APPROACH FOR MINIMIZATION OF PHONEME GROUPS IN AUTHORSHIP ATTRIBUTION

Iryna Khomytska [1], Vasyl Teslyuk [1], Iryna Bazylevych [2], Inna Shylinska [3]

[1] Lviv Polytechnic National University, 12 S. Bandery str., Lviv, 79000, Ukraine, vasyl.m.teslyuk@lpnu.ua
[2] Ivan Franko National University of Lviv, 1 Universytetska str., Lviv, 79000, Ukraine, i_bazylevych@yahoo.com
[3] Ternopil National Economic University, 11 Lvivska str., Ternopil, 46009, Ukraine, inna.shylinska2012@gmail.com

Abstract: The developed mathematical support for authorship attribution software includes a combination of statistical methods (Student's t-test, Kolmogorov-Smirnov's test) and a statistical model for determining significant differences between styles. The combination of statistical methods allows us to enhance test validity of authorship attribution by obtaining the same results by the two methods applied. The model developed makes it possible to identify a consonant phoneme group with high style identification capability. The phoneme position in a word is taken into account. The greater number of significant differences is, the higher authorship identification capability of the phoneme group is. The developed system software is based on the algorithms of the used combination of methods and statistical model. The Java programming language provides platform independence. The minimized number of consonant phoneme groups makes the process of style and authorship attribution more automated. The obtained results of comparisons of the scientific, belles-lettres, conversational and newspaper styles are presented. The data obtained allows us to assert that the used combination of methods and the developed statistical model improve test validity of style and authorship attribution.

1. INTRODUCTION

We are living in an epoch of implementation of information technologies on a large-scale in all fields of human activity. In cases of an unknown author of a text authorship attribution is performed. An author is identified on different language levels. The differentiation of the studied texts is carried out on the phonological level which has an advantage of an unaltered number of elements. The authorship attribution of a text on the phonological level can be performed according to the following criteria of differentiation: absolute frequency of phoneme groups; relative frequency of phoneme groups; average frequency of phoneme groups.

The texts by different authors can be differentiated according to the following positions of phonemes in a word: arbitrary position in a word; at the beginning of a word; at the end of a word.

In addition, the texts can be differentiated by one group of phonemes, two or three phoneme groups and all phoneme groups.

The most effective approach to solve the above-mentioned problem is text differentiation by one, two or three of eight phoneme groups in which significant differences are found when comparing the texts by different authors. Such simplification of process of authorship attribution of a text can be regarded as minimization of a number of phoneme groups.

The frequency of occurrence of phonemes is influenced by such factors as language, style and authorial style. When a text is significantly marked with stylistic and authorial elements, it can be revealed by the phoneme group which is characterized by a high authorship identification capability.

In addition, the reliability and accuracy of the results of the authorship attribution of a text is higher on the phonological level than on other language levels due to the fact that this level is a closed system with an unaltered number of elements.

Consequently, the differentiation of texts by different authors on the phonological level with the help of modern information technologies (IT) is an important task of research.

