Machine Learning and Classical Methods Combined for Text Differentiation

Iryna Khomytska [1], Vasyl Teslyuk [1], Iryna Bazylevych [2] and Yuliia Kordiiaka [1]

[1] *Lviv Polytechnic National University, Lviv, 79013, Ukraine*
[2] *Ivan Franko National University of Lviv, Lviv, 79000, Ukraine*

Abstract

The novelty of the research is an offered combination of the machine learning method – the data clustering and the classical method – the Student's t-test to differentiate English and Ukrainian texts. The efficiency of the two methods has been proved to be high for determining the style factor effect and the authorial style factor effect. The research allows us to conclude that the data clustering is a simpler method than the Student's t-test, but it ensures essential differences in fewer cases than the Student's t-test. The use of the Student's t-test is more complicated as it can be performed only after the Pearson's normality test. However, with the help of the Student's t-test, the essential differences have been established in most cases with a test validity of 95%. The research shows that the proposed combination of methods ensures reliable results. The obtained results may be used for text analysis and authorship attribution.

Keywords

Data clustering, Student's t-test, Style factor effect, Authorial style factor effect, Authorship attribution

1. Introduction

The problem raised in the research is closely connected with text analysis. Text differentiation implies identifying the text distinctive features. There are different approaches to text analysis. They can be classified according to the language level (phonological, lexical, syntactic) and language units. All the approaches aim at characterizing specificity of the researched functional style or authorial style. The machine learning methods are widely used for text analysis [1, 2]. However, classical methods also give good results [3]. Distribution of language units on every language level has its particular character. It is different for every style and text. This particular distribution of language units has a differentiating capability. The established degree of similarity between the compared texts has its practical application. This way we can attribute a text to an author. In other words, we can perform authorship attribution. The problem is not easy to solve, as several linguistic factors may overlap. These are: the style factor, the topic related factor and the authorial style factor. The texts of two different authors should have the same topic. Only in this case, the authorial style peculiarities can be identified. Otherwise, the differences will be topic related. Text differentiation is successfully done by the machine learning method – the data clustering. This method consists in grouping language units according to some common feature. The language units of one cluster are different from those of the other cluster. The difference between the clusters reflects the difference between the authorial styles. The data clustering is used for psychological portrait formation of social networks users [4]. Emotional coloring of news headlines is also detected by the data clustering [5, 6]. The method of data clustering is widely used along with the other methods for solving linguistic tasks on different language levels.

The quantitative approach is used for feminism studies in Ukraine [7], for researching the semantic nature of the community Reddit feed post [8], for mapping emotional dislocation of translational fiction

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: Iryna.khomytska@ukr.net (I. Khomytska); vasyl.m.teslyuk@lpnu.ua (V. Teslyuk); i_bazylevych@yahoo.com (I. Bazylevych); yuliia.m.kordiiaka@lpnu.ua (Yu. Kordiika)
ORCID: 0000-0003-3470-7191 (I. Khomytska); 0000-0002-5974-9310 (V. Teslyuk); 0000-0002-5391-0556 (Yu. Kordiika)

[9], for characterizing peculiarities of Lucy Montgomery's literary style [10], for analyzing the distribution of meiosis and litotes in The Catcher in the Rye by Jerome David Salinger [11], for studying anthropocentrism as implementation of a testator/testatrix's communicative goal [12]. The analysis of the mentioned research allows us to state that the quantitative approach gives valuable results for linguistics. However, we recommend to combine the machine learning methods with the classical ones.

The purpose of our research is to determine an efficient combination of the machine learning and the classical methods which ensures high test validity results for text differentiating. The novel approach consists in offering a combination of the data clustering and the Student's t-test for differentiating English and Ukrainian texts. In our previous research, the Student's t-test proved to be efficient on the phonological level. The authors were differentiated by consonant phoneme groups [13-15]. This method was also successfully applied on the lexical level [16]. The data clustering method is efficient on the same language levels – the phonological and lexical levels [4, 5, 17]. Consequently, the Student's t-test and the data clustering method can be combined for text differentiation. The combination of the two methods ensures more reliable results.

