Lexical Peculiarities of British Blogs

Peculiaridades léxicas de los blogs británicos

ELENA VALERIEVNA VARLAMOVA
https://orcid.org/ 0000-0002-1448-0215
el-var@mail.ru
Kazan Federal University. Russia

DARYA DMITRIEVNA KORNILOVA
https://orcid.org/ 0000-0002-5522-6897
darya.d.wizard@gmail.com
Kazan Federal University. Russia

OLGA SERGEEVNA SAFONKINA
https://orcid.org/ 0000-0002-5802-1897
olga.safonkina@gmail.com
National Research Mordovia State University. Russia

ABSTRACT

This work is devoted to identify the specifics of lexical component of texts in the Internet discourse and study the general pattern in creating an English-language blog. The relevance of this study lies in the fact that lexical features and their functions are comprehensively studied in the popularized sphere of the Internet. The purpose is to analyze and compare the lexical characteristics of English texts, identify their interconnection and significance in the design of the blog, as well as, applying to concrete examples, by using a consistent analysis algorithm based on material of the British site British Library Blogs.

Keywords: Frequency of lexical units, lexical diversity, qualitative and quantitative parameters, readability, text complexity.

RESUMEN

El trabajo presentado está dedicado a identificar los detalles del componente léxico de los textos en el discurso de Internet y estudiar el patrón general en la creación de un blog en inglés. La relevancia de este estudio radica en el hecho de que las características léxicas y sus funciones se estudian exhaustivamente en la esfera popularizada de Internet. El propósito es analizar y comparar las características léxicas de los textos en inglés, identificar su interconexión y significado en el diseño del blog, así como también aplicar a ejemplos concretos, usando un algoritmo de análisis consistente basado en material del sitio británico Blogs de la Biblioteca Británica.

Palabras clave: Complejidad del texto, diversidad léxica, frecuencia de unidades léxicas, legibilidad, parámetros cualitativos y cuantitativos.

Utopía y Praxis Latinoamericana; ISSN 1316-5216; ISSN-e 2477-9555
Año 25, n° extra 10, 2020, pp. 122-127
123

INTRODUCTION

The work is dedicated to identify the specifics of the lexical component of texts in the Internet discourse and study the patterns of creating an English-language blog. In this study, the text is considered as a unit of language, possessing a unity of form, content and function. The pragmatic potential of the text is the result of the choice not only of the content of the message, but also of the way the language is expressed in the message (Солнышкина & Кисельников: 2015, pp. 99-107).

The features of the quantitative and qualitative parameters of the text in the Internet discourse have not been studied by scientists yet. That is why the scientific novelty of the work is in the study of the characteristics of blogs, which are a constantly developing sphere. This work demonstrates the trends used by bloggers in text creation, which can be used to write and evaluate English-language blogs (Loginova: 2017; Viera: 2017).

In this regard, the purpose of this study is to analyze and compare the lexical characteristics of English texts, to identify their relationship and significance in the design of the blog, as well as to apply a sequential analysis algorithm. This goal requires achieving the following tasks: to distinguish between the concepts of difficulty, complexity and readability of the text; to master the program Text Inspector to obtain quantitative parameters of texts; to examine the existing readability formulas and formulas for calculating the lexical diversity of the text; to master the program Word and Phrase to get the qualitative parameters of the texts; to analyze the quantitative and qualitative parameters of 50 blogs and systematize the material for further comparative analysis; to trace the patterns and features of the lexical organization of the English-language blog. The following research methods were used in the process of completing these tasks: the method of component analysis, the method of contextual analysis, the method of statistical analysis, and the comparative method (Laufer & Ravenhorst-kalovski: 2010; Mcbeath: 2011; Villalobos et al.: 2018; Ramírez et al.: 2019).

