su svojstvene pojedinim jezicima. U četvrtom dijelu na velikom uzorku data je i skala međusobne udaljenosti pojedinih jezika, jednog od drugog u pogledu učestalosti pojedinih slova. Pokazano je da učestalost pojedinih slova može biti iskorištena za potrebe automatskog prepoznavanja nepoznatog jezika.*

Ovakva istraživanja pripadaju lingvistici, kao nauci o jeziku ali se rezultati mogu upotrijebiti kao mali segmenti u razvoju danas veoma aktuelne vještačke inteligencije.

Ključne riječi: *računarska lingvistika, jezik, analiza teksta, prepoznavanje jezika, vizualizacija podataka*

1. UVOD

Svako biće ima potrebu za komunikacijom. To je ujedno i uslov opstanka jer primljene informacije omogućavaju spoznaju okoline, pribavljanje hrane, bijeg od neprijatelja ili hvatanje plijena. Sposobnost verbalnog i neverbalnog komuniciranja je jedna od najvažnijih čovekovih odlika koje ga čine homo sapiensom. Ljudska komunikacija je dinamičan i veoma složeni proces koji traje od rođenja do kraja života. Komuniciranje može biti jednosmjerno, dvosmjerno, verbalno, neverbalno, ciljano, masovno, povremeno, trajno, prikriveno, anonimno, javno ili od svega toga kombinovano. U ljudskoj komunikaciji nemoguće je izbjeći neverbalne znake, jer čak i mirno sjedenje i šutnja su neka poruka. Komunikacija riječima može biti govorna ili pisana, a u oba slučaja podrazumijeva se da se odvija na nekom od jezika, što znači razumijevanje simbola, odnosno riječi i njihovog značenja. Jezik se može opisati kao sistem znakova koji služe za sporazumijevanje među ljudima... Lingvisti su sročili više definicija jezika. Prema Jahiću „To je forma kojom se čovjek ispoljava kao misaono biće, otkrivajući tako svoju suštinu i svoju razlikovnost prema drugim živim bićima." [1]. Različiti izvori navode i druge definicije jezika. U

rječniku Merriam Webster stoji: „Jezik je sistematsko sredstvo komuniciranja ideja ili osjećaja upotrebom konvencionalizovanih znakova, zvukova, pokreta, ili oznaka da bi se shvatilo značenje." [15]

Alfabetom se naziva uređeni skup slova ili drugih znakova kojima se piše jedan ili više jezika. To može biti i naziv za sistem znakova ili signala koji služe kao ekvivalenti za slova. Znakovi alfabeta se stapaju u riječi, uzimajući u obzir određena gramatička pravila kojima se formira pisani jezik.

Jezikom se bavi lingvistika (franc. linguistique, prema lat.lingua: jezik) koja se dalje može podijeliti na: fonetiku (nauku o glasovima), fonologiju (nauku o funkciji glasova), morfologiju (nauku o oblicima jezičkih jedinica), sintaksu (nauku o organizaciji rečenice), semantiku (nauku o značenju u jeziku) itd. Lingvistika je ujedno i multidisciplinarna, pa se u 20. vijeku javljaju i njene specijalizirane grane kao što su: matematička lingvistika, psiholingvistika, sociolingvistika, neurolingvistika...

U svakodnevnoj komunikaciji jezikom se služimo automatski, bez mnogo razmišljanja o pravilnosti njegove upotrebe. On se uči od prvog do zadnjeg dana života i može se reći da postaje sastavni dio ličnosti i jedna od važnih karakteristika svakog pojedinca. U naučnom svijetu proučavanju pisma i govora posvećuje se velika pažnja. Prostor za istraživanje je gotovo beskrajan jer relevantni izvori potvrđuju da živimo u svijetu u kojem preko 7,7 milijardi ljudi [18]. komunicira, govor i piše na preko 7.100 jezika. Taj broj je teško tačno utvrditi jer su često isprepletene granice među pojedinim jezicima i dijalektima. Jezik može i umrijeti, pa je samo u 20. vijeku ukupno 110 jezika proglašeno izumrlim. Organizacija Ujedinjenih naroda za obrazovanje, nauku i kulturu (UNESCO) u svojim izvještajima procjenjuje da će, ako se ništa ne učini, do kraja ovog vijeka nestati polovina jezika koji se danas govore. Nestankom nepisanih i nedokumentovanih jezika čovječanstvo bi izgubilo ne samo kulturno bogatstvo, nego i važno znanje o precima ugrađeno, posebno, u autohtone jezike [12].

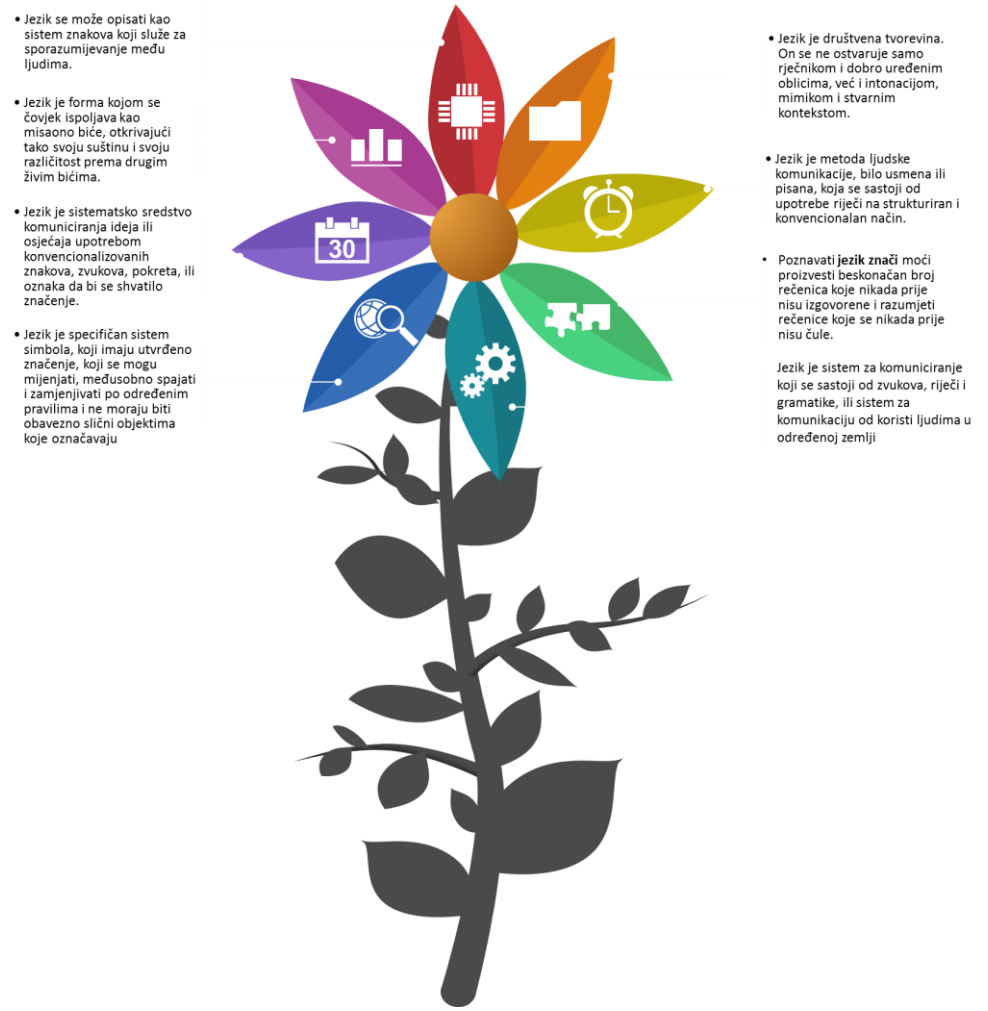
U savremenom svijetu sve većih i bržih promjena u svim segmentima života i jezik se mijenja. Neke riječi nestaju, neke dobijaju novo značenje, a neke potpuno nove nastaju.

Iz svakodnevne upotrebe kod nas nestaju riječi: granap, buđelar, puce, izba, japija...

Riječ banda nekad se upotrebljavala u smislu strana,snimiti je značilo skinuti, a riječ haljine se koristila u smislu odjeća.

Na pitanje novinara: „Kakvi su vam rezultati poslovanja?", upitani je na TV odgovorio: „Trudimo se, ali bi nam bilo mnogo lakše da se mi nojari u Bosni i Hercegovini udružimo." U Bosni i Hercegovini nikad se do sada nisu gajili nojevi, ali, kada se to desilo, morala je nastati nova riječ „nojari". Skovana je u duhu jezika, kao npr. mesari, pekari, stočari i sl.

U takvom ambijentu jezik opstaje, prilagođava se i ostaje najmoćnije sredstvo komunikacije među ljudima.



Slika 1 *Neke od mnogobrojnih definicija jezika*

O ljepoti jezika veoma lijep sud izriče Isidora Sekulić, prva žena akademik Srpske Akademije nauka, davne 1941. godine, dočekujući bosanskohercegovačke pisce u Beogradu, riječima: "Bosanski jezik i književnost to je jedna ogromna livada koja se guši od rasta, cveća i mirisa. Livada ostaje blizu, zemlju krasi i preliva, u zemlju otresa seme."

Na inicijativu Vijeća Evrope od 2001. godine na dan 26. septembra svake godine obilježava se Evropski dan jezika.

Slika 2 *Evropski dan jezika*

(Izvor: https://edl.ecml.at/Home/WhyaEuropeanDayofLanguages/tabid/1763/language/bs-Latn-BA/Default.aspx)

Ciljevi obilježavanja Evropskog dana jezika su:

1. Upozoriti javnost na važnost učenja jezika i raznolikost raspona naučenih jezika kako bi se povećala višejezičnost i međukulturalno razumijevanje;
2. Podsticati, njegovati i čuvati bogatu jezičnu i kulturnu raznolikost Evrope;
3. Podsticati cjeloživotno učenja jezika u školi i izvan nje, bilo u svrhe školovanja, za profesionalne potrebe, za potrebe mobilnosti ili užitka i razmjene.

U Evropi ima preko 220 autohtonih jezika, a još više ih se govori. Mnoge porodice su dvojezične ili višejezične, a dosta ih ima porijeklom sa drugih kontinenata. Najčešći neevropski jezici tu su arapski, kineski i hindi, koji imaju i svoj poseban sistem pisanja.

2. STATISTIČKA ANALIZA TEKSTA U ODABRANIM KOLUMNAMA

Za svaku analizu potreban je veći ili manji uzorak. Uzorci za analizu u ovom radu su uzeti iz tekstova kolumnista četiri balkanska elektronska medija: Bosna i Hercegovina - Dnevni avaz (Muhamed Filipović), Srbija - Politika (Aleksandar Apostolovski), Hrvatska - Jutarnji list (Miljenko Jergović) i Crna Gora - Vijesti (Miodrag Lekić). [6] [5] [14] [17] Kolumne su preuzete iz nedavnih online izdanja, pa se može reći da analiza obuhvata savremeni jezik.

Da bi se mogla raditi poredbena analiza jezika pomenutih kolumnista potrebno je porediti uzorke jednake dužine, pa su tekstovi uzimani i objedinjavani da budu jednaki po broju karaktera (bez razmaka). U svim primjerima zbirni skup kolumni svakog kolumniste ima tačno 148.232 karaktera (bez razmaka). Među ovim karakterima su i ona slova koja se ne koriste u abecedi kolumnista (w, q i sl.). Za taj broj karaktera trebalo je: 23 kolumne autora Jergovića, 32 kolumne autora Filipovića, 34 kolumne autora Apostolovskog i 25 kolumni autora Lekića. U nekim analizama je analiziran i zbirni tekst svih kolumnista što čini uzorak od preko pola miliona karaktera.

Tabela 1 *Okvirna statistika uzoraka teksta*

	SVI	JERGOVIĆ	FILIPOVIĆ	APOSTOLOVSKI	LEKIĆ
Broj stranica	**163**	40	46	41	36
Broj riječi	**109.707**	27.976	28.857	27.509	25.365
Broj karaktera bez razmaka	**592.928**	148.232	148.232	148.232	148.232
Broj karaktera sa razmacima	**701.032**	175.893	176.607	175.417	173.115
Broj paragrafa	**1.936**	323	565	425	623
Broj linija	**7.997**	1.953	1.946	2.161	1.937
Broj rečenica	**5.519**	1.301	1.140	1.503	1.575

Važno je napomenuti da ovim radom nisu analizirani zaključci, mišljenja i stavovi kolumnista niti redakcija njihovih elektronskih medija, već je riječ isključivo o statističkoj analizi korištenog teksta.

