Yuri Tambovtsev, Ludmila Tambovtseva, and Juliana Tambovtseva
Novosibirsk Pedagogical University

The Frequency of Linguistic Units as an Indicator of Typological Similarity between Texts

Introduction

1.   Goal

The goal of this report is to determine the degree of stylistic typological similarity between texts, based on the frequency of occurrence of certain linguistic units. In this case such an indicator of the typological similarity between the texts in the prose of several American and British writers is the frequency of different functions of the Gerund, Participle 1 and Verbal Noun in the sentence.

2.   The Topicality and Actuality of the Report.

The stylistic and typological differences between the authors, male and female British and American writers, taken for this study, have never before been investigated quantitatively.

3.   Scientific Novelty.

Thus the results of the paper are scientifically new. The occurrence of the functions of the Gerund and the other chosen grammatical features allow us to distinguish every author from any other author quite precisely.

4. Theoretical and Practical Importance.

The results of the research are important since they may help to place certain British and American authors accurately from the standpoint of their stylistic typological peculiarities.

5. The corpus used the Research.

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

Christie 1 - Agatha Christie. "Sanctuary" and "Strange Jest," from *Miss Marple's Final Cases*. - Glasgow: William Collins Sons Ltd, 1979.

Cristie 2 - Agatha Christie. *Curtain*. Agatha Christie. Glasgow: William Collins Sons Ltd, 1979.

Colebatch - Hal Colebatch *The earthquake lands*. London: Angus and Robertson, 1990.

Defoe - Daniel Defoe. *Robinson Crusoe*. Gutenberg electronic version.

Dickens - Charles Dickens. *Great Expectations*. Penguin books, 1977.

Dreiser - Theodore Dreiser. *Sister Carrie*. New York: Signet classic, 1980.

Fitzgerald - F. Scott Fitzgerald. *The Great Gatsby*. New York: Charles Scribner's sons, 1953.

Hemingway - Ernest Hemingway. *The Sun Also Rises*. New York: Collier books. Macmillan Publishing Company, 1986.

Maugham 1 - William Somerset Maugham. "The force of circumstance"
Maugham 2 - William Somerset Maugham "Flotsam and Jetsam"
Maugham 3 - William Somerset Maugham "The Creative impulse"
    from "Selected Stories". Moskva: Menedzher, 2000.

McBain - Ed McBain. *Fuzz.* - London: Doubleday and Co. Inc., 1974.

Rowling - Joan K. Rowling. *Harry Potter and the Chamber of Secrets*. New York: Scholastic Inc., 2000.

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

Thackeray - William Thackeray *Vanity Fair*. Oxford University Press, 1983.

Updike - John Updike. *Rabbit, Run*. New York: Fawcett Crest, 1988.

6. The Method of the Research.

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 like myself 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 following formula:

$$\chi^2 = \Sigma \; \frac{(\,O - E)^2}{E}$$

where O is the observed value (frequency)and E the expected value.

We took the 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. Where we compare two series it is equal to the number of the members in it minus one. Actually, 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 non-parametric. It 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 explain how we use the Chi-square test here. The procedure is described in detail elsewhere (Tambovtsev, 2003: 26 - 27). Let's take the third line of Table 1: the function of the gerund as an attribute in the preposition to a noun. We found 18 occurrences of it in the first 10 thousand words of *Sister Carrie* by Theodore Dreiser and 2 occurrences in Agatha Christie "Sanctuary". In order to calculate the value of Chi-square test, we must find the expected value,

i.e. the mean. Here the mean is 10. Next we must subtract it from each value, i.e. 18 - 10 = 8, and raise the remainder to the square, 8 x 8 = 64. Then we must divide it by the mean i.e., 10. So, 64/10 = 6.4. After that we must multiply it by 2, i.e., 6.4 x 2 = 12.80. The same procedure must be done with the other values in the other lines, and so on. We have chosen the level of significance 0.10 or 5%. The degrees of freedom are calculated by n-1, where n is the number of the features compared (here we have two authors, thus n = 2). Thus, we must look the Critical Value in the table of Critical Values of Chi-square test with 1 degree of freedom: it is 3.84 (Butler, 1985: 176). If the received value of the Chi-square is less than 3.84, then there is no statistical difference between these two authors according to this particular function. In this case, the obtained value is much greater than the theoretical value (c.f. 12.80 and 3.84). However, it is not recommended to use the Chi-square criterion when one of the values is less than 5. In this case it is 2 for Agatha Christie. In such a case the result may be too liberal, and may show no difference in cases where differences may exist.

As we can observe from the tables, it is possible to add all the values of the Chi-square in order to receive the total value of similarity of one writer to the other. Thus, we can obtain the total between Theodore Dreiser and F. Scott Fitzgerald, which is equal to all the values of Gerund (17.70), Participle-1 (4.81) and Verbal Noun (1.50): the total is 24.01 (c.f. Table 112). As we can see further, this is the closest distance between two American writers. However, the smallest distance is between Agatha Christie and Ed McBain (19.95), although the former is a British writer while the latter is an American writer.

Main Part.   Discussion.

Before describing our model it is necessary to point out that all models which use quantitative features fall under the classification of German philosopher Immanuil Kant (1724 – 1804), who

stated that everything in this world has a number embodiment. A. V. Jakushev, developing the ideas of I. Kant, calls upon the researchers to take into consideration the numerical relationships which can reflect a system fully (Jakushev, 2005: 25). 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, 1994-a; 1994-b; 2003: 7 - 23).