2. RELATED WORK

The problem of determining the features of author's individual style is widely represented in current research in the field of IT [1, 2, 3]. The analysis of recent publications makes it possible to conclude that usually, for authorship attribution, short words (prepositions, conjunctions) are selected, which are characterized by relative stability of use. The most common methods are statistical tests. Thus, the statistical method – the discriminant analysis is used for differentiating functional words with accuracy of 81.5 % [4]. The method applied is less powerful than the Kolmogorov-Smirnov test. The linear regression method and the support vector machine algorithm are used to differentiate functional words and punctuation marks in texts by different authors. The accuracy is 90%, 95% and 97% [5, 6, 7]. This combination of methods is more powerful and gives better results, but the lexical level, it is used on, is less strictly arranged than the phonological one, on which still better results could be obtained. To establish authorship of documents the support vector machine algorithm is used. The letters and stop words are differentiated with accuracy of 97.22 % [8]. A combination of methods could be more efficient. Uni-grams and bi-grams are differentiated with the help of the Kolmogorov-Smirnov test for author identification. The accuracy is 95% and 97 % [9, 10]. Phoneme based research could demonstrate higher efficacy. Functional words, prepositions, pronouns, nouns and verb forms from two authorial styles are differentiated by the chi-square test with accuracy 90% and 95% in the texts on judiciary [11]. The chi-square test is less powerful than the Kolmogorov-Smirnov's test [12]. The Student's t-test is used for revealing difference between two compared authors. The first words of sentences and punctuation marks are differentiated with accuracy 66.7% [13]. The Student's t-test and the chi-square test are less powerful than the Kolmogorov-Smirnov's test [12]. The rank-frequency relations for phonemes of a text depend on its author and can be described by the Dirichlet distribution [14]. The ranking method is less powerful than the hypothesis method [15]. In stylometric studies, it is proposed to use linguametric coefficients when dealing with the corpus of Ukrainian language texts of the scientific style. The results are obtained with accuracy 77% and 87%. In this research, the openness of the lexical and syntactic systems causes difficulties in describing language phenomena with a high degree of accuracy [16, 17]. Determining the characteristics of author's style in a piece fiction is a more complicated task than identifying author of a scientific article. Components of lexical-semantic groups of the belles-lettres style are more numerous and vary in terms of emotionality and expressiveness. Therefore, the level of accuracy of authorship attribution of literary works may be lower than the level of accuracy of scientific articles [18, 19].

Modern IT allows us to differentiate the stylistic characteristics of author's style, but this task is rather complicated because the stylistic devices are not as frequently used as words of common word-stock [20].

Online messages have their own specific genre, which allows us to distinguish between authorial distinctive features. Due to penetration of a great number of conversational elements, typical of this genre, the level of accuracy is lower than in the scientific style [21].

The analysis of the above-mentioned research has shown that differentiation of functional styles and texts by different authors on the phonological level is a topical task nowadays. The problem of enhancing test validity of authorship attribution and the problem of minimizing the number of phoneme groups by which styles differ essentially still remain to be solved. In order to increase the reliability of results of authorship attribution, it is necessary to combine the most efficient methods [12, 15] and to differentiate texts on the most strictly arranged language level – the phonological one. The most optimal level of significance is a classical level – 0.05. On this level the greatest amount of results is obtained. On the level of significance of 0.01 some part of results is lost. The use of modern information technology tools makes the authorship attribution fast.

3. PURPOSE AND OBJECTIVES OF THE RESEARCH

The purpose of the study is to minimize the number of consonant phonemes in which the texts by different authors differ essentially on the phonological level. To attain the goal, the following tasks must be fulfilled:

– a combination of statistical methods and a statistical model, which allow us to improve test validity of authorship attribution must be proposed;

– the phoneme group with the highest style identification capability in a certain phoneme position in a word must be determined;

– the system software for automating style and authorship attribution must be developed;

4. MATHEMATICAL SUPPORT FOR SYSTEM SOFTWARE

Mathematical support of the system software for authorship attribution includes a combination of statistical methods and a statistical model for identifying the consonant phoneme group with the highest author and style identification capability. The combination of methods includes the Student's t-test and the Kolmogorov-Smirnov's test. The test validity of style and authorship attribution can be improved, if the same results are obtained by the two statistical tests applied. The Student's t-test is a parametric method and can be used only if consonant phoneme group frequencies follow normal distribution. The Kolmogorov-Smirnov's test is a nonparametric test and consonant phoneme group frequencies may not follow normal distribution.

The algorithm for applying the proposed combination of two statistical tests includes the following steps:

Step 1. The average frequency of occurrence of consonant phoneme groups is determined:

$$\overline{X} = \frac{\sum_{i=1}^{m} x_i n_i}{N}, \tag{1}$$

where m – a number of intervals, x_i – middle of an interval, n_i – a number of frequencies getting into an interval, N – a number of portions of a sample.

