

SOME POSSIBILITIES OF COMPUTER LINGUISTICS ON AN EXAMPLE OF ANALYSIS OF NOVELS

Nedim Smailović, Zoran Ž. Avramović

Pan-European University APEIRON, Banja Luka, Republic of Srpska, Bosnia and Herzegovina

nedim.i.smailovic@apeiron-edu.eu, zoran.z.avramovic@apeiron-edu.eu

Contribution to the State of the Art

<https://doi.org/10.7251/JIT2001005S>

UDC: 004.82:81`322, 821.163.4`322:004.9

Abstract: This paper shows some aspects of statistical analysis of well-known novels: *Death and the Dervish* by Meša Selimović (1966), *Autobiography* by Branislav Nušić (1924) and *In the Registrar's Office* by Ante Kovačić (1888). The goal of the analysis is to point to mutual similarities and differences of statistical data in those texts and to compare them with the up to date findings in that field. A part of the analysis relates to comparison of languages of these writers with today's language, used by column authors in electronic media. These kinds of researches belong to linguistics, as a science on language, but the results may be used in the contemporary development of artificial intelligence.

Keywords: computer linguistics, language, text analysis, visualization of data.

INTRODUCTION

Linguistics deals with language (French *linguistique*, after the Latin *lingua*: language) and it may be divided into: phonetics, phonology, morphology, syntax, semantics, stylistics and pragmatics. Linguistics is also multidisciplinary, so its specialized branches emerged in the 20th century, such as: mathematical linguistics, computer linguistics, psycholinguistics, sociolinguistics, neurolinguistics, etc.

A language may be described as a system of signs that serve for understanding among men. According to Jahic, "It is a form of human expression as a thinking being, disclosing its essence and its differentiation from other living beings." [18]. Different sources also offer other definitions of a language. Merriam Webster dictionary gives: "A language is a systemic means of communicating ideas or feelings through use of conventionalized signs, sounds, movements or markings in order to comprehend a meaning." [14]

In everyday communication we utter language automatically, without much thinking about the correctness of its use. Language is acquired and learned

from the birth throughout a lifetime, and we could say it becomes an integral part of and a significant characteristic of every individual. Nevertheless, the world of science pays due attention to studying speech and systems of writing. The research area is practically endless as relevant sources confirm that we live in a world in which over 7.7 billion people [17] communicate, speak and write in over 7.100 languages. The number is not easy to ascertain due to intertwining among languages and dialects. A language can also die out, so in the 20th century alone a total of 110 languages were proclaimed extinct. In its reports, UNESCO assesses that, unless something is done about it, by the end of this century half of languages spoken today might disappear. With disappearance of uncodified and undocumented languages, mankind would lose not only cultural wealth, but also important knowledge about ancestors that is installed, in particular, into autochthonic languages. [11]

In a contemporary world of large and swift changes in all aspects of life, language is changing as well. Some words vanish, some get new meanings,

and some completely new emerge. In such an ambient language survives, adjusts and remains the most powerful means of human communication.

STATISTICAL ANALYSIS OF TEXT IN SELECTED NOVELS

Language is alive and is changing, but there are foundations of which those changes stand. An answer to the question: *Are the foundations susceptible to changes?* may be given by modern information technology with its powerful tools and statistical analyses. Results of one such analysis are presented in this paper. The analysis encloses the following novels: *Death and the Dervish* by Meša Selimović (1910 – 1982), *Autobiography* by Branislav Nušić (1864 – 1938) and *In the Registrar's Office* by Ante Kovačić (1854 – 1899).

In quality electronic versions of these novels [6], [9], [4] text processor statistics was as follows:

Table 1. General statistics of a text processor

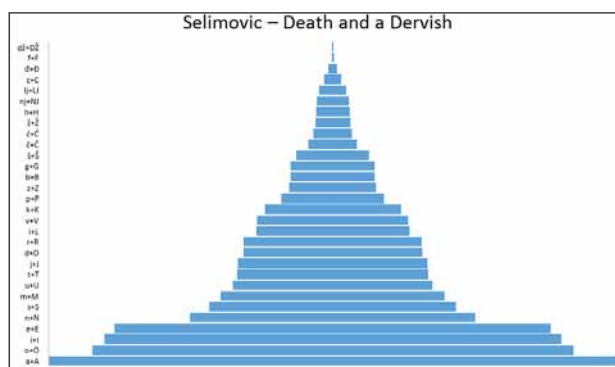
	Selimović	Nušić	Kovačić
Number of pages	248	126	210
Number of words	119,045	63,898	147,323
Number of characters (no blanks)	541,162	297,749	685,059
Number of characters (with blanks)	662,322	363,026	832,585
Number of paragraphs	3,465	1,511	2,752
Number of lines	9,435	4,924	10,521

Considering that the analyzed texts are of different length, for some analyses we took percentage of participation of certain elements, while in some other examples we were satisfied with absolute number.

Linguistics is a science that studies internal order among language units. Table 2 shows results of analysis of total number of particular letters (both capital and small) in the analyzed novels, with graphical interpretation of obtained results on Pictures 1-3. Below each graph, we have given an order of letters in that language in a descending order, left to right.

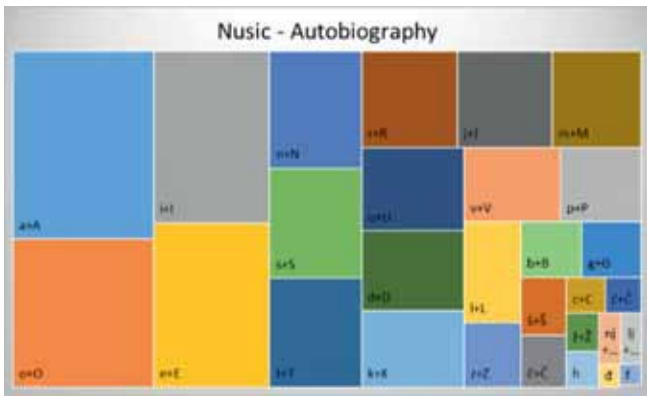
Table 2. Order of the total number of letters in novels

Selimović		Nušić		Kovačić	
Letter	Total	Letter	Total	Letter	Total
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