Gustav Herdan wondered in his famous book *The Advanced Theory of Language as Choice and Chance* 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 instances 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 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 goal of this article is to establish a stylistico-grammatical typological similarity or distance on the one hand between the prose of some American writers (Theodore Dreiser F. Scott Fitzgerald, Ernest Hemingway), and some British writers (J. K. Rowling , Agatha Christie and

William Somerset Maugham) on the other. These grammatical features may be a good clue for obtaining stylistic typological distances, which are measured on the basis of the frequency of occurrence of the English (-ing) forms: Verbal noun, Gerund, and Participle 1 in different functions in the sentence.

It should be interesting to find out if American writers (Theodore Dreiser, F. Scott Fitzgerald and Ernest Hemingway) would form one group, while the British writers (Daniel Defoe, Charles Dickens s, Joan K. Rowling, Agatha Christie and William Somerset Maugham) should form the other group (taxon). We must consider the American taxon and the British one to be formed if and only if the distances between the American authors are less than those between American and British authors. Similarly, the total distance between the British authors would less than that between British and American writers.

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. Consequently, the texts of two different authors must be more different 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 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 material of this study is described in detail at the end of this article. Nevertheless it is worth while mentioning that we took for our study the first ten thousand words of texts by each writer. One must observe the rules of commensurability: for this reason the sample is advised to

be fixed at the volume of not less than 10,000 words. The same sample volume of 10,000 words was also taken from the texts by Daniel Defoe, Charles Dickens, Theodore Dreiser, F. Scott Fitzgerald, Ernest Hemingway, Joan K. Rowling, Agatha Christie, William Somerset Maugham, Ed McBain, and Hal Colebatch. We must 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 less the difference is, the greater the similarity between the texts. One can see that we use the terms "similarity" and "distance" in their usual meaning reflected in a dictionary, but not in the strict mathematical definition. So, we understand similarity and distance as 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 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 don't 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.

In this case the selected features are the values of the frequency of occurrence of some functions of the Gerund, Participle 1 and the Verbal noun. We believe that the frequencies of the (-ing) forms show certain concrete manifestations of the tendencies of thinking of an author. We assume the modes of grammar to be stable enough to produce an underlying linguistic stability. In our study we must distinguish between insignificant fluctuations of the selected features which do not affect the basic character of the author's style on the one hand, and the significant fluctuations which indicate stylistic differences on the other hand.

Our study uses the methods of computational linguistic. 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 the interests of linguists and mathematicians to yield new models. The statistical investigation of texts can explain the intrinsic features inherent in the text (Dolezhel et al., 1969: VII, etc.). 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 (Dolezhel, 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).

Dolezhel is not right to assume that a numerical stylistic theory can only be said to account for stylistic differences between texts (Dolezhel, 1969: 11). In fact, it can provide different measures, which allow us to construct a taxonomy of styles. Nevertheless, we agree with Dolezhel 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 (Dolezhel, 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, it 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; they express themselves in different ways (Mealand, 1999: 479). Thus, authorship attribution methods allow us to define the degree of

affinity between the 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 less 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; Rudman, 1998: 352 - 353). However, we can't help agreeing with Joseph Rudman that for every paper announcing an authorship attribution method that "works", there is a counter paper that points out real or imagined crucial shortcomings (Rudman, 1998: 352). 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 Gerund, Participle 1 and Verbal noun 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. Nevertheless, our study is only a small step to solve the problem of grammatical features, which can differentiate the styles of different authors. It needs deeper and further studies if there are some more stable grammar phenomena than those mentioned above in texts of an author.

One can see that to differentiate texts stylistically it is possible to use many characteristics, among them, for instance, the length and types of syllables, of words, sentences, the richness of vocabulary, as presented by G. U. Yule (1944), or concentration of vocabulary, by P. Guiraud,

M. Teshitelova, J. Mistrik, etc. (Brainerd, 1974: 248 - 252; Fucks, 1968; Teshitelova, 1992: 160 - 176).

Let us consider in detail the features taken for our study: Gerund, Participle 1 and Verbal noun. M. A. Belyaeva, V. L. Kaushanskaya, R. L. Kovner, and others consider the Gerund, formed by adding the suffix (-ing) to the stem of the verb and coinciding in form with Participle 1, to be a form which developed from the Verbal noun. In course of time, the gerund became verbalized, preserving at the same time its nominal character (Belyaeva, 1971: 165; Kaushanskaya et al., 1967: 170).

Let us consider the functions of the gerund in the sentence. The gerund may be a subject, a part of a compound verbal predicate, an object, an attribute and an adverbial modifier.

1. **The gerund as the subject**.

"Minnie answered "yes", but Carrie could feel that going to the theatre was poorly advocated here." (Theodore Dreiser. *Sister Carrie*. p. 14)

"Reserving judgements is a matter of infinite hope"
(F. Scott Fitzgerald. *The Great Gatsby*, p. 1.)

"Listen, Robert, going to another country doesn't make any difference."
(Ernest Hemingway. *The Sun Also Rises*, p. 31.)

"Thinking about Ron was painful, too."
 (Joan K. Rowling *Harry Potter and the Chamber of Secrets*, p.15.)

2. **The gerund as a part of the compound verbal predicate**.