Autori imaju svoj stil pisanja u kojem se neke kombinacije riječi češće javljaju. U slijedećoj tabeli date su kombinacije od tri riječi koje su kod svakog autora zastupljene više od pet puta.

U objedinjenom tekstu svih kolumnista "ono što je" je najviše puta ponovljena kombinacija od tri riječi.

Kod pojedinih autora to su:

Jergović: "ono što je" (13 puta), Filipović: "Bosne i Hercegovine" (24 puta), Apostolovski: "da li je" (12 puta) i Lekić: "u Crnoj Gori" (25 puta).

Tabela 2 *Najviše puta ponovljena kombinacija od tri riječi*

3 riječi zajedno	SVI	3 riječi zajedno	JERGOVIĆ	3 riječi zajedno	FILIPOVIĆ	3 riječi zajedno	APOSTO-LOVSKI	3 riječi zajedno	LEKIĆ
ono što je	33	ono što je	13	Bosne i Hercegovine	24	da li je	12	u Crnoj Gori	25
i da se	30	koji su se	12	Bosni i Hercegovini	23	se da je	10	s druge strane	10
s druge strane	29	i to je	11	a to je	19	kako bi se	10	o Crnoj Gori	9
u Crnoj Gori	27	a onda i	11	u Bosni i	18	da li će	9	u isti mah	9
Bosne i Hercegovine	26	ne samo da	10	ono što je	18	kao da je	8	Crnoj Gori i	8
ono što se	25	u vrijeme kada	8	s druge strane	18	ne može da	7	i da se	6
Bosni i Hercegovini	25	je riječ o	8	i da se	15	kao što je	6	SAD i Rusije	6
kao što je	23	prije nego što	8	na taj način	13	i da se	6	radi se o	6
tako da je	23	ono što se	7	kao što je	12	je da je	6	na međunarodnom planu	6
a to je	22	tako da se	7	da je to	12	je reč o	6	u svakom slučaju	6
koji su se	22	tako da je	7	tako da je	12	saveza za Srbiju	6	i da je	5
kao što su	21	kao što su	6	Bosna i Hercegovina	12	u kojoj je	6	u vezi sa	5
da je to	20	a zatim i	6	da se ne	11	pravoslavne nove godine	6	u kojem je	5

U objedinjenom tekstu svih kolumnista najviše je puta ponovljena kombinacija od dvije riječi je "da je" – 418 puta. Kod pojedinih autora dvije riječi jedna do druge najčešće su: Jergović: "da je" (86 puta), Filipović: "da se " (171 puta), Apostolovski: "da je" (127 puta) i Lekić: "da je" (63 puta).

Tabela 3 *Najviše puta ponovljena kombinacija od dvije riječi*

2 riječi zajedno	SVI	2 riječi zajedno	JERGOVIĆ	2 riječi zajedno	FILIPOVIĆ	2 riječi zajedno	APOSTO-LOVSKI	2 riječi zajedno	LEKIĆ
da je	418	da je	86	da se	171	da je	127	da je	63
da se	395	što je	78	da je	140	da se	100	da se	59
što je	210	da se	63	i da	90	da su	48	pa i	44
koji je	174	je u	60	što je	86	koji je	43	crnoj gori	39
je u	167	i u	47	koji je	73	su se	36	je u	36
i da	157	su se	39	koji su	55	je u	36	koji su	28
koji su	134	kao i	37	to je	54	da će	36	i u	27
i u	127	što se	37	ono što	45	koji se	36	koji je	27
su se	127	koji su	32	koja je	43	da li	32	u crnoj	26
da su	126	ono što	32	je bio	42	je to	30	je bio	26
to je	113	koji je	31	tako da	41	je da	30	i to	23
je to	107	bi se	30	zbog toga	41	se u	30	je i	21
što se	102	je to	29	da će	40	se da	29	su se	20

Tabela 4 *Najfrekventnije riječi*

Najčešća riječ	SVI	Najčešća riječ	JERGOVIĆ	Najčešća riječ	FILIPOVIĆ	Najčešća riječ	APOSTO-LOVSKI	Najčešća riječ	LEKIĆ
i	4603	i	1389	i	1153	je	1033	i	1137
je	3889	je	1058	je	1098	i	924	u	873
u	3390	u	827	da	1088	da	847	je	700
da	2796	se	520	u	859	u	831	da	407
se	1986	da	454	se	580	se	539	se	347
na	1551	na	416	na	389	na	412	na	334
su	1223	su	323	sam	319	su	365	su	243
za	768	što	297	su	292	za	235	sa	184
što	692	ne	224	to	262	kao	200	za	174
to	684	a	196	a	241	od	176	o	144

ne	678	kao	196	koji	232	koji	165	od	132
a	673	s	195	što	230	a	160	iz	116
koji	663	od	190	ne	182	ne	156	ne	116
kao	652	nije	188	za	176	ali	147	koji	115
od	622	za	183	s	165	će	145	to	111
o	562	to	181	kao	161	s	138	kao	95
s	528	bi	171	o	158	bi	137	nije	81
nije	515	o	164	mi	141	to	130	a	76
bi	452	koji	151	bio	130	nije	128	treba	72
sam	440	ali	136	od	124	iz	114	dakle	69
iz	439	ili	134	koja	119	što	111	do	67
će	421	će	117	nije	118	o	96	pa	65
ali	401	iz	115	on	108	kako	91	bio	63
ili	353	tako	96	tako	108	ili	79	istorije	62
sa	337	samo	95	odnosno	102	kada	76	već	61
bio	330	ni	94	će	101	–	74	koje	60
koja	277	bilo	94	bi	96	bio	72	će	58
kako	273	nego	88	koje	95	već	70	bez	57
koje	272	po	83	iz	94	sa	66	sve	57
tako	268	sve	82	bilo	92	godine	66	ili	56
bilo	257	te	79	smo	92	ga	65	što	54
samo	257	kada	73	ja	88	po	65	rata	52
kada	245	kako	68	bih	86	li	62	ali	51
sve	242	koja	68	jer	86	mu	60	uz	49
do	236	biti	66	ili	84	jer	60	koja	49
po	228	koje	65	kako	79	ako	58	zemlje	49
pa	223	bio	65	nego	78	samo	58	bi	48
ni	218	pa	63	do	75	sve	56	još	48
mi	214	ga	62	toga	75	posle	55	po	45
jer	212	do	57	tome	71	koje	52	kada	45

on	201	jer	54	kad	71	ni	50	poslije	41
nego	193	onda	51	rekao	70	dok	50	sam	41
već	190	bila	51	ali	67	pa	49	dvije	40
ga	189	ono	49	države	67	još	49	između	40
te	180	on	48	samo	65	jedan	47	crnoj	40
biti	167	vrijeme	47	ljudi	65	tako	44	gori	39
smo	167	godina	47	šta	62	više	43	samo	39
ako	165	može	46	vrlo	62	koja	41	godine	39
li	158	ona	46	način	61	on	40	kako	35
bila	158	ništa	45	prema	59	sam	39	države	35

Kada se promatraju pojedinačne riječi, uočljivo je da među najfrekventnijim u zbirnom tekstu nema imenica, glagola, pridjeva, brojeva... Dominiraju veznici, pomoćni glagoli, prijedlozi (neleksičke, funkcijske riječi)... Zanimljivo je da se: i, je, u, da, se, na kod svih autora nalaze među prvih šest najfrekventnijih riječi.

U vezi s tim u kolumnama autora bi se mogla uraditi posebna analiza leksičke gustine kojom se mjeri koliko je tekst informativan.

Napomena:

Leksička gustina se definiše kao broj leksičkih riječi (ili sadržajnih riječi) podijeljenih sa ukupnim brojem riječi. Leksičke riječi daju tekstu značenje To su imenice, pridjevi, glagoli i prilozi. Druge vrste (funkcijskih) riječi kao što su pomoćni glagoli, čestice, prijedlozi, veznici … više su gramatičke prirode i daju malo ili nimalo informacija o čemu se radi u tekstu.

Kod pojedinih autora najfrekventnija imenica je: Jergović: “vrijeme” i “godina” (47 puta), Filipović: “države” (67 puta) i “ljudi” (65 puta), Apostolovski: “godine” (66 puta) i Lekić: “istorije” (62 puta) i “rata” (52 puta).

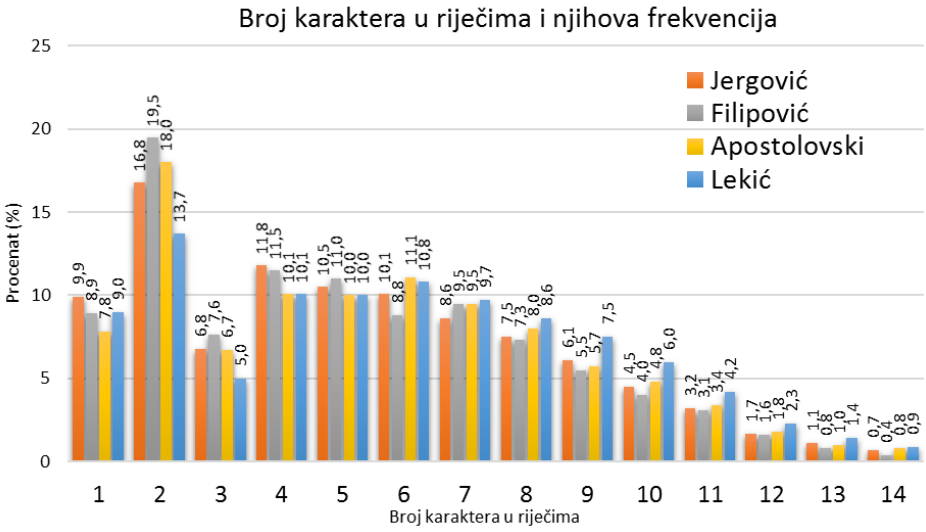
Digrami ili digrafi, iz grčkog jezika: δίς dís, "dvostruko, dva puta" i γράφω gráphō, "pisati") su kombinacije dvije najmanje pisane jedinice, slova (grafema) kojima se označava jedan glas (fonem) u nekom jeziku. Jedan digram nije isto što i dva karaktera izgovorena jedan za drugim. Digrami su često prisutni u stranim jezicima, npr: qu, ch, ph, ee, cs, dz, dzs, gy, ly, ny, sz, ty, zs, dh, gj, ll, nj, rr, sh, th, xh, zh. Kada se digram piše velikim slovima, oba znaka se pišu velikim slovima.

U bosanskom, srpskom, hrvatskom i crnogorskom jeziku, koji su tretirani kao jedinstveni srpsko-hrvatski jezik prije raspada SFR Jugoslavije primjeri digrama su dž, lj i nj.

Svaki autor ima svoj stil pisanja koji se odlikuje različitim osobinama. Kod svih autora u analiziranim tekstovima riječi dužine do 14 karaktera čine preko 99% teksta, a njihova učestalost prema broju karaktera data je u slijedećoj tabeli i grafikonu.

Tabela 5 *Frekvencija riječi prema broju karaktera*

svi		Jergović		Filipović		Apostolovski		Lekić	
Broj karaktera u riječima	Frekvencija (%)	Broj karaktera u riječima	Frekvencija (%)	Broj karaktera u riječima	Frekvencija (%)	Broj karaktera u riječima	Frekvencija (%)	Broj karaktera u riječima	Frekvencija (%)
1	8,9	1	9,9	1	8,9	1	7,8	1	9,0
2	17,1	2	16,8	2	19,5	2	18,0	2	13,7
3	6,6	3	6,8	3	7,6	3	6,7	3	5,0
4	10,9	4	11,8	4	11,5	4	10,1	4	10,1
5	10,4	5	10,5	5	11,0	5	10,0	5	10,0
6	10,1	6	10,1	6	8,8	6	11,1	6	10,8
7	9,3	7	8,6	7	9,5	7	9,5	7	9,7
8	7,8	8	7,5	8	7,3	8	8,0	8	8,6
9	6,2	9	6,1	9	5,5	9	5,7	9	7,5
10	4,8	10	4,5	10	4,0	10	4,8	10	6,0
11	3,5	11	3,2	11	3,1	11	3,4	11	4,2
12	1,8	12	1,7	12	1,6	12	1,8	12	2,3
13	1,1	13	1,1	13	0,8	13	1,0	13	1,4
14	0,7	14	0,7	14	0,4	14	0,8	14	0,9



Slika 3 *Frekvencija riječi prema broju karaktera*

U tekstu se javljaju češće i rjeđe kombinacije dva susjedna slova. Neke čak predstavljaju i riječ, kao npr.: iz, na, li, on, to, mi i sl. U uzetim uzorcima teksta ima čak i kombinacija dva ista slova, ali to je najčešće zbog toga što autori navode izvorno pisana neka imena ili riječi stranog porijekla.