The latest methodologies and approaches aim at an optimal solution of the problem of text differentiation. The solution must be simple and it must ensure high accuracy. The problem is not easy to solve as the authorial features are not often clear-cut. The degree of clarity of authorial style features must be sufficient. An author may use the vocabulary common for certain sphere of communication. Because of this, the authorial style lacks the distinctive individual features, by which the manner of writing of one author can be differentiated from that of another author. In fiction, the author's writing is peculiar and can be easily characterized. In scientific papers and formal documents, the author's manner of writing can hardly be noticeable. In this case, different approaches are used to define the differentiating features of this piece of writing. Therefore, we propose a combination of the machine learning and the classical methods. The data clustering ensures a simple solution of text differentiation. The Student's t-test ensures reliable results.

The research is done on the lexical level (function words) and the phonological level (consonant phoneme groups) [18]. The texts from Ukrainian emotive prose, English poetry and the colloquial style are researched with the help of the data clustering method and the Student's t-test.

2. Mathematical support of software system

2.1. The Proposed Combination of Methods

A combination of the machine learning and the classical methods – the data clustering and the Student's t-test is proposed for text differentiation on the lexical level and the phonological level. The research is done according to the following algorithm.

1. Change uppercase to lowercase of all the letters in the researched Ukrainian and English texts of equal size
2. Remove all the punctuation marks
3. Leave only one space between the words
4. Put a space at the beginning and at the end of the text
5. Calculate the absolute frequency of occurrence of function words
6. Use the method of hierarchical clustering [19]
7. Transcribe the English texts
8. Form samples of equal size for consonant phonemes
9. Calculate the absolute and the mean frequency of occurrence for consonants
10. Form eight consonant phoneme groups
11. Perform the Pearson's normality test for eight consonant phoneme groups:

$$\hat{\chi}_n^2 = \sum_{i-1}^{N} \frac{(\upsilon_i - np_i)^2}{np_i}, \quad (1)$$

where N is a number of intervals [20 – 21].

12. Perform the Student's t-test:

$$t = (\bar{\xi} - \bar{\eta}) / s\sqrt{\frac{n+m}{nm}} \geq t_{\alpha;(n+m-2)}, \quad (2)$$

where $\bar{\xi}$ and $\bar{\eta}$ are the mean frequencies of occurrence of consonant phoneme groups for the compared samples n and m [22 – 24].

2.2. The Developed Software

A combination of the data clustering and the Student's t-test is the basis of the program for text differentiation. The structure of the program includes the following modules [25].

- Module of data input/output
- Module of forming samples of Ukrainian function words
- Module of calculating the absolute frequencies of function words
- Module of performing the data hierarchical clustering
- Module of forming samples of English consonant groups
- Module of calculating the absolute and the mean frequencies for consonants
- Module of performing the Pearson's test
- Module of performing the Student's t-test

The structure of the classes of the software is the following: Main, SampleProcessor, TranscriptionProcessor, ConsonantProcessor, ConsonantUtils, StatisticProcessor.

In the class Main, the text files are downloaded and the sequence of operations is controlled.
In the class SampleProcessor, all unnecessary symbols are removed.
In the class TranscriptionProcessor, the English texts are transcribed.
In the class ConsonantProcessor, the samples of consonants are formed.
In the class Consonant Utils, the absolute and the mean frequencies for consonants are calculated.
In the class StatisticProcessor, the Pearson's test and the Student's t-test are performed.
The program code is the following:

```
>library(readxl)
>x=read_excel("C:/Users/Камя/Desktop/mag/clust.xlsx")
>z=c("London (Before Adam)","Henry(The Sea-Wolf)","Henry(The last leaf)","London(White fang)","Henry(The furnished room)","London(Advanture)")
>rownames(x) = z
>XDYST=dist(x, method="euclidean")
>tree=hclust(XDYST, method="single")
>plot(tree)
>tree=hclust(XDYST, method="complete")
```

The Python program code for the literary work "Tsyklon" by O. Honchar is presented in Figure 1. *Single Linkage* and *Complete Linkage* are used for a distance between the clusters. *Euclidean distance* is used for a distance between the objects of the clusters. *Complete Linkage* is used for the texts of Ukrainian emotive prose in the case Single Linkage is not successful.

The algorithm of the program functioning for the text differentiation by the data clustering and the Student's t-test is shown in Figure 2.