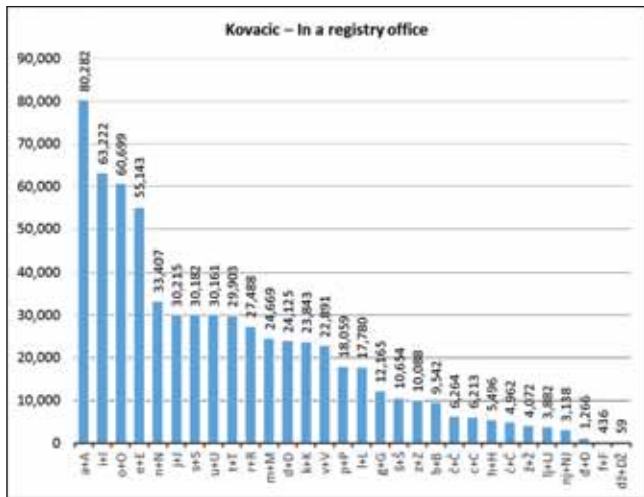


Picture 1. Number of particular letters: Selimović – Death and the Dervish

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.



Picture 2. Number of particular letters: Nušić - Autobiography
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.



Picture 3. Number of particular letters: Kovačić - In the Registrar's Office

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.

In the paper [19], on sample of contemporary texts [6], [13], [5], [16] this analysis gave very similar results. It was noticed that the order of letter presence with particular authors was as in Table 3.

Table 3. Order of letter presence with particular authors

Author	Letter order
Filipović	aioenstrjudmklvlpgzbšččcnjžljđfdž
Jergović	aioentsrjukvmdlpgzbcšchnjžljđfdž
Apostolovski	aioensrutkdvljmgpzbcščcnjžljđfdž
Lekić	aieonrstujkmvdlpgzbcčcnjshljžćfdđž

Table 4 shows unified statistical data from this research and paper [19]. Starting from the highest to lowest frequency of letter appearance, the order is as follows:

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

The column *Average* shows average percentage presence of particular letters in observed texts of all authors altogether. In other columns we gave absolute amounts of discrepancy for each author in relation to the average (average of an author minus summary average for all). Discrepancies are very small. The biggest positive discrepancy is with Lekić and a letter R (the difference is 1.04%), while the largest negative discrepancy is with Selimović and letter R (difference is -0.84%).

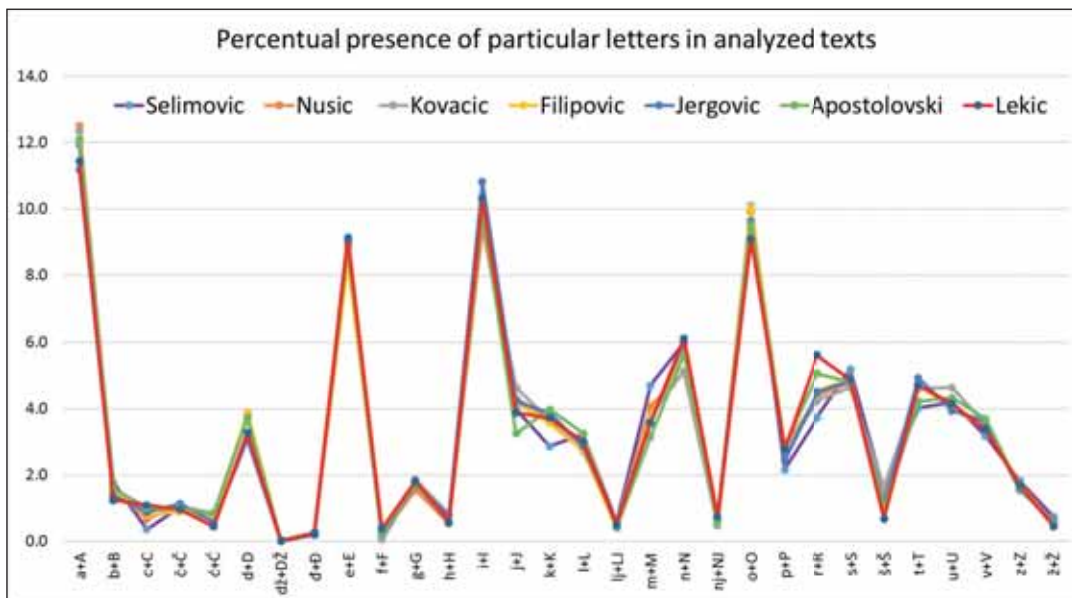
Table 4. Percentage of presence of every letter in all the texts altogether and individual discrepancies for each of the authors

Letter	Average	Filipovic	Jergovic	Apostolovski	Lekic	Selimovic	Nusic	Kovacic
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

Based on performed analysis and data presented in the above table we can draw several conclusions, and one of the most important one is that in the overall text of 2.026.250 letters almost half 931.227, or 46.10%, are the five most frequently used: **a i o e n**.

Looking from another angle, 17 letters: **v l p g z b š č h ć nj ž lj đ f dž** make just above one fifth of the text, 20.54%).

Picture 4 graphically presents previously described analysis results.



Picture 4. Graphical presentation of percentual presence of particular letters in the analyzed texts

The significance of five most frequent letters can be seen in the next example. If we take the first three sentences of the novel *Death and the Dervish*:

“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.”

And then we delete and leave blanks in the spots where letters **a i o e n** were:

“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 r j 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.”

The possibility of recognizing (comprehending) the content after that is extremely lowered, yet we are missing only five letters from the whole alphabet.

FREQUENCY OF LETTERS IN EUROPEAN LANGUAGES

Presented results relate to the languages used by the authors in the area of former Yugoslavia. However, other languages have, in most cases, other phonemes, and thus also different sets of letters used to mark them. Each language has its own relative frequency of particular letters. For some of the languages, the data is given in the following table.

