

Yuri Tambovtsev
Novosibirsk Pedagogical University, Russia

The imitation of style: Frequency of occurrence of linguistic elements

Introduction

The goal of this report is to determine the degree of stylistic typological similarity between texts one of which was written in imitation of the original text, and intended to be similar. Our study compares the texts of different authors based on the frequency of occurrence of certain linguistic elements. In this case, the chosen typological features are English prepositions. We analyzed *Gone with the Wind*, by Margaret Mitchell, and the sequel, written in conscious imitation by Alexandra Ripley, *Scarlett*. We computed the frequency of occurrence prepositions in Ripley's sequel *Scarlett* in order to compare it to the same features in the text by Margaret Mitchell.

The stylistico-typological differences between these American writers, as well as other authors taken for this study, have never been investigated quantitatively from this point of view. This is the first attempt to study it from the point of view of spatial orientation of their texts, expressed by the prepositions.

The results obtained in this study are quite new. For one, it shows how well Alexandra Ripley imitated the style of Margaret Mitchell. It was proved elsewhere that the frequency of occurrence of the chosen grammatical features allows us to distinguish every author from every other author quite precisely (**Tambovtsev et al., 2008**).

Establishing identity or distinction is not possible in any other way; only by this method may the exact distances between the texts be obtained. The distances calculated in this research are important because they constitute the basis for placing certain British and

American authors accurately and exactly, from the standpoint of their spatial stylistic typological peculiarities.

Typological and Quantitative Foundations of the Research

Before describing our model it is necessary to point out that our approach is both typological and quantitative. By text typology we mean common characteristics of some texts which helps us put the texts in different groups or classes. Thus we must choose features which share formal quantitative characteristics in the texts and construct our comparison on them. It is desirable that the chosen features should be basic and expressed numerically. Lindsay Whaley proposes to call it partial typology since it takes into consideration only some features, while holistic typology should consider all the features (**Whaley, 1997: 23**). Indeed, our typology is partial because we took only 25 features, though it is possible to take an unlimited number of features. Our typology is numerical, since every feature is expressed quantitatively. In this case, typology of texts is very close to the notion of the text classification.

In fact, all the models in the Humanities as well as in Natural Sciences which use quantitative features fall under the classification of German philosopher Immanuel Kant (1724 – 1804), who stated that everything in this world has a numerical embodiment. He meant that every linguistic or other object can be substituted by its measurements. A. V. Jakushev, developing the ideas of I. Kant, calls upon researchers to take into consideration the numerical relationships which can reflect a system fully (**Jakushev, 2005: 25**).

Unfortunately, in linguistics numerical relationships are not given the proper attention. A linguist must not forget that quantity goes over into quality as G.W.F. Hegel pointed out in his law, “The Transition of Quantity into Quality”. This law states that the accumulation of some entity may turn this entity into some other when the accumulated number is big enough (**Hegel, 1978: 219**). In this case, when the number of some features (here, prepositions) in

some text is big enough, while in the other text it is not so big, the text may be seen as a different text. The task of the linguist is to detect this transition by some criterion of mathematical linguistics. Here, we do it with the help of the Chi-square test.

If the texts of one and the same author show less distance in comparison to those of the other authors, then one can say that the features which we have chosen are reliable and invariant. Our hypothesis is that one and the same writer has his own style, which depends upon his thinking process. Therefore, the assumption is that the text of some author is identical to some other text written by this author, i.e, they are samples of the same text. Consequently, the texts of two different authors must be more different, i.e, different to a higher degree, than those of one and the same author. Two authors whose style is similar must have little distance between them. If two different authors are closer than one and the same author, then our tool is sharp enough to distinguish the two texts. In this case, the topic of the description may influence the style more than the manner of the description. If one writer wants to imitate the style of the other writer, then his text may be closer to the author in question. The data in the tables given below show that our assumption is correct, i.e. the texts of some authors are identical while the texts written by other authors are different in the incidence of the chosen features. Tables 2 – 9 compare the results for works of given authors with other of their texts and with other authors.

As we have already stated elsewhere, our chosen features characterise a given text as a real object. In this case we consider the text of a certain author to be such an object, to which statistical methods may be applied. These objects may be similar or different to some certain degree. Our model reflects this object in the wholeness of all its features from the point of view of their frequency of occurrence. Therefore our model reflects both the degree of similarity and the degree of difference by using quantitative characteristics (**Tambovtsev, 2003: 7 — 23; 2007; 2008**).