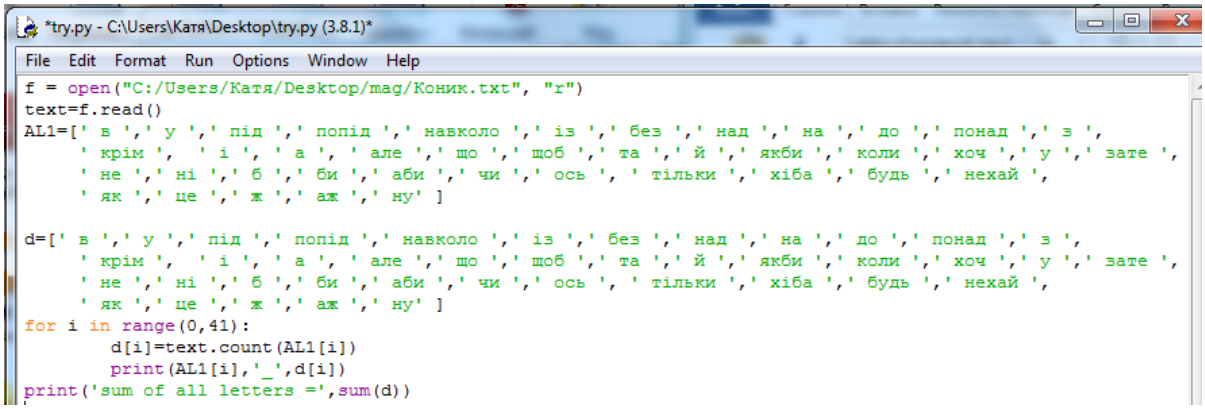

Figure 1: Calculations for the literary work "Tsyklon" by O. Honchar

3. Results of the Study

The data clustering has been performed in eight samples from Ukrainian emotive prose. These are the texts from the following literary works: "Tsyklon" by O. Honchar, "Sobor" by O. Honchar, "Lev ta mysha" by L. Hlibov, "Konyk strybunets" by L. Hlibov, "Malyy Myron" by I. Franko, "Na loni pryrody" by I. Franko and "Zakhar Berkut" by I. Franko. In these comparisons, the authorial style effect is determined. The results of the data clustering for the mentioned literary works are shown in Figure 3.

In Figure 3, we see that the results of the data clustering are successful for "Tsyklon" and "Sobor" by O. Honchar, "Lev ta mysha" and "Konyk strybunets" by L. Hlibov, "Malyy Myron" and "Na loni pryrody" by I. Franko, but not very successful for "Zakhar Berkut" by I. Franko. All the researched literary works by I. Franko are not in the same cluster. Therefore, we change the used *Single Linkage* for *Complete Linkage* (Figure 4).

The matrix of distances is shown in Table 1. In this Table, we can see that there is a little distance between two literary works by I. Franko – "Malyy Myron" and "Na loni pryrody". This result proves that the two literary works have the same author.

Table 1.

The matrix of distances between the researched texts

Variable	Euclidean distances (Spreadsheet1)						
	Гончар (Циклон)	Гончар (Собор)	Франко (Малий Мирон)	**Франко (На лоні природи)**	Глібов (Лев та Миша)	Глібов (Коник стрибунець)	Франко (Захар Беркут)
Гончар (Циклон)	0,0	16,6	17,4	17,7	17,5	20,5	24,9
Гончар (Собор)	16,6	0,0	18,9	18,9	22,6	20,7	19,2
Франко (Малий Мирон)	17,4	18,9	0,0	13,4	18,8	23,4	17,9
Франко (На лоні природи)	17,7	18,9	13,4	0,0	25,2	28,8	17,4
Глібов (Лев та Миша)	17,5	22,6	18,8	25,2	0,0	11,9	27,7
Глібов (Коник стрибунець)	20,5	20,7	23,4	28,8	11,9	0,0	27,5
Франко (Захар Беркут)	24,9	19,2	17,9	17,4	27,7	27,5	0,0

The analysis of the comparisons of literary works "Lev ta mysha" and "Konyk strybunets" by L. Hlibov shows a little distance – 11,9. A greater distance is for the comparison "Tsyklon" and "Sobor" by O. Honchar – 16,6. The greatest distance – 28,8 is for the comparison of literary works by different authors – "Na loni pryrody" by I. Franko and "Konyk strybunets" by L. Hlibov.

In Figure 4, we see that the use of Complete Linkage has given a better result, as all the literary works by one author ("Malyy Myron", "Na loni pryrody" and "Zakhar Berkut" by I. Franko) are in one cluster. Consequently, the use of Complete Linkage is more efficient for solving this task.