Table 5. Relative frequencies of letters in some European languages (descending order according to the English alphabet)

Letter	English	French	German	Spanish	Esperanto	Italian	Turkish	Swedish	Polish	Danish	Czech
e	12.702%	14.715%	16.396%	12.181%	8.995%	11.792%	9.912%	10.149%	7.352%	15.453%	7.562%
t	9.056%	7.244%	6.154%	4.632%	5.276%	5.623%	3.314%	7.691%	2.475%	6.862%	5.727%
a	8.167%	7.636%	6.516%	11.525%	12.117%	11.745%	12.920%	9.383%	10.503%	6.025%	8.421%
o	7.507%	5.796%	2.594%	8.683%	8.779%	9.832%	2.976%	4.482%	6.667%	4.636%	6.695%
i	6.966%	7.529%	6.550%	6.247%	10.012%	10.143%	9.600%*	5.817%	8.328%	6.000%	6.073%
n	6.749%	7.095%	9.776%	6.712%	7.955%	6.883%	7.987%	8.542%	6.237%	7.240%	6.468%
s	6.327%	7.948%	7.270%	7.977%	6.092%	4.981%	3.014%	6.590%	5.224%	5.805%	5.212%
h	6.094%	0.737%	4.577%	0.703%	0.384%	0.636%	1.212%	2.090%	1.015%	1.621%	1.356%
r	5.987%	6.693%	7.003%	6.871%	5.914%	6.367%	7.722%	8.431%	5.243%	8.956%	4.799%
d	4.253%	3.669%	5.076%	5.010%	3.044%	3.736%	5.206%	4.702%	3.725%	5.858%	3.475%
l	4.025%	5.456%	3.437%	4.967%	6.104%	6.510%	5.922%	5.275%	2.564%	5.229%	3.802%
c	2.782%	3.260%	2.732%	4.019%	0.776%	4.501%	1.463%	1.486%	3.895%	0.565%	0.740%
u	2.758%	6.311%	4.166%	2.927%	3.183%	3.011%	3.235%	1.919%	2.062%	1.979%	2.160%
m	2.406%	2.968%	2.534%	3.157%	2.994%	2.512%	3.752%	3.471%	2.515%	3.237%	2.446%
w	2.360%	0.049%	1.921%	0.017%	0	0.033%	0	0.142%	5.813%	0.069%	0.016%
f	2.228%	1.066%	1.656%	0.692%	1.037%	1.153%	0.461%	2.027%	0.143%	2.406%	0.084%
g	2.015%	0.866%	3.009%	1.768%	1.171%	1.644%	1.253%	2.862%	1.731%	4.077%	0.092%
y	1.974%	0.128%	0.039%	1.008%	0	0.020%	3.336%	0.708%	3.206%	0.698%	1.043%
p	1.929%	2.521%	0.670%	2.510%	2.755%	3.056%	0.886%	1.839%	2.445%	1.756%	1.906%
b	1.492%	0.901%	1.886%	2.215%	0.980%	0.927%	2.844%	1.535%	1.740%	2.000%	0.822%
v	0.978%	1.838%	0.846%	1.138%	1.904%	2.097%	0.959%	2.415%	0.012%	2.332%	5.344%
k	0.772%	0.074%	1.417%	0.011%	4.163%	0.009%	5.683%	3.140%	2.753%	3.395%	2.894%
j	0.153%	0.613%	0.268%	0.493%	3.501%	0.011%	0.034%	0.614%	1.836%	0.730%	1.433%
x	0.150%	0.427%	0.034%	0.215%	0	0.003%	0	0.159%	0.004%	0.028%	0.027%
q	0.095%	1.362%	0.018%	0.877%	0	0.505%	0	0.020%	0	0.007%	0.001%
z	0.074%	0.326%	1.134%	0.467%	0.494%	1.181%	1.500%	0.070%	4.852%	0.034%	1.503%

The source table [8] contains also data on frequency of other letters used in each of the listed 11 languages.

The observed languages have 84 letters in use, and those are:

a, á, à, â, ä, ã, å, ą, æ, b, c, ç, ċ, ć, d, d', ð, e, ë, é, è, ê, ì, ï, j, k, l, ł, m, n, ñ, ó, ò, ô, õ, ø, œ, p, q, r, ř, s, ś, ş, ß, š, t, ť, þ, u, ů, ú, ù, û, ü, v, w, x, y, ý, z, ź, ž, ž.

When we show the frequency of these letters for each language in a descending order, we get the following table.