"You can't stop going with an old friend on account of rumors …"
(F. Scott Fitzgerald. *The Great Gatsby*, p. 20.)

"They would look on and on after everyone else's eyes in the world would have stopped looking." (E. Hemingway. *The Sun Also Rises*, p. 26.)

She's used to flying around outside.
(J. K. Rowling. *Harry Potter and the Chamber of Secrets*, p.16).

The gerund as the object.

a) **The gerund as the direct object**.
    One must bear in mind that the gerund as the direct object is used after the following verbs:
to avoid , to mind, to fancy, to mention, to deny, to put off, to delay, to postpone,
to excuse, to propose, to give up, to require, to leave off, to need, to suggest, to enjoy, to forget,
to hate, to like, to dislike, to neglect, to omit, to detest, to prefer, to prevent, to refuse, to regret,
to remember, to resist, and some other verbs.

"She could not help smiling as he told her of some popular actress of whom she reminded him".
(Theodore Dreiser. *Sister Carrie*. p. 3)

"I enjoyed looking at her." (F. Scott Fitzgerald. *The Great Gatsby*, p. 11)

"I stepped aside to avoid walking into the thread with which his girl assistant manipulated the
boxers." (E. Hemingway. *The Sun Also Rises*, p. 35)

"… the photograph was morning; wizard, who Harry supposed was Gilderoy Lockhart kept
winking cheekily at them all."
(J. K. Rowling. *Harry Potter and the Chamber of secrets*. p.36)

b) **The gerund as the object of a preposition**

    It is important to remember that the gerund as the object of prespositions is used after the

following verbs:

to agree (to), to congratulate (on), to depend (on), to dream (of), to succeed (in), to rely (on), to

be proud (of), to insist (on), to be fond (of), to object (to), to be sure (of), to be pleased -

(at/with), to hear (of), to be surprised (at), to thank (for), to be interested (in), to prevent (from),

to be connected (with), to think (of, about) and some others.

"Some time she spent in wandering up and down…" (Th. Dreiser. *Sister Carrie*. p. 9)

"I wouldn't think of changing the light"
(F. Scott Fitzgerald. *The Great Gatsby*, p. 32.)

"Did you ever think about going to British East Africa to shoot?"
(E. Hemingway. *The Sun Also Rises*. p. 10.)

"Julian felt rather like a dog who has been congratulated on performing a clever trick." (A.
Christie "Sanctuary". p.20)

3. **The gerund as an attribute**.

The gerund as an attribute may be either in preposition or postposition to the noun.

The Gerund as an attribute in preposition to the noun.

"The huge railroad corporations which had long before recognised the prospects of the place had seized upon vast tracts of land for transfer and shipping purposes."
(Theodore Dreiser. *Sister Carrie*. p. 7)

"The apartment was on the top floor - a small living room, a small dining room, a small bedroom and a bath." (F. Scott Fitzgerald. *The Great Gatsby*. p. 29)

"She was built with curves like the hull of a racing yacht, and you missed none of it with that wool jersey."   (E. Hemingway. *The Sun Also Rises*. p. 22)

Often, it is felt that the gerund should be marked to show it is the gerund, not participle 1, in

this function. Then it is written with a hyphen, e.g. "dancing-club"

"Frances was a little drunk and would have liked to have kept it up but the coffee came, and Lavigne with the liqueurs, and after that we all went out and started Braddock's dancing-club".
(E. Hemingway. *The Sun Also Rises*. p. 19.)

However, it is not clear why Ernest Hemingway writes the gerund in this function and the noun

after it, sometimes as two words and sometimes as a hyphenated word.

"The other letter was a wedding announcement"
(E. Hemingway. *The Sun Also Rises*, p.30.)

"I put on a dressing-gown and went to the door."
(E. Hemingway. *The Sun Also Rises*, p. 31.)

In our investigation on this function of the gerund and the participle, we came to the

conclusion that if we can see that it represents something (i.e. preserving more feature of a

noun), then it is the gerund. If it acts (i.e. preserving more features of a verb), then it is the

participle. For instance, *a sporting type*, *a firing squad*, *a spooling disc*, *an answering fire*, etc."

could be either the gerund or participle 1. Since they can be distinguished only from the sentence

(Tambovtsev et al., 1980: 55 - 59), sometimes a wider context is required.

"He was not a sporting type. Neither his mother, nor his father had been heavily built, but Lewis started out with impressive proportions and continued to develop. His mind was agile enough, but his body was sluggish. He was in fact a fat boy." (J. Burke. "The Boys".)

The Gerund as an attribute in postposition to the noun.

"She abandoned the thought of appealing to the other department stores…"
(Theodore Dreiser. *Sister Carrie*. p. 11)

"And I had the high intention of reading many other books besides."
(F. Scott Fitzgerald. *The Great Gatsby*. p. 4)

"The Catholic Church had an awfully good way of handling all that."
(E. Hemingway. *The Sun Also Rises*. p. 31.)

"I've always wanted to have the chance of looking for buried treasure."
(A. Christie "Strange Jest". p.41)

4. **The gerund as an adverbial modifier**.