Every scholar must observe the rules of commensurability, that one must compare only commensurable data. For this reason the text sample is advised to be fixed at the same volume. In fact, this volume should not be less than 10,000 words. The same ideal sample volume of 10,000 words was taken from texts by the chosen British and American writers to maintain the principle of commensurability.

We must again underline the importance of the samples being equal. When they are equal, the linguistic conclusions are more liable to be correct, even if a linguist does not pay attention to the type of distribution of the language units. However, we have noticed that when the sample is big enough, the distribution is always normal (**Tambovtsev, 2003**). At the same time, one must use the statistical criteria in the strict manner in which they should be used in mathematical statistics.

We have used "Chi-square" criterion values as the measure of similarity. Similarity is inverse to the distance. The more similar two texts are, the less the distance between them. On the contrary, when the texts are not similar, then the distance between them is great. The distance is directly proportional to the value of the "Chi-square", since it measures the difference. This is the reason one can state that the lesser the difference, the greater the similarity between the texts.

One can see that we use the terms "similarity" and "distance" in their usual meaning reflected in any dictionary, but not in the strict mathematical definition. So, we understand similarity and distance as a measure of space between two points, places or objects (**Hornby, 1984: 177**). It is the degree or amount of separation between two objects (**Webster, 1965: 242**). Thus, one can say that the objects (in this case - texts) are closer if they are more similar.

The other main assumption in this work is that every text sample of an author is a separate object with certain values of the selected features which can be measured to

determine the distances from other texts. One can assume that the wholeness of all the objective features of some text, which may be called the style of the text. In this case we do not speak of the meaning of the text. Our approach is more formal, i.e., our approach takes into consideration the concrete forms of the text, embodied in some forms of word. We measure the numerical characteristics of a certain text to compare it with some other texts by the same or different authors. These numerical characteristics are the frequencies of occurrence of certain chosen forms, or features. The values of these features construct the stylistic mosaic of this or that text.

Gustav Herdan posed in his famous book *The Advanced Theory of Language as Choice and Chance* the question why it is customary in linguistics to consider literary texts only with regard to the message they are meant to transmit but not with regard to their quantitative characteristics (**Herdan, 1966: 2 - 3**). One should agree with him that the text is an individual creation of the mind, which is comparable with other such creations and thus can be regarded as an instance of a production process. He correctly states that what were regarded before as quite unique events, the products of willful creation, appear now, when studied quantitatively, as mere variants of the typical expenditure of linguistic material (**Herdan, 1966: 3 – 4**). When reading a book, we may feel that the style of this author is closer to the style of some other. More often than not, we are not aware why we think so. We can try to formalise our intuition to realise why we feel two authors are different.

The corpus of texts used for the research

The text of *Gone with the Wind* by Margaret Mitchell has been compared to those of Alexandra Ripley, Agatha Christie, Daniel Defoe, Charles Dickens, Theodore Dreiser, F. Scott Fitzgerald, Jack London, George Orwell, Jonathan Swift, Virginia Woolf, and others.

The list of all the writers and their texts is provided in the appendix.

The research method

During our research we used methods of computational linguistics which are described in detail below. To measure the similarity of the frequency values, we use the criterion of Chi-square. The usual approach was taken, which can be found in all the books dealing with the applications of mathematical statistics (e.g., **Brownlee, 1949: 50; Butler. 1985: 112 – 114; Herdan, 1966: 37 – 38; Tambovtsev, 2003**). It is well-known that the Chi-square criterion allows us to compare the frequencies one actually observes with the expected ones. Thus, the requirement for the size of the samples is quite strict: they must be equal. As is mentioned below, we take the same sample size, equal to 10 000 words, for each writer. For a linguist it is hard to understand how to calculate the theoretically expected value, but here, the expected value is believed to be the mean between two values of the frequencies of the grammar forms found in the texts of a pair of writers. Thus, we calculated the Chi-square by the usual formula which we also used in our previous works (**Tambovtsev, 2003; 2007; 2008-a; 2008-b**).

We took advantage of the property of the Chi-square which allows us to add all the values. One should remember then that the number of the degrees of freedom could change. When we compare two series, the degree of freedom is equal to the number of the members in it minus one. In fact, we may use the Chi-square test when we do not know or when we are unwilling to assume that the distribution of the feature values are normal. The Chi-square distribution is nonparametric.