U bosanskom, srpskom, hrvatskom i crnogorskom jeziku rijetki su primjeri dva ista slova zaredom. Javljaju u složenicama kao npr.: najjači, najjasniji, narodnooslobodilački, prekookeanski, kooperativan itd.

Tabela 6 *Najčešće kombinacije dva susjedna slova u analiziranim kolumnama*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
je	na	ra	ko	ni	st	ij	no	an	li	ti	da	po	ta	ka	re	ov	ne	ja	to	pr	ma	en	va	ri	im	oj	al	nj	vi

U slijedećoj tabeli prikazana je učestalost svih kombinacija susjednih slova u objedinjenom tekstu svih autora kolumni. Tabela se čita tako što se prvo gleda kolona, a zatim red. Na presjeku se nalazi broj tih kombinacija u tekstu. Npr. kombinacija „do“ je pronađena 1.966 puta. Polja sa brojem uočenih kombinacija su uslovno obojena tako što su u svakoj vertikali najrjeđe crvene, a što je broj veći boja prelazi u bijelu, a zatim u zelenu. U tabeli se jasno po crvenoj boji prepoznaju kombinacije dva slova koje se ne pojavljuju nijednom ili rijetko i one zeleno obojene koje su najfrekventnije (u jednoj koloni).

Tabela 7 *Pregled učestalosti kombinacija dva susjedna slova (prvo je u koloni, a drugo u redu tabele).*

	a	b	c	č	ć	d	dž	đ	e	f	g	h	i	j	k	l	lj	m	n	nj	o	p	r	s	š	t	u	v	z	ž		rank
a	5	1114	755	743	591	5644	33	314	168	225	1626	261	72	4495	4816	3603	613	4430	8525	882	74	1576	7689	2787	562	5079	178	4378	3019	746	a	65.003	1
b	290	0	5	0	0	164	10	0	610	2	2	2	110	55	0	36	1	154	10	0	1533	0	314	10	0	55	378	0	402	75	b	4.218	22
c	1248	0	6	2	0	1	0	0	287	0	0	0	1467	29	168	32	2	27	460	3	277	37	172	128	0	10	167	81	0	0	c	4.604	19
č	1083	0	0	0	0	0	0	0	375	0	0	0	1707	16	13	2	0	6	45	0	571	8	34	4	29	2	632	11	0	0	č	4.535	20
ć	196	0	0	0	0	0	0	0	963	0	0	0	947	0	8	0	0	8	2	0	203	57	13	0	122	0	598	1	0	0	ć	3.118	23
d	2560	15	1	0	0	1	0	0	3034	0	144	5	607	58	0	48	19	40	456	0	3682	1	210	15	0	21	827	79	173	103	d	12.099	15
dž	30	0	0	0	0	0	0	0	6	0	0	0	16	17	0	0	0	0	2	0	5	0	6	0	0	0	17	0	0	0	dž	99	30
đ	378	0	0	0	0	0	0	0	542	0	0	0	25	0	0	0	0	0	0	0	216	0	75	0	0	0	82	0	0	2	đ	1.320	28
e	50	860	790	938	1143	1728	39	450	7	231	389	201	51	14848	1517	1983	868	2621	4725	1862	37	639	4759	2975	771	2689	71	3062	577	932	e	51.813	2
f	156	0	0	0	0	0	0	0	153	3	0	0	100	4	1	49	0	13	109	0	119	0	16	39	0	6	16	0	0	0	f	784	29
g	605	0	9	0	0	174	0	0	1460	3	2	0	494	10	2	23	0	0	161	0	3592	0	219	1	0	0	795	13	202	0	g	7.765	16
h	229	3	59	0	0	0	0	0	84	0	1	0	2054	5	1	1	0	0	5	0	95	19	95	87	0	0	137	0	0	0	h	2.875	25
i	166	2831	2652	1368	829	2570	40	91	50	480	530	230	2	2755	1750	5927	331	1743	6820	700	103	934	4253	1141	512	5702	27	3845	1127	846	i	50.355	3
j	2522	211	60	30	0	433	0	0	205	0	0	0	6186	4	2	2677	0	353	3873	0	4077	194	88	268	1	92	516	616	49	21	j	22.478	7
k	3007	2	36	1213	23	0	0	0	1868	1	0	5	2309	126	0	62	21	8	343	1	1212	13	272	3040	357	327	557	50	0	0	k	14.853	13
l	4054	401	10	84	0	212	0	0	2184	32	684	25	2797	20	591	24	0	328	4	0	2092	388	237	1751	251	13	688	752	331	12	l	17.965	10
lj	302	59	1	0	0	21	0	0	375	0	2	0	227	2	108	0	0	230	2	0	273	76	25	132	80	36	20	389	1	12	lj	2.373	26
m	2625	1	0	2	0	101	0	25	2575	0	45	18	4179	72	15	102	0	4	2	0	3802	2	277	680	12	25	650	3	319	9	m	15.543	12
n	5939	193	0	640	116	1970	0	0	4386	2	62	25	3535	478	253	715	122	346	21	0	2822	58	881	982	364	695	673	1241	817	247	n	27.583	5
nj	1267	3	0	13	4	165	0	0	490	0	10	0	268	12	195	0	1	38	0	0	26	7	20	0	276	28	20	4	0	37	nj	2.884	24
o	1975	1173	70	199	48	1966	21	53	488	309	2770	446	1876	614	7181	2702	23	2740	6007	93	39	5292	2959	568	55	4445	130	3454	302	44	o	48.042	4
p	772	0	0	0	0	6	0	0	458	0	0	0	337	47	4	13	0	239	15	0	1008	1	248	894	18	112	706	0	0	0	p	4.878	18
r	2950	797	366	0	0	1538	0	0	2578	207	1101	309	1251	8	1076	0	0	175	30	0	3392	4432	13	574	2	2129	613	1237	130	34	r	24.942	6
s	1697	3	0	0	0	406	0	0	1732	20	2	1	2802	448	306	84	57	110	974	28	3703	383	500	30	0	79	1189	109	0	0	s	14.663	14
š	1022	1	0	0	0	1	0	0	554	0	0	0	812	15	7	3	6	11	9	0	721	104	202	0	0	1	837	88	0	0	š	4.394	21
t	3037	0	21	0	0	2	0	0	2334	6	11	63	2926	47	549	267	2	15	1245	0	1351	80	436	6661	1782	22	837	5	1	0	t	21.700	8
u	289	433	316	304	218	778	7	475	140	104	536	44	0	2687	1592	897	523	871	1113	298	76	568	1670	2002	110	1456	5	553	382	89	u	18.540	9
v	3605	14	18	9	0	424	0	0	1274	0	17	160	1708	116	442	11	0	1	23	0	4728	0	562	1766	27	1203	293	0	450	6	v	16.857	11
z	1275	35	0	0	0	12	0	0	806	2	20	0	2213	47	1	2	0	1	101	0	662	0	100	0	0	24	433	0	7	0	z	5.741	17
ž	528	0	0	0	0	0	0	0	236	0	0	0	188	0	0	2	0	0	12	0	514	0	541	0	0	0	305	0	0	0	ž	2.326	27
	a	b	c	č	ć	d	dž	đ	e	f	g	h	i	j	k	l	lj	m	n	nj	o	p	r	s	š	t	u	v	z	ž		
	43.862	8.149	5.175	5.545	2.972	18.317	150	1.412	30.422	1.627	7.954	1.795	41.266	27.035	20.598	19.265	2.589	14.512	35.092	3.867	41.005	14.869	26.886	26.532	5.331	24.251	12.397	19.971	8.289	3.215		474.350
rank	1	18	22	20	25	13	30	29	5	28	19	27	2	6	10	12	26	15	4	23	3	14	7	8	21	9	16	11	17	24		

Može se uočiti da dva susjedna slova najčešće na prvom mjestu imaju slovo a (43.862 puta), zatim i, o, n, e, j, itd. a najrjeđe slova: dž (150 puta), đ, f, h, lj, itd.

U kombinaciji dva susjedna slova drugo slovo je najčešće a (65.003 puta), zatim e, i, o, n, r, itd., a najrjeđe slova: dž (99 puta), f, đ, ž, lj, h itd.

ANALIZA UČESTALOSTI POJEDINIH SLOVA U TEKSTU

Učestalost pojedinih slova u tekstu može zavisiti od više faktora. Među njima su: jezik na kojem se piše, vrsta teksta (proza, pjesma, stručni tekst...), tema o kojoj se piše, autorov stil i uobičajeni fond riječi kojima se koristi itd. Na internetu se mogu pronaći stranice [11] i programi kojima se može saznati broj pojedinih slova i datom tekstu, ali dobar dio njih ne prepoznaje naša slova č, ć, dž, đ, lj i nj. U tom slučaju mogu se iskoristiti mogućnosti koje nude tekst procesori.

Slijedi analiza učestalosti pojedinih slova za svakog autora kolumni posebno.

Tabela 8 *Frekvencija slova pojedinih autora kolumni - po abecedi i u opadajućem nizu*

			Filipović	Jergović	Apostolovski	Lekić	SVI
1	a	A	17.213	15.795	17.140	16.067	66.215
2	b	B	2.168	1.952	2.266	1.748	8.134
3	c	C	1.132	1.248	1.445	1.549	5.374
4	č	Č	1.269	1.610	1.419	1.355	5.653
5	ć	Ć	961	809	1.227	630	3.627
6	d	D	5.506	4.359	5.271	4.616	19.752
7	dž	Dž	53	21	56	28	158
8	đ	Đ	371	293	404	345	1.413
9	e	E	11.923	12.619	12.539	12.807	49.888
10	f	F	422	341	372	545	1.680
11	g	G	2.426	2.412	2.504	2.553	9.895
12	h	H	1.025	1.129	800	823	3.777
13	i	I	14.615	15.252	14.051	14.503	58.421
14	j	J	6.000	5.960	4.600	5.451	22.011
15	k	K	5.062	5.471	5.646	5.239	21.418
16	l	L	3.913	4.070	4.624	4.241	16.848
17	lj	Lj	612	804	584	701	2.701
18	m	M	5.333	5.036	4.469	5.047	19.885
19	n	N	8.048	8.497	8.019	8.605	33.169
20	nj	Nj	949	1.027	841	1.056	3.873
21	o	O	14.294	13.591	13.509	12.784	54.178
22	p	P	3.465	3.432	4.075	3.911	14.883
23	r	R	6.426	6.356	7.173	7.896	27.851
24	s	S	6.973	6.818	6.822	6.853	27.466
25	š	Š	1.412	1.528	1.399	971	5.310
26	t	T	6.651	6.965	5.974	6.615	26.205
27	u	U	5.806	5.564	6.141	5.843	23.354
28	v	V	4.974	5.102	5.233	4.748	20.057
29	z	Z	2.238	2.203	2.291	2.386	9.118
30	ž	Ž	913	885	752	642	3.192
			142.153	141.149	141.646	140.558	**565.506**

sort		Filipović
a	A	17.213
i	I	14.615
o	O	14.294
e	E	11.923
n	N	8.048
s	S	6.973
t	T	6.651
r	R	6.426
j	J	6.000
u	U	5.806
d	D	5.506
m	M	5.333
k	K	5.062
v	V	4.974
l	L	3.913
p	P	3.465
g	G	2.426
z	Z	2.238
b	B	2.168
š	Š	1.412
č	Č	1.269
c	C	1.132
h	H	1.025
ć	Ć	961
nj	Nj	949
ž	Ž	913
lj	Lj	612
f	F	422
đ	Đ	371
dž	Dž	53
		142.153