The following were submitted for proof in the work: quantitative and qualitative indicators work for the text stylistically; the most complex texts will be recognized as those texts in which the sentences are the longest, and those texts in which there is a greater number of polysyllabic words since in such sentences information is frames in a more complicated manner and there is a more challenging task for the reader to understand and interpret it; both of the above indicators do not have to be high, only one of them can have a high value (either the length of a sentence or the number of polysyllabic words); quantitative indicators are largely correlated with the substantive parameters of the texts – long sentences or the presence of polysyllabic words correspond to the subject of blogs; the lexical diversity of the text also stems from its thematic focus; in blogs, the number of nouns prevails over the verbal component of texts; in the Internet discourse, bloggers purposely set the goal of using the vocabulary of high frequency (Babashamsi et al.: 2013; Pavlovna et al.: 2015).

METHODS

The modern scientific paradigm considers readability formulas as tools for creating correspondence of texts to readers, thereby providing the basis for a better understanding (Тулдава: 1975; Mare-anne: 2012). They are widely used throughout the world for almost a century, but their accuracy, as well as the effectiveness of ranking the cognitive abilities of readers, have been criticized since their initial appearance. One of the most common arguments shared by experts against existing readability formulasis that they evaluate only a limited number of textual characteristics, and neglect important parameters such as lexical diversity, complexity of grammatical structures, and the general logic of textual information (Solnyshkina et al.: 2015; Vetoshchuk: 2018). «Currently, the quantitative characteristics of the text are not considered as sufficiently reliable parameters when calculating the degree of complexity of the text, since they do not take into account the referential and deep connectivity of the text as a whole (Anesa: 2018; Tarasova et al.: 2019). It is the qualitative parameters of the text that most likely allow us to calculate its complexity» (Сунгатуллина: 2018). "Comparing studies on text readability with research in the field of psychology and research in linguistics in the field of

language processing, i.e. Going beyond formulas and expanding the boundaries of text complexity assessment methods, one can come up with effective results" (Dickson et al.: 1995). We believe that for the most reliable analysis and a complete understanding, in addition to quantitative parameters, qualitative parameters should be studied. Thus, we calculated the lexical variety of the text, as well as the frequency of lexical units (Alderson: 2000; Chen: 2014).

The analysis of the quantitative parameters of text complexity assessment is carried out at the level of studying the following indicators: the type-token ratio (the indicator is based on the ratio of the number of individual lexical units (lemmas, types) and the number of their uses in the text (text forms, tokens)), average sentence length, the number of multi-syllable words in the text (more than 2 syllables), the number of verbs in the text, the number of nouns in the text, indicators of Flesch Reading Ease, Flesch-Kincaid Grade, Gunning Fox, and indicators of the lexical diversity of the text VOCD and MTLD (McLaughlin: 1969; Laufer: 1989; Fulcher: 1997).

For the analysis and comparison of data, as well as the derivation of the general patterns governing the construction of an English-language blog, we studied 50 texts. The volume of all texts studied by us in the online discourse is 350-400 words, but the length and number of sentences vary.

We have already mentioned that the complexity of a text is a complex category, determined not only by qualitative parameters, but also by quantitative ones, such as the frequency of lexical units of text. High-frequency vocabulary is extremely quickly processed in the mind of the reader, since it is associated with the corpus of facts that make up knowledge of the world. We analyzed the frequency of lexical units on the basis of the online Word and Phrase platform. In English, 1–500 lexical units are attributed to high-frequency ones, 500–3,000 to medium-frequency ones, and 3,000 to low-frequency ones.

RESULTS

A comparative analysis of the obtained values led to the following: in the texts with the largest number of sentences (maximum – 24) the length of the sentence decreases. For example, in the text number 36, consisting of 24 sentences, the average length of the sentence is 16 words: The future of great High Street retail will be around experience, knowledge and incredible service. We see similar results in the analysis of text No. 41, which consists of 24 sentences, where the average length is 16 words. The text number 24 consists of 21 sentences, the average length of which is 18 words: Every April the stockholders of the East India Company elected 24 men to serve as directors for the following year; Matters were either dealt with immediately or referred to one of the specialized committees of directors. In the text number 35, consisting of 10 sentences, the average length of each of them reaches the value of 37: Anne-Marie is vocal about the importance of diversity, not just for its own sake, but diversity of thought as a necessary part of innovation: only by introducing the female perspective in industries that are severely lacking one can we create a genuinely equal future. We concluded that the more sentences the presented blog text contains, the shorter they are and, accordingly, the opposite: the smaller the sentences in the text, the longer they are. It should be added that when composing texts, the authors of these blogs clearly try not to overload them, staying within 350 to 400 words, selecting one or another way of shaping their thoughts – numerous large sentences or few common ones.