"She felt the flow of the tide of effort and interest - felt her own helplessness without quite realising the wisp on the tide that she was." (Th. Dreiser. *Sister Carrie*. p. 12)

"Then she wet her lips and without turning around spoke to her husband in a soft, coarse voice.
(F. Scott Fitzgerald. *The Great Gatsby*. p. 26)

"At one time or another I had probably considered it from most of its various angles, including the one that certain injuries or imperfections are a subject of merriment while remaining quite serious for the person possessing them."
(E. Hemingway. *The Sun Also Rises*. p. 27.)

"She hesitated before using the knocker for the third time."
(A. Christie "Strange Jest" p.53).

Let us consider Participle 1. First of all, it is necessary to define the participle. M. A.

Belyaeva, V. L. Kaushanskaja, R. L. Kovner, and others consider the participle a non-finite form

of the verb which has a verbal and an adjectival or an adverbial character. There are two

participles in English: Participle 1 and Participle 2, traditionally called the Present Participle and

the Past Participle. Participle 1 is formed by adding the suffix (-ing) to the stem of the verb

(Belyaeva, 1971: 177; Kaushanskaya et al., 1967: 155). We take for our study only Participle 1,

which coincides in form with the gerund. Actually, it is rather hard to distinguish between them.

Nevertheless, it is possible to distinguish between them if we consider their exact environments.

We can see this from the following examples and discussion.

Participle 1 may have different syntactic functions in the sentence.

1. **Participle 1 as a part of a compound verbal predicate**.

"She noticed that men and women were smiling." (Th. Dreiser. *Sister Carrie*. p.13)

"I've been trying to get you to New York all afternoon."
(F. Scott Fitzgerald. *The Great Gatsby*. p. 11)

"I can't stand it to think my life is going so fast and I'm not really living it."
(E. Hemingway. *The Sun Also Rises*. p. 11)

"Harry wasn't listening."
(J. K. Rowling. *Harry Potter and the Chamber of secrets*. p.18)

2. **Participle 1 as a part of the complex object.**

"She could see them moving about now and then upon the upper floors".
(Theodore Dreiser. *Sister Carrie*. p. 8)

"At high tide in the afternoon I watched his guests diving from the tower of his raft…
(F. Scott Fitzgerald. *The Great Gatsby*. p. 39)

"I went back upstairs and from the open window watched Brett walking up the street to the
big limousine drawn up to the curb under the arc light."
(E. Hemingway. *The Sun Also Rises*. p. 34.)

"I've heard Dad talking about him", said George.
(J.K.Rowling. *Harry Potter and the Chamber of secrets*. p.29)

3. **Participle 1 as an attribute in preposition to the noun (i.e. before the noun).**

"From all the more pleasing or imposing places she was turned away abruptly with the most
chilling formality". (Th. Dreiser. *Sister Carrie*. p. 12)

"For a moment the last sunshine fell with romantic affection upon her glowing face…" (F. Scott Fitzgerald. *The Great Gatsby*. p.14)

"He was from New York by way of Chicago, and was a rising new novelist."
(E. Hemingway. *The Sun Also Rises*. p. 21.)

"The vicar came up to the sofa and looked down at the dying man.
He made his way towards a waiting car."
(A. Christie. "Sanctuary" p.15)

4. Participle 1 as an attribute in postposition to a noun (i.e. after a noun).

"She looked at the little slip bearing her sister's address and wondered."
(Theodore Dreiser. *Sister Carrie*. p. 1)

"On week-ends his Rolls-Royce became an omnibus, bearing parties to and from the city
between nine in the morning and long past midnight…"
 (F. Scott Fitzgerald. *The Great Gatsby* ". p. 39)

"The policemen standing by the door looked at me and smiled."
(E. Hemingway. *The Sun Also Rises*. p. 20.)

"Her fingers went to his pulse - a pulse so feeble and fluttering that it told its own
story, as did the almost greenish pallor of his face."
(A.Christie. "Sanctuary." p. 26.)

5. **Participle 1 as an adverbial modifier**.

"Many of them had their sleeves rolled up, revealing bare arms, and in some cases, owing to
the heat, their dresses were open at the neck".
(Theodore Dreiser. *Sister Carrie*. p. 11)

"She laughed again, as if she said something very witty, and held my hand for a moment,
looking up into my face, promising that there was no one in the world she so much wanted to
see." (F. Scott Fitzgerald. *The Great Gatsby*. p.9)

"I lay awake thinking and my mind jumping around."
(E. Hemingway. *The Sun Also Rises*. p. 31.)

"I know what day it is, - sang Dudley, waddling toward him."
(J. K. Rowling. *Harry Potter and the Chamber of secrets*. p. 8.)

The Verbal Noun

The verbal noun, which also has the suffix (-ing), should not be confused with the gerund. Kaushanskaya and her co-authors warn that the verbal noun has only a nominal character while the gerund and Participle 1 have a double character, nominal and verbal. The verbal noun may be used with an article, has the plural form, and does not take a direct object (Kaushanskaya et al., 1967: 185).

"…the unspoken shade of disapproval to the doing of those things… slightly affected the atmosphere of the table." (Th. Dreiser. *Sister Carrie*. p. 14)

"…he strained the top lacing… (F. Scott Fitzgerald. *The Great Gatsby*. p. 7)

"But I guess you have pretty big doings yourself over in London."
  (E. Hemingway. *The Sun Also Rises*. p. 28.)

"It was Eccles's revolver, but I rather fancy it was Moss who did the shooting."
  (A. Christie. "Sanctuary". p.31).

The Results Obtained with the Chi-Square Criterion.