The whole process of the data clustering is presented in Table 2.

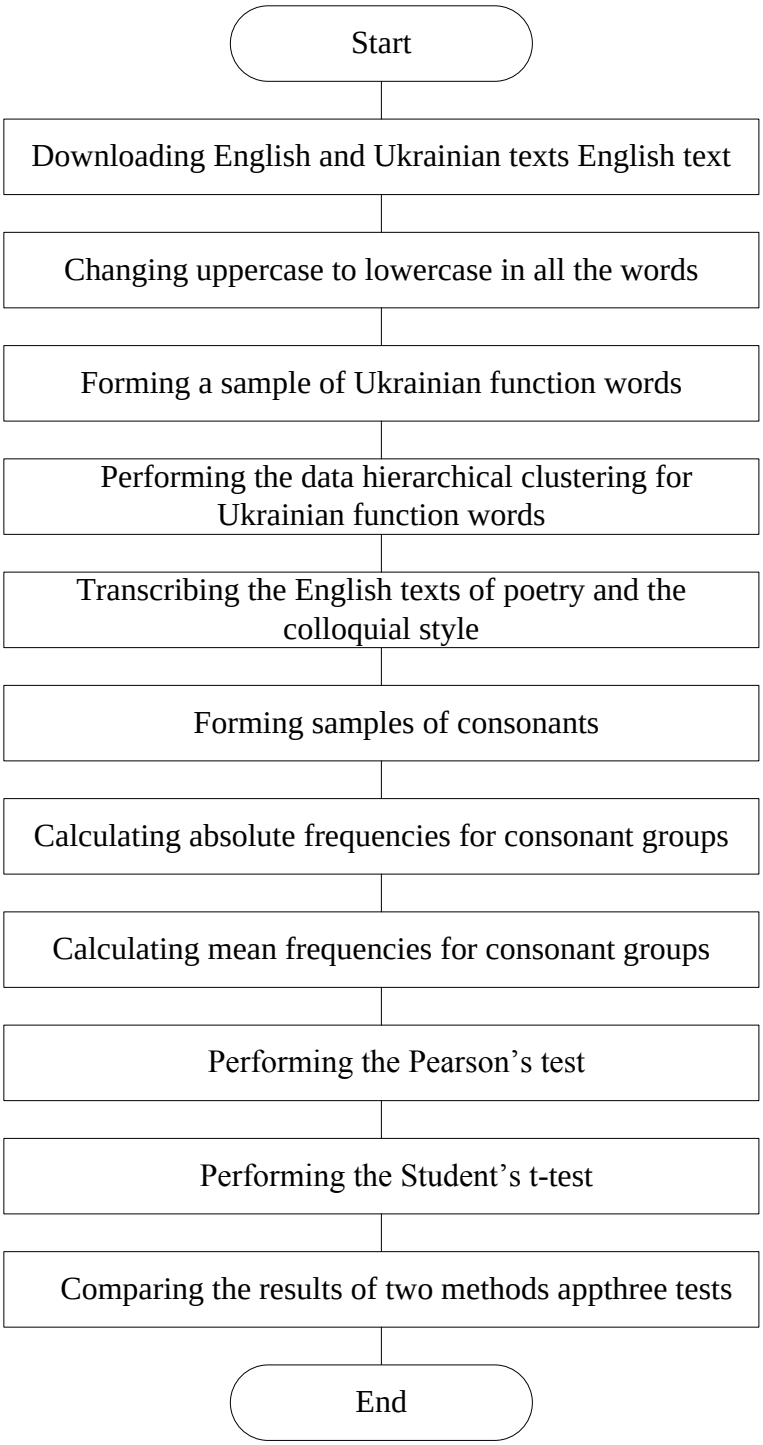

Figure 2: A block-scheme of the algorithm of the program functioning for the text differentiation by the data clustering and the Student's t-test

