On Variability of Authors’ Style under the Influence of the Socio-Cultural Environment in the Context of Dialogue of Natural Scientific and Humanitarian Cultures

Svetlana Nikolaevna Dvoryatkina
Anjelica Aleksandrovna Dyakina

Yelets State Bunin University, 28 Kommunarov Str., Yelets, Lipetsk Region, 399770, Russian Federation
E-mail: nauka@elsu.ru

Doi:10.5901/mjss.2015.v6n5s4p167

Abstract

The article covers the problem of dialogue between the humanitarian and natural scientific cultures in solving crucial interdisciplinary issues relating to the study of variability of authors’ styles under the influence of socio-cultural conditions. The introduction highlights the problem of the study of the author’s style variability based on the dialogues of culture, its methodological principles. The principal refrain is as follows: a combination of different ways of reality cognition, including rational natural scientific and irrational humanitarian contributes to the preservation of national identity, original traditions, language, customs and spiritual and moral values of the Russian people, the basics of cultural variety in the era of globalization. The main part of the research determines the role of the dialogue of cultures in obtaining quantitative and qualitative conclusions in literature studies regarding text evaluation for stylistic identification and determination of the authorship authenticity in disputable cases; the hypothesis for the research has been suggested covering the possibility of the author’s style variability under the influence of the socio-cultural environment; socio-cultural conditions for the variability of the author’s style have been developed; theoretical aspects of the literary text attribution problem have been analyzed; the spectrum of the developed multilateral linguistic model and the parameters of the analyzed text have been expanded, the importance of the model has been proved; the position of the authors of the research has been justified highlighting the need to apply the aggregate of mathematic methods for the statistic accuracy in identifying the author’s style and the development of an integrative mathematic model based on modern computer technologies. The conclusion suggests presumable findings and perspectives related to the heuristic potential of the interrelation of the natural scientific and humanitarian sciences in solving inter-disciplinary problems of the authors’ styles variability under the influence of a socio-cultural potential for the interrelation of natural scientific and humanitarian sciences in solving inter-disciplinary problems of the authors’ style variability under the influence of a socio-cultural environment.

Keywords: interrelation of natural scientific and humanitarian knowledge, comparison of styles, statistical methods of text identification.

1. Introduction

1.1 Introduce the Problem

The era of the global culture formation and the inter-cultural dialogue development preconditions a weighted and constructive approach to the solution of tasks that Russia and the global community need to implement in the 21st century, i.e. integration of science and culture, technical and humanitarian knowledge, combination of modern achievements of the technical progress with the basic spiritual and moral values. Today’s world should rely on the integral systematic knowledge able to perceive a dialogue of different cultures and communication of their bearers.

1.2 Explore Importance of the Problem

M.M. Bakhtin developed the methodology of the dialogue of cultures. He wrote that “A dialogue is always a development and cooperation. It is always a union, not a breakdown. A dialogue is not a means; it is a goal in itself. To exist means to communicate with the help of a dialogue. When the dialogue is over, everything is over. Therefore, a dialogue cannot and should not stop (Bakhtin, 1972). While analyzing the problem of a “dialogue between Me and Other”, M.M. Bakhtin developed a theory of a “dialogue of cultures”. All great events in culture result from the dialogue of different cultures and only in the point of their intersection. Ability of one culture to master the achievements of the other is one of the sources of its existence.
We perceive a dialogue of naturally scientific and humanitarian cultures as their contact, mutual penetration and enrichment. A combination of different ways of perception of reality, different ways of perception of reality, natural scientific and irrational humanitarian, contributing to the preservation of national identity, original traditions, language, customs and spiritual and moral values of the Russian people, the basics of cultural variety in the era of globalization.

The problem of analysis and comparison of different text styles became inter-disciplinary long ago due to effective implementation of probability statistical methods. Such “non-conventional” scientific connections allowed making not only quantitative but also qualitative conclusions in literature researches devoted to stylistic peculiarities of different genres and authors. In literature studies, the test of texts for stylistic identification is required to determine the authenticity of authorship in disputable cases. Discussions on certain works by M.Yu. Lermontov, F.M. Dostoyevsky, M.E. Sallowy-Shchedrin, and M.A. Sholokhov are well known. In all such cases, the literature proofs not supported by an archive and facts base are unconvincing, subjective and excessively emotional. The case is different when mathematically accurate data is involved to support the authorship authenticity with a greater degree of possibility.