The method used here is to count the occurrence values of the (-ing) forms in the prose of the British and American authors mentioned above, using the criterion of Chi-square. Actually, 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 non-parametric, so 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. For this reason we preferred the Chi-square criterion.

Let us explain how we use the Chi-square test here. The procedure is described in detail elsewhere (Tambovtsev, 2003: 26 - 27). Let's take the third line of Table 1: the function of gerund as an attribute in preposition to a noun. We found 18 occurrences of it in the first 10,000

words of *Sister Carrie* by Theodore Dreiser and two occurrences in Agatha Christie's "Sanctuary". In order to calculate the value of Chi-square test, we must find the expected value, i.e., the *mean*. Here the *mean* is 10. Then we must distract it from every value, i.e. 18 - 10 = 8 and raise the remainder to the square, 8 x 8 = 64. Then we must divide this by 10, i.e. the *mean*, 64/10 = 6.4. The same procedure must be done with the second value and so on. We have chosen the level of significance 0.10 or 5%. The degrees of freedom are calculated by n-1, where n is the number of the features compared. Here we have two authors, thus n = 2. Thus we must look the critical value in the table of critical values of Chi-square test with 1 degree of freedom. It is 3.84 (Butler, 1985: 176). If the received value of the Chi-square is less than 3.84, then there is no statistical difference between these two authors by this particular function.

In this case (6.4), it is much greater than 3.84. It means that Theodore Dreiser and Agatha Christie are quite different by this function of the gerund. We compare the pairs to obtain the Chi-square values in every line. It is possible to add all Chi-square values to obtain the total. We can also try to consider if the total is in the limits of the critical values shown in the "Chi-square" table with 10 degrees of freedom 18.31. In all the cases, it is greater than that. It may mean that in every case the total of the use of Gerund, Participle 1 and Verbal noun can differentiate the texts even of one and the same author (c.f. 12.66) or different authors (82.02).
However, in our case, it is more important to regard the total as the total distance between the authors. Therefore, we consider the total of the Chi-square values as the distances between the authors in question. We need not prove the possibility of using the values of Chi-square criterion as metrics as proven in mathematical statistics by A. A. Borovkov (Borovkov, 1984: 194).

We placed the frequency of occurrence of gerund of *Sister Carrie* by Theodore Dreise (TD) in the first column and Agatha Christie 1 (AC-1) in the second column of Table 1. The third

column shows the expected value, i.e. the mean (M). The 4th column shows the Chi-square

values. The 5th column shows if the difference between the critical value (3.84) and the received

value is statistically significant.

Functions of Gerund in the Sentence on the material of *Sister Carrie* by Theodore Dreiser

(TD) and Christie 1 (AC-1) - Stories: "Sanctuary" and "Strange Jest" from Agatha Christie "Miss

Marple's Final Cases":

Table 1

Functions of gerund	TD	AC-1	Mean	Chi-square	Significance
Subject	1	0	0.5	1.00	no
Predicate	0	2	1.0	2.00	no
Attribute in preposition to noun	18	2	10.0	12.80	yes
Attribute in postposition to noun	6	6	6.0	0.00	no
Adverb	22	7	14.5	7.80	yes
Object	13	6	9.5	2.60	no
Total				26.20	

Functions of Participle 1 in the Sentence on the material of *Sister Carrie* by Theodore Dreiser

(TD) and Christie 1 (AC-1) - Stories: "Sanctuary" and "Strange Jest" from Agatha Christie "Miss

Marple's Final Cases":

Table 2

Functions of Participle 1	TD	AC-1	Mean	Chi-square	Significance
Complex Object	3	0	1.5	3.00	no
Predicate	40	50	45.0	1.10	no
Attribute in preposition to noun	44	19	31.5	9.90	yes
Attribute in postposition to noun	29	15	22.0	4.50	yes
Adverb	76	57	66.5	2.70	no
Total				21.20	

Verbal Noun in the texts of *Sister Carrie* by Theodore Dreiser (TD) and Christie 1 (AC-1) -

Stories: "Sanctuary" and "Strange Jest" from Agatha Christie "Miss Marple's Final Cases":

Table 3

	TD	AC-1	Mean	Chi-square	Significance
Verbal Noun	26	22	24.0	0.30	no

Functions of Gerund in the Sentence on the material of *Sister Carrie* by Theodore Dreiser (TD)

and Christie 2 (AC-2) - Story: "Curtain" by Agatha Christie:

Table 4

Authors		Gerund	Participle 1	Verbal noun	Chi-square Total
American authors					
Dreiser – Fitzgerald		17.70	4.81	1.50	24.01
Dreiser - Christie 1		26.20	21.20	0.30	47.70
Dreiser - McBain		22.8	30.6	5.2	58.60
Dreiser - Christie 2		19.80	39.50	9.50	59.30
Dreiser - Colebatch		25.20	36.72	0.02	59.34
Dreiser - Rowling		20.03	42.60	5.20	67.83
Dreiser - Maugham 3		15.80	54.00	9.50	69.80
Dreiser - Maugham 2		16.50	65.60	2.90	85.00
Dreiser - Hemingway		16.49	62.61	8.30	87.40
Dreiser - Maugham 1		24.30	77.30	7.10	101.60
Fitzgerald - Christie 1		15.70	9.30	0.80	25.40
Fitzgerald - Christie 2		12.80	25.80	3.85	38.60
Fitzgerald - McBain		13.30	29.90	1.20	44.40
Fitzgerald - Maugham 3		35.80	29.91	15.40	47.41
Fitzgerald - Hemingway		3.23	44.20	3.00	50.43
Fitzgerald - Maugham 2		8.30	44.80	0.30	51.00
Fitzgerald - Rowling		8.98	41.60	1.20	51.78
Fitzgerald - Colebatch		15.40	55.20	1.80	67.10
Fitzgerald - Maugham 1		17.70	51.70	4.60	69.40
Hemingway - Maugham 2		9.06	14.04	3.00	24.60
Hemingway - Maugham 1		10.78	16.10	0.06	26.88
Hemingway - Christie 2		13.78	15.00	0.06	28.86
Hemingway - Rowling		10.54	33.78	0.42	44.74
Hemingway - Christie 1		11.95	27.81	10.90	47.94
Hemingway - Maugham (3)		14.60	35.50	0.06	50.10
Hemingway - McBain		13.16	40.58	0.80	54.14
Hemingway - Colebatch		9.30	145.30	9.00	163.60
T. Wolf - Hemingway		13.34	9.68	1.50	23.77
T. Wolf - Maugham 1		15.09	9.94	1.00	25.53
T. Wolf - Thackeray		9.20	15.42	1.25	25.87
T. Wolf - Maugham 2		10.42	18.19	0.00	28.61
T. Wolf - Maugham 3		11.56	20.58	4.00	36.14
T. Wolf - Rowling		12.71	24.59	0.16	37.46
T. Wolf - Christie 2		16.98	20.8	1.07	38.13
T. Wolf - Fitzgerald		13.00	26.19	0.14	39.33
T. Wolf - Christie 1		25.80	17.04	0.67	40.61
T. Wolf - Swift		20.33	27.93	0.24	48.50
T. Wolf - Updike		12.42	37.32	0.36	50.10

T. Wolf - McBain	15.05	35.21	0.17	50.43
T. Wolf - Dreiser	18.11	36.79	1.48	56.38
T. Wolf - Dickens s	13.54	41.02	5.68	59.73
T. Wolf - Defoe	22.09	52.09	1.00	74.68
T. Wolf - Colebatch	18.36	78.52	1.71	98.59
British Authors				
Rowling - Christie (1)	25.80	35.26	5.80	63.96
Rowling - McBain	19.30	53.42	0.00	72.72
Rowling - Christie (2)	22.35	59.60	1.60	77.10
Rowling - Maugham (1)	31.86	62.93	0.80	94.79
Rowling - Maugham (2)	24.25	66.37	0.30	90.62
Rowling - Maugham (3)	35.30	85.40	1.60	118.05
Rowling - Colebatch	24.30	131.44	8.30	164.04
Christie (1) - Christie (2)	5.68	19.63	6.50	25.31
Christie (1) - McBain	5.20	26.87	5.80	34.97
Christie (1) - Maugham (2)	12.59	27.34	1.32	41.25
Christie (1) - Maugham (3)	22.45	18.92	10.67	52.04
Christie (1) - Maugham (1)	16.88	38.93	9.00	60.31
Christie (2) - McBain	1.95	114.40	1.60	19.95
Christie (2) - Maugham (2)	11.34	7.13	2.13	21.20
Christie (2) - Maugham (3)	22.80	17.99	0.00	39.39
Christie (2) - Maugham (1)	25.13	24.36	0.22	49.71
Christie (1) - Colebatch	18.80	93.16	1.00	103.06
Christie (2) - Colebatch	11.70	101.00	10.30	122.50
Maugham (1) - Maugham (2)	10.80	13.79	1.00	24.19
Maugham (1) - Maugham (3)	5.41	27.10	0.80	28.96
Maugham (1) - McBain	23.160	63.80	0.80	87.16
Maugham (1) - Colebatch	6.20	156.71	7.80	170.71
Maugham (2) - McBain	11.14	37.44	0.30	48.88
Maugham (2) - Colebatch	10.30	142.21	6.80	155.91
Maugham (3) - McBain	23.34	50.47	1.60	74.67
Maugham (3) - Colebatch	12.00	114.84	10.30	137.14
Colebatch - McBain	16.20	81.08	5.80	103.08

The ordered total distances between some selected American and British authors based on the

Chi-square values of the gerund, participle 1 and verbal noun:

Table 5

Authors	Gerund	Participle 1	Verbal noun	Chi-square Total
Christie (2) - McBain	1.95	114.40	1.60	19.95
Dreiser - Fitzgerald	115.40	4.81	3.00	24.01
Maugham (1) - Maugham (2)	10.80	13.79	1.00	24.19
Hemingway - Maugham 2	9.06	14.04	3.00	24.60
Christie (1) - Christie (2)	5.68	19.63	6.50	25.31