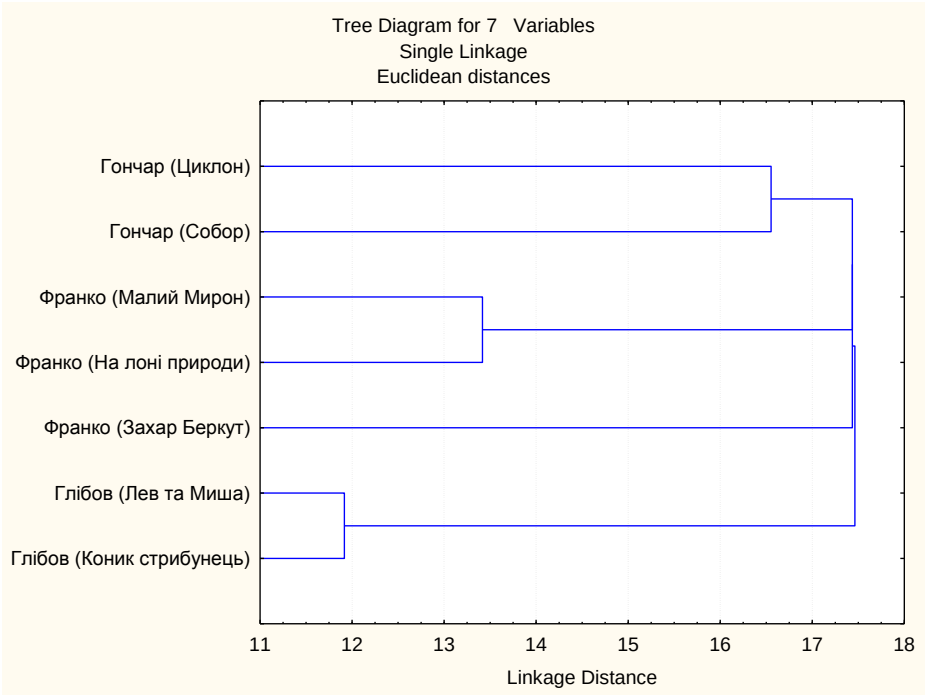

Figure 3: The results of the data clustering for “Tsyklon” by O. Honchar, “Sobor” by O. Honchar, “Lev ta mysha” by L. Hlibov, “Konyk strybunets” by L. Hlibov, “Malyy Myron” by I. Franko, “Na loni pryrody” by I. Franko and “Zakhar Berkut” by I. Franko

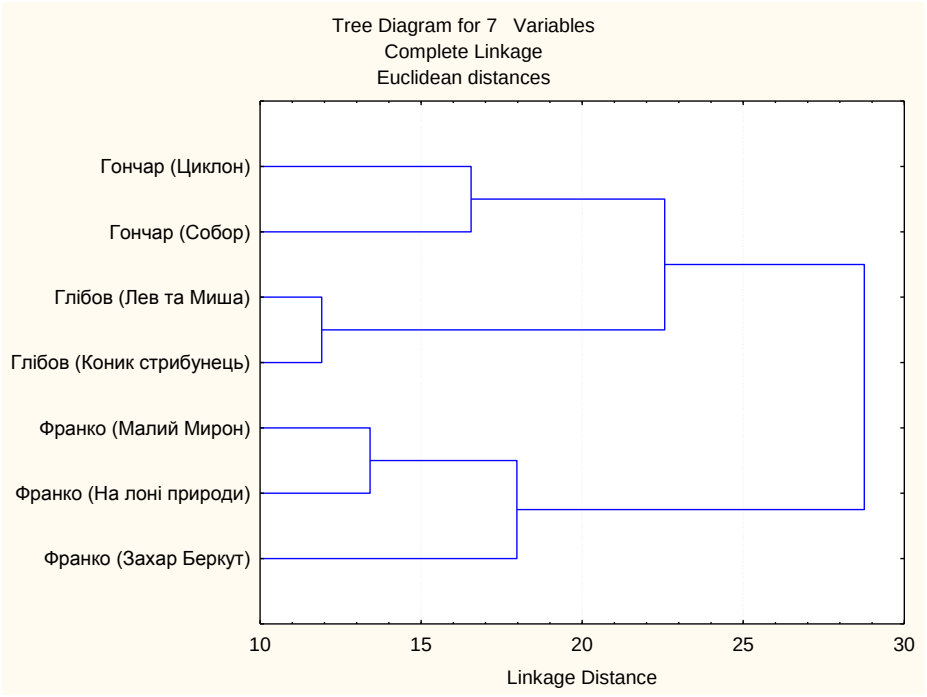

Figure 4: The use of complete linkage for data clustering

The task of text differentiating has also been done on the phonological level with the help of the classical method – the Student’s t-test. The English texts – Th. Moore’s poetry and the colloquial style have been differentiated in eight consonant groups. The essential differences between the text compared are shown in Tables 3, 4.