The theoretical basis for the application of the statistical methods for the study of written speech is the probability model for the occurrence of the speech act. Due to recording of the writing skills and occurrence in the brain cortex of the systems of dynamic stereotype, the efforts applied in the process of the text writing decrease and the opportunity for “automatic” writing appears. This factor allows determining the authorship of the created works.

1.3 Describe Relevant Scholarship

A.A. Markov, M.A. Marusenko, N.A. Morozov, V.P. Fomenko, T.G. Fomenko, W. Fucks, D.V. Khmelev, G. Kjetsaa, O.G. Shevelev and others worked on the issues of text tests for similarity of styles with the application of formal quantitative methods in Russian and foreign science.

A.A. Markov was one of the first researchers who applied the probabilistic statistical methods for the analysis of literature texts (Markov, 1913). In 1913, he conducted an analysis for the distribution of vowels and consonants among the first 20,000 letters of "Evgeny Onegin", a novel by A.S. Pushkin not including the hard and soft signs.

In 1915, N.A. Morozov (Morozov, 1915) suggested the methodology for calculation of the application frequency of the auxiliary parts of speech by different writers to determine the features of their individual style. The results of his calculations were presented in the form of "linguistic spectrums" (diagrams). To determine the author’s style, N.A. Morozov applied the following quantitative characteristic: a combination of a frequency of prepositions “in”, “on” and particle “not”. Morozov’s method became the basis for many researches on lexical composition of the authors’ language. However, in 1916 it was criticized by A.A. Markov in his article “On Application of a Statistical Method” (Markov, 1216). The method suggested by N.A. Morozov did not exceed the framework of the lexical and syntactical analysis of the sentence; therefore, it could not be used as a basis for a complete stylistic study.

The range of the problem under consideration further expanded due to the use of precise method. Wilhelm Fucks (Fucks, 1975) developed the transition frequency matrix for syntactic classes of words in scriptures and after the reference to the difference index indicators he managed to clearly define stylistic peculiarities of four canonic evangelic texts. However, the application of a similar coefficient for the evaluation of Old Russian works proved to be inefficient.

Development of information technologies resulted in a raised interest in the technological analysis of texts. In 1978, Norwegian Slavicist Geir Kjetsaa (Kjetsaa, 1978) suggested the methodology for authorship determination based on the text analysis with the automated development of frequency vocabularies and obtaining of statistical data. Regrettably, the development contained a mistake implicating the application of a single parameter – average length of a sentence.

In 1990, M.A. Marusenko (Marusenko, 1990) based on the characteristics of the author’s style applied the method of “recognition of images” in text attribution. The method was based on a multidimensional statistical analysis presented in the theory of “recognition of images”. Application of mathematical apparatus developed by M.A. Marusenko to the real historical and literature material demonstrated its high efficiency. Such system of recognition is resistant to fluctuations of a small volume of texts and to temporary evolution of author’s style parameters.

One of the last works based on automatic text processing is the dissertation of O.V. Shevelev (Shevelev, 2006) that suggested a new approach to the comparison of the text styles based on a two-way Fisher criteria and $\chi^2$-criteria of Pearson by frequency features, aggregation of features and their distribution. The author suggested the “Style Analyzer” program complex ensuring a full cycle for the text style analysis.

1.4 State Hypotheses and Their Correspondence to Research Design

In spite of the available range of researches on the test of text styles similarity there are still a lot of uninvestigated or
under-investigated areas. If we go deeper to the history of text attribution with the application of formal methods and analysis, we may notice that none of the existing works has covered the issue of application of a probabilistic and statistical method for the analysis of a variability (preservation) of the author's style, with the texts created in different periods of life under the influence of objective socio-cultural factors.