This means that we can use it when the distribution is not normal (**Owen et al., 1990: 386**). In fact, we may use the Chi-square criterion in any case, without bothering about the type of distribution, while one can use the Student's criterion only if the distribution is normal. This is the reason we prefer the Chi-square criterion. Let us not explain how we use

the Chi-square test here. The procedure is described in detail elsewhere (**Tambovtsev, 2003: 26 — 27; 2007; 2008-a; 2008-b**).

Results of the investigation.

One can't help agreeing with Lubomir Dolezhal and Richard W. Bailey, who think that not only linguistics, but also many other fields in the humanities, have survived and profited from the application of mathematical models and techniques. Stylistics has become the crossroads for linguists and mathematicians to yield new models. The statistical investigation of texts can explain the intrinsic features inherent in the text (**Dolezhal et al., 1969: VII**). We agree with the definition of style given by Wermer Winter, who believes that style may be said to be characterized by a pattern of recurrent selections from the inventory of optional features of a language (**Winter, 1969: 3**).

Along with Lubomir Dolezhal, we adhere to the foundations of the statistical theory of style, which considers style to be a probabilistic phenomenon. We can regard style as a preference for one or another mode of expression. The overall character of style is called forth by the degree of presence (or absence) of a certain mode of expression, rather than by its exclusive use or complete suppression (**Dolezhal, 1969: 10 - 11**). In other words a probabilistic approach takes into account the frequency with which this or that feature is used in the text (**Tambovtsev, 1997: 171 - 172**).

Dolezhal is not right to assume that a numerical stylistic theory can only be said to account for stylistic differences between texts (**Dolezhal, 1969: 11**). In fact, it can provide different measures, which allow us to construct a taxonomy of styles. Nevertheless, we agree with Dolezhal that style-forming processes are to be considered a fundamental component of linguistic performance: style originates in the process of producing (encoding) a text by a certain author (**Dolezhal, 1969: 11**). In our opinion, a probabilistic approach may reveal a substantial part of linguistic performance.

Speaking about style as a statistical concept, Gustav Herdan defines style as the general characteristics of a person's way of expressing himself in language. What Herdan's definition implies is that no matter what somebody who is said to have a style of his own is writing about, the text will bear the imprint of the personality of the writer. Thus, for G. Herdan, "style" is used in the sense of a subconscious factor, which the writer must obey. It implies that the linguistic performance of a person is much a matter of particular mental make-up of the individual (**Herdan, 1966: 70**).

A specialist in the studies of authorship attribution, David L. Mealand, remarks that anyone familiar with literature knows that there are differences that we sense almost instantly and instinctively when we move between texts by different authors. We just know that their style is different. Different authors write in different ways, i.e., they express themselves in different ways (**Mealand, 1999: 479**). Thus, authorship attribution methods allow us to define the degree of affinity between authors. So, the method of authorship attribution studies is close to our study method if we look at the problem of the distance between the authors from the following angle: the lesser the distance, the more similar the texts under study, i.e., the more affinity between the authors. If the texts are similar enough, one can claim that the author is the same. For this reason we took texts of different authors and the texts of the same author. We try to contribute another attribution method to the many other methods which already exist. The use of these methods may be found elsewhere (**Holmes, 1985: 328 - 329; Tambovtsev et al, 2003; 2007; 2008-a; 2008-b**). However, we can't help agreeing with Joseph Rudman and other linguists that for every paper announcing an authorship attribution method that "works", there is a counter paper that points out real or imagined crucial shortcomings (**Tambovtsev et al., 2003; 2007; 2008-a; 2008-b**). It happens so, because it is possible to find some better features as the basis for analysis. Nevertheless, we believe that the quantitative values of the text (e.g. Gerund, Participle 1, Verbal noun, or prepositions)

may be reliable features to judge the thinking process of different authors. The result of this thinking process may be some special features of the text, which have different values. The features which we selected do not seem to depend on this or that situation being described in the text, but they seem to show the peculiar way this or that author depicts the reality. Our selected features do not intersect.