Step 2. The theoretical normal distribution is determined: $\rho_i = (\beta_i - \alpha_i) f(x)$, where $f(x)$ – density of normal distribution:

$$f(x) = -\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}, \tag{2}$$

where μ and σ^2 are respectively average value and variance of a general population. These values are unknown and are changed for the known – $\overline{x}$ and S^2, where S^2 – unbiased estimator of dispersion:

$$S^2 = \frac{\sum_{i=1}^{m} (x_i - \overline{x})^2 n_i}{N-1}. \tag{3}$$

Step 3. The theoretical frequency is determined:

$$n_i' = \frac{N\Delta x}{S} \varphi(z), \ \varphi(z) = -\frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}}, \ z = \frac{|x - \overline{x}|}{S}. \tag{4}$$

In the given expressions: x – middle of an interval, $\Delta x = \beta_i - \alpha_i$ – width of an interval, where α_i – a lower limit of an interval, β_i – an upper limit of an interval, $\beta_i = \alpha_{i+1}$.

Step 4. Compliance of frequencies of all eight consonant phoneme groups with the normal distribution law is determined by the Pearson's test [22]; the level of significance g is 0.05 and 0.03; the number of degrees of freedom $\nu = n - 1$ where n – a number of portions in a sample:

$$x^2 = \sum_{i=1}^{k} \frac{(n_i - n_i')^2}{n_i'}, \tag{5}$$

where i – a serial number of an interval; k – a number of intervals; n_i – the empirical frequency of occurrence of phonemes; n_i' – the theoretical frequency of occurrence of phonemes; x^2 – statistics value of the Pearson's test.

Step 5. Significant differences between the styles in all eight consonant phoneme groups are determined by the Student's t-test:

$$t = \frac{\overline{x}_1^a - \overline{x}_2^a}{S} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}, \tag{6}$$

where t – value of Student's t-test; $\overline{x}_1^a - \overline{x}_2^a$ – difference between the average frequencies of two samples compared in a fixed phoneme group α ; n_1 and n_2 – a number of portions of the first and second samples; $S = \sqrt{S^2}$, S^2 – an unbiased estimator of dispersion [22].

Step 6. Significant differences between the styles in all eight consonant phoneme groups are determined by the Kolmogorov-Smirnov's test:

$$\lambda_{n,m} = \sqrt{\frac{nm}{n+m}} D_{n,m} = \sqrt{\frac{nm}{n+m}} \sup_{-\infty < z < \infty} |F_n(z) - F_m(z)|, \tag{7}$$

where $F_n(z)$ and $F_m(z)$ – the empirical functions of distribution of two samples of the sample-size n and m respectively, $D_{n,m}$ – the Kolmogorov-Smirnov's statistics, $\lambda_{n,m}$ – value of the Kolmogorov-Smirnov's statistics [9].

Step 7. The consonant phoneme group, in which significant differences are determined by the Student's t-test and the Kolmogorov-Smirnov's test is considered the group with a high style and author identification capability.

Step 8. The texts are differentiated by the definite phoneme group (with a high style and author identification capability) in the definite phoneme position in a word.