sort		Jergović
a	A	15.795
i	I	15.252
o	O	13.591
e	E	12.619
n	N	8.497
t	T	6.965
s	S	6.818
r	R	6.356
j	J	5.960
u	U	5.564
k	K	5.471
v	V	5.102
m	M	5.036
d	D	4.359
l	L	4.070
p	P	3.432
g	G	2.412
z	Z	2.203
b	B	1.952
č	Č	1.610
š	Š	1.528
c	C	1.248
h	H	1.129
nj	Nj	1.027
ž	Ž	885
ć	Ć	809
lj	Lj	804
f	F	341
đ	Đ	293
dž	Dž	21
		141.149

sort		Apostolovski
a	A	17.140
i	I	14.051
o	O	13.509
e	E	12.539
n	N	8.019
r	R	7.173
s	S	6.822
u	U	6.141
t	T	5.974
k	K	5.646
d	D	5.271
v	V	5.233
l	L	4.624
j	J	4.600
m	M	4.469
p	P	4.075
g	G	2.504
z	Z	2.291
b	B	2.266
c	C	1.445
č	Č	1.419
š	Š	1.399
ć	Ć	1.227
nj	Nj	841
h	H	800
ž	Ž	752
lj	Lj	584
đ	Đ	404
f	F	372
dž	Dž	56
		141.646

sort		Lekić
a	A	16.067
i	I	14.503
e	E	12.807
o	O	12.784
n	N	8.605
r	R	7.896
s	S	6.853
t	T	6.615
u	U	5.843
j	J	5.451
k	K	5.239
m	M	5.047
v	V	4.748
d	D	4.616
l	L	4.241
p	P	3.911
g	G	2.553
z	Z	2.386
b	B	1.748
c	C	1.549
č	Č	1.355
š	Š	971
nj	Nj	823
h	H	701
lj	Lj	642
ć	Ć	630
f	F	545
đ	Đ	345
dž	Dž	28
		140.558

sort		SVI
a	A	66.215
i	I	58.421
o	O	54.178
e	E	49.888
n	N	33.169
r	R	27.851
s	S	27.466
t	T	26.205
u	U	23.354
j	J	22.011
k	K	21.418
v	V	20.057
m	M	19.885
d	D	19.752
l	L	16.848
p	P	14.883
g	G	9.895
z	Z	9.118
b	B	8.134
c	C	5.374
č	Č	5.653
š	Š	5.310
nj	Nj	3.873
h	H	3.777
ć	Ć	3.627
ž	Ž	3.192
lj	Lj	2.701
f	F	1.680
đ	Đ	1.413
dž	Dž	158
		565.506

Redoslijed slova kod pojedinih autora:

Filipović: **a i o e n s t r j u d m k v l p g z b š č c h ć nj ž lj f đ dž**

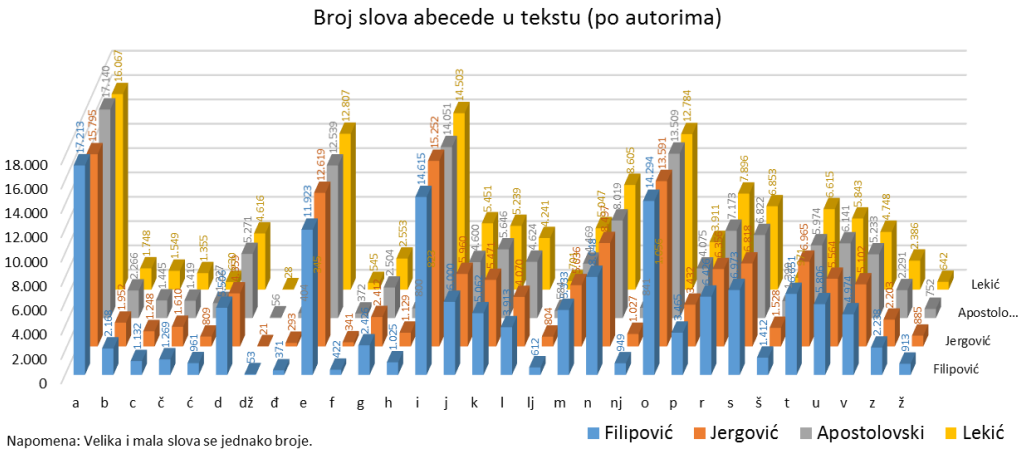
Jergović **a i o e n t s r j u k v m d l p g z b č š c h nj žć lj f đ dž**

Apostolovski **a i o e n r s u t k d v l j m p g z b c č š ć nj h ž lj đ f dž**

Lekić **a i e o n r s t u j k m v d l p g z b c č nj š h lj ž ć f đ dž**

Svi **a i o e n r s t u j k v m d l p g z b č c š nj h ć ž lj f đ dž**

Indikativno je uporediti ove redoslijede sa redoslijedima datim u Tabeli 12: Frekvencija svih slova u jednom jeziku – ilustrativni primjer



Slika 4 *Broj pojedinih slova abecede u tekstovima autora kolumni*

Četiri istaknuta kolumnista elektronskih medija pišu o različitim temama, različitim jezičkim stilovima, ali su statistički gledano jasno uočljive velike sličnosti u nekim segmentima. Ovakvi rezultati ostavljaju prostora za zaključak da se radi o jednom policentričnom jeziku, što nije rijetkost u savremenom svijetu. Naravno, konačan sud o tome trebaju reći lingvisti. Policentrični jezici su i engleski jezik (Velika Britanija, SAD), njemački (Njemačka, Austrija, Švicarska), francuski (Francuska, Kvebek, Belgija), španjolski (Španija, Argentina, Meksiko), persijski (Iran, Avganistan, Tadžikistan), portugalski (Brazil i Portugalija), arapski (Saudijska Arabija, Iran, Irak, Tunis, Egipat…), itd. Svaka od varijanti policentričnog jezika ima svoju standardnu nacionalnu varijantu, gramatiku i pravopis prepoznatljive po nekim jezičnim razlikama. [9]

Postoje i monocentrični jezici kakav je npr. japanski ili ruski koji nemaju više standardiziranih varijanti.

3. STATISTIČKA ANALIZA TEKSTA U ODABRANIM KNJIŽEVNIM DJELIMA

Jezik je živ i svakodnevno se mijenja ali postoje temelji na kojima te promjene stoje. Odgovor na pitanje: „Da li su i temelji podložni promjenama?“ može dati savremena informaciona tehnologija svojim moćnim alatima i statističkim analizama. Rezultati jedne takve analize prikazani su u ovom radu i u analizi poznatih djela: „Derviš i smrt“ - Meše Selimovića (1910-1982), „Autobiografija“ – Branislava Nušića (1864-1938) i „U registraturi“ – Ante Kovačića (1854-1889).

U dostupnim elektronskim verzijama ovih djela [7], [10], [3] statistika tekstprocesora je slijedeća:

Tabela 9 *Generalna statistika tekstprocesora*

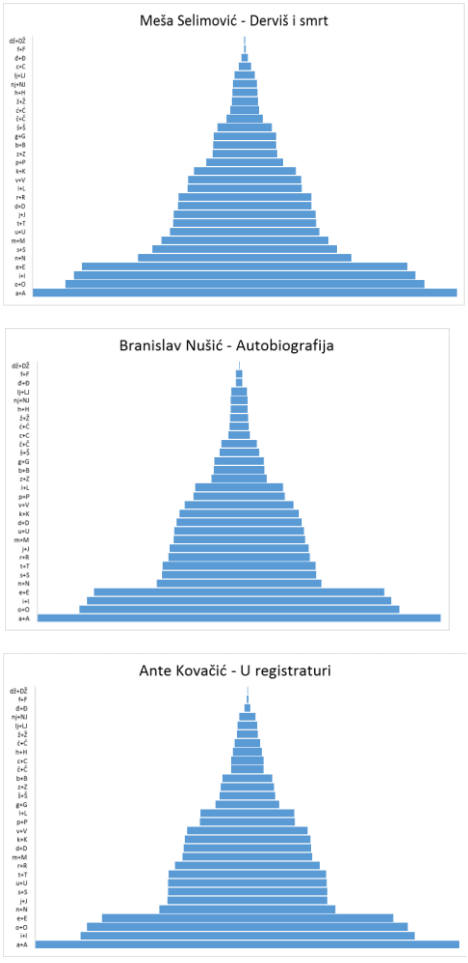
	Selimović	**Nušić**	**Kovačić**
Broj stranica	248	126	210
Broj riječi	119.045	63.898	147.323
Broj karaktera (bez razmaka)	541.162	297.749	685.059
Broj karaktera (sa razmakom)	662.322	363.026	832.585
Broj paragrafa	3.465	1.511	2.752
Broj linija	9.435	4.924	10.521

S obzirom da su ovi tekstovi različite dužine, za neke analize računato je procentualno učešće pojedinih elemenata, a u drugim primjerima dovoljni su bili apsolutni iznosi.

Lingvistika je nauka koja proučava unutarnji red među jezičkim jedinicama. Na početku pogledajmo rezultate analize ukupnog broja pojedinih slova (i mala i velika) u datim djelima i grafičku interpretaciju dobijenih rezultata.

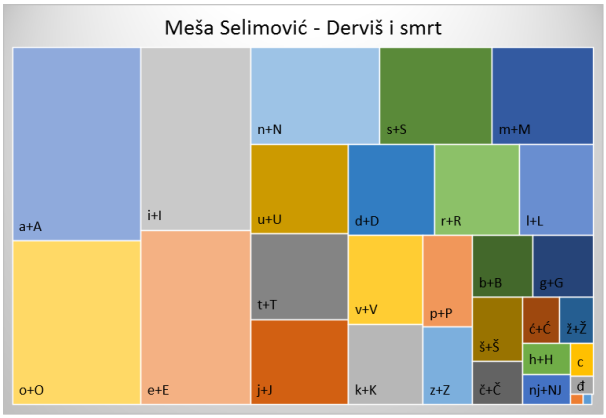
Tabela 10 *Redoslijed ukupnog broja slova u djelima*

Selimović		Nušić		Kovačić	
Slovo	Ukupno	Slovo	Ukupno	Slovo	Ukupno
a+A	62.258	a+A	36.018	a+A	80.282
o+O	52.725	o+O	28.568	i+I	63.222
i+I	50.102	i+I	27.193	o+O	60.699
e+E	47.781	e+E	25.942	e+E	55.143
n+N	31.335	n+N	14.681	n+N	33.407
s+S	27.107	s+S	13.830	j+J	30.215
m+M	24.555	t+T	13.643	s+S	30.182
u+U	21.939	r+R	12.618	u+U	30.161
t+T	21.012	j+J	12.434	t+T	29.903
j+J	20.850	m+M	11.772	r+R	27.488
d+D	19.662	u+U	11.617	m+M	24.669
r+R	19.536	d+D	11.161	d+D	24.125
l+L	16.831	k+K	10.691	k+K	23.843
v+V	16.619	v+V	9.786	v+V	22.891
k+K	14.989	p+P	8.175	p+P	18.059
p+P	11.319	l+L	7.858	l+L	17.780
z+Z	9.577	z+Z	5.014	g+G	12.165
b+B	9.284	b+B	4.576	š+Š	10.654
g+G	9.240	g+G	4.471	z+Z	10.088
š+Š	8.061	š+Š	3.578	b+B	9.542
č+Č	5.407	č+Č	3.202	č+Č	6.264
ć+Ć	4.329	c+C	1.926	c+C	6.213
ž+Ž	3.900	ć+Ć	1.795	h+H	5.496
h+H	3.740	ž+Ž	1.666	ć+Ć	4.962
nj+NJ	3.580	h+H	1.578	ž+Ž	4.072
lj+LJ	3.089	nj+NJ	1.542	lj+LJ	3.882
c+C	1.901	lj+LJ	1.430	nj+NJ	3.138
đ+Đ	1.038	đ+Đ	654	đ+Đ	1.266
f+F	328	f+F	629	f+F	436
dž+DŽ	237	dž+DŽ	59	dž+DŽ	59