It was proved that the texts of different authors can be distinguished by different linguistic units. In our previous research we distinguished between the authors on the basis of the frequency of occurrence of the functions of Gerund and Participle 1 in the English sentence (**Tambovtsev et al., 2003**). It was also proved that prepositions can serve as distinctive features, to distinguish between the authors (**Tambovtsev et al., 2007; 2008**). It is necessary to agree with Boris A. Il'ish, who emphasised that prepositions are an important element of the structure of English, Russian, and many other world languages. However, they are of vital importance in those world languages which have no case system for their nominal parts of speech, like Modern English (**Il'ish, 1971: 149- 155**). In any language so far studied, spatial expressions are expressed in a few prepositions but show a wide range of uses (**Landau et al., 1993**). The small number of prepositions makes it simpler to study the spatial distinctions in language (**Tambovtsev, 2008-a: 1-8**). On the one hand, the intuitive simplicity of the chosen prepositions, “*above, about, across, against, after, along, at, away, before, between, down, for, from, in, into, off, on, out, over, to, under, upon, with, without*”, would indicate a correspondingly simple semantic analysis. On the other hand, however, their wide range of usage seems to take into account their meaning (**Garrod et al., 1999**). Nevertheless, the prepositions preserve their basic meanings, which can be found in any dictionary (MEDFAL, 2002; OALDCE, 1998). These are also reflected in different grammars (**Gude et al., 2002: 138; Murphy, 1995: 240 — 241**).

In the light of the quantitative and typological approaches the present paper addresses the question **whether** the representation of the frequency of use of the prepositions (*at, before, in, on, under, etc.*) is geometric enough. How (i.e. what prepositions) and how often (i.e. their frequency) an author reflects spatial functional relations between the objects in the situations described in **their** texts. In attempting to answer this question we report the data on the frequency of occurrence of prepositions that measure the authors' outlook on the surrounding world. This compares approaches that concentrate on a geometric characterisation of the core meaning of the chosen prepositions. Accounts of the semantics of locatives, such as *at, above, in, on, under, etc.* differ in terms of the nature of the representation assumed to underlie their meaning. Strictly speaking there are two kinds of the accounts of space: geometric accounts, which treat the underlying representation in terms of basic geometric relations, and functional accounts, which assume that the prepositions reflect functional or physical relations between objects in the world (i.e. *with, without, etc.*), but we shall not distinguish between them. Besides, the traditional approach to the semantics of spatial prepositions is to treat them as expressing geometric relations (**Crangle et al., 1989: 400**).

Thus, each preposition is associated with a spatial feature, which may be expressed in terms of geometric or topological relations such as enclosure and spatial contact. The definition given for the preposition *in* points to the fact that its meaning shows that something must be included or enclosed by something. For the preposition *on*, the assumed representation is one of spatial contact of the surfaces. Thus the semantic representation of the prepositions is primarily geometrical, expressing enclosure or spatial contact. In the same manner we can find the meanings of the other prepositions taken for this study in different dictionaries (e.g. see: **Hornby, 1984; MEDFAL, 2002; OALDCE, 1998; Webster, 1965**).

The data in the tables provided below (c.f. Tables 1-9) show that prepositions in the texts of one and the same author are distributed more similarly. Thus, the distances are less than

between the texts of different writers. This indicates that the chosen 25 English prepositions may be considered reliable features to distinguish different authors. The data in Tab. 1 reflect the distribution of each of the 25 prepositions in selected texts.

First of all, we shall concentrate on measuring the distances between the text of *Gone with the Wind* by Margaret Mitchell and the text of *Scarlett* by Alexandra Ripley (Tab. 2). The similarity between them is great because the distance is little, i.e. 0.39. One can see the distances between the text in question by Margaret Mitchell and the texts by some other authors. The data of Tab. 1 indicates that the second most similar to it is the text of *The Titan* by Theodore Dreiser (1.06). Judging by the distances we can state that they are more than twice as distant, i.e., less similar. The text by Charles Dickens is farther away – 1.59. This may be explained by the fact that Margaret Mitchell is a female writer. The text of *Robinson Crusoe* by Daniel Defoe is also far away – 1.74, as well as that of *Gulliver's Travel* by Jonathan Swift – 2.45. These distances may be explained by the time span. Differences may be caused also by the style or genre of the text. Thus, for instance, the political text by Winston Churchill is also quite far away – 1.96. Oral dialogues in the films also diverge much (1.89 and 1.94). For some reason the style of the fairy tale *Ali Baba* has the greatest distance (2.95). Looking at the text of *Ali Baba*, we realized that it is rather obsolete and picturesque. It followed the original “Arab” eloquent style.

In order to see if it is not just a coincidence that the text of *Scarlett* by A. Ripley is so similar to that of M. Mitchell, we compared it to some other authors (Tab. 3). One can see that the distances between the text of *Scarlett* by A. Ripley is much more similar to that of *Gone with the Wind* by M. Mitchell. At the same time the distances with the other authors are bigger (c.f. Tab. 3).