The algorithm for determining style and author identification capability of consonant phoneme groups is presented in Fig. 1.
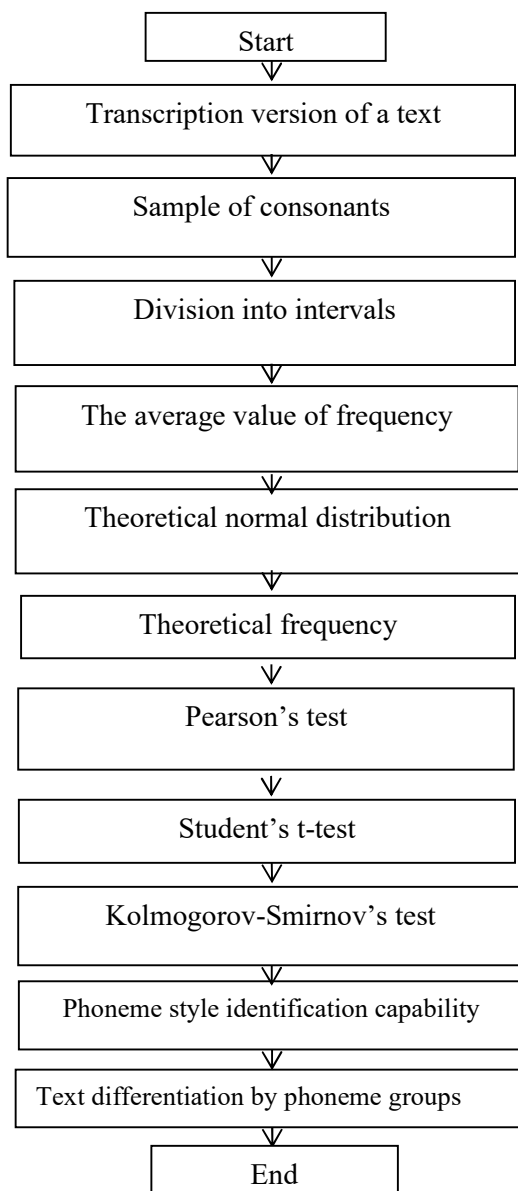

Figure 1 – Block-scheme of algorithm of a combination of the methods applied.

A statistical model is developed on the basis of the applied combination of statistical methods. The model shows the dependence of style and author identification capability of a certain consonant phoneme group on a number of obtained significant differences between compared styles.

This model is based on the expression: $n \times sc = const$, where n – a number of significant differences between the compared texts;

sc – style identification capability of a certain phoneme group. If the maximum number of significant differences is 10, it corresponds to 1, and the following expression is true: $10 \times 1 = const$.

According to the developed model it is possible to determine the following dependence: the greater number of the obtained significant differences is, the higher style and author identification capability of this phoneme group is, as it is shown in Table 1.

Table 1. Dependence of the style and author identification capability of consonant phoneme groups on the number of significant differences found between the styles

Number of significant differences	Indicator of style identification capability
18	1
12	2
6	3

18 – the maximum number of significant differences; 1 – the highest indicator of the style and author identification capability of the phoneme group.

Consequently, the developed model based on the proposed combination of statistical methods allows us to minimize the number of phoneme groups by which the samples differ essentially and to make style and authorship attribution more automated. The level of test validity is 95%, 97%.

5. SYSTEM SOFTWARE

The structure of the developed system is based on a modular principle and includes four main modules, namely: user interface, data input module, statistical data processing module and data output module. The data input module is aimed at converting the text into transcription and forming consonant phoneme samples. Data output module secures presentation of the statistically processed results in a format convenient for a user. The user interface provides a connection of the system software with a user, the work with files, a sequence of data processing, etc. The statistical data processing module secures implementation of the methods and model applied.

The use of the modular principle makes it possible to quickly modify and improve the developed software product.

Information support is of particular importance in the process of software development. The constructed scheme of data conversion in the developed system is shown in Fig. 2.
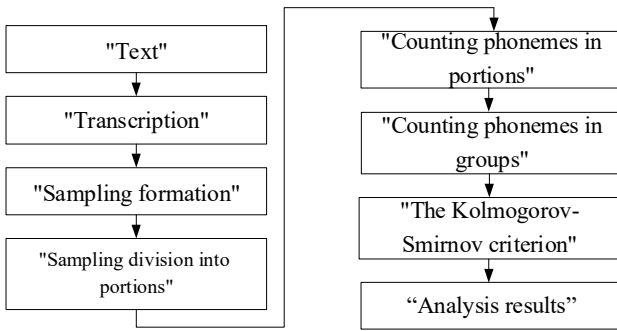
Figure 2 – Scheme of data conversion in the system for determining style identification capability of the consonant phoneme groups for authorship attribution

The first step of the system performance is to download two files with texts. Then, the texts are transcribed. The next steps involve creating a sample of consonant phonemes, dividing the sample into portions and counting the number of phonemes in each portion. Then, statistical processing of data is performed.