Table 2.

The whole process of the data clustering for the researched texts

linkage distance	Amalgamation Schedule (7ознак.sta) Complete Linkage Euclidean distances Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6	Obj. No. 7
11,91638	Глібов (Лев та Миш	Глібов (Коник стриб					
13,41641	Франко (Малий Ми	Франко (На лоні при					
16,55295	Гончар (Циклон)	Гончар (Собор)					
17,94436	Франко (Малий Ми	Франко (На лоні при	Франко (Захар Бер				
22,56103	Гончар (Циклон)	Гончар (Собор)	Глібов (Лев та Миш	Глібов (Коник стриб			
28,75761	Гончар (Циклон)	Гончар (Собор)	Глібов (Лев та Миш	Глібов (Коник стриб	Франко (Малий Ми	Франко (На лоні при	Франко (Захар Бер

Table 3.

The results of the calculations for the comparison between Moore's poetry and the colloquial style in an unidentified position

CG	MP $\bar{x}$	MP $\Sigma(x_i-\bar{x})^2$	CS $\bar{x}$	CS $\Sigma(x_i-\bar{x})^2$
Lb	137,9	4156,56	131,9	7611,48
Dr	425,0	8178,00	362,9	32500,3
Cr	5,9	143,58	18,6	5175,36
Vl	59,3	3242,26	72,6	5157,36
Ns	82,9	1902,71	76,8	4202,84
Sn	233,9	4890,01	226,9	14575,5
Fr	210,3	8529,18	158,9	7948,71
St	182,7	10670	226,5	5725,74

Table 4.

The essential differences between Moore's poetry and the colloquial style in an unidentified position

CG	S	t	$2Q$	$\bar{x}_1-\bar{x}_2$
Lb	14,00	1,69	> 5%	Unessential
Dr	26,04	9,39	< 0.1%	Essential
Cr	9,42	5,31	< 0.1%	Essential
Vl	11,83	4,43	< 0.1%	Essential
Ns	10,09	2,38	< 5%	Essential
Sn	18,01	1,53	> 10%	Unessential
Fr	16,57	12,21	< 0.1%	Essential

In Tables 3, 4, 5 and 6 the following designations are used: CG – consonant groups; MP – Moore's poetry; CS – the colloquial style; Lb – labials; Dr – dorsals; Cr – coronals; Vl – velars; Ns – nasals; Sn – sonorous; Fr – fricatives; St – stops; S is a dispersion; t is the Student's statistic; $2Q$ is a significance level; $\bar{x}$ is the mean value of frequencies of occurrence of consonant groups; $\Sigma(x_i-\bar{x})^2$ is a sum of squares of difference of the value of middle of the interval and the mean value of frequencies of occurrence of consonant groups, $\bar{x}_1-\bar{x}_2$ is the value of difference between the two compared samples.

In an unidentified position, the applied Student's t-test has given a very good result: the essential differences have been established in six out of eight consonant groups. For the groups of the labial and sonorous consonants the differences are statistically insignificant. The mentioned degree of similarity can be explained by the use of words from the colloquial style in the researched Moore's poetry.

The results have been obtained with a test validity of 95% in the comparisons presented in Tables 3, 4, 5 and 6.

Table 5.
The results of the calculations for the comparison between Moore's poetry and the colloquial style at the beginning of a word

CG	MP $\bar{x}$	MP $\Sigma(x_i - \bar{x})^2$	CS $\bar{x}$	CS $\Sigma(x_i - \bar{x})^2$
Lb	71,8	1842,24	72,0	5077,57
Dr	115,7	3421,99	100,0	7930,97
Cr	2,3	130,39	11,6	3409,36
Vl	31,8	1128,24	31,3	2880,77
Ns	355,99	355,99	5,7	336,77
Sn	63,1	2009,91	68,4	5383,10
Fr	101,9	6095,91	69,3	3813,94
St	57,1	2217,51	78,2	5504,19

Table 6.
The essential differences between Moore's poetry and the colloquial style at the beginning of a word

CG	S	t	$2Q$	$\bar{x}_1 - \bar{x}_2$
Lb	10,74	0,07	> 80%	Unessential
Dr	13,76	4,49	< 0,1%	Essential
Cr	7,68	4,77	< 0,1%	Essential
Vl	8,17	0,24	> 80%	Unessential
Ns	3,40	2,32	< 5%	Essential
Sn	11,10	1,88	> 5%	Unessential
Fr	12,85	9,99	< 0,1%	Essential
CG	S	t	$2Q$	$\bar{x}_1 - \bar{x}_2$

In the position at the beginning of a word, the results are also good (Tables 5, 6). Statistically significant differences have been revealed in five out of eight consonant groups. In addition to the labial and sonorous consonants, the differences are statistically insignificant for the nasals.