Table 6. Eleven languages and sorted order of letters used in them

Letter	English	Letter	French	Letter	German	Letter	Spanish	Letter	Esperanto	Letter	Italian	Letter	Turkish	Letter	Swedish	Letter	Polish	Letter	Danish	Letter	Czech
e	12.70%	e	14.72%	e	16.40%	e	12.18%	a	12.12%	e	11.79%	a	11.92%	e	10.15%	a	10.50%	e	15.45%	a	8.42%
t	9.06%	s	7.95%	n	7.27%	a	11.53%	i	10.01%	a	11.75%	e	8.91%	a	9.38%	i	8.33%	r	8.96%	e	7.56%
a	8.17%	a	7.64%	s	7.27%	o	8.68%	e	9.00%	i	10.14%	i	8.60%	n	8.54%	e	7.35%	n	7.24%	o	6.70%
o	7.51%	i	7.53%	r	7.00%	s	7.98%	o	8.78%	o	9.83%	n	7.49%	r	8.43%	o	6.67%	t	6.86%	n	6.47%
i	6.97%	t	7.24%	i	6.55%	r	6.87%	n	7.96%	n	6.88%	r	6.72%	t	7.69%	n	6.24%	a	6.03%	i	6.07%
n	6.75%	n	7.10%	a	6.52%	n	6.71%	l	6.10%	l	6.51%	l	5.92%	s	6.59%	w	5.81%	i	6.00%	t	5.73%
s	6.33%	r	6.69%	t	6.15%	i	6.25%	s	6.09%	r	6.37%	i	5.11%	i	5.82%	r	5.24%	d	5.86%	v	5.34%
h	6.09%	u	6.31%	d	5.08%	d	5.01%	r	5.91%	t	5.62%	d	4.71%	l	5.28%	s	5.22%	s	5.81%	s	5.21%
r	5.99%	o	5.80%	h	4.58%	l	4.97%	t	5.28%	s	4.98%	k	4.68%	d	4.70%	z	4.85%	l	5.23%	r	4.80%
d	4.25%	l	5.46%	u	4.17%	t	4.63%	k	4.16%	c	4.50%	m	3.75%	o	4.48%	c	3.90%	o	4.64%	l	3.80%
l	4.03%	d	3.67%	l	3.44%	c	4.02%	j	3.50%	d	3.74%	y	3.34%	m	3.47%	d	3.73%	g	4.08%	d	3.48%
c	2.78%	c	3.26%	g	3.01%	m	3.16%	u	3.18%	p	3.06%	t	3.31%	k	3.14%	y	3.21%	k	3.40%	k	2.89%
u	2.76%	m	2.97%	c	2.73%	u	2.93%	d	3.04%	u	3.01%	u	3.24%	g	2.86%	k	2.75%	m	3.24%	m	2.45%
m	2.41%	p	2.52%	o	2.59%	p	2.51%	m	2.99%	m	2.51%	s	3.01%	v	2.42%	l	2.56%	f	2.41%	u	2.16%
w	2.36%	v	1.84%	m	2.53%	b	2.22%	p	2.76%	v	2.10%	b	2.84%	h	2.09%	m	2.52%	v	2.33%	p	1.91%
f	2.23%	é	1.50%	w	1.92%	g	1.77%	v	1.90%	g	1.64%	o	2.48%	f	2.03%	t	2.48%	b	2.00%	í	1.64%
g	2.02%	q	1.36%	b	1.89%	v	1.14%	g	1.17%	z	1.18%	ü	1.85%	u	1.92%	p	2.45%	u	1.98%	z	1.50%
y	1.97%	f	1.07%	f	1.66%	y	1.01%	f	1.04%	f	1.15%	ş	1.78%	p	1.84%	ł	2.11%	p	1.76%	j	1.43%
p	1.93%	b	0.90%	k	1.42%	q	0.88%	b	0.98%	b	0.93%	z	1.50%	ä	1.80%	u	2.06%	h	1.62%	h	1.36%
b	1.49%	g	0.87%	z	1.13%	ó	0.83%	c	0.78%	h	0.64%	g	1.25%	b	1.54%	j	1.84%	ä	1.19%	ě	1.22%
v	0.98%	h	0.74%	ü	1.00%	í	0.73%	ğ	0.69%	à	0.64%	h	1.21%	c	1.49%	b	1.74%	ø	0.94%	y	1.04%
k	0.77%	j	0.61%	v	0.85%	h	0.70%	č	0.66%	q	0.51%	ç	1.16%	â	1.34%	g	1.73%	æ	0.87%	ý	1.00%
j	0.15%	à	0.49%	p	0.67%	f	0.69%	ž	0.52%	è	0.26%	ğ	1.13%	ö	1.31%	ó	1.14%	j	0.73%	á	0.87%
x	0.15%	x	0.43%	ä	0.58%	á	0.50%	z	0.49%	ú	0.17%	c	0.96%	y	0.71%	ę	1.04%	y	0.70%	b	0.82%
q	0.10%	z	0.33%	ö	0.44%	j	0.49%	š	0.39%	w	0.03%	v	0.96%	j	0.61%	h	1.02%	c	0.57%	c	0.74%
z	0.07%	è	0.27%	ß	0.31%	z	0.47%	h	0.38%	í	0.03%	p	0.89%	x	0.16%	ś	0.81%	w	0.07%	ž	0.72%
		ê	0.22%	j	0.27%	é	0.43%	ĵ	0.06%	y	0.02%	ö	0.78%	w	0.14%	ć	0.74%	z	0.03%	š	0.69%
		y	0.13%	y	0.04%	ñ	0.31%	ĥ	0.02%	j	0.01%	f	0.46%	z	0.07%	ź	0.71%	x	0.03%	é	0.63%
		ç	0.09%	x	0.03%	x	0.22%			k	0.01%	j	0.03%	q	0.02%	ą	0.70%	q	0.01%	č	0.46%
		k	0.07%	q	0.02%	ú	0.17%			x	0.00%					ń	0.36%			ř	0.38%
		ü	0.06%			w	0.02%									f	0.14%			ů	0.20%
		ù	0.06%			ü	0.01%									ź	0.08%			g	0.09%
		â	0.05%			k	0.01%									v	0.01%			f	0.08%
		w	0.05%																	ú	0.05%
		î	0.05%																	x	0.03%
		ó	0.02%																	ó	0.02%
		œ	0.02%																	w	0.02%
		ë	0.01%																	ď	0.02%
		ï	0.01%																	ň	0.01%
																				ť	0.01%

For easier remembering of order of letter frequency, bibliography offers the first twelve, most frequent letters, presented in the form of two (non-existing) six-letter words. For the eleven languages of the above table, where words are shown in Table 7.

Table 7. The most frequent letters in languages

Language	Most frequent letters
English	etaoin shrdlc
French	esaitn ruoldc
German	ensria tdhulg
Spanish	eaosrn idltcm
Esperanto	aieonl srtkju
Italian	eaionl rtscdp
Turkish	aeinrl idkmyt
Swedish	eanrts ildomk
Polish	aieonw rszcdy
Danish	erntai dslogk
Czech	aeonit vsrldk

For joined text of novels from this and mentioned paper [19], which treats texts of columnists, the two words would be:

aioens rtujmk

N-GRAM WORD ANALYSIS

In computer linguistics and statistics, there is a term of n-gram. It is a continued line of n elements in the observed sample. The elements may be consecutive letters, words or selected sets of marks to be analyzed. N-grams are put in tables for the comparing of a number of repetitions of the same linguistic elements in different languages. N-grams are studied in several areas of informatics, computer linguistics and applied mathematics.

Each author has its own style in writing and some word patterns are more often found, some seldomly or not at all.

Tables 8-12 contain the most frequent bi-grams, three-grams, the most frequent words, nouns and verbs in the analyzed novels by Selimović, Nušić and Kovačić.

The most frequent word three-gram with Selimović is *kao da je* (51 times), with Nušić *Tako na primer* (32 times), and with Kovačić *kano da je* (68 times).