The system software is developed with the use of the Java language providing platform independence. An example of the created diagram of software classes is shown in Fig. 3.
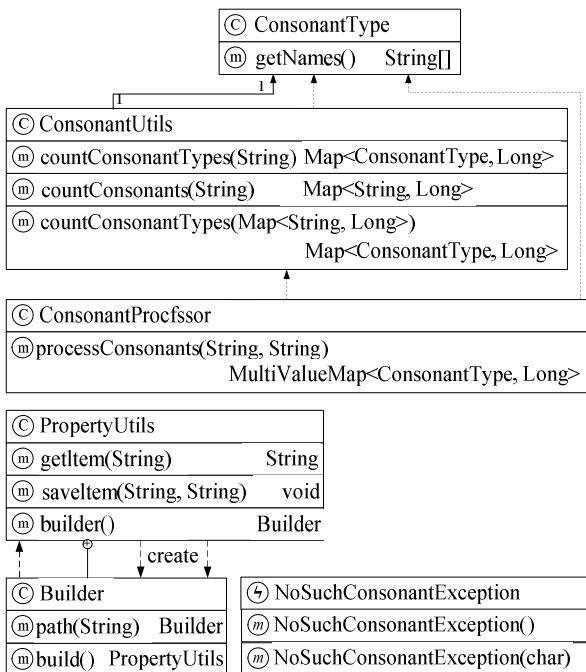

Figure 3 – Diagram of the system classes for determining style identification capability of the consonant phoneme groups for authorship attribution

The developed interface includes 7 tabs, the names of which correspond to the steps of algorithm of the system functioning: "Text", "Transcription", "Sample formation", "Sample division into portions", "Sample division into groups", "Counting phonemes in portions", "Counting phonemes in groups", "The Kolmogorov-Smirnov's criterion".

The proper organization of the system provides the user with a user-friendly and intuitive interface.

Two jPanels are created in the "Text" tab; two text fields jTextAreaLeft and jTextAreaRight are placed on the panels; a file is selected using jFileChooser; text fields are filled with the contents of the files using the SetText method.

In the "Transcription" tab, we work with Word.txt and Transcription.txt files. The files are filled with new texts and therefore the sample is dynamic. ArrayList <String> list structures allow us to work with dynamic data.

In the "Sample formation" tab, transcription symbols for consonants are singled out from the symbolic stream of transcription symbols. The sample is formed from the first 51,000 consonant phonemes.

Class lists from the java.util library are used to create "Sample division into portions", "Counting phonemes in portions" and "Counting phonemes in groups" tabs. The number of phonemes in portions and groups is counted in the tabs.

Mathematical Library Colt is used for the Pearson's, Student's and Kolmogorov-Smirnov's criteria. Consequently, the developed system software allows us to automate the process of identifying a group of phonemes with the highest style and authorship identification capability.

6. TESTING RESULTS OF THE SYSTEM SOFTWARE FOR DETERMINING STYLE IDENTIFICATION CAPABILITY OF A PHONEME GROUP FOR AUTHORSHIP ATTRIBUTION OF A TEXT

The combination of methods, model and system software based on them are used to perform style and authorship attribution. The scientific style is compared with the conversational, belles-lettres and newspaper styles. In this comparison, a high style identification capability of the occlusive phoneme group is determined. For style differentiation by the Student's t-test, it is necessary to perform the Pearson's test. The obtained by the Pearson's test results of verification of the compliance of frequencies of the occlusive phoneme group with the law of normal distribution are given in Table 2.