Having analyzed the results of this research, we can state that both the data clustering method and the Student's t-test are efficient for text differentiation on the phonological and lexical levels. However, the former is simpler, the latter is more reliable.

4. Conclusions

The use of the machine learning method – the data clustering and the classical method – the Student's t-test has solved the task of text differentiation, the practical application of which is authorship attribution. The proposed combination of the data clustering method and the Student's t-test is the novelty of the research. The text differentiation task has been successfully done on the lexical level. The texts by I. Franko, O. Honchar and L. Hlibov have been analyzed. The established little distance between the researched texts has proved the fact that they are written by the same author. Consequently, the authorial style effect has been revealed. A good example is the comparison of "Malyy Myron" and "Na loni pryrody" by I. Franko in which the distance is equal to 13,4. The applied classical method – the Student's t-test has given a good result for determining the style factor effect on the phonological level. The texts of Th. Moore's poetry and the colloquial style differ in 6 out of 8 consonant groups for an unidentified position in a word and in 5 out of 8 – for the position at the beginning of a word. The results of the research have shown that the data clustering is a simpler method if compared to the Student's t-test. It shows better results if Complete Linkage is used. However, the Student's t-test ensures more reliable data with a test validity of 95%. The practical application of the results is the style

and authorship attribution. In our future research, another combination of the machine learning methods and the classical methods will be tested for text differentiation.

5. References

[1] J. Bevendorff, B. Ghanem, A. Giachanou, M. Kestemont, E. Manjavacas, M. Potthast, F. Rangel, P. Rosso, G. Specht, E. Stamatatos, B. Stein, M. Wiegmann, E. Zangerle, Shared Tasks on Authorship Analysis at PAN 2020. In book: Advances in Information Retrieval, 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II, pp. 508–516. (2020) DOI: 10.1007/978-3-030-45442-5_66.

[2] M. Kestemont, M. Tschuggnall, E. Stamatatos, W. Daelemans, G. Specht, B. Stein, M. Potthast, Overview of the author identification task at PAN-2018: cross-domain authorship attribution and style change detection. In Working Notes Papers of the CLEF 2018 Evaluation Labs. CEUR Workshop Proceedings, vol. 2125, pp. 1–25. (2018).

[3] Th. S. Gries, Statistics for Linguistics with R: A Practical Introduction (Trends in Linguistics: Studies & Monographs), Mouton de Gruyter, p. 348. (2009).

[4] V. Lytvyn, V. Vysotska, A. Rzheuskyi, Technology for the psychological portraits formation of social networks users for the IT specialists recruitment based on Big Five, NLP and Big Data Analysis, in CEUR Workshop Proceedings, vol. 2392, 2019, pp. 147–171. E-ISSN: 1613-0073. http://ceur-ws.org/Vol-2392/paper12.pdf

[5] M. Zanchak, V. Vysotska, S. Albota, The Sarcasm Detection In News Headlines Based on Machine Learning, in Proceedings of the IEEE 16th International Conference on Computer Sciences and Computer technologies, CSIT 2021, 22 – 25 Sept., Lviv, Ukraine, vol. 1, 2021, pp. 131 – 137.

[6] O. Mulesa, F. Geche, A. Batyuk, Information technology for determining structure of social group based on fuzzy c-means, in Proceedings of the IEEE Xth International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2015, Lviv, 2015, pp. 60-62. doi: 10.1109/STC-CSIT.2015.7325431

[7] O. Levchenko, M. Dilai, Attitudes Toward Feminism in Ukraine: A Sentiment Analysis of Tweets. In: Shakhovska N., Medykovskyy M. (eds) Advances in Intelligent Systems and Computing III. CSIT, 2018. Advances in Intelligent Systems and Computing, vol. 871. Springer, Cham, pp. 119-131. (2019).

[8] S. Albota, Linguistically manipulative, disputable, semantic nature of the community Reddit feed post. CEUR Workshop Proceedings. 2021, vol. 2870, in Proceedings of the 5th International conference on computational linguistics and intelligent systems, COLINS 2021, Lviv, Ukraine, April 22–23, vol. I: main conference, pp. 769–783. (2021).