DISCUSSION

The following thesis – the most complex texts, calculated by the formulas Flesch Reading Ease, Flesch-Kincaid Grade, Gunning Fox Index, will be recognized those texts in which the sentences are the longest, and those texts in which the greater number of complex words is found. Thus, the text number 45, where the average length of sentences – 16, received the following complexity rating: 64.89 (Flesch Reading Ease), 8.38 (Flesch-Kincaid Grade), 11.18 (Gunning Fox Index). These results can be interpreted as follows: 8th and 9th grades – ordinary English, easily understood by 13-15-year-old students. A text №35 with an average sentence of 37 words has the following indicators: 37.35, 17.24, 20.53. These results can be interpreted as follows: the level of higher education institutions, it is difficult to read. We see that the average sentence length directly affects the coefficient of text complexity. In text No. 15, the average length of sentences is not so great, but in this case the number of complex words, of which 114 are here, plays a crucial role (international, agreement, scientific, government, fundamental, ministries, collaborative, partnerships, etc.). The evaluation of the complexity of this text is as follows: 18.07, 15.94, 19.84. We interpret it in the following way: the level of a graduate of a higher educational institution, it is very difficult to read, university graduates understand best of all.

Thus, the average sentence length and the number of complex words in the text directly affect its complexity factor, since these values are present in the calculation formulas. However, the most important thing is that both indicators do not have to be high; only one of them can have a high value (either the length of a sentence or the number of polysyllabic words). These quantitative indicators correlate in many ways with the substantive parameters of the texts – long sentences or the presence of polysyllabic words correspond to the subject of blogs.

CONCLUSION

After a comparative analysis of the obtained data and considering them as components of the whole, and not separately, we came to a set of conclusions, some of which have theoretical value, and some of which have practical value.

1. Modern formulas for assessing the complexity of a text are not comprehensive enough, they lack the meanings associated with the lexical diversity of a text;
2. All the obtained indicators work on the text stylistically;
3. The pragmatic potential of the text is the result of a choice not only of the content of the message, but also of the way the language of this message is expressed;
4. The most complex texts, calculated according to the Flesch Reading Ease, Flesch-Kincaid Grade, Gunning Fox Index formulas, will be recognized those texts in which the sentences are the longest, and those texts in which a greater number of polysyllabic words is known.
5. Both of the above-mentioned indicators do not have to be high, only one of them can have a high value (either the length of a sentence or the number of polysyllabic words);

The following conclusions were also made regarding the design of texts in the online discourse:

1. When composing texts, the authors of blogs obviously try not to overload them, keeping them within a certain number of words, selecting one or another way of formalizing their thoughts – numerous succinct sentences or few common ones;
2. Quantitative indicators are largely correlated with the substantive parameters of the texts - long common sentences or the presence of polysyllabic words correspond to the subject of blogs;
3. The lexical diversity of the text also comes from its subject orientation;
4. In blogs, the number of nouns prevails over the verbal component of texts;
5. In the Internet discourse, bloggers use high-frequency vocabulary.

BIBLIOGRAPHY

ALDERSON JC (2000). Assessing reading, New York: Cambridge University Press.

ANESA P (2018). Lexical Innovation in World Englishes: Cross-fertilization and Evolving Paradigms. Routledge.

BABASHAMSI P, BOLANDIFAR S, SHAKIB N (2013). "Various models for reading comprehension process." International Journal of Applied Linguistics and English Literature, 2(6), pp. 150-154.

CHEN KY (2014). "Vocabulary, syntactic knowledge and reading comprehension: the perspective of college EFL Students." An Interdisciplinary Journal, 66, pp. 39-52.