We must consider the texts of one and the same author to find out if the distances between the texts of his books are less than to the other authors. We can illustrate it by the

texts of Theodore Dreiser (Tab. 4), F.Scott Fitzgerald (Tab. 5), Jack Eden (Tab. 6), George Orwell (Tab. 7), Oscar Wilde (Tab. 8) and Virginia Woolf (Tab. 9). In all the tables we can observe the common tendency: the texts of one and the same author are more similar to each other than to other authors.

If we measure the distances between text of *Tales from the Jazz Age* by F. Scott Fitzgerald (Tab. 5) and his text *Beautiful and Damned* we receive the value 0.16. This is the shortest distance among all the texts measured. The next closest text is *This Side of Paradise* with the value 0.37. The greatest distance is the one between *Tales from the Jazz Age* by F. Scott Fitzgerald and the text of the fairy tale “Ali Baba” (2.17). This may be because the style of this fairy tale is rather old fashioned and obsolete. The political text of W.S. Churchill “Malakand Field Force” with the value – 1.57, also shows a big distance from *Tales from the Jazz Age* by F. Scott Fitzgerald.

The shortest distance between the books by Dreiser is 0.27 (see Tab. 4) while the greatest distance is 1.42. It is between *The Titan* by Theodore Dreiser and *The Picture of Dorian Gray* by Oscar Wilde. In the same way it is possible to see the shortest and the longest distances between *Martin Eden* by Jack London (Tab. 6), *A Clergyman’s Daughter* by George Orwell (Tab. 7), and *The Picture of Dorian Gray* by Oscar Wilde (Tab. 8), or *Jacob’s Room* by Virginia Woolf (Tab. 9). It is interesting to consider the distances between one of the female authors, Virginia Woolf, and the male authors. The nearest to her text is the text of Theodore Dreiser, with the texts by D. Defoe and O. Wilde to follow. The least similar to her text is the text by J. Swift. Let us again point out that the distance is measured by the criterion “Chi-square” divided by the critical value of it taken from the statistical tables (Tambovtsev, 2003). It can easily be found elsewhere.

The data in the tables (Tab 1 – 9) help us to make the following conclusions:

- 1) The imitation of Mitchell's style by A. Ripley was a success. By the critical value of Chi-square, they enter the same general sample as if written by one and the same author.
- 2) The frequency of prepositions is a reliable tool to differentiate between the authors;
- 3) The texts of one and the same author are closer to each other than to the texts of other authors;
- 4) The old and modern authors are in general farther away from each other than those of the authors who live approximately at the same time span;
- 5) The political style (like that of W.S. Churchill) is usually farther away.
- 6) The oral dialogues use prepositions quite differently

Tables

Table 1.

Samples of the frequency of occurrence of prepositions in some texts of different writers, in % to all the words in the text.

MMG — Margaret Mitchell «Gone with the Wind», 422900 words.

ARS — Alexandra Ripley «Scarlet», 268181 words.

DDR — Daniel Defoe «Robinson Crusoe», 117960 words.

JSG — Jonathan Swift «Gulliver's Travel», 105344 words.

CDC — Charles Dickens «A Christmas Carol», 28485 words.

TDT — Theodore Dreiser «The Titan», 196526 words.

FAB — Film: American Beauty. Dialogue, 20111 words.

ABF — Ali Baba. Fairy Tale, 12127 words.