The calculations are obtained by the formulae given in part 4: $\overline{X} = 198{,}102$ – the average frequency value of the occlusive phoneme group (1);

$$S = \sqrt{\frac{47242{,}58}{123}} = 19{,}60 \, , \; S = \sqrt{S^2} \, , \; S^2 - \text{the value}$$

of an unbiased estimator for dispersion (3); $\Delta X = 7{,}636$ – the value of the width of the interval

(4); $\dfrac{N\Delta X}{S} = 48,31$ – the value of the theoretical frequency (4); $\nu = 10 - 3 = 7$ – the value of the number of degrees of freedom; $\chi^2_{0,10;\,7} = 12,02$; $\chi^2 = 11,06$; the level of significance is 5%, $P(\chi^2) > 10\%$ – the value of the Pearson's test (5). The obtained by the Pearson's test results prove that frequencies of the occlusive phoneme group comply with the law of normal distribution.

The results obtained by the Pearson's test make it possible to apply the Student's t-test for style differentiation. The average frequency of occurrence of consonant phoneme groups is a style differentiation criterion. Two styles are compared in a fixed phoneme group by the values of average frequencies (6). The levels of significance are 5% and 3%. If the difference of the values of average frequencies of two styles compared is significant, the styles differ essentially.

As the results show, significant differences have been obtained in 15 of 18 comparisons in the occlusive phoneme group. The same results have been established by the Kolmogorov-Smirnov's test (7). The results of the comparisons are shown in Table 3. The symbol + means essential difference, the symbol – stands for unessential difference.

Table 2. Values for the occlusive phoneme group. Pearson's test

№ intervals	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \overline{X})^2 n_i$	z_i	$\varphi(z_i)$	n_i'	n_i' rounded	$\dfrac{(n_i - n_i')^2}{n_i'}$
1	150,50–158,14	154,32	5 ⎫ 9	9585,19	-2,23	0,0332	1,60 ⎫ 5,15	2	2,83
2	158,14–165,77	161,95	4 ⎭	5226,71	-1,84	0,0734	3,55 ⎭	4	
3	165,77–173,41	169,59	8	6503,47	-1,45	0,1394	6,74	7	0,23
4	173,41–181,04	177,23	9	3922,27	-1,06	0,2275	10,99	11	0,36
5	181,04–188,68	184,86	9	1577,68	-0,68	0,3166	15,29	15	2,58
6	188,68–196,32	192,50	13	408,26	-0,29	0,3825	18,48	18	1,62
7	196,32–203,95	200,13	25	103,23	0,10	0,3970	19,18	19	1,76
8	203,95–211,59	207,77	19	1775,93	0,49	0,3538	17,09	17	0,19
9	211,59–219,22	215,41	15	4491,43	0,88	0,2709	13,08	13	0,28
10	219,22–226,86	223,04	10	6220,04	1,27	0,1781	8,61	9	0,22
11	226,86–234,50	230,68	7	7428,37	1,66	0,1006	4,86	5	
$\sum$			124	47242,58			119,47	120	0,94

The analysis of the obtained results, given in Table 3, allows us to conclude that the occlusive phoneme group has the highest style and authorship identification capability in the arbitrary position in a word and at the end of a word. On the basis of the obtained results, a statistical model for determining style identification capability of the occlusive phoneme group has been developed (Fig. 4).

According to the results of the study, it can be concluded that style and authorship attribution can be performed in the occlusive phoneme group. The number of consonant phoneme groups minimized from 8 to 1 secures more automated style and authorship attribution.

Table 3. Significant differences between the styles compared

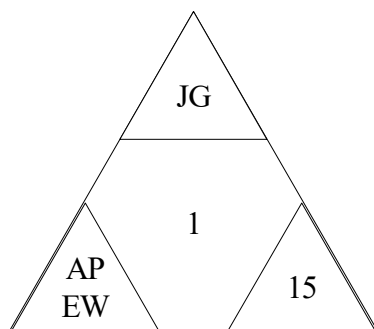| Styles	Scientific style		
	Arbitrary position in a word	The beginning of a word	The end of a word
Byron's poetry	+	+	+
Moore's poetry	+	-	+
Byron's prose	+	-	+
Shaw's drama	+	+	+
Newspaper style	+	-	+
Conversational style	+	+	+

Figure 4 – Schematic representation of a statistical model for determining style identification capability of the occlusive phoneme group

In the model in Fig. 4 the following designations are used: JG – the occlusive phoneme group; AP– arbitrary position in a word; EW – position at the end of a word; 15 – a number of significant differences revealed in a comparison of the scientific, belles-lettres, conversational and newspaper styles; 1 – the highest index of style identification capability.