Having analyzed the creative biographies of the Russian writers of the 19th-20th century we noted that the authors had to travel quite often. The impressions of the trips encouraged creation of new literature works describing the picturesque images of the visited places. Let us remember the influence of Caucasus and Oriental countries on the word artists, starting from A.S. Pushkin, M.Yu. Lermontov and finishing with I.A. Bunin and S.A. Yesenin. All theorists of literature would say that the surroundings, environment, the locus influenced the creative personality and were reflected in images, stylistic units of the works; they would also give numerous examples supporting such conclusions. However, they would not demonstrate qualitative indicators relying on the mathematically accurate calculation considering such numeric frequency as the sphere of other scientific interests. At this stage, the humanitarian knowledge simply requires support from the natural scientific culture.

Apart from the trips, Russian writers had to carry out forced moves, for example under the influence of revolutions and various public uprisings. Thus, we need to determine whether the changing socio-cultural environment could affect the author's stylistic peculiarities. Such research is relevant in the cases when the writer's social and cultural environment changes significantly and forever. In this case we are talking about the representatives of the so-called "white" emigration who left Russia after the October 1917 events. The first couple of years of their life abroad Russian refugees hoped to come back and did not try to assimilate in a new environment. It is quite possible that the style of the authors of that period (most of them continued working, i.e. I.A. Bunin, B.K. Zaytsev, A.M. Remizov, I.S. Shmelev) preserved in a habitual state, well known for the readers. Later, in 1921, the situation suddenly changed and return to the motherland became impossible. This fact with a high degree of possibility could influence the condition of the author's style and the literature atmosphere of the Russian immigrant society in general.

2. Method

Comparison of certain text styles is possible based on the aggregate principles reflecting the significant features of the author's style. The following may be referred to the identifying principles: statistical characteristics (frequency of words, letters, their combinations, quantitative application of specific parts of speech, syntactical constructions, etc.). Science determined that the probability of occurrence of combination of different symbols in the text of specific authors is governed by certain consistent patterns. These may include:

a) letter pairs in their natural sequences in the text – in words (in the form they were used in the text) and gaps between them;
b) letter pairs in letter sequences in the presented word forms (vocabulary, lemmatic or initial);
c) pairs of the most generalized grammatical categories (parts of speech, conditional categories, "end of sentence" type, "abbreviation" type, etc.) in their sequences in the text sentences;
d) pairs of less generalized grammatical word classes (animate - inanimate nouns, adjectives categories, etc.).

Herewith, the main problem of formal methods for the authorship styles analysis lies in the choice of the required characteristics. According to A.A. Markov (Markov, 1913), the characteristic parameters should comply with certain requirements, such as statistical stability, large scale, distinguishing power and, consequently, may be formalized only with the help of a qualitative analysis of text units with the application of probabilistic and statistical methods.

Development of a multidimensional linguistic model for the analyzed text parameters based on frequency indicators is an open issue for now, although certain steps in this area have already been taken (Shatalova, 2012).

Based on the developed model, we suggested the selection (continuous sampling method) of the most accurate linguo-statistical parameters for the experimental processing allowing obtaining important information on the vocabulary applied by the author as well as the information on the preferred grammatical constructions. The accuracy of results will be ensured by the developed integrative mathematical model implemented based on modern computer technologies. The most widely spread statistical methods of text attribution include: t-criteria of Student, $\chi^2$-criteria of Pearson, $\lambda$-criteria of Kolmogorov-Snimov, Markov's chains, entropy approach, etc. Each method has its advantages and the area of application. It is planned that they will be applied integrally and each separately as and when necessary during the study of specific stylistic indicators.

For instance, the essence of Markov's chains technology application for the analysis of the letter pairs in their natural sequences in the text to determine the stability of the author's style is as follows. We assume that we have quite long fragments $T=t_1t_2…t$ of prose works by the same author in Russian written in different periods of his/her life. We
choose one control work among earlier works of a specific author. Furthermore, we calculate the number of inclusions of each \( n^2 \) of possible letter pairs (bigrams) in longer open texts \( T = t_1 t_2 \ldots t_l \) composed of the alphabet letters \( \{a_1, a_2, \ldots, a_n\} \). Herewith, we analyze the consecutive text bigrams, calculate the transition frequency matrix of the applied letter pairs. Based on the matrix, we develop a theoretically assumed transition frequency matrix to be applied as an evaluation of the matrix for probabilities of transition form letter to letter for the experimental literature work of the later period. The application of \( \chi^2 \)-criteria of Pearson allows determining the significance of differences between the transition frequencies of the theoretical matrix and the matrix of an experimental work of the later period. If the determined probability is high, the author’s style has not changed under the influence of external factors and vice versa. Such method proved to be accurate for natural linguistic texts. It is likely to be applied for the analysis of fiction and publicistic works. If this proves to be true, the researcher has wide opportunities for comprehension of peculiarities of the author’s style of certain creative individuals and a literature process in the specific period in the specific historical and cultural atmosphere.