Table 8. Number of repeated three-grams of words

Three-gram of words	SELIMOVIC	Three-gram of words	NUSIC	Three-gram of words	KOVACIC
kao da je	51	Tako na primer	32	kano da je	68
mi se da	39	da je to	25	kano da se	40
da se ne	36	Razume se da	24	si ga neka	34
sam da je	33	da bi se	21	kano da si	33
ono što je	33	šta je to	20	Vrag si ga	32
kao da sam	31	da sam ja	19	da mu je	28
da je to	27	kao što je	16	i kano da	26
mi je bilo	25	na taj način	14	kano da će	22
Htio sam da	25	ja sam se	14	kano da mu	19
zato što je	25	ne bi li	13	da mi je	19
kao da se	24	docnije u životu	12	Ha ha ha	18
znao sam da	23	da sam se	12	kano da ga	18
je u meni	23	Ja ne znam	12	da se ne	17
nisam mogao da	23	Ono se smejalalo	11	da je u	17
ono što sam	22	sam se u	11	i da se	17
ne mogu da	22	s tim da	11	se kano da	17
i ono što	21	ne može da	11	da mu se	16
Mislio sam da	21	sam se ja	10	da ga je	16
sve što je	21	ima ih koji	10	li je to	15
a ja sam	20	u tom pogledu	9	je sve to	14
što mi je	20	ne mogu da	9	ne bi li	14

Table 9. The most frequent words in text

The most frequent word	SELIMOVIC	The most frequent word	NUSIC	The most frequent word	KOVACIC
je	4,933	i	2,550	i	7,343
i	4,505	je	1,969	se	3,829
da	4,452	da	1,886	u	2,935
se	3,067	se	1,559	je	2,853
sam	2,753	u	1,334	da	2,697
u	2,239	na	900	a	1,931
ne	1,903	sam	808	na	1,854
što	1,439	to	629	ne	1,462
na	1,250	a	583	ja	866
a	1,218	ne	553	ti	844
me	975	ja	540	to	823
nije	964	bi	488	za	820
kao	938	kao	431	te	810
su	901	su	406	pa	804
to	865	pa	356	što	790
mi	837	koji	351	mi	762
ga	826	mi	350	tako	700
ali	728	tako	349	ga	649
bi	688	za	339	će	634
od	677	sa	336	sve	602
ni	677	od	334	od	596
sve	668	kad	319	mu	579
nisam	577	ali	308	li	573
za	569	te	244	on	566
kad	540	iz	240	kano	563
mu	474	nije	232	kako	562
ili	465	već	218	s	537
o	433	smo	214	iz	534
li	430	sve	200	bi	523
s	418	ti	195	sam	509

Table 10. Number of repeated bi-grams of words

bi-gram of words	SELIMOVIC	bi-gram of words	NUSIC	bi-gram of words	KOVACIC
da se	512	da je	216	kano da	358
da je	424	da se	181	da je	349
mi je	320	da sam	108	da se	325
što je	299	što je	108	mi se	141
sam se	277	bi se	104	da će	121
sam da	247	na primer	103	se i	114
kao da	204	sam se	101	se u	114
da ga	193	je to	94	i ne	113
ono što	176	koji je	91	mu se	112
da sam	157	u životu	84	je to	112
što se	155	mi je	81	da mu	99
što sam	152	da bi	78	što je	96
mu je	149	ja sam	77	mali kanonik	92
da me	146	sam ja	77	mi je	92
se ne	141	se u	73	je i	90
je to	137	to je	65	i da	90
i da	135	i da	62	da mi	86
je da	134	nam je	60	da ga	84
je u	129	kao što	58	se ne	83
je i	128	da su	55	ne bi	79
se da	128	kao da	55	te se	77
da ne	127	i u	55	a onda	73
mi se	117	ne bi	54	kako je	72
mogao da	116	je u	53	mu je	72
ne bi	116	se da	53	kao da	71
su se	109	me je	49	ti je	71
je bilo	105	je i	48	da si	70
to je	105	što sam	47	se na	69
da mu	103	koja je	47	sve to	67
i ne	100	što se	45	je u	66

Table 11. The most frequent nouns in the text

The most frequent noun	SELIMOVIC	The most frequent noun	NUSIC	The most frequent noun	KOVACIC
čovjek	201	primer	100	kumordinar	233
ljudi	190	životu	95	lvica	186
Hasan	155	čovjek	86	otac	176
riječi	114	reči	82	kanonik	172
misao	109	dana	82	Miha	170
život	83	profesor	76	oči	167
oči	83	život	65	jožica	154
vrijeme	82	reč	63	ruke	148
čovjeka	79	ljubav	59	bog	139
riječ	78	stvar	51	glavu	137
noć	77	godina	41	Laura	129

Table 12. The most frequent verbs in the text

The most frequent verb	SELIMOVIC	The most frequent verb	NUSIC	The most frequent verb	KOVACIC
bilo	411	bio	136	bijaše	334
bio	379	ima	117	stade	144
rekao	300	bilo	114	biti	136
mogao	262	bila	108	vidiš	123
znao	233	može	106	uhvati	94
može	195	biti	97	budi	91
mislio	190	razume	76	mogao	76
treba	157	mogao	63	ima	76
bila	149	imao	60	znam	70
mogu	144	mora	59	skoči	60
biti	142	treba	59	može	58
bude	140	znam	49	znao	56

Table 13. The most frequent words according to number of letters

Selimovic

	1 letter	Broj	2 letters	Broj	3 letters	Broj	4 letters	Broj	5 letters	Broj	6 letters	Broj	7 letters	Broj	8 letters	Broj	9 letters	Broj	10 letters	Broj
1	i	4,505	je	4,933	sam	2,753	nije	964	nisam	577	čovjek	201	sigurno	111	svejedno	99	zaboravio	51	neprestano	35
2	u	2,239	da	4,452	što	1,439	bilo	411	možda	318	mislio	190	govorio	109	razgovor	60	izgledalo	30	mula-jusuf	31
3	a	1,218	se	3,067	kao	938	samo	403	rekao	300	toliko	128	trebalo	104	drukčije	45	prijatelj	27	oslušivao	14
4	o	433	ne	1,903	ali	728	zbog	399	ništa	272	uvijek	126	poslije	85	pogledao	45	gledajući	24	moogućnosti	13
5	s	418	na	1,250	sve	668	kako	361	mogao	262	riječ	114	vrijeme	82	očekivao	37	odgovorio	21	pogriješio	12
6	e	8	me	975	kad	540	više	306	zašto	211	učinio	98	čovjeka	79	odjednom	34	prijatelji	21	razgovaram	11
7			su	901	ili	465	mene	293	nešto	174	njegova	97	osjećao	75	prijatelj	33	uznemiren	16	razgovarao	11
8			to	865	šta	413	koji	290	treba	157	koliko	93	izgleda	67	trenutak	32	zadovoljan	16	zaboravili	11
9			mi	837	bio	379	tako	288	hasan	155	desilo	92	siguran	51	potrebno	27	razgovora	15	zaustaviti	10
10			ga	826	bih	343	znam	278	vidio	127	gledao	88	svijetu	49	zatvoren	27	sigurnost	15	osjećajući	10
11			bi	688	bez	341	njega	260	prema	122	zaista	76	učiniti	46	vjerovao	27	pokušavao	15	uznemireno	10
12			od	677	ako	336	znao	233	dobro	119	jednom	73	ponekad	44	vjerujem	27	prijatelj	15	primijetio	10
13			ni	677	jer	323	onda	214	mnogo	115	svijet	70	razloga	44	slučajno	25	Sinanudin	15	neprijatelj	10
14			za	569	sad	270	meni	208	nikad	111	suviše	69	vremena	40	nasmijao	25	nesiguran	14	sigurnosti	9
15			mu	474	smo	233	može	195	svoju	109	uzalud	65	osjetio	40	riječima	22	zaustavio	14	zaboraviti	8
16			li	430	još	229	pred	193	misao	109	njegovo	65	svijeta	39	govoriti	21	moogućnost	13	zavorenih	8
17			ja	395	ono	217	ljudi	190	teško	105	otišao	63	sjećanje	38	potrebna	20	opasnosti	13	ravnodušan	8
18			će	329	već	189	zato	185	svoje	99	ljudima	60	govorim	38	nemoguće	19	izmijenio	13	pokazujući	8
19			on	329	nas	150	sebe	177	jedan	99	sasvim	57	Muhamed	34	naročito	19	unaprijed	13	vrijednost	7
20			ti	309	dok	120	htio	157	kažem	95	između	57	postoji	33	dvadeset	18	razgovoru	12	oduševljenje	7