Consequently, the results of testing by the Student's t-test and the Kolmogorov-Smirnov's test show that significant differences have been obtained in most cases of style comparisons. This proves high reliability of this combination of statistical methods for style and authorship attribution.

7. CONCLUSION

The proposed combination of statistical methods: the Student's t-test and the Kolmogorov-Smirnov's test allows us to improve test validity of style and authorship attribution by obtaining the same results by the two tests applied. The level of test validity is 95% and 97%.

The developed statistical model makes it possible to determine a group of phonemes with the highest style and authorship identification capability. In the research this group is the occlusive phoneme group in the arbitrary position in a word and at the end of a word in a comparison of the scientific, belles-lettres, conversational and newspaper styles. The research results allow us to minimize the number of phoneme groups in which the compared styles differ essentially, simplifying the process of style and authorship attribution and making it more automated.

The developed system software allows us to carry out style and authorship attribution more efficiently – in one instead of eight consonant phoneme group.

In further research, another combination of statistical methods will be tested for improving efficacy of style and authorship attribution.

8. REFERENCES

[1] E. A. Stamatatos, "Survey of modern attribution methods," *Journal of the Association for Information Science and Technology,* vol. 60, pp. 538-556, 2009.

[2] *Neoneuro. Authorship attribution,* [Online]. Available at: https://neoneuro.com/ru/products/authorship-attribution.

[3] *Authorship attribution,* [Online]. Available at: http:www.aicbt.com/authorship-attribution/online-software.

[4] R. H. Baayen, *Word Frequency Distributions*, *Springer Netherlands eBook.* DOI 10.1007/978-94-010-0844-0, 2001.

[5] P. Juala, "Authorship attribution," *Foundations and Trends(R) in Information Retrieval*, vol. 3, Boston-Delft, pp. 233-334, 2008.

[6] Sh. Argamon, M. Koppel, J. Pennebaker, J. Schler, "Automatically profiling the author of an anonymous text," *Communications of the ACM*, vol. 52, issue 2, pp. 119-123, 2009.

[7] M. Koppel, J. Schler, Sh. Argamon, "Computational methods in authorship attribution," *Journal of the Association for Information Science and Technology*, vol. 60, issue 1, pp. 9-26, 2009.

[8] R.N.V. Ganapathi, Ch. Sadhvi, P. Tejaswini, Y. Mounica, "Style based authorship attributionon English editorial documents," *International Journal of Computer Applications*, vol. 159, no. 4, pp. 5-8, 2017.

[9] O. Granichin, L. Klebanov, D. Shalymov, Z. Volkovich, "Authorship attribution method based on KNN re-sampling approach," *Proceedings of the Elmar – International Symposium Electronics in Marine,* 2015. [Online]. Available at: http://www.elmar-zadar.org/2015/technical_program/.

[10] V. Keselj, F. Peng, N. Cercone, C. Thomas, "N-gram-based author profiles for authorship attribution," *Proceedings of the Conference Pacific Association for Computational Linguistics, PASLING*, vol. 3, 2003, pp. 255-264.

[11] V. Guillén Nieto, C. Vargas Sierra, M. Pardiño Juan, P. Martínez Barco, A. Suárez Cueto, "Exploring state-of-the-art software for forensic authorship identification," *International Journal of English Studies*, vol. 8, no. 12008. pp. 1-28. [Online]. Available at: https://revistas.um.es/ijes/article/view/49071.