3. Results

As a result of an experimental test of significance of the developed methodology, we chose an outstanding Russian prose writer, the Noble Prize winner, I.A. Bunin as the object for our research. To develop the transition frequency matrix, we selected two works by I.A. Bunin, one from his early works, “The Village” (1910), and one from his later creative stage, “The Life of Arsenyev” (1927-1929). The analysis of the letters co-occurrence (bigram of the type vowel-consonant, vowel-vowel, consonant-consonant) was carried out with the application of the transition probability matrix with the 33×33 dimension as we should bear in mind the dependence of letters occurrence on the number of preceding letters, i.e. the notion of conditional probability. The theoretically suggested transition frequency matrix is developed in accordance with the following rule: for the homogeneous Markov’s chain with the transition probability matrix transition matrix \( P \) in \( m \) steps is determined by \( P^{(m)} = P^m, m \geq 1 \) (Dvoryankina, 2013; Ventsel, 2013). Since the difference between the years of writing comprises 17 years, \( m=17 \). Herewith, we should calculate the theoretical transition frequency matrix \( P(17) \). We compare the obtained matrix with the experimental matrix developed based on the text of the later work with the help of \( \chi^2 \)-criteria of Pearson. The divergence between the frequencies at a 10% level of significance allows talking about discrepancy between the theoretical and experimental frequencies. Consequently, the author’s style has changed. In this relation, the question arises regarding the reason for the above. The answer may be found in the aggregate factors of the socio-cultural environment reflected in the author’s work and environment.

4. Discussion

The components of the socio-cultural environment affecting the author may be generalized and analyzed by the scientists-philologists to the fairly full extent (Smirnova, 1991). Moreover, there are many original sources allowing fulfilling such work (Kuznetsova, 1995; Muromtseva-Bunina, 2007). For instance, I. Bunin worked on “The Village” story under the influence of the events of the first Russian revolution (1905) and reflected the anticipation of new social disturbances. The writer who lived at that time in the central Russia perfectly knew the severities of the peasants’ life; his perception of Russian character correlated with the gloomy picture of the Russian reality. Hence, the writer’s style is determined by the most frequently applied bigrams \{ga, ha, da, ta, zs, bp... \} representing the uneasiness of the wordsmith.

The novel “The Life of Arsenyev” was created in the different socio-cultural atmosphere. In summer 1927, Galina Kuznetsova came to visit I. Bunin in Grasse. She became the last muse of the writer. The state of amorousness and magnificence of the surrounding nature (the sea, the mountains) resulted in the condition of appeasement that was reflected in the author’s style. The novel was devoted to the author’s memories of childhood, youth, and first love. Positive emotions led to development of a levelled, almost lyrical style. As a result, the following letter combinations became the most efficient in the representation of the author’s style \{le, lo, me, mi, ro, rl\}.

Such observations, coupled with qualitative indicators determined by the statistical analysis allow making the most accurate conclusions regarding the variability of the author’s style. Hence, the dialogue of natural scientific and humanitarian knowledge enriching our perception of Russian word artists, peculiarities of the creative process as a spiritual event becomes quite efficient.

5. Conclusion

The conducted research is an initial stage in the experimental study of the variability or preservation of the author’s style
based on the dialogue of cultures. We focused on only one component of our multi-parametric linguistic model – binary letter combination. The obtained results enable to make certain intermediate conclusions supporting our scientific hypothesis on the influence of socio-cultural environment on the author's style. The next stages of the study will focus on the development of an integral mathematical model with the application of a wide spectrum of statistical methods and modern programming languages and network technologies.

References