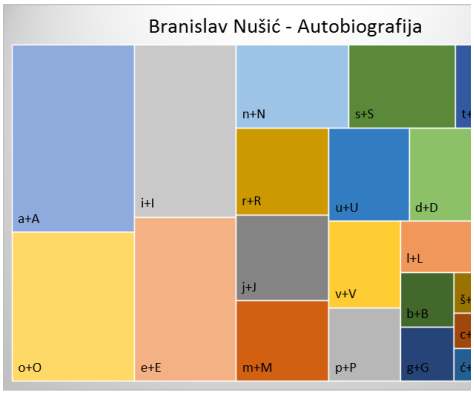


Slika 5 *Broj pojedinih slova: Selimović: Derviš i smrt*

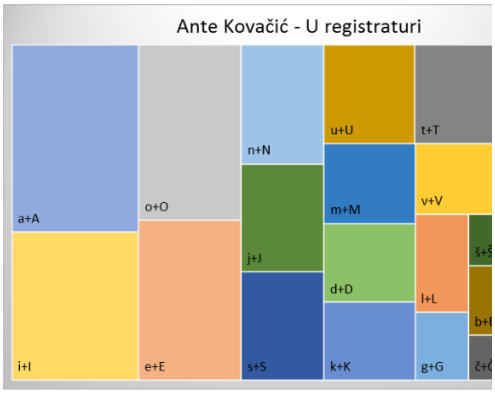
Slijedi prikaz istih ovih veličina u drugom tipu grafikona tipa Treemap. Frekvencije pojedinih slova proporcionalne su veličinama njima odgovarajuće površine na grafikonu. Ovakvim prikazom jasno se vidi dominacija (veća frekvencija) pojedinih slova u tekstu u odnosu na druga slova. Ispod svakog grafikona prikazan je redoslijed slova u tom jeziku u opadajućem nizu slijeva nadesno.



a+A, o+O, i+I, e+E, n+N, s+S, m+M, u+U, t+T, j+J, d+D, r+R, l+L, v+V, k+K, p+P, z+Z, b+B, g+G, š+Š, č+Č, ć+Ć, ž+Ž, h+H, nj+NJ, lj+LJ, c+C, đ+Đ, f+F, dž+DŽ.



a+A, o+O, i+I, e+E, n+N, s+S, t+T, r+R, j+J, m+M, u+U, d+D, k+K, v+V, p+P, l+L, z+Z, b+B, g+G, š+Š, č+Č, c+C, ć+Ć, ž+Ž, h+H, nj+NJ, lj+LJ, đ+Đ, f+F, dž+DŽ.



a+A, i+I, o+O, e+E, n+N, j+J, s+S, u+U, t+T, r+R, m+M, d+D, k+K, v+V, p+P, l+L, g+G, š+Š, z+Z, b+B, č+Č, c+C, h+H, ć+Ć, ž+Ž, lj+LJ, nj+NJ, đ+Đ, f+F, dž+DŽ.

Slika 6 *Zastupljenost pojedinih slova u analiziranim djelima*

Kada se objedine podaci i promatranih kolumni (Filipović, Jergović, Apostolovski, Lekić) i književnih djela (Selimović, Nušić, Kovačić), počevši od najfrekventnijeg slova naniže, redoslijed je slijedeći:

a i o e n s r t u j m k d v l p g z b š č c h ć nj ž lj đ f dž.

Tabela 11 *Procentualna zastupljenost svakog slova u svim tekstovima zbirno i pojedinačna odstupanja kod svakog autora.*

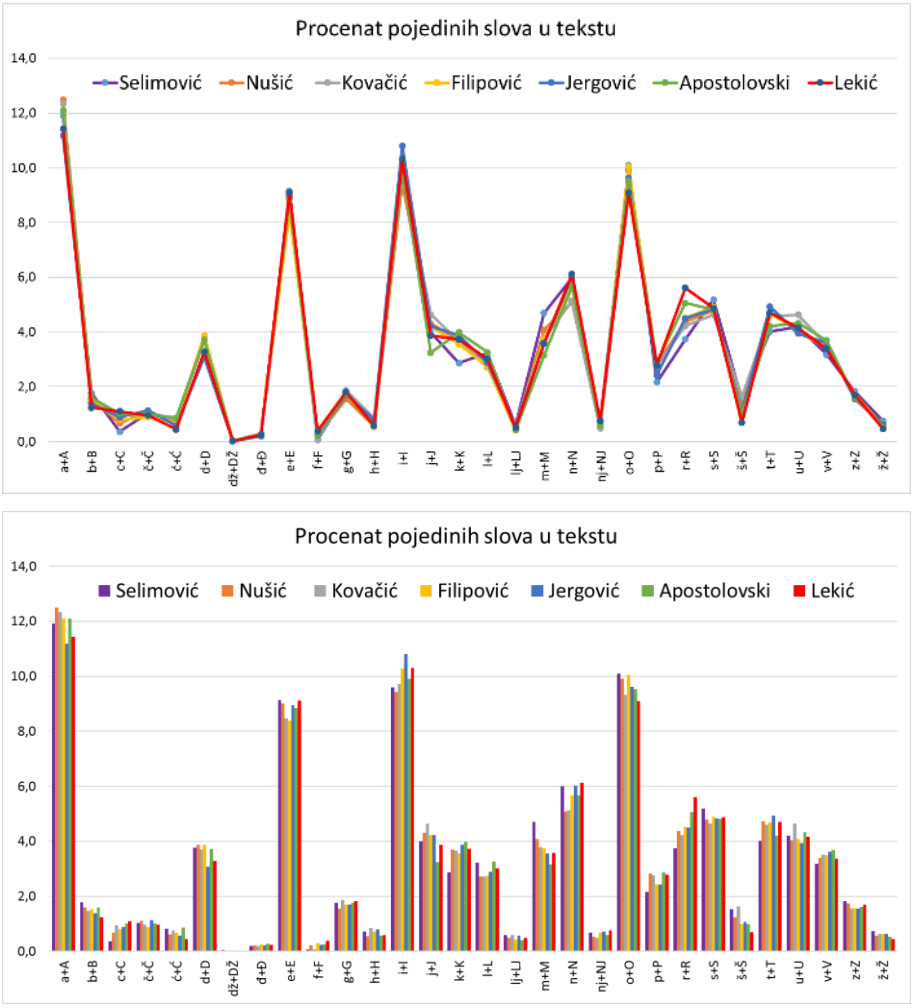
Slovo	Prosjek	Filipović	Jergović	Apostolovski	Lekić	Selimović	Nušić	Kovačić
a+A	11,94	0,17	-0,75	0,16	-0,51	-0,02	0,56	0,40
b+B	1,51	0,01	-0,13	0,09	-0,27	0,27	0,08	-0,04
c+C	0,83	-0,03	0,06	0,19	0,27	-0,46	-0,16	0,13
č+Č	1,02	-0,12	0,13	-0,01	-0,05	0,02	0,10	-0,05
ć+Ć	0,68	-0,01	-0,11	0,18	-0,23	0,15	-0,06	0,08
d+D	3,62	0,26	-0,53	0,10	-0,33	0,15	0,26	0,09
dž+DŽ	0,03	0,01	-0,01	0,01	-0,01	0,02	-0,01	-0,02
đ+Đ	0,23	0,03	-0,02	0,05	0,01	-0,03	0,00	-0,04
e+E	8,85	-0,46	0,09	0,01	0,27	0,30	0,16	-0,37
f+F	0,22	0,08	0,02	0,04	0,17	-0,16	0,00	-0,15
g+G	1,74	-0,03	-0,03	0,03	0,07	0,03	-0,19	0,13
h+H	0,68	0,04	0,12	-0,12	-0,10	0,03	-0,14	0,16
i+I	10,01	0,27	0,79	-0,09	0,31	-0,42	-0,57	-0,29
j+J	4,07	0,15	0,15	-0,83	-0,20	-0,08	0,24	0,57
k+K	3,63	-0,07	0,25	0,36	0,10	-0,76	0,08	0,04
l+L	2,94	-0,19	-0,06	0,32	0,07	0,28	-0,22	-0,21
lj+LJ	0,51	-0,08	0,06	-0,10	-0,01	0,08	-0,02	0,08
m+M	3,81	-0,05	-0,24	-0,65	-0,22	0,89	0,28	-0,01
n+N	5,67	-0,01	0,35	-0,01	0,45	0,33	-0,58	-0,53
nj+NJ	0,63	0,03	0,09	-0,04	0,12	0,05	-0,10	-0,15
o+O	9,67	0,39	-0,04	-0,13	-0,57	0,43	0,25	-0,33
p+P	2,62	-0,18	-0,18	0,26	0,17	-0,45	0,22	0,16
r+R	4,58	-0,06	-0,08	0,49	1,04	-0,84	-0,20	-0,35
s+S	4,87	0,04	-0,04	-0,05	0,01	0,32	-0,07	-0,22
š+Š	1,17	-0,17	-0,09	-0,18	-0,48	0,38	0,07	0,47
t+T	4,56	0,12	0,38	-0,34	0,15	-0,53	0,18	0,04
u+U	4,20	-0,11	-0,26	0,14	-0,04	0,00	-0,17	0,44
v+V	3,47	0,03	0,15	0,23	-0,09	-0,29	-0,07	0,05
z+Z	1,65	-0,08	-0,09	-0,04	0,04	0,18	0,09	-0,10
ž+Ž	0,60	0,04	0,03	-0,07	-0,14	0,15	-0,02	0,03

U stupcu *Prosjek* nalazi se prosječna procentualna zastupljenost pojedinih slova u promatranim tekstovima svih autora zajedno. U ostalim stupcima su za svako slovo dati apsolutni iznosi odstupanja kod svakog autora u odnosu na prosjek (prosjek kod autora minus prosjek kod svih). Odstupanja su veoma mala. Najveće pozitivno odstupanje je kod autora Lekića i slova R (razlika procenata 1,04), a najveće negativno odstupanje je kod autora Selimovića i slova R (razlika procenata -0,84).

Na osnovu urađene analize i podataka u prikazanoj tabeli može se izvesti više zaključaka, a jedan od indikativnih je da u ukupnom tekstu od 2.026.250 slova gotovo polovinu 931.227 ili 46,1% čini 5 najzastupljenijih: a i o e n.

S druge strane posmatrano, 17 slova: v l p g z b š č c h ć nj ž lj đ f dž čini tek malo više od jedne petine teksta – 20,54%).

Koliko se učestalost pojedinih slova malo razlikuje između svih 7 autora, najbolje pokazuju ujednačene linije na slijedećim grafikonima:



Slika 7 *Grafikoni procentualnog učešća pojedinih slova u analiziranim tekstovima.*

Koliki je značaj pet najfrekventnijih slova može se vidjeti u slijedećem primjeru. Uzmimo tri rečenice na početku romana *Derviš i smrt*:

„Počinjem ovu svoju priču, nizašto, bez koristi za sebe i za druge, iz potrebe koja je jača od koristi i razuma, da ostane zapis moj o meni, zapisana muka razgovora sa sobom, s dalekom nadom da će se naći neko rješenje kad bude račun sveden, ako bude, kad ostavim trag mastila na ovoj hartiji što čeka kao izazov. Ne znam šta će biti zabilježeno, ali će u kukama slova ostati nešto od onoga što je bivalo u meni, pa se više neće gubiti u kovitlacima magle, kao da nije ni bilo, ili da ne znam šta je bilo. Tako ću moći da vidim sebe kakav postajem, to čudo koje ne poznajem, a čini mi se da je čudo što uvijek nisam bio ono što sam sad.“

Kada se samo 5 najfrekventnijih slova: *a, i, o, e, n,* zamijeni praznim prostorom, dobija se:

„P č j m vu sv ju pr ču, z št , b z k r st z s b z drug , z p tr b k j j j č d k r st r zum , d st z p s m j m , z p s muk r zg v r s s b m, s d l k m d m d ć s ć k rj š j k d bud r ču sv d , k bud , k d st v m tr g m st l v j h rt j št č k k z z

v. z m št ć b t z b lj ž , l ć u kuk m sl v st t št d g št j b v l u m , p s v š ć gub t u k v tl c m m gl , k d j b l , l d z m št j b l . T k ću m ć d v d m s b k k v p st j m, t čud k j p z j m, č m s d j čud št uv j k s m b št s m s d.“