Fitzgerald - Christie 1		15.70	9.30	0.80	25.40
Hemingway - Maugham 1		10.78	16.10	0.06	26.88
Hemingway - Christie 2		13.78	15.00	0.06	28.86
Maugham (1) - Maugham (3)		5.41	27.10	0.80	28.96
Christie (1) - McBain		5.20	26.87	5.80	34.97
Fitzgerald - Christie 2		12.80	25.80	15.40	38.60
Christie (2) - Maugham (3)		22.80	17.99	0.00	39.39
Christie (1) - Maugham (2)		12.59	27.34	1.32	41.25
Fitzgerald - McBain		16.60	219.80	1.20	44.40
Hemingway - Rowling		10.54	33.78	0.42	44.74
Dickens s - Updike				6.66	46.04
Fitzgerald - Maugham 3		35.80	29.91	15.40	47.41
Dreiser - Christie 1		26.20	21.20	0.30	415.40
Hemingway - Christie (1)		11.95	27.81	10.90	47.94
Maugham (2) - McBain		11.14	37.44	0.30	48.88
Christie (2) - Maugham (1)		25.13	24.36	0.22	49.71
Hemingway - Maugham (3)		14.60	35.50	0.06	50.10
Fitzgerald - Hemingway		3.23	44.20	3.00	50.43
Fitzgerald - Maugham 2		8.30	44.80	0.30	51.00
Fitzgerald - Rowling		8.98	41.60	1.20	51.78
Christie (1) - Maugham (3)		22.45	18.92	10.67	52.04
Hemingway - McBain		13.16	40.58	0.80	54.14
Dreiser - McBain		22.8	30.6	5.2	58.60
Dreiser - Christie 2		19.80	39.50	9.50	59.30
Dreiser - Colebatch		25.20	36.72	0.02	59.34
Christie (1) - Maugham (1)		16.88	38.93	9.00	60.31
Rowling - Christie (1)		25.80	35.26	5.80	63.96
Fitzgerald - Colebatch		15.40	55.20	1.80	67.10
Dreiser - Rowling		20.05	45.20	5.20	67.83
Fitzgerald - Maugham 1		115.40	51.70	4.60	69.40
Dreiser - Maugham 3		15.80	54.00	9.50	69.80
Rowling - McBain		19.30	53.42	0.00	72.72
Maugham (3) - McBain		23.34	50.47	1.60	74.67
Rowling - Christie (2)		22.35	59.60	1.60	77.10
Dreiser - Maugham 2		16.50	611.20	5.80	85.00
Maugham (1) - McBain		23.160	63.80	0.80	87.16
Dreiser - Hemingway		16.49	62.61	8.30	87.40
Rowling - Maugham (2)		24.25	66.37	0.30	90.62
Rowling - Maugham (1)		31.86	62.93	0.80	94.79
Dreiser - Maugham 1		24.30	77.30	7.10	101.60
Christie (1) - Colebatch		18.80	93.16	1.00	103.06
Colebatch - McBain		16.20	81.08	5.80	103.08
Rowling - Maugham (3)		35.30	85.40	1.60	118.05
Christie (2) - Colebatch		11.70	101.00	10.30	122.50
Maugham (3) - Colebatch		12.00	114.84	10.30	137.14

Maugham (2) - Colebatch		10.30	142.21	6.80	155.91
Hemingway - Colebatch		9.30	145.30	9.00	163.60
Rowling - Colebatch		24.30	131.44	8.30	164.04
Maugham (1) - Colebatch		6.20	156.71	7.80	170.71

The mean total distance between the particular writer and the group of writers:

Table 6

Group		American	British
American			
Hemingway		68.92	37.19
Fitzgerald		37.22	47.26
Dreiser		55.71	71.87
Total mean		53.95	52.11
British			
Christie		41.30	50.55
Maugham		58.42	62.97
Rowling		54.78	88.91
Total mean		51.50	67.48

Conclusions

In conclusion we can state that:

1. The Gerund, Participle 1 and Verbal Noun are used in the same 12 functions and with more or less frequency of occurrence in the texts of Charles Dickens *Great Expectations* and John Updike *Run, Rabbit*. The distance is 46.039. This may show the influence of the style of Dickens on Updike.

2. Theodore Dreiser has a tendency to use the (-ing) forms discussed in his prose in the same way as does F. Scott Fitzgerald. Thus the distance is minimal (24.01). Dreiser's prose is also close to the prose of Christie 1 (47.70). Dreiser's prose by these 12 features is far away from that of Maugham 1 (101.60). Our task is to find out the distances between the authors, but not to explain those distances; one can, however, see that the prose F. Scott Fitzgerald and Agatha Christie might have influenced the prose of Theodore Dreiser.

3. Ernest Hemingway has a tendency to use the Gerund in the same functions as Joan. K. Rowling (10.54), Participle 1 - as William Somerset Maugham in "Flotsam and Jetsam" (14.04) and Verbal noun - as Maugham in "The force of circumstance" and "The Creative impulse" or as Christie in "Curtain" (0.06).

4. The prose of *The Sun Also Rises* by Ernest Hemingway, in the total use of the (-ing) forms in all their functions, is closer to that of " Flotsam and Jetsam " by William Somerset Maugham (24.60), while Maugham's - "The force of circumstance" (26.88), "Curtain" by Agatha Christie (28.86), Rowling's *Harry Potter* (44.74) and Christie 1 - "Sanctuary" (47.94) follow it rather tightly. At the same time Maugham's - "The Creative impulse" (65.13) is as far away form Ernest Hemingway as Joan K. Rowling from Agatha Christie (c.f. 63.96 and 65.13).

5. The prose of Joan K. Rowling, in the total use of the (-ing) forms, is closer to the story "Sanctuary" and "Strange Jest" (63.96), than to "Curtain" (65.13) by Agatha Christie. The prose of Joan K. Rowling is closer to the prose of Agatha Christie (63.96 - 65.13), than to the prose of William Somerset Maugham (82.12 - 93.63 - 118.55).