The longest word with Selimović: neprikosnovenošću – 17 letters.

Nusic

	1 letter	Broj	2 letters	Broj	3 letters	Broj	4 letters	Broj	5 letters	Broj	6 letters	Broj	7 letters	Broj	8 letters	Broj	9 letters	Broj	10 letters	Broj
1	i	2,550	je	1,969	sam	808	koji	351	jedan	129	primer	100	jednoga	59	profesor	76	profesora	29	matematika	19
2	u	1,334	da	1,886	što	501	tako	349	nisam	112	životu	95	docnije	42	neobično	49	verovatno	27	roditeljima	7
3	a	583	se	1,559	kao	431	nije	232	čovjek	86	razume	76	izgleda	33	prilikom	40	gimnazije	25	upotrebiti	7
4	o	192	na	900	kad	319	koje	186	posle	85	odista	71	misli	33	nekoliko	36	gospodine	24	geografije	6
5	s	66	to	629	ali	308	samo	179	svoje	84	koliko	67	dovoljno	28	ministar	32	otprilike	23	pozdraviti	6
6	e	33	ne	553	već	218	kako	171	život	65	toliko	55	predmet	24	naročito	31	pantalone	23	biografiju	5
7			ja	540	smo	214	koja	142	mogao	63	jednom	51	lekciju	22	uostalom	29	profesori	22	francuskog	5
8			bi	488	sve	200	onda	141	prema	63	najzad	51	mišljenje	21	dvadeset	25	zamislite	19	neprestano	5
9			su	406	jer	166	vrlo	138	jedna	61	nekako	42	međutim	18	jedanput	24	zaboravio	15	podjednako	5
10			pa	356	ili	151	bilo	114	zanim	60	godina	41	potpuno	18	odgovara	22	latinskog	15	odgovaramo	5
11			mi	350	još	151	bila	108	treba	59	stvari	41	zapumem	18	razgovor	18	pantalona	12	policijski	5
12			za	339	ona	143	može	106	ljubav	59	ljubavi	41	izvesno	17	direktor	17	nepoznate	12	generacije	5
13			od	334	ako	131	biti	97	mnogo	54	četiri	37	razreda	17	istorije	16	finansija	7	količnima	5
14			te	244	taj	131	toga	95	dobro	54	sasvim	36	slučaju	17	pogdekad	15	pokušavao	7	zajednički	5
15			iz	240	šta	122	mene	93	jedno	53	razred	36	zajedno	16	ministra	15	odgovorio	7	latinskoga	5
16			ti	195	ima	117	više	92	jednu	53	života	34	trebalo	16	latinski	14	činovnika	6	praktikant	5
17			mu	183	nam	107	reći	82	možda	53	čovjeka	32	odgovor	16	količina	14	stihovima	6	književnika	4
18			po	172	kod	95	dana	82	ništa	51	između	32	smejalo	15	drukčije	13	zoologije	6	biografija	4
19			on	170	vam	94	tome	75	stvar	51	recimo	31	gospođa	15	mladosti	13	profesija	6	odgovorila	4
20			me	165	ono	87	pred	72	nešto	48	kojima	29	mišljenja	15	priznaje	13	izmislili	6	strahovito	4

The longest word with Nušić: šezdesetogodišnjicu and devetnaestogodišnji – 18 letters.