Mogućnost razumijevanja sadržaja znatno je umanjena, gotovo do neprepoznavanja

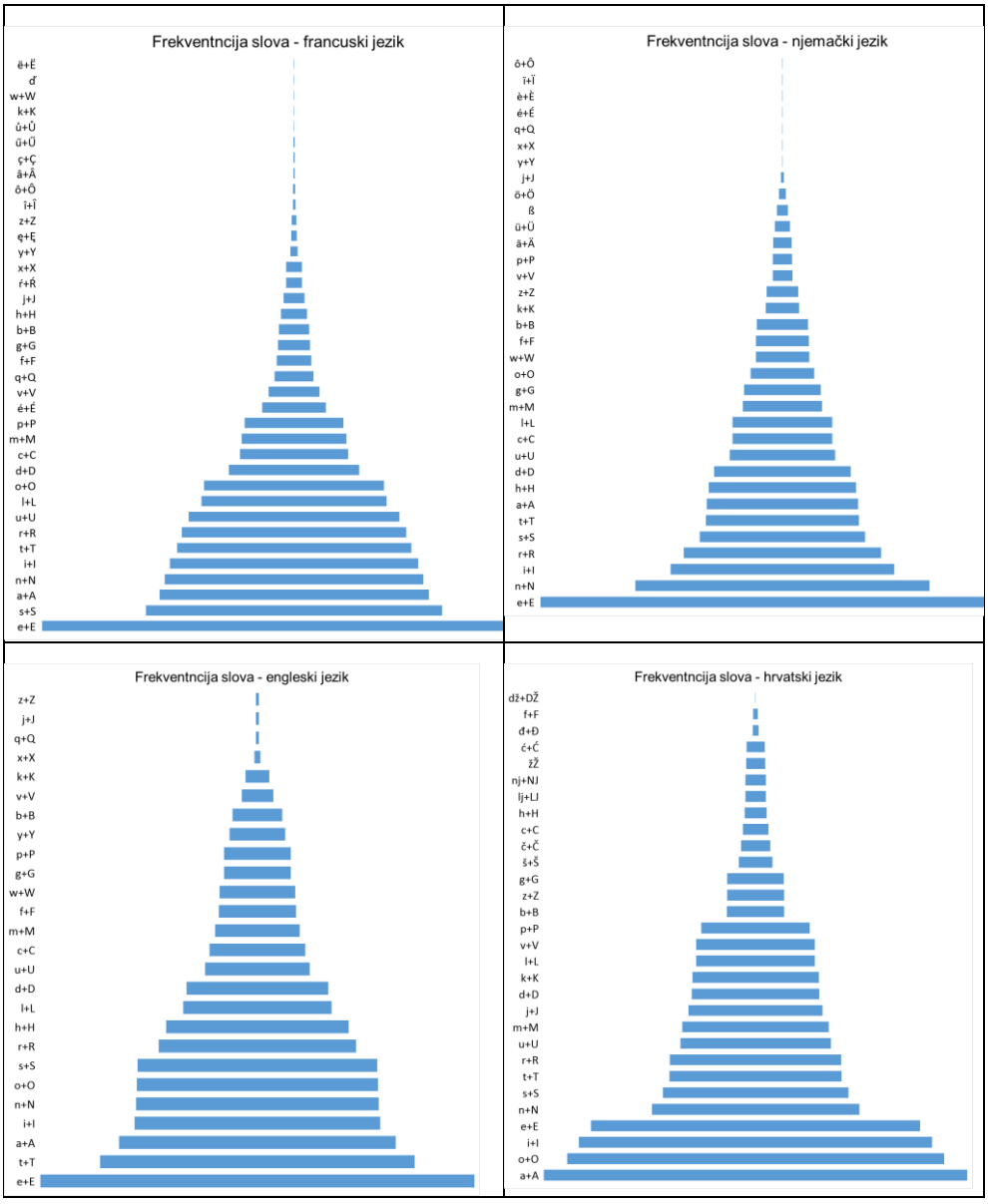
U prethodnoj analizi promatrana su djela tri poznata pisca Meše Selimovića, Branislava Nušića i Ante Kovačića i četiri savremena kolumnista. Svi autori su porijeklom iz raznih geografskih prostora (Bosna i Hercegovina, Srbija, Hrvatska, Crna Gora). Djela su različita po sadržaju (psihološki i filozofski roman, komedija sa autobiografskim sadržajem, roman iz doba realizma, savremene kolumne). Vremenski raspon između objavljivanja ovih djela je preko sto trideset godina (od 1888. do danas). Pa, ipak, računarskom analizom jasno su uočljive velike sličnosti u nekim elementima.

4. STATISTIČKA ANALIZA TEKSTA U ROMANU „20.000 MILJA POD MOREM“

Jezik je živ i svakodnevno se mijenja ali postoje temelji na kojima te promjene stoje. Odgovor na pitanja: *Da li se i temelji vremenom mijenjaju? Koliko se ti temelji razlikuju u pojedinim jezicima?* i sl. - može dati savremena informaciona tehnologija svojim moćnim alatima i statističkim analizama. Rezultati jedne takve analize prikazani su u ovom radu. Analizom su obuhvaćeni prijevodi romana „20.000 milja pod morem“ na četiri jezika: njemački, francuski, engleski i hrvatski jezik.

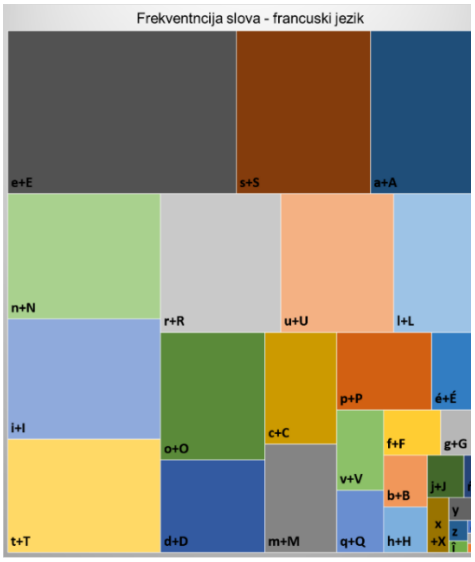
S obzirom da su ovi tekstovi različite dužine, za neke analize računato je procentualno učešće pojedinih elemenata, a u drugim primjerima dovoljni su bili apsolutni iznosi.

Lingvistika je nauka koja proučava unutarnji red među jezičkim jedinicama. Različiti jezici imaju najčešće i različite skupove glasova koji se u njima nalaze, pa i različite skupove slova kojima se oni bilježe. Pogledajmo taj red putem analize procentualnog učešće pojedinih slova (mala i velika) u romanu „20.000 milja pod morem“ u prijevodu na četiri jezika: njemački, francuski, engleski i hrvatski jezik.

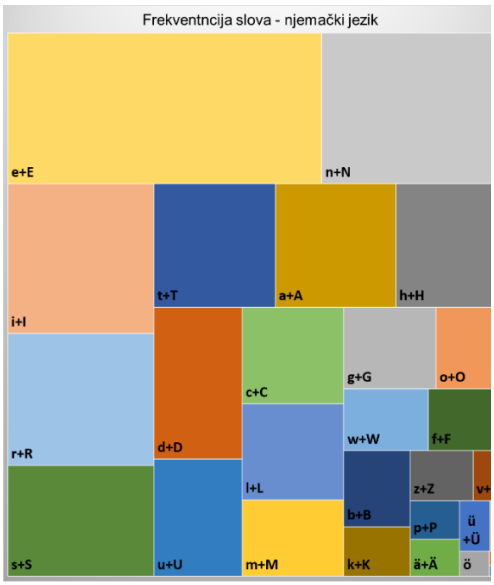


Slika 8 *Frekvencija slova u romanu „20.000 milja pod morem“ na raznim jezicima (Funnel grafikon).*

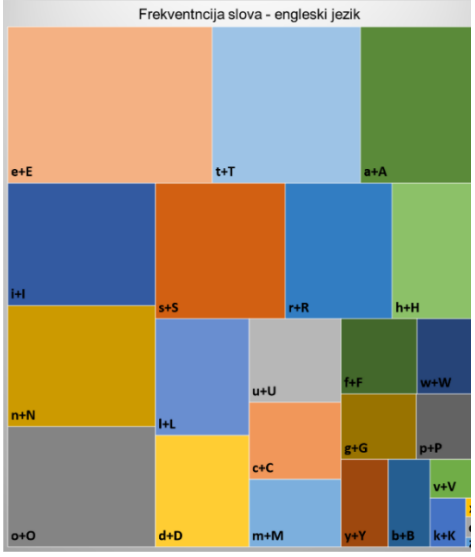
Slijedi prikaz istih ovih veličina u drugom tipu grafikona tipa Treemap. Frekvencije pojedinih slova proporcionalne su veličinama njima odgovarajuće površine na grafikonu. Ovakvim prikazom jasno se vidi dominacija (veća frekvencija) pojedinih slova u tekstu u odnosu na druga slova. Ispod svakog grafikona prikazan je redoslijed slova u tom jeziku u opadajućem nizu slijeva nadesno.



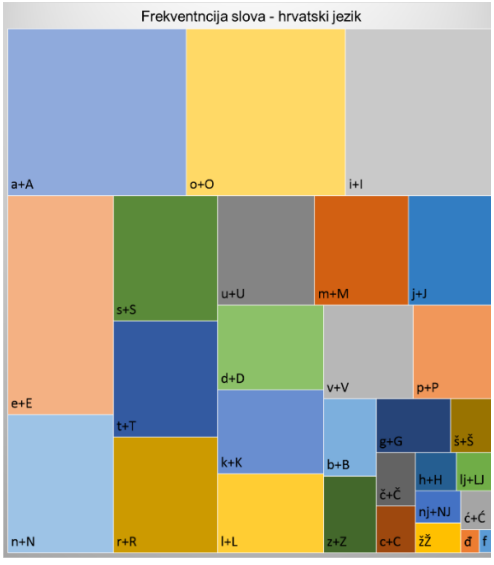
Francuski jezik: e+E; s+S; a+A; n+N; i+I; t+T; r+R; u+U; l+L; o+O; d+D; c+C; m+M; p+P; é+É; v+V; q+Q; f+F; g+G; b+B; h+H; j+J; ŕ+Ŕ; x+X; y+Y; ę+Ę; z+Z; î+Î; ô+Ô; â+Â; ç+Ç; ű+Ű; ů+Ů; k+K; w+W; ď; ë+Ë.



Njemački jezik: e+E; n+N; i+I; r+R; s+S; t+T; a+A; h+H; d+D; u+U; c+C; l+L; m+M; g+G; o+O; w+W; f+F; b+B; k+K; z+Z; v+V; p+P; ä+Ä; ü+Ü; ß; ö+Ö; j+J; y+Y; x+X; q+Q; é+É; è+È; ï+Ï; ô+Ô.



Engleski jezik: e+E; t+T; a+A; i+I; n+N; o+O; s+S; r+R; h+H; l+L; d+D; u+U; c+C; m+M; f+F; w+W; g+G; p+P; y+Y; b+B; v+V; k+K; x+X; q+Q; j+J; z+Z.



Hrvatski jezik: a+A; o+O; i+I; e+E; n+N; s+S; t+T; r+R; u+U; m+M; j+J; d+D; k+K; l+L; v+V; p+P; b+B; z+Z; g+G; š+Š; č+Č; c+C; h+H; lj+LJ; nj+NJ; žŽ; ć+Ć; đ+Đ; f+F; dž+DŽ.

Slika 10 *Frekventni raspored slova u romanu „20.000 milja pod morem“ na raznim jezicima (grafikon tipa Treemap).*

Na osnovu urađene analize može se izvesti više zaključaka, a jedan od indikativnih je da u pojedinim prijevodima istog romana blizu jedne polovine teksta čini samo 5 slova i to:

e, n, i, r, s - u njemačkom jeziku (48,7%)

e, s, a, n, i - u francuskom jeziku (46,1%)

e, t, a, i, n - u engleskom jeziku (44,3%)

a, o, i, e, n - u hrvatskom jeziku (46,6%)

e, a, i, n, s - u svim jezicima (44,2%)

S druge strane posmatrano, npr. u engleskom prijevodu 15 najmanje frekventnih slova: u, c, m, f, w, g, p, y, b, v, k, x, q, j, z zajedno čini tek malo više od jedne petine teksta – 22,0%).