6. The texts of one and the same author by the occurrence of Gerund, Participle 1 and Verbal noun are closer to each other (26.64 - 35.40 - 35.63), than the texts of different authors (35.92 - 118.55). The difference may be nearly 5 times greater (c.f. 24.19 vs 118.05).

7. The chosen 12 (-ing) features can distinguish different texts very well. The texts of the same author have the tendency of being rather close. The closeness of Agatha Christie 2 and Ed McBain may be explained by the fact that the style in his detective stories willingly or unwilingly resembles that of Agatha Christie. It may be that McBain subconciencously immitated Christie's style.

8.Hemingway seems to be closer to the chosen (Christie, Maugham, Rowling) British authors (37.19), than to the chosen American ones (68.92), i.e. Fitzgerald and Dreiser. On the contrary, Fitzgerald is closer to Hemingway and Dreiser, than to Christie, Maugham or Rowling (cf. 37.22 vs 47.26). Dreiser is both far away from the chosen American (55.71) or British (71.87) writers. It is a surprise that all British writers are closer to the American writers, than to the British ones (cf. 53.00 vs 67.48). At the same time, in general, the chosen American writers are more or less equally close to the American or the British writers (cf. 53.95 vs 52.11). It may lead us to the conclusion that the writer's choice of the Gerund, Participle 1 and Verbal noun in different functions is not very much influenced by whether the author is American or British. In this case, some individual tendencies are more important.

References

Belyaeva, 1971 - Belyaeva, Mariya Aleksandrovna. English Grammar. - Lenengrad: Prosveshchenie, 1971.

Borovkov, 1984 - Borovkov, Aleksey Aleksandrovich. Matematicheskaja statistika. Otsenka parametrov. [Mathematical statistics. Estimation of Parameters]. - Moskva: Nauka, 1984. (in Russian).

Brainerd, 1974 - Brainerd, Barron. Weighing evidence in language and literature: a statistical approach. - Toronto: University of Toronto Press, 1974.

Brownlee, 1947 – K. A. Brownlee. Industrial Experimentation. – Oxford: Oxford University Press, 1947.

Butler, 1985 - Butler, Christopher. Statistics in Linguistics. - Oxford: Basil Blackwell, 1985.

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.

Fucks, 1968 - Fucks, Wilhelm. Nach allen Regeln der Kunst. Diagnosen ueber Literatur, Musik, bildende Kunst - die Werke, ihre Autore und Schoepfer. - Stuttgart: Deutsche Verlags-Anstalt, 1968. (in German).

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.

Kaushanskaja et al,, 1967 - Kaushanskaya V. L., Kovner R. L., Kozhevnokova O. N., Prokof'eva E. V., - - Raynes Z. M. Skvirskaya S. E. Tsyrlina F. Ja.   A Grammar of the English Language. - Leningrad: Prosveshchenie, 1967.

Ledger, 1989 - Ledger, Gerard. Re-Counting Plato: A Computer Analysis of Plato's Style. - Oxford: Clarendon Press, 1989.

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.

Owen, 1990 - Owen F., Jones R. Statistics. - London: Pitman, 1990.

Tambovtsev et al, 1980 - Tambovtsev, Yuri Alekseevich - Tambovtseva, Irina Vladislavovna. Gerund and participle 1 in preposition to noun. (in Russian). -In: V pomoshch prepodavatel'am inostrannyh jazykov. Vol.11. - Novosibirsk: Nauka, 1980, p. 55 - 65.

Tambovtsev, 1994-a - Tambovtsev, Yuri. A. Dynamics of Functioning of Phonemes in Sound Chains of Languages of Different Structure (in Russian). - Novosibirs: Novosibirsk University, 1994. (in Russian).

Tambovtsev, 1994-b - Tambovtsev, Yuri. A. Stylistic differences in Persian texts from phonological viewpoint. - In: Archiv orientalni (Praha). Quarterly Journal of African and Asian Studies, Volume 65, # 2, p. 171 - 174.

Tambovtsev, 2003 - Tambovtsev, Yuri. A. Typology of Functioning of Phonemes in Sound Chain of Endo-European, Paleo-Asiatic, Ural-Altaic and Other World Languages: Compactness of Subgroups, Groups, Families and Other Language Taxa (in Russian). - Novosibirsk: Sibirskij Nezavisimyj Institut, 2003.

Tambovtsev et al., 2003-a - Tambovtsev, Yuri - Tambovtseva, Ludmila - Tambovtseva, Juliana. Some stylistic typological distances between the prose of some British writers. - Studia Anglica Posnaniensia, # 39, 2003 (Poznan'), p. 247 - 261.

Teshitelova, 1992 - Teshitelova, Maria.   Quantitative Linguistics. - Amsterdam: John Benjamins publishing company, 1992.

Webster, 1965 - Webster's Seventh New Collegiate Dictionary. - Springfield, Massachusetts: Merriam Company Publishers, 1965.

Winter, 1969 - Winter, Werner. Styles as dialects. - In: Statistics and style. - New York: Elsevier Publishing Company, Inc., 1969, p. 3 - 9

Yuri Tambovtsev, Ludmila Tambovtseva, and Juliana Tambovtseva
Novosibirsk Pedagogical University
Novosibirsk, Russia
Email: yutamb@mail.ru