Kovacic

	1 letter	Broj	2 letters	Broj	3 letters	Broj	4 letters	Broj	5 letters	Broj	6 letters	Broj	7 letters	Broj	8 letters	Broj	9 letters	Broj	10 letters	Broj
1	i	7,343	se	3,829	što	790	tako	700	svoje	305	bijaše	334	kanonik	172	gospodin	114	gospodine	57	kumordinar	233
2	u	2,935	je	2,853	sve	602	kano	563	poput	254	nikada	172	muzikaš	110	iznenada	64	gospodina	51	dobrotvora	16
3	a	1,931	da	2,697	sam	509	kako	562	nešto	195	Jožica	154	nijesam	90	kanonika	59	milostivi	41	gospodstvo	15
4	s	537	na	1,854	ali	497	nije	482	ivica	186	svojim	139	svijetu	87	nekoliko	54	zgbudane	32	rajhercera	15
5	o	326	ne	1,462	već	447	sada	382	svoga	180	svijet	103	ferkonja	87	djevojka	51	providnik	26	djevojčica	13
6	k	204	ja	866	moj	383	samo	312	svoju	177	uhvati	94	gospoda	64	muzikaša	46	dobrotvor	26	poniznosti	13
7	e	71	ti	844	ona	370	kada	297	danas	157	lijepo	88	napokon	62	naprijed	40	razumijem	21	trepetljika	12
8	nj	20	to	823	još	331	opet	291	ništa	155	rukama	88	Medonić	59	najprije	37	gospodara	21	strastveno	12
9			za	820	šta	294	onda	247	stade	144	našega	78	vrijeme	58	obrazima	37	ferkonjina	21	kažiprstom	11
10			te	810	ili	266	koji	236	glavu	137	daleko	73	poslije	54	najednom	33	Margarita	21	trepetljike	10
11			pa	804	joj	260	moja	235	Laura	129	riječ	69	starica	45	gospodar	32	gospodara	20	strastveno	12
12			mi	762	taj	216	neka	234	vidiš	123	negoli	69	svijeta	44	medonića	30	zadovoljno	19	kažiprstom	11
13			ga	649	kao	204	mene	191	ovdje	123	njegova	67	stadoše	43	ramenima	29	uhvatiši	19	trepetljike	10
14			će	634	kad	202	više	187	vrata	123	Jožice	67	odijelo	37	djevojče	29	drugачije	19	neprestano	9
15			od	596	baš	180	gdje	183	dobro	120	čovjek	66	vremena	36	dječarac	26	gospodaru	19	Kičmanović	9
16			mu	579	oko	176	otac	176	prije	111	jednom	62	zajedno	36	polagano	24	Zorkovića	16	kojekavih	9
17			li	573	dok	169	Miha	170	preko	109	veliki	60	svakoga	36	djevojke	24	Medonićev	16	providnika	9
18			on	566	oči	167	svoj	168	pravo	101	njegove	59	drugoga	35	zgbuidan	21	porugljivo	14	pogledavši	9
19			iz	534	tko	156	pred	163	drugi	97	nekuda	59	osobito	33	desnicom	21	plemeniti	14	družinskoj	9
20			bi	523	oni	155	mali	160	ivice	92	natrag	59	vratima	33	rodakinja	20	gospodske	14	djevojčicu	9

The longest word with Kovačić: pedagoški kod daktičkih – 22 letters.

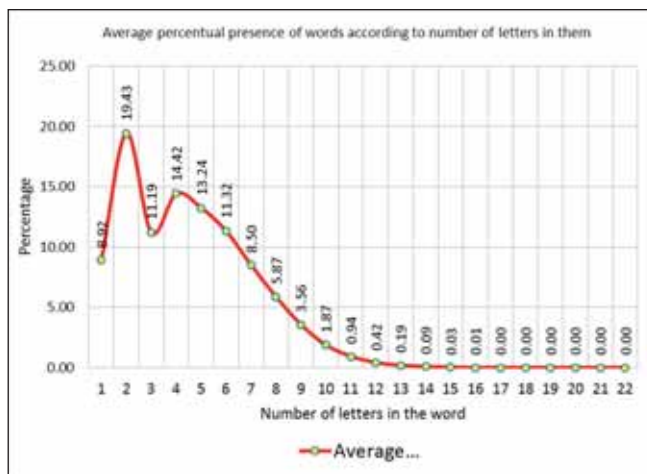
Observing individual words, it is noticeable that with all three authors there are no nouns, adjectives and numbers among the most frequent words. The dominant ones are conjunctions, auxiliary verbs, prepositions (non-lexical, function words).

With Selimović, the most frequent noun is čovjek (**man**), and it is found only at the 70th place of used words. With Nušić it is a noun primer (**example**), at the 64th place, and with Kovačić **kumordinar** at 63rd place. It is similar with the use of verbs.

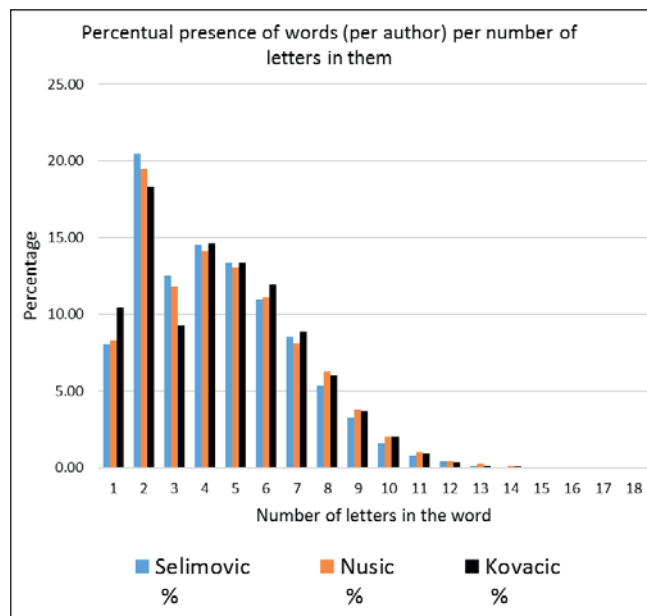
Considering that analyzed novels are of different size in terms of number of paragraphs, sentences, words and characters, the frequency of words per number of characters in them is best compared in percentages. The results are given in Table 14 and Pictures 5 and 6.

Table 14. Absolute and percentual presence of words according to number of letters in them

Number of letters in the word	Selimovic total	Selimovic %	Nusic total	Nusic %	Kovacic total	Kovacic %	Prosjek %
1	8,712	8.04	4,732	8.30	13,198	10.41	8.92
2	22,184	20.48	11,116	19.49	23,216	18.31	19.43
3	13,557	12.52	6,731	11.80	11,716	9.24	11.19
4	15,687	14.48	8,060	14.13	18,566	14.64	14.42
5	14,426	13.32	7,456	13.08	16,906	13.33	13.24
6	11,900	10.99	6,307	11.06	15,097	11.91	11.32
7	9,227	8.52	4,629	8.12	11,253	8.87	8.50
8	5,759	5.32	3,584	6.29	7,625	6.01	5.87
9	3,503	3.23	2,155	3.78	4,663	3.68	3.56
10	1,715	1.58	1,150	2.02	2,557	2.02	1.87
11	912	0.84	590	1.03	1,186	0.94	0.94
12	483	0.45	254	0.45	458	0.36	0.42
13	167	0.15	138	0.24	225	0.18	0.19
14	45	0.04	85	0.15	95	0.07	0.09
15	13	0.01	22	0.04	33	0.03	0.03
16	9	0.01	8	0.01	9	0.01	0.01
17	1	0.00	4	0.01	0	0.00	0.00
18	0	0.00	2	0.00	4	0.00	0.00
19	0	0.00	0	0.00	0	0.00	0.00
20	0	0.00	0	0.00	0	0.00	0.00
21	0	0.00	1	0.00	1	0.00	0.00
22	0	0.00	0	0.00	1	0.00	0.00



Picture 5. Average percentual presence of words according to number of letters in them



Picture 6. Percentual presence of words (per author) per number of letters in them

Each author has its own style of writing, characterized with different features. One of them is the length of used words. On the presented charts, we can see the similarity of distribution of percent of single-letter words, words with two letters, three letters, etc., with all three authors.

