

Ponaida Ivanna

Student,

Lviv Polytechnic National University

**USING AUTOMATIC TEXT ANALYSIS PROGRAMS
IN RESEARCHING THE PECULARITIES
OF THE AUTHOR'S IDIOSTYLE
(BASED ON THE NADINE GORDIMERS'S NOVEL)**

Due to the increasing availability and distribution of the text content, the degree of importance of using automatic methods of text content analysis is increasing. The tasks of content analysis include the problems of classification and clustering of text-based publications according to various criteria, for example, genre, epoch of writing, format (a novel, an essay, and a scientific article), emotional coloration, style of speech, as well as the problem of text authorship attribution. The peculiarities of the application of linguo-statistics technologies for the identification of the style of the author of text content are considered.

The issue of language of the personality and its unique stylistic features always has been one of the fundamental in the linguistic analysis. Consequently, the notion of the author's individual style became the object of investigation of many great scientists such as N. Bolotnova, V. Vinogradov, G. Vinokur, N. Golovchenko, V. Kucharenko, B. Larin, I. Tarasova. The extensive research on author's style has led to the emergence of different approaches.

Firstly the author's style meaning as a categorical unit of the fiction science was presented in the works of V. Vinogradov. Under individual style he understands «the unique, historically caused, complex component, but such which represents structurally unity of means and forms of verbal expression system in its development» [2, p. 135].

N. Bolotnova regards idiostyle as a system of associative-semantic fields that characterize the cognitive level of linguistic personality [1, p. 7]. According to V. Samokhina. idiostyle is defined as a set of socio-historical, national, individual-psychological, moral-aesthetic norms, interwoven with the author's worldview features [3, p. 15].

For the purpose of clarifying the notion of the author's style, we have generated its definition. Idiostyle is a complex multifaceted system of language means, applied with the aim of reflecting the author's reality perception.

As a material for this linguistic research the novel «Good Climate, Friendly Inhabitants» written by an outstanding South African writer Nadine Gordimer has been chosen.

In order to characterize the literary richness of the text, we used Internet site Victana. With the help of this site we were able to analyze the linguistic diversity of the text and calculate such coefficients as lexical diversity, the degree of syntactic complexity, speech coherence, indices of exclusivity and concentration of the text.

According to the results of the research (provided in the Table 1), we can see that the variety of the lexic used by the author is not very diverse. At the same time, the coefficient of syntactic complexity is high. This indicates that Nadine Gordimer uses a large number of complex sentences. Speech coherence coefficient is higher than 1, which means that our text is syntactically coherent and contains a sufficient number of functional words.

Table 1

Quantitative parameters of the text

№	Coefficient	Data-in	Calculation
1.	Coefficient of lexical diversity: $Kl = W / N$	W = 408 N = 937	$Kl = 0.43543223053395$
2.	Coefficient of syntactic complexity: $Ks = 1 - P / W$	P = 42 W = 408	$Ks = 0.89705882352941$
3.	Coefficient of speech coherence: $Kz = (Z + S) / (3 * P)$	Z = 91 S = 76 P = 42	$Kz = 1.3253968253968$
4.	Index of exclusivity: $Iwt = W1 / W$	W1 = 264 W = 408	$Iwt = 0.64705882352941$
5.	Index of concentration: $Ikt = W10 / W$	W10 = 11 W = 408	$Ikt = 0.026960784313725$

The level of vocabulary variability in our text is also very high, as the index of exclusivity is almost 65% according to the results from the site Victana. The fraction of words used in the text only once is 264.

The concentration index displays the number of words that are found in the text more than 10 times. In our case it is equal to 0,02 or 2%, that proves once again that the author resorted to usage of a variety of lexic.

In order to investigate the most frequent words we used the AntConc program – the platform for corpus linguistics research. The results (given in the Table 2) showed that such words as «one», «alone», «nobody», which belong to linguistic category of loneliness often occur in the text.

The most frequent words also include the following words: «boy», «fellow», «boys», «man». This emphasizes that the narrator is the only woman in the male environment. In addition, the main character often calls her colleagues «ducktails» or «duckies».

In the list of frequently used words there are also lexemes «black», «white», «native(s)», which indicates that the author raises the problem of racial prejudice.

In general, the link between frequent words and linguistic categories is not very visible in the text, but we managed to trace it with the help of the AntConc program.

Table 2

The most frequent lexical units concerning the main issues raised in the text calculated by the AntConc program

Frequency	Word
25	<i>one</i>
19	<i>boy</i>
13	<i>fellow</i>
12	<i>man</i>
7	<i>alone</i>
7	<i>white</i>
5	<i>native</i>
5	<i>natives</i>
4	<i>nobody</i>

By means of the statistical method we have determined the most frequent words and revealed their connection with the problems raised in

the story. Due to the Internet site Victana we have investigated the literary richness of the text and calculated five coefficients for its evaluation.

With the help of the program for corpus linguistic research AntConc we established the connection between frequency of lexical units and main topics of the novel – loneliness, social isolation and racial prejudice.

The relevance of the topic is determined by the necessity to determine the individual characteristics of the author's style using programs for linguistic research that calculate the coefficients of linguistic diversity and frequency of use of certain language units in order to have a better understanding of the text.

The work has theoretical value in terms of studying the unique features of the author's individual style, and practical value for linguistic analyzes in using the automatic text analysis programs.

References:

1. Болотнова Н. С. Изучение идиостиля в современной коммуникативной стилистике художественного текста / Н. С. Болотнова // Материалы II Международного конгресса русистов-исследователей «Русский язык: исторические судьбы и современность». – Москва, 2004. – 520 с.
2. Виноградов В. В. Стилистика. Теория поэтической речи. Поэтика. – Москва: Наука, 1963. – 255 с.
3. Самохина В. А. Современная англоязычная шутка: монография / В. А. Самохина. – Харьков: ХНУ, 2008. – 356 с.
4. AntConc program. URL: <http://laurenceanthony.net/software/antconc/>
5. Gordimer N., Good Climate, Friendly Inhabitants. *Selected Stories*. New York: Viking Press, 1976. Pp. 234–45.
6. Linguometry. URL: <http://victana.lviv.ua/nlp/linhvometriia>