[12] I. Khomytska, V. Teslyuk, "Mathematical methods applied for authorship attribution on the phonological level," *Proceedings of the XIVth Scientific and Technical Conference CSIT*, Lviv, 2019, pp. 7–11.

[13] T. Petmanson, "Authorship identification using correlations of frequent features," *Notebook for PAN at CLEF. Institute of Computer Science*, University of Tartu, 2013. [Online]. Available a: http://ceur-ws.org/Vol-1179/CLEF2013wn-PAN-Petmanson2013.pdf.

[14] W. Deng, A. Allahverdyan, "Stochastic model for phonemes uncovers an author-dependency of their usage," 2016. [Online]. Available at: https://arxiv.org/pdf/1510.01315.pdf.

[15] I. Khomytska, V. Teslyuk, "Authorship and style attribution by statistical methods of style differentiation on the phonological level," *ebook: Advances in Intelligent Systems and Computing III*, Natalia Shakhovska editor, vol. 871, Lviv, 2018, pp. 105-118.

[16] V. Vysotska, O. Kanishcheva, Y. Hlavcheva, "Authorship identification of the scientific text in Ukrainian with using the lingvometry methods," *Proceedings of the 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies*, 2018, pp. 34-38

[17] A. Abbasi, H. Chen, "Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace," *ACM Transactions on Information Systems*, vol. 26, no. 2, article 7, 2008.

[18] E. Stamatatos, N. Fakotakis, G. Kokkinakis, "Computer-based authorship attribution without lexical measures," *Computers and the Humanities*, vol. 35, no. 2, pp. 193-214, 2001.

[19] Y. Zhao, J. Zobel, "Searching with style: Authorship attribution in classic literature," *Proceedings of the thirtieth Australasian Conference on Computer Science*, Australian Computer Society, Inc., vol. 62, 2007, pp. 59-68.

[20] J. Burrows, "Delta: A measure of stylistic difference and a guide to likely authorship," *Literary and Linguistic Computing*, vol. 17, issue 3, pp. 267-287, 2002.

[21] R. Zheng, J. Li, H. Chen, Z. Huang, "A framework for authorship identification of online messages: Writing-style features and classification techniques," *Journal of the Association for Information Science and Technology*, vol. 57, no. 3, pp. 378-393, 2006.

[22] A. K. Mitropolskiy, *Statistical Computation Technique*, Moscow: Nauka, 1971, 576 p. (in Russian)

***Iryna Khomytska**, an Assistant Teacher at Lviv Polytechnic National University, Applied Linguistics Department, Lviv, Ukraine. She has published 48 papers in international and national scientific journals and she is an author of 1 textbook. The main research interests include mathematical linguistics, statistical linguistics.*

***Vasyl Teslyuk, a** Professor at Lviv Polytechnic National University, Lviv, Ukraine. PhD (1995), Doct. of Sc. (2009) from the Lviv Polytechnic National University; Assoc. Prof. (2001), Professor (2011) of computer aided design department.*

He has published more than 450 papers in international and national scientific issues and journals and he is an author of 4 monographs. The main research interests include design automation and systems modeling, Smart Home, Smart City and MEMS.

***Iryna Bazylevych**, an Associate Professor at Ivan Franko National University of Lviv, Ukraine. PhD (1992), Assoc. Prof. (2005) at Theoretical and Applied Statistics Department. She has published 53 papers in international and national scientific issues and journals and she is an author of 1 mono-monograph, 1 textbook, 1 lecture course. The main research interests include mathematical statistics, branching process, Markov process.*

***Inna Shylinska**, an Associate Professor at Ternopil National Economic University, Department of Foreign Languages, Ternopil, Ukraine. PhD (2012) in Pedagogy, specialty 13.00.04 – Theory and Methods of Professional Education. She has published 35 papers in scientific issues and journals and she is an author of 2 textbooks. The main research interests include methods of teaching English for Specific Purposes (ESP), subject field lexicon, translation skills development, applied linguistics.*