U hrvatskom prijevodu čak 17 najmanje frekventnih slova: l, v, p, b, z, g, š, č, c, h, lj, nj, ž, ć, đ, f, dž zajedno čini malo manje od jedne petine teksta – 19,9%).

U ovom radu statistički je analiziran jedan tekst preveden na četiri jezika. Francuski jezik pripada grupi romanskih jezika, engleski i njemački grupi germanskih jezika, a hrvatski grupi (južno)slavenskih jezika. S obzirom da se radi o istom djelu na četiri jezika, a ne o različitim tekstovima, izbjegnut je utjecaj koji na statističke rezultate koje mogu imati razni faktori kao npr. stil pisanja, tema, vrijeme nastanka djela, vrsta teksta (književni žanr) i sl.

5. JEZICI U EVROPI – POREĐENJE I PREPOZNAVANJE

Mada ne postoji jedinstveno stajalište u definiciji kontinenta (lat. continens sc. terra, continere – držati zajedno, sadržavati) Evropa je kontinent koji zauzima zapadni dio većeg prostora nazvanog Evroazija. Podjela na Evropu i Aziju je više stvar tradicije i dogovora i više kulturni nego geografski pojam, jer nema prirodne granice na mnogim mjestima. Evropa (semit. ered – zapad) zauzima površinu od 10.180.000 km^2. Tu živi oko 750 miliona stanovnika ili približno 73 na jednom kvadratnom kilometru. U Evropi ima 50 suverenih, međunarodno priznatih država (članica UN). Neke od njih su i evroazijske. U tako složenoj društvenoj zajednici raznih naroda ljudi se sporazumijevaju na mnogim jezicima, a pišu latinicom, ćirilicom i grčkim pismom. Za oko 90% stanovništva maternji jezik je neki od indoevropskih jezika, a to su slavenski, romanski i germanski jezici.

Grupi slavenskih jezika pripadaju: bjeloruski, bosanski, bugarski, crnogorski, hrvatski, češki, makedonski, poljski, ruski, srpski, slovački, slovenski, ukrajinski ...

Grupi romanskih jezika pripadaju: francuski, italijanski, portugalski, rumunski, španski ...

Grupi germanskih jezika pripadaju: danski, holandski, engleski, njemački, norveški, škotski, švedski ...

Slika 11. *Okvirna karta evropskih jezika*

Mada jezici pripadaju pojedinim većim grupama, svaki od njih ima svoje specifične skupove glasova, pa i (najčešće) sopstvene skupove slova kojima se oni bilježe. Svaki jezik ima i svoju relativnu frekvenciju pojedinih slova. U naučnoj i stručnoj literaturi mogu se pronaći razni izvori podataka o frekvenciji slova u pojedinim jezicima. Podaci u slijedećem pregledu su preuzeti sa Interneta na adresi *Letter frequency* (Sa Wikipedije, besplatne enciklopedije) [8]. Na Internet adresi *How often is which letter?* (Ova stranica navodi najčešće slova na različitim jezicima Wikipedije) se nalazi još jedan obiman izvor [4]. Tu su analizirani i neevropski jezici, a o obimnosti izvora govori navedeni podatak da je za analizu samo engleskog jezika korišteno 830.180.525 znakova.

Na slici 12. prikazana je, u opadajućem poretku, frekvencija pojedinih slova za neke od evropskih jezika.

Letter	e	t	a	o	i	n	s	h	r	d	l	c	u	m	w	f	g	y	p	b	v	k	j	x	q	z
English	12.70%	9.06%	8.17%	7.51%	6.97%	6.75%	6.33%	6.09%	5.99%	4.25%	4.03%	2.78%	2.76%	2.41%	2.36%	2.23%	2.02%	1.97%	1.93%	1.49%	0.98%	0.77%	0.15%	0.15%	0.10%	0.07%

Letter	e	s	a	i	t	n	r	u	o	l	d	c	m	p	v	é	q	f	b	g	h	j	à	x	z	è	ê	y	ç	k	û	ù	â	w	î	ô	œ	ë	ï
French	14.72%	7.95%	7.64%	7.53%	7.24%	7.10%	6.69%	6.31%	5.80%	5.46%	3.67%	3.26%	2.97%	2.52%	1.84%	1.50%	1.36%	1.07%	0.90%	0.87%	0.74%	0.61%	0.49%	0.43%	0.33%	0.27%	0.22%	0.13%	0.09%	0.07%	0.06%	0.06%	0.05%	0.05%	0.05%	0.02%	0.02%	0.01%	0.01%

Letter	e	n	s	r	i	a	t	d	h	u	l	g	c	o	m	w	b	f	k	z	ü	v	p	ä	ö	ß	j	y	x	q
German	16.40%	9.78%	7.27%	7.00%	6.55%	6.52%	6.15%	5.08%	4.58%	4.17%	3.44%	3.01%	2.73%	2.59%	2.53%	1.92%	1.89%	1.66%	1.42%	1.13%	1.00%	0.85%	0.67%	0.58%	0.44%	0.31%	0.27%	0.04%	0.03%	0.02%

Letter	e	a	o	s	r	n	i	d	l	t	c	m	u	p	b	g	v	y	q	ó	í	h	f	á	j	z	é	ñ	x	ú	w	ü	k
Spanish	12.18%	11.53%	8.68%	7.98%	6.87%	6.71%	6.25%	5.01%	4.97%	4.63%	4.02%	3.16%	2.93%	2.51%	2.22%	1.77%	1.14%	1.01%	0.88%	0.83%	0.73%	0.70%	0.69%	0.50%	0.49%	0.47%	0.43%	0.31%	0.22%	0.17%	0.02%	0.01%	0.01%

Letter	a	i	e	o	n	l	s	r	t	k	j	u	d	m	p	v	g	f	b	c	ĝ	ĉ	ü	z	ŝ	h	ĵ	ĥ
Esperanto	12.12%	10.01%	9.00%	8.78%	7.96%	6.10%	6.09%	5.91%	5.28%	4.16%	3.50%	3.18%	3.04%	2.99%	2.76%	1.90%	1.17%	1.04%	0.98%	0.78%	0.69%	0.66%	0.52%	0.49%	0.39%	0.38%	0.06%	0.02%

Slika 12. *Relativne frekvencije slova u nekim evropskim jezicima*

U svim posmatranim jezicima u upotrebi su ukupno 84 slova, a to su:

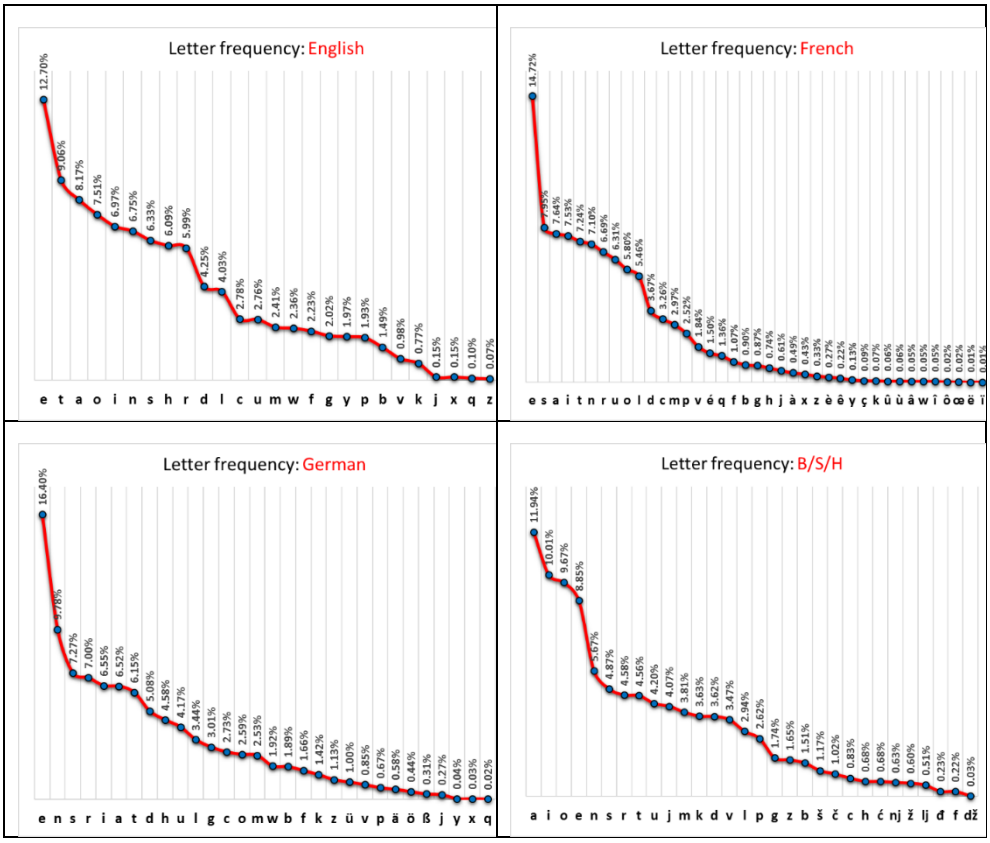
a, á, à, â, ä, ã, å, ą, æ, b, c, ĉ, ç, č, ć, d, ď, ð, e, ě, é, è, ê, ë, ę, f, g, ğ, ĝ, h, Ĥ, i, ı, í, ì, î, ï, j, ĵ, k, l, ł, m, n, ň, ń, ñ, o, ó, ò, ô, ö, õ, ø, œ, p, q, r, ř, s, ś, ŝ, ş, ß, š, t, ť, þ, u, ŭ, ú, ù, û, ü, ů, v, w, x, y, ý, z, ź, ż, ž.

Među ovim podacima, objavljenim na Internetu, nema podataka koji se odnose na bosanski, srpski i hrvatski jezik. Ta analiza je urađena u radovima: *Statistical analysis of texts of the Balkans electronic media columnists* i S*ome possibilities of computer linguistics on An example of analysis of novels* [19] [20]. Počevši od najfrekventnijeg slova naniže, redoslijed je slijedeći:

Letter	a	i	o	e	n	s	r	t	u	j	m	k	d	v	l	p	g	z	b	š	č	c	h	ć	nj	ž	lj	đ	f	dž
B/S/H	11.94%	10.01%	9.67%	8.85%	5.67%	4.87%	4.58%	4.56%	4.20%	4.07%	3.81%	3.63%	3.62%	3.47%	2.94%	2.62%	1.74%	1.65%	1.51%	1.17%	1.02%	0.83%	0.68%	0.68%	0.63%	0.60%	0.51%	0.23%	0.22%	0.03%

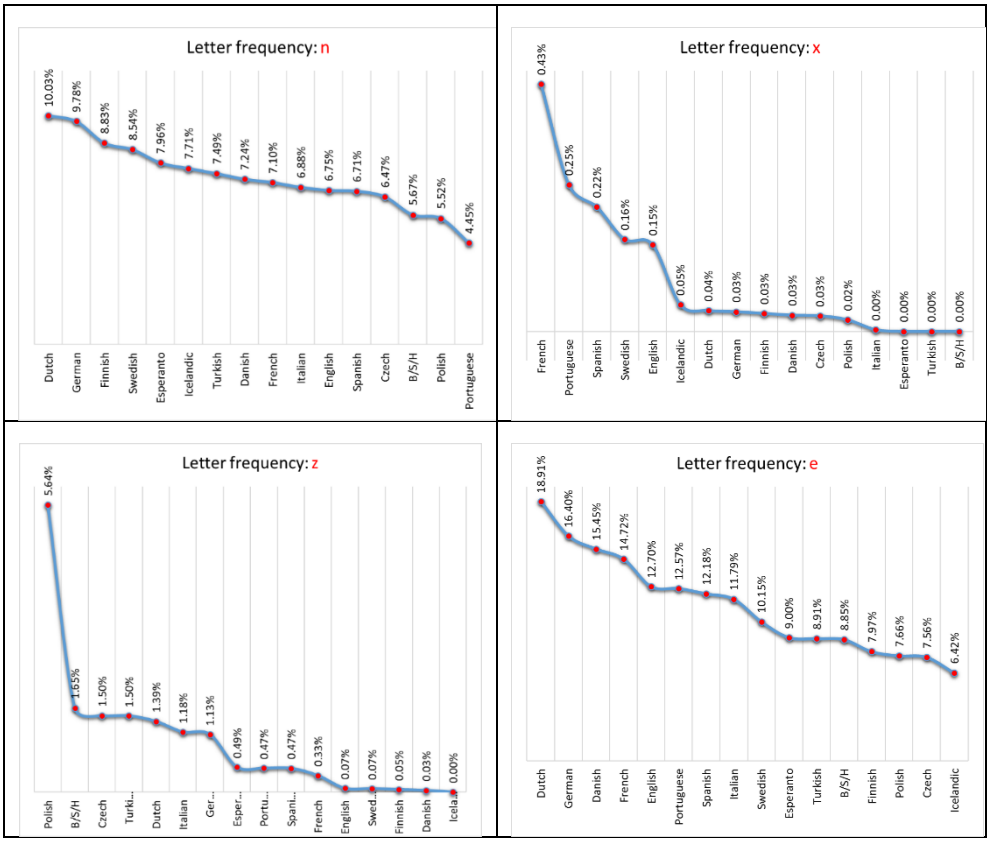
U navedenim dvjema analizama, na uzorku od preko 2 miliona slova, konstatovana je velika sličnost, gotovo istovjetnost, frekvencijske raspodjele pojedinih slova u bosanskom, srpskom i hrvatskom jeziku, pa se u ovom radu ta frekvencija označava kao frekvencija B/S/H jezika. Sa ovim podacima ostvarena je mogućnost da se statistički međusobno porede prethodno navedeni evropski jezici, ali sada i sa bosanskim, srpskim i hrvatskim jezikom (Tabela 12 i 13 i Slika 14).