#	Preposition	MMG	ARS	DDR	JSG	CDC	TDT	FAB	ABF
1.	in	1.39	1.19	1.49	1.76	1.47	1.66	1.16	1.32
2.	with	0.79	0.67	0.94	0.85	0.80	0.75	0.45	1.10
3.	for	0.76	0.68	0.86	0.73	0.58	0.69	0.40	0.58
4.	at	0.56	0.56	0.50	0.57	0.46	0.62	0.96	0.42
5.	on	0.54	0.56	0.43	0.42	0.44	0.49	0.75	0.33
6.	from	0.38	0.33	0.37	0.42	0.22	0.23	0.35	0.31
7.	about	0.30	0.24	0.18	0.21	0.14	0.16	0.19	0.11
8.	out	0.23	0.21	0.20	0.18	0.20	0.19	0.41	0.22
9.	by	0.22	0.17	0.52	0.73	0.33	0.37	0.14	0.42
10.	to	0.21	0.25	0.42	0.41	0.16	0.24	0.21	0.73
11.	into	0.20	0.14	0.21	0.25	0.16	0.13	0.17	0.21
12.	down	0.16	0.10	0.10	0.08	0.14	0.04	0.16	0.10
13.	over	0.14	0.13	0.08	0.07	0.03	0.15	0.10	0.10
14.	before	0.11	0.10	0.14	0.10	0.11	0.07	0.05	0.11
15.	after	0.11	0.11	0.17	0.15	0.05	0.13	0.05	0.28
16.	off	0.08	0.05	0.06	0.03	0.07	0.01	0.15	0.11
17.	under	0.07	0.05	0.05	0.07	0.02	0.05	0.05	0.07
18.	away	0.06	0.05	0.06	0.01	0.03	0.02	0.03	0.04
19.	against	0.06	0.04	0.06	0.08	0.02	0.04	0.06	0.03
20.	without	0.06	0.05	0.09	0.11	0.05	0.05	0.01	0.05
21.	across	0.04	0.03	0.00	0.00	0.01	0.00	0.12	0.01
22.	upon	0.04	0.00	0.37	0.33	0.32	0.06	0.00	0.41
23.	between	0.04	0.02	0.05	0.05	0.02	0.04	0.01	0.00
24.	above	0.02	0.02	0.03	0.06	0.02	0.01	0.02	0.01
25.	along	0.02	0.03	0.03	0.09	0.02	0.02	0.03	0.02

Tab.2

The Distances between the text of the book “Gone with the Wind” by Margaret Mitchell and some other texts by different authors.

#	Author	Distance
1.	Alexandra Ripley. Scarlett.	0.39
2.	Th. Dreiser. The Titan.	1.06
3.	Ch. Dickens. A Christmas Carol.	1.59

4.	D. Defoe. Robinson Crusoe.	1.74
5.	Film: American Beauty. Dialogues.	1.89
6.	Film: Cruel Intentions. Dialogues.	1.94
7.	W. Churchill. Malakand Field Force.	1.96
8.	J. Swift. Gulliver's Travel.	2.45
9.	Ali Baba. Fairy Tale.	2.95

Tab.3

The Distances between the text of the book “Scarlet” by Alexandra Ripley and some other texts by different authors.

#	Author	Distance
1.	Margaret Mitchell. Gone with the Wind.	0.39
2.	Theodore Dreiser. The Titan.	1.03
3.	Jack London. Martin Eden.	1.06
4.	Virginia Woolf. Jacob's Room.	1.20
5.	Film: American Beauty. Dialogues.	1.45
6.	Film: Cruel Intentions. Dialogues.	1.53
7.	Winston Churchill. Malakand Field Force.	1.82
8.	Daniel Defoe. Robinson Crusoe.	2.28
9.	Jonathan Swift. Gulliver's Travel.	3.04
10.	Ali Baba. Fairy Tale.	3.17

Tab.4

The Distances between the text of the book “The Titan” by Theodore Dreiser and some other texts by the same and some other different authors.

#	Author	Distance
1.	Th. Dreiser. The Financier	0.27
2.	Th. Dreiser. Twelve Men	0.38
3.	Th. Dreiser. Sister Carrie	0.77
4.	A. Ripley. Scarlett.	1.03
5.	M. Mitchell. Gone with the Wind.	1.06
6.	V. Woolf. Jacob's Room	1.13
7.	O.Wilde. The Picture of Dorian Grey	1.42
8.	Ch.Dickens. A Christmas Carol.	1.42
9.	W. Churchill. Malakand Field Force.	1.45
10.	D.Defoe. Robinson Crusoe.	1.58
11.	J. Swift. Gulliver's Travel.	1.79
12.	Ali Baba. Fairy Tale.	2.70
13.	Film: American Beauty. Dialogues.	2.92
14.	Film: Cruel Intentions. Dialogues.	3.29

Table 5.

The Distances between the text of the book “Tales from the Jazz Age” by F. Scott Fitzgerald and some other texts by the same and different authors.

#	Author	Distance
1.	Fitzgerald. Beautiful and Damned	0.16
2.	Fitzgerald. This Side of the Paradise	0.37
3.	Fitzgerald. Flapper and Philosophers	0.39
4.	Fitzgerald. The Great Gatsby	0.77
5.	Ch.Dickens. A Christmas Carol	1.09
6.	D.Defoe. Robinson Crusoe	1.54
7.	W.S.Churchill. Malakand Field Force.	1.57
8.	O.Wilde. The Impotence of Being Ernest	1.71
9.	Film: American Beauty. Dialogues.	1.72
10.	Film: Cruel Intentions. Dialogues.	2.01
11.	J.Swift. Gulliver’s Travel	2.01
12.	Ali Baba. Fairy Tale.	2.17

Table 6.

