



<https://doi.org/10.22363/2687-0088-35817>

EDN: NAZBGA

Research article / Научная статья


## Cognitive complexity measures for educational texts: Empirical validation of linguistic parameters

Roman V. KUPRIYANOV<sup>1,2</sup>  , Olga V. BUKACH<sup>2</sup>   
and Oksana I. ALEKSANDROVA<sup>3</sup> 

<sup>1</sup>*Kazan National Research Technological University, Kazan, Russia*

<sup>2</sup>*Kazan (Volga Region) Federal University, Kazan, Russia*

<sup>3</sup>*RUDN University, Moscow, Russia*

 [kroman1@mail.ru](mailto:kroman1@mail.ru)

### Abstract

The article presents a study conducted within the framework of discourse complexology – an integral scientific domain that has united linguists, cognitive scientists, psychologists and programmers dealing with the problems of discourse complexity. The issue of cognitive complexity of texts is one of the central issues in discourse complexology. The paper presents the results of the study aimed to identify and empirically validate a list of educational texts' complexity predictors. The study aims to identify discriminant linguistic parameters sufficient to assess cognitive complexity of educational texts. We view text cognitive complexity as a construct, based on the amount of presented information and the success of reader–text interactions. The idea behind the research is that text cognitive complexity notably increases across middle and high schools. The research dataset comprises eight biology textbooks with the total size of 219,319 tokens. Metrics of text linguistic features were estimated with the help of automatic analyzer RuLingva ([rulingva.kpfu.ru](http://rulingva.kpfu.ru)). Linguistic and statistical analysis confirmed the hypothesis that text syntactic and lexical parameters are discriminative enough to classify different levels of cognitive complexity of educational texts used in middle and high schools. Text parameters that manifest variance in cognitive complexity include lexical diversity (TTR); local argument overlap; abstractness index; number of polysyllabic words, Flesch–Kincaid Grade Level; number of nouns