Tabela 12 *Frekvencija svih slova u jednom jeziku – ilustrativni primjer*



Kompletna tabela sadrži grafikone učestalosti svih slova pojedinih jezika. Broj slova je različit, od 26 u engleskom i holandskom jeziku, do 39 u portugalskom i 41 u češkom jeziku. Sabiranjem svih učestalosti pojedinih slova dolazi se do opadajućeg niza zastupljenosti slova: **e a i n r o t s l d u m k c p g v h b f** .

Tabela 13 *Frekvencija slova (jednog slova u evropskim jezicima)*

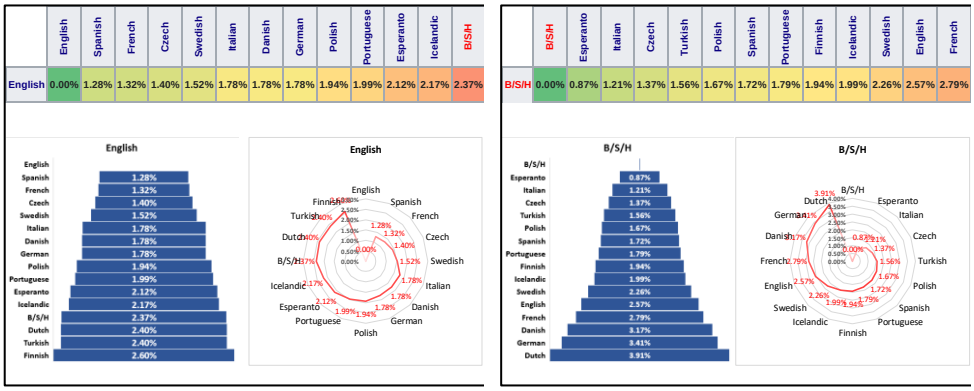


Iz prethodnih grafikona može se zaključiti da se učestalost pojave pojedinih slova u različitim jezicima mnogo razlikuje. Najveća razlika je kod slova „e“ koje u holandskom jeziku ima učestalost 18,91%, a u islandskom 6,42%. Najmanja razlika je kod slova „x“, kojeg u nekim jezicima i nema (esperanto, turski, B/S/H), a najviše ga ima u francuskom jeziku 0,43%.

Za prepoznavanje jezika kojim je pisan neki tekst, lingvisti koriste frekvencijsku analizu pojave pojedinih slova. Uočena frekvencija u relativno kratkim tekstovima može da varira, ali postaje jasno izražena njena konvergencija stvarnoj zastupljenosti u dužim tekstovima.

Za potrebe ovoga rada ispitivanje frekvencije pojedinih slova u datom tekstu urađeno je pomoću alata koji je dostupan na Internet stranici *Frequency Analysis* [13]. Poređenjem frekvencije pojedinih slova iz prethodno opisane referentne tabele i frekvencije slova datog teksta, dobijene na Internet stranici *Frequency Analysis*, može se pronaći minimalno ukupno odstupanje, što sa velikom vjerovatnoćom ukazuje na kojem jeziku je pisan dati tekst. Višestrukom provjerom na uzorcima teksta pisanog raznim jezicima, ova metoda je potvrđena. Sve analize su rađene u programu *MS Excel*.

Kada se saberu razlike učestalosti pojedinih slova, može se primijetiti da su neki jezici po tom kriterijumu međusobno bliži, a neki dalji. Slijedeći grafikoni sa procentima ukupnih razlika to jasno pokazuju.



Slika 13 *Razlike ukupne učestalosti pojedinih slova u nekim evropskim jezicima*

Iz prethodne analize može se izvesti više zaključaka. Ako gledamo pojedine jezike, tada je npr. po učestalosti slova od engleskog jezika najudaljeniji finski jezik, od njemačkog i francuskog B/S/H jezik, od poljskog holandski jezik itd. Ukupno gledano, u međusobnom poređenju je najčešće holandski jezik po učestalosti pojedinih slova najmanje sličan drugim jezicima.

6. ZAKLJUČAK

Noam Chomsky, američki lingvist i politički pisac, smatra da poznavati jezik znači biti sposoban proizvesti beskonačan broj rečenica koje nikada prije nisu izgovorene i razumjeti rečenice koje se nikada prije nisu čule. Chomsky ovu sposobnost naziva "stvaralačkim aspektom" jezika[16]. U svakodnevnoj komunikaciji u svojoj rodnoj sredini ljudi najčešće govore i čuju riječi svoga maternjeg jezika. Međutim, nisu rijetki susreti i sa riječima napisanim ili izgovorenim na nekom drugom jeziku. Takve su situacije pri slušanju stranih pjesama, gledanju stranih filmova, čitanju strane literature ili pri posjeti nekim Internet stranicama. Već na prvi pogled, ili čim se čuje nekoliko riječi, može se zaključiti da li se radi o maternjem ili o nekom drugom jeziku. Ako slušalac ili čitalac poznaje druge jezike, može zaključiti o kojem stranom jeziku se radi. Česte su situacije da jezik nije maternji, ali se ne može prepoznati koji je to jezik. Prethodno životno iskustvo može biti korisno, pa se jezik može sa povećanom vjerovatnoćom prepoznati na osnovu karakteristične melodike ili specifičnih glasova (slova). I kada se ne razumije tekst neke pjesme, teško da bi neko slušajući francusku originalnu šansonu zaključio da je otpjevana na njemačkom ili turskom jeziku.

U naučnom svijetu proučavanju pisma i govora posvećuje se velika pažnja. U ovom radu za prepoznavanje jezika iskorištene su mogućnosti kojima raspolažu informacione tehnologije. Statistički precizno na velikom uzorku analizirano je šesnaest evropskih jezika, među njima i „neživi“ esperanto jezik. Pokazalo se da je B/S/H jeziku baš esperanto jezik najsličniji po ukupnoj učestalosti pojedinih slova. Šesnaest jezika je tek kap u moru od preko 7.100 današnjih jezika u svijetu, pa bi bili veoma interesantni rezultati istraživanja u kojima bi se analizom obuhvatilo mnogo više jezika kojima se ljudi sporazumijevaju i izvan Evrope. S obzirom na veliku ugroženost mnogih jezika u današnjem svijetu i na činjenicu da neki jezici izumiru, dragocjen je svaki, pa i najmanji doprinos izučavanju jezika, a naročito jezika malih naroda. Objavljeni stručni

lingvistički radovi, štampani i elektronski, na mnogo mjesta analiziraju jezik i njegove karakteristike. Naravno, analize su najčešće za dominantne svjetske jezike: kineski, španski, engleski, hindi, arapski, bengalski, portugalski, ruski, japanski, ... [2] ali rijetke su ovakve detaljne i vizuelnim jezikom popraćene analize u koje su uključeni jezici sa područja Balkana.

7. LITERATURA

[1] Dževad Jahić, Trilogija o bosanskom jeziku, knjiga 3, Školski rječnik bosanskog jezika, Sarajevo: Ljiljan biblioteka Linguos, 1999.
[2] Ethnologue: Jezici svijeta, sedamnaesto izdanje. Dallas, Teksas: SIL International. Online verzija: http://www.ethnologue.com. (5.7.2020)
[3] http://gimnazija-sb.com/portal/wp-content/uploads/2015/02/kovacica_uregistraturi.pdf
[4] http://simia.net/letters/#explanation (5.7.2020)
[5] http://www.politika.rs/scc/authors/texts/901
[6] https://avaz.ba/tag/4975/muhamed-filipovic
[7] https://biblioteka.elektronskaknjiga.com/dervis_i_smrt.php
[8] https://en.wikipedia.org/wiki/Letter_frequency (5.7.2020)
[9] https://hr.wikipedia.org/wiki/Policentri%C4%8Dni_standardni_jezik
[10] https://klubcitalaca.files.wordpress.com/2010/12/branislav-nusic-autobiografija.pdf
[11] https://norvig.com/mayzner.html
[12] https://www.ardahan.edu.tr/CUAConference2014/ (29.7.2020)
[13] https://www.dcode.fr/frequency-analysis. (5.7.2020)
[14] https://www.jutarnji.hr/autori/miljenko-jergovic
[15] https://www.merriam-webster.com/dictionary/language (28.7.2020)
[16] https://www.sk.com.br/sk-chom.html
[17] https://www.vijesti.me/autor/miodrag-lekic
[18] https://www.worldometers.info/ (24.8.2020)
[19] Nedim Smailović, Statistical Analysis of Texts of the Balkans Electronic Media Columnists, JITA – Journal of Information Technology and Applications Banja Luka, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 9(2019) 1:5-16, (UDC: 659.3/.4:316.776]:004.738.5), (DOI: 10.7251/JIT1901005S), Volume 9, Number 1, Banja Luka, june 2019 (1-48), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004.
[20] Nedim Smailović, Zoran Ž. Avramović, Some Possibilities of Computer Linguistics on an Example of Analysis of Novels, JITA – Journal of Information Technology and Applications Banja Luka, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 10(2020) 1:5-16, (UDC: 004.82:81`322, 821.163.4`322:004.9), (DOI: 10.7251/JIT2001005S), Volume 10, Number 1, Banja Luka, June 2020 (1-68), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004.

CIP - Каталогизација у публикацији
Народна и универзитетска библиотека
Републике Српске, Бања Лука

37.018.43:004.738.5(082)(0.034.4)

МЕЂУНАРОДНИ научно-стручни скуп Информационе технологије за е-Образовање ITeO (11 ; 2019 ; Бања Лука)
Zbornik radova [Електронски извор] = Proceedings / XI međunarodni naučno-stručni skup Informacione Tehnologije za e-Obrazovanje ITeO, 6 - 7. 12. 2019. Banja Luka ; urednici Gordana Radić, Zoran Ž. Avramović. - Banja Luka : Panevropski univerzitet Apeiron, 2019 (Banja Luka : CD izdanje). - 1 elektronski optički disk (CD-ROM) : tekst ; 12 cm. - (Edicija Informacione tehnologije = Information technologies ; knj. br. 25)

Системски захтјеви нису наведени. - Насл. са насловног екрана. - Лат. и ћир. - Радови на срп. и енгл. језику. - Тираж 200. - Библиографија уз све радове. - Резимеи на енг. језику уз већину радова.

ISBN 978-99976-34-61-0

COBISS.RS-ID 8554008

SPONZORI:

PROINTER
IT SOLUTIONS AND SERVICES

teleklik.d.o.o
Usluge u informatici i telekomunikacijama

Microsoft

ISBN 978-999-76-34-13-9