The Distances between the text of the book “Martin Eden” by Jack London and some other texts by the same and some other different authors.

#	Author	Distance
1.	Jack London. White Fang	0.54
2.	Jack London. The Call of the Wild	0.74
3.	Ch.Dickens. A Christmas Carol	1.06
4.	Agatha Christie. The Mysterious	1.20
5.	Swift. Gulliver’s Travel	1.34
6.	Churchil. Malakand Field Force.	1.37
7.	Ali Baba. Fairy Tale	1.98

Table 7.

The Distances between the text of the book “A Clergyman’s Daughter” by George Orwell and some other texts by the same and some other different authors.

#	Author	Distance
1.	Orwell. Down and out in Paris and London	0.45
2.	Orwell. 1984	0.77
3.	Orwell. Animal Farm	0.86
4.	Ch. Dickens. A Christmas Carol	1.18
5.	D.Defoe. Robinson Crusoe	1.63
6.	Film : American Beauty. Dialogues.	1.74
7.	Film:Cruel Intentions. Dialogues.	1.75
8.	W.S.Churchil.Malakand Field Force	1.81
9.	J.Swift. Gulliver’s Travel	2.31
10.	Ali Baba. Fairy Tale	2.54

Table 8.

The Distances between the text of the book “The Picture of Dorian Gray” (PDG) by Oscar Wilde and some other texts by the same and some other different authors. Chapters 1 – 3.

#	Author	Distance
1.	O. Wilde. PDG. Chapters 4 - 6	0.77
2.	O. Wilde. PDG. Chapers 7 - 9	0.82
3.	Th. Dreiser. The Titan	1.42
4.	V. Woolf. Jacob’s Room	1.56
5.	Martin Eden	2.48
6.	D. Defoe. Robinson Crusoe	2.76
7.	J. Swift. Gulliver’s Travel	3.38

Table 9.

The Distances between the text of the book “Jacob’s Room” by Virginia Woolf and some other texts by the same and some other different authors.

#	Author	Distance
1.	V. Woolf. Orlando	0.39
2.	V. Woolf. To the Lighthouse	0.43
3.	V. Woolf. Mrs. Dalloway	0.60
4.	V. Woolf. Night and Day	0.99
5.	Dreiser. The Titan	1.13
6.	D. Defoe. Robinson Crusoe	1.53
7.	O. Wilde. Picture of D. Grey	1.56
8.	J. Swift. Gulliver’s Travel	2.75

Appendix

The material used in this research was taken from the following books:

Ali Baba. A Fairy Tale

Americal Beauty: A film Script

Christie - Agatha Christie. 1) "Sanctuary" and "Strange Jest," from Miss Marple's Final Cases
2) Curtain 3) The Mysterious Affair at Styles, Glasgow: William Collins Sons Ltd, 1979

Cruel Intentions: A film Script

Churchill – W.S. Churchill: Malakand Field Force

Defoe - Daniel Defoe. Robinson Crusoe. Gutenberg electronic version

Dickens - Charles Dickens: A Christmas Carol, London: Penguin books, 1977

Dreiser - Theodore Dreiser: 1) Sister Carrie 2) The Titan 3) The Financier 4) Twelve Men,
New York: Signet classic, 1980

Fitzgerald - F. Scott Fitzgerald: 1) The Great Gatsby 2) Tales from the Jazz Age 3) The
Beautiful and Damned 4) Flappers and Philosophers 5) This Side of Paradise, New York:
Charles Scribner's sons, 1953

London – Jack London: Martin Eden

Orwell – George Orwell: 1) A Clergyman's Daughter 2) 1984 3) Animal Farm 4) Down and
Out in Paris and London

Swift - Jonathan Swift: Gulliver's Travels. Wordsworth Classics, 1992

Wilde – Oscar Wilde: 1) The Picture of Dorian Gray 2) The Importance of Being Ernest

Woolf – Virginia Woolf: 1) Jacob's Room 2) Mrs. Dalloway 3) Night and Day 4) Orlando: A
biography