DICKSON SV, SIMMONS DC, KAMEENUI EJ (1995). Text Organization and Its Relation to Reading Comprehension: a Synthesis of the Research, University of Oregon.

FULCHER G (1997). "Text difficulty and accessibility: Reading formula and expert judgement", System, 25(4), pp. 497–513.

LAUFER B (1989) "What percentage of lexis is essential for comprehension", From Humans Thinking to Thinking Machines, Clevedon: Multilingual Matters, pp. 316-323.

LAUFER B, RAVENHORST-KALOVSKI GC (2010). "Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension." Reading in a foreign language, 22(1), pp. 15-30.

LOGINOVA AA (2017). Ways of lexical and conceptual lacunae compensation in media texts translation. Young scholars'research in the humanities.

MARE-ANNE L (2012). Readability of research writing and text variables in readability formulas. In 11th International Symposium Parnu.

MCBEATH N (2011). Common European Framework of Reference for Languages: Learning, Teaching, Assessment, Cambridge University Press.

MCLAUGHLIN GH (1969). "SMOG Grading – A new readability formula", Journal of Reading, 12(8), pp. 639-646.

PAVLOVNA SN, ARSENTYEVA EF, SAFINA RA (2015). "Expanded metaphor and double actualization of phraseological units in advertising texts", Journal of Language and Literature, 6(1), pp. 282-286.

SOLNYSHKINA MI, HARKOVA EV, KISELNIKOV AS (2014). "Comparative Coh-Metrix Analysis of Reading Comprehension Texts: Unified (Russian) State Exam in English vs Cambridge First Certificate in English", English Language Teaching, 7(12), pp. 65-76.

TARASOVA FH, VARLAMOVA EV, GURYANOV IO (2019). "Syntactic Complexity of Russian Unified State Exam Texts in English: A Study on Reliability and Validity", Journal of Research in Applied Linguistics, 7, pp. 228 – 240.

VETOSHCHUK V (2018). Linguistic peculiarities of neologisms and ways of their translation.

VIERA RT (2017). "Vocabulary knowledge in the production of written texts: a case study on EFL language learners." Revista Tecnológica-ESPOL, 30(3).

VILLALOBOS ANTÚNEZ, J & RAMÍREZ MOLINA, R (2018). "El derecho a la autobiografía: dimensión ius-filosófica desde la perspectiva de H. Arendt y P. Ricoeur". Opción. Revista de Ciencias Humanas y Sociales, 34(18), pp. 1012-1587.

RAMÍREZ MOLINA, R., MARCANO, M., RAMÍREZ MOLINA, R., LAY RABY, N & HERRERA TAPIAS, B (2019). "Relationship Between social intelligence and resonant leadership in public health Institutions". Opción. Revista de Ciencias Humanas y Sociales, 35(90), pp.2477-9385.

BIODATA

ELENA VALERIEVNA VARLAMOVA, Born 1982, candidate of philological sciences. In 2004 she graduated from Kazan State University. IN AND. Ulyanov-Lenin. Specialty - "Philologist. Teacher". In 2010, she defended her thesis entitled "The Image of America in US Literature in the First Half of the 20th Century (Sherwood Anderson, John Dos Passos, Thornton Wilder)." Associate Professor, Department of Romano-Germanic Philology Institute of Philology and Intercultural Communication KFU. Research interests: lexicology, lexicography, sociolinguistics, phraseology, text complexity.

DARYA DMITRIEVNA KORNILOVA, Born in 1998, graduate of the Department of Romano-Germanic Philology 2020, direction "Philology", profile "Foreign philology: English language and literature, translation studies". Research interests: lexicology, lexicography, sociolinguistics, phraseology, text complexity.

OLGA SERGEEVNA SAFONKINA, Born in 1980, candidate of philosophical sciences. Graduated from the National Research Mordovian State University. N. P. Ogaryova in 2002. Ph.D. thesis on "Women's subjectivity in the problem field of feminism: textual practices" in 2008. Associate Professor of the Department of English Philology, Deputy Dean of the Department of Foreign Languages for International Cooperation. Research interests: linguodidactics, project management in the field of education, intercultural communication, cognitive linguistics.