In the analyzed novels, we could perform a separate analysis of lexical density to measure how informative the text is. Lexical density is defined as number of lexical words (or content words) divided by total number of words. Lexical words give meaning to a text. Those are nouns, adjectives, verbs and adverbs. Other types of (function) words, such as

auxiliary verbs, prepositions, conjuncts, are more of grammatical nature and give little or none information about what is going on in a text.

CONCLUSION

Encyclopedia Britannica on language, inter alia, gives the following: "The most important individual characteristic of human language (i.e. each individual language) in relation to all other known ways of communication among animals, is its infinite productivity and creativity." [12]

Noam Chomsky, an American linguist and political author, believes that to know a language means to be capable to produce an infinite number of sentences that have never been uttered before and understand sentences that have never been heard before. Chomsky calls this feature a "creationist aspect" of language. [15]

Published expert linguistic works, printed and electronic, have analyzed language and its characteristics in many places. Naturally, analyses have most often been done for dominant world languages, Chinese, Spanish, English, Hindi, Arab, Bengali, Portuguese, Russian, Japanese [3], but works like this are rare, detailed and visually illustrated analyses for languages in the area of former Yugoslavia.

In this paper, we statistically analyzed the language of three famous writers, Mesa Selimović, Branislav Nušić and Ante Kovačić, used in their best-known novels. These authors originate from different geographical areas (Bosnia and Herzegovina, Croatia, Serbia). Novels differ in content (psychological and philosophical novel, a comedy with autobiographical content, a novel from the time of realism). The time span between publishing of the novels is almost eighty years. Nevertheless, computer analysis showed great similarities in certain elements of text.

Contemporary informatics and its branch, computer linguistics, enable analyses that were unthinkable until recently, and it will probably influence the creation of new branches of linguistics and bring about new insights of language. Considering large endangerment of languages in present world and the fact that many languages are dying out, every contribution, even the smallest one, to the study of language, particularly language of small nations, is precious.

REFERENCES

- [1] Avramović Zoran Ž., Marinković Dražen, Lastrić Igor, Use of Computer Search Algorithms in the Research of Statistical, Semantic and Contextual Rules of Language in Digital Information Space, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropean University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 7(2017) 1:5-8, Volume 7, Number 1, Banja Luka, June 2017 (1-56), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004
- [2] Čubelić, T. *Književni leksikon*. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu, 1972.
- [3] Ethnologue: *Jezici svijeta*, sedamnaesto izdanje. Dallas, Teksas: SIL International. Online verzija: <http://www.ethnologue.com>. (24.10.2019)
- [4] http://gimnazija-sb.com/portal/wp-content/uploads/2015/02/Kovačića_uregistraturi.pdf (23.9.2019)
- [5] <http://www.politika.rs/scc/authors/texts/901> (25.3.2019)
- [6] <https://avaz.ba/tag/4975/muhamed-filipovic> (28.3.2019)
- [7] https://biblioteka.elektronskaknjiga.com/dervis_i_smrt.php (23.9.2019)
- [8] https://en.wikipedia.org/wiki/Letter_frequency (26.10.2019)
- [9] <https://klubcitalaca.files.wordpress.com/2010/12/branislav-Nušić-autobiografija.pdf> (23.9.2019)
- [10] <https://norvig.com/mayzner.html> (27.4.2019)
- [11] <https://www.ardahan.edu.tr/CUACConference2014/> (27.10.2019)
- [12] <https://www.britannica.com/topic/language> (24.10.2019)
- [13] <https://www.jutarnji.hr/autori/miljenko-jergovic> (28.3.2019)
- [14] <https://www.merriam-webster.com/dictionary/language> (23.9.2019)
- [15] <https://www.sk.com.br/sk-chom.html> (24.10.2019)
- [16] <https://www.vijesti.me/autor/miodrag-lekic> (28.3.2019)
- [17] <https://www.worldometers.info/> (24.10.2019)
- [18] Jahić, Dževad. *Trilogija o bosanskom jeziku*. Knjiga 3, Školski rječnik bosanskog jezika, Sarajevo: Ljiljan biblioteka Linguos, 1999.
- [19] Smailović, Nedim. Statistical Analysis of Texts of the Balkans Electronic Media Columnists, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropean 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 00

Submitted: November 21, 2019

Accepted: May 17, 2020.

ABOUT THE AUTHORS



Nedim Smailović was born in Tuzla. He has been living in Banja Luka since 1973. He graduated from the Faculty of Electrical Engineering, department of Telecommunications. Since 1982 he has worked in RO PTT traffic of Bosnia and Herzegovina, and a series of organizational transformation it is now called Mtel doo Banja Luka. His first work experience was in designing and maintaining the PTT capacities. He obtained his Master's degree from Pan European University

'Aperion' Banja Luka, in 2005. There he also defended his doctoral thesis titled: Computer information graphics in presenting Bosnia and Herzegovina on the road to accessing the European Union. He was elected Associate Professor in 2013 and he has been teaching since in three universities in Bosnia and Herzegovina subjects relating to computer technology. He is an author and co-author of several books from the field of information technology and mathematics. He is married, father of two daughters.



Zoran Ž. Avramović is the Rector of Pan-European University APEIRON Banja Luka. He was born in Loznica (Serbia). He graduated from the University of Belgrade - Faculty of Electrical Engineering, with an average grade of 9.72 (out of 10). He received his master's degree at this faculty (average grade 10 out of 10), and then received a PhD in technical

sciences (1988). He studied mathematics at the Mathematical faculty in Belgrade. He was the champion of Serbia in mathematics ("first prize") and Yugoslavia in electrical engineering ("gold medal"). He is member of the Russian Transport Academy, of the Russian Academy of Natural Sciences, of the Engineering Academy of Serbia, of the Academy of Electrotechnical Sciences of the Russian Federation.

FOR CITATION

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