UDC 81-112.2

THE TRANSFORMATION OF ENGLISH POPULAR SCIENCE PROSE IN THE CHANGING COMMUNICATIVE ENVIRONMENT

Stavrova V.V. (Saint Petersburg National Research University of Information Technologies, Mechanics and Optics)

Research supervisor: Puchkovskaya A.A, PhD, Director of International Digital Humanities Center, Associate Professor (Petersburg National Research University of Information Technologies, Mechanics and Optics)

Summary: The paper examines the lexical and grammatical features of English popular science prose in the diachronic aspect (based on the books "The Selfish Gene" by Richard Dawkins and "The Vital Question" by Nick Lane); it analyzes the main external and internal factors of linguistic evolution that determine the language features of modern popular science prose. Beyond that it identifies the opportunities of using data mining and machine learning techniques in diachronic stylistics.

Popularization of scientific knowledge is one of the key characteristics of the modern literary process. Due to intersection of information technologies in all spheres of state and public life, special scientific knowledge becomes relevant not only for specialists who know the language of science, but also for the general reader. Hence, the need arises for texts, communicating scientific information to a wide audience.

In this paper, we applied the methods of continuous sampling and statistical analysis, which allowed us to draw the necessary conclusions and generalizations based on the analysis of quantitative indicators of the studied characteristics of the texts of the popular science style. Thus we identified the main lexical and grammatical features of English popular science style in the time period from 1976 to 2016. For the most part, the stylistic features of the popular science works considered by us coincide; only the frequency of using certain stylistic devices varies.

On the lexical level the usage of terms changes: the terminology is mainly introduced without explanation, while the reader is expected to actively study the glossary, as noted by the author of "The Vital Question" in the preface of the book. This is undoubtedly connected with the transition to active forms of gaining scientific knowledge and the transition from the "school of reproduction" to the "school of understanding" and the "school of thinking".

In addition, the modern popular science prose has a tendency to implement to the maximum the principle of visualization as a necessary condition for fulfilling a communicative task. On the lexical level, there is a notable decrease in the percentage of terminology used (from 20.61% to 14.44%), which is countervailed by an increase in the frequency of figurative means used: metaphors (from 8,38% to 10,7%), similes (from 7,22% to 9,99%), periphrases (from 5,58% to 7,53%), personifications (from 9,43% to 12,27%), epithets (from 12,29% to 16,59%). Thereby, a short and vivid image becomes a characteristic form of the scientific knowledge embodiment in the modern popular science prose.

The syntax of the text has also undergone a series of changes. Despite the fact that the communication of scientific knowledge requires syntactic accuracy and detailed statements, the usage of complex sentences has been reduced. Thus, the usage of subordinate conjunctions and linking words has been reduced from 40.5% to 29.46%.
The ubiquitous use of simple and compound sentences, coordinating conjunctions is also notable. The frequency of the usage is also increased for imperative mood (from 1,03% to 2,34%),

rhetorical questions (from 0,94% to 2,64%) and rhetorical exclamations (from 1,34 % to 2,49%). This allows enhancing the emotional impact and reducing the distance between the author and the reader, creating the effect of a direct dialogue.

The emergence of such trends is presumably due to the existence of the so-called “clip thinking”, the ability of a person to perceive the world through short vivid images and messages: a short clip, a news feed, a short article, etc. The prevalence of this type of thinking among modern readers has become a prerequisite for the conciseness and simplification of the text organization.

Therefore, over the past 40 years of the development of popular science literature, the key feature that enables us to achieve the implementation of the communicative task — an accessible and clear communication of scientific knowledge — have remained unchanged. However, we have identified a number of stylistic changes that reflect modern social trends in science and education, namely the transition to active forms of communicating scientific knowledge based on a short and vivid image conveyed in a laconic form.

This research has thrown up many questions in need of further investigation. While the qualitative data analysis conducted manually allows us to claim a high level of its accuracy, still it may not be always appropriate to extrapolate this interpretation based on analysis of two authors to all texts related to popular science prose. Further study requires to extend the sample to a large corpus of texts and to identify less time consuming and more effective methods of performing data analysis. Recent research in computational stylistics has developed some sophisticated learning algorithms, which can use combinations of thousands of features to classify texts.

However, despite the achieved results in authorship attribution and sentiment analysis, these algorithms are of little value in the case of diachronic stylistics. This is primarily due to the fact that the stylistic features of texts related to the same style tend to coincide; only the frequency of using certain stylistic devices varies. It results in the necessity not only for style classification, but also for counting stylistic devices frequencies.

We suggest the following model that could be useful for organizing research in diachronic stylistics. For the moment it is possible to perform the qualitative data analysis in this sphere semi-automatically. At the first stage of the analysis we suggest sorting out a limited sample and compiling a list of targeted stylistic (both lexical and grammatical) features. At the second stage we suggest counting their frequencies using data mining techniques.

The text mining with the programming language R has proven quite effective regarding some lexical features: terms, quantitative adjectives, logical emphasis, similes, and evaluative vocabulary. In relation to the grammatical features, the most quantifiable are the following: complex sentences, introductory words, the subjunctive mood, rhetorical questions and rhetorical exclamations. In this connection, it should be pointed out that such text mining is primarily recommended for analysis of the scientific, official, newspaper and publicistic functional styles. It should be also noted that this analysis may be less accurate than the manual one, but at the same time it allows covering the larger set of information and nevertheless shows the main tendencies. Further research regards the role of machine learning techniques and artificial neural networks in analysis of figurative and contextual meanings (concerning metaphors, periphrases, and epithets) and allusions.