[9] I. Bekhta, N. Hrytsiv, Computational Linguistics Tools in Mapping Emotional Dislocation of Translated Fiction, in Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems, COLINS 2021, vol. I: Workshop. Kharkiv, Ukraine, April 22-23, CEUR-WS.org, pp. 685-699. (2021).

[10] N. Hrytsiv, T. Shestakevych, I. Shyyka, Quantitative parameters of Lucy Montgomery's literary style. CEUR Workshop Proceedings. 2021, vol. 2870, in Proceedings of the 5th International conference on computational linguistics and intelligent systems, COLINS 2021, vol. I: main conference. Kharkiv, Ukraine, April 22-23, pp. 670–684. (2021).

[11] M. Karp, N. Kunanets, Y. Kucher, Meiosis and litotes in The Catcher in the Rye by Jerome David Salinger: text mining. CEUR Workshop Proceedings. 2021, vol. 2870, in Proceedings of the 5th International conference on computational linguistics and intelligent systems, COLINS 2021, vol. I: main conference. Kharkiv, Ukraine, April 22-23, pp. 166–178. (2021).

[12] O. Kulyna, Anthropocentrism as implementation of a testator/testatrix's communicative goal. CEUR Workshop Proceedings, 2021, vol. 2870, in Proceedings of the 5th International conference on computational linguistics and intelligent systems, COLINS 2021, Lviv, Ukraine, April 22–23, vol. I : main conference, pp. 845–854. (2021).

[13] I. Khomytska, V. Teslyuk, N. Kryvinska, I. Bazylevych, Software-Based Approach towards Automated Authorship Acknowledgement—Chi-Square Test on One Consonant Group. Electronics, 9, 1138, (2020). https://doi.org/10.3390/electronics9071138.
[14] A. V. Doroshenko, Application of global optimization methods to increase the accuracy of classification in the data mining tasks, in CEUR Workshop Proceedingsthis link is disabled, 2019, 2353, p. 98–109.
[15] A. Doroshenko, R. Tkachenko, Classification of imbalanced classes using the committee of neural networks, in International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018, 1, pp. 400–403. (2018).
[16] V. S. Perebyjnis, Statystychni metody dlia lingvistiv. Nova Knyha: Vinnytsia, Ukraine, (2013). (in Ukrainian).
[17] M. A. Boukhaled, J.-G. Ganascia, Using function words for authorship attribution: Bag-of-words vs. sequential rules, in the 11th International Workshop on Natural Language Processing and Cognitive Science, Oct 2014, Venice, Italy, De Gruyter, Natural Language Processing and Cognitive Science Proceedings, pp. 115–122. (2015). Available online: https://hal.sorbonne-universite.fr/hal-01198407/document
[18] I. Khomytska, V. Teslyuk, Authorship Attribution by Differentiation of Phonostatistical Structures of Styles, in International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018, 2, pp. 5–8. (2018).
[19] V. O. Klymchuk, Klasternyy analiz.: vykorystannia u psykholohichnyh doslidzhenniah, Praktychna psykhologia ta sotsialna robota, №4, p. 30–36. (2006). (in Ukrainian).
[20] P. C. Gomez, Statistical Methods in Language and Linguistic Research. University of Murcia, Spain (2013).
[21] A. Kornai, Mathematical Linguistics. Springer (2008).
[22] V. M. Turchyn, Matematychna statystyka. Navch. Posib. Vydavnychyj tsentr "Akademia": Kyiv, Ukraine, (1999). (in Ukrainian).
[23] I. Khomytska, V. Teslyuk, I. Bazylevych, I. Shylinska, Approach for minimization of phoneme groups in authorship attribution, International Journal of Computing 19(1), 2020, pp. 55–62.
[24] G. I. Ivchenko, Yu. I. Medvedev, Matematicheskaya statistika. Moskva: Vyssh. Shk., p. 248 (1984).
[25] A. Batyuk, V. Voityshyn, V. Verhun, Software Architecture Design of the Real-Time Processes Monitoring Platform, in: Proceedings of the IEEE Second International Conference on Data Stream Mining & Processing, DSMP 2018, Lviv, Ukraine, 2018, pp. 98-101. doi: 10.1109/DSMP.2018.8478589