References

- Brownlee, 1947 – K. A. Brownlee. *Industrial Experimentation*. – Oxford: Oxford University Press, 1947.
- Butler, 1985 - Butler, Christopher. *Statistics in Linguistics*. - Oxford: Basil Blackwell, 1985.
- Crangle et al., 1989 - Crangle C. and Suppes, P., 1989. Geometric Semantics for Spatial Prepositions. - In: *Midwest Studies in Philosophy* 14, p. 399–421.
- Dolezhel et al., 1969 - Dolezhel, Lubomir, Bailey, Richard W. (editors) *Statistics and Style*. - New York: Elsevier Publishing company, Inc., 1969.
- Dolezhel, 1969 - Dolezhel Lubomir. A framework for the statistical analysis of style. - In: *Statistics and Style*. - New York: Elsevier Publishing Company, Inc., 1969, p. 10 – 25.
- Garrod et al., 1999 - Simon Garrod, Gillian Ferrier and Siobhan Campbell “*In*” and “*on*”: investigating the functional geometry of spatial prepositions. - *Cognition* 72 (2, 30 September 1999), p. 167-189.
- Hegel, 1978 – Georg Wilhelm Friedrich Hegel. *Gesammelte Werke*. Bd 11. - Hamburg: Felix Meiner Verlag, 1978.
- Herdan, 1966 - Herdan, Gustav. *The Advanced Theory of Language as Choice and Chance*. - Berlin: Springer-Verlag, 1966.
- Holmes, 1985 - Holmes, David. The analysis of literary style - a review. - In: *The Journal of the Royal Statistical Society, Series A*, 148 (4), 1985, p. 328 - 341.
- Hornby, 1984 - Hornby A. S. with the assistance of Christina Ruse. *Oxford Student's Dictionary of Current English*. - Oxford: Oxford University Press, 1984.
- Jakushev, 2005 - Jakushev, A. V. *Philosophy. Lectures*. – Moskva: Prior-Izdat, 2005.
- Landau et al., 1993 - Landau B. and Jackendoff, R. ‘What’ and ‘where’ in spatial language and cognition. - In: *Behavioural and Brain Sciences* 16, 2, 1993, p. 217–265.
- Il'ish, 1971 — Il'ish, Boris Aleksandrovich. *The Structure of Modern English*. - Leningrad: Prosveshchenie, 1971. - 366 pages.
- Mealand, David. Style, genre, and authorship in Acts, the Septuagint, and Hellenistic historians. - In: *Literary and Linguistic Computing*, Vol.14, # 4, 1999, p.479 - 505.
- MEDFAL, 2002 — *Macmillan English Dictionary For Advanced Learners*. - Oxford: Macmillan Publishers Limited, 2002.
- OALDCE, 1998 – *Oxford Advanced Learner's Dictionary of Current English*. - Oxford: Oxford University Press, 1998.

- Owen, 1990 -Owen F., Jones R. Statistics. - London: Pitman, 1990.
- Tambovtsev et al., 2003 - Tambovtsev, Yuri. Ludmila Tambovtsev. Juliana Tambovtsev. Some stylistic typological distances between the prose of some British writers. - In: *Studia Anglica Posnaniensia*, # 39, 2003 (Poznan'), p. 247 - 261.
- Tambovtsev et al., 2007 - Tambovtsev, Yuri, Ludmila Tambovtseva and Juliana Tambovtseva. The frequency of linguistic units as an indicator of typological similarity between texts. - In: *California Linguistic Notes (USA)*, Vol. XXXII, # 2, Spring, 2007, p. 1 - 28.
- Tambovtsev et al., 2008 - Tambovtsev, Yuri, Yuliana Tambovtseva and Ludmila Tambovtseva. Krylov and the Silf letters: The question of authorship. - In: *California Linguistic Notes*, Vol. XXXIII, # 2, Spring, 2008, p.1 – 34.
- Tambovtsev et al, 2008-b - Тамбовцев Ю.А., Тамбовцева А.Ю., Тамбовцева Л.А. Типология распределения некоторых лингвистических единиц в тексте как показатель авторства текста. - *Вестник омского университета*, № 2 (48), 2008, с. 88 - 96.
- Webster, 1965 - Webster's Seventh New Collegiate Dictionary. - Springfield, Massachusetts: Merriam Company Publishers, 1965.
- Whaley, 1997 — Whaley, Lindsay J. Introduction to Typology: the Unity and Diversity of Language. - London: SAGE Publications, 1997.
- Winter, 1969 - Winter, Werner. Styles as dialects. - In: *Statistics and style*. - New York: Elsevier Publishing Company, Inc., 1969, p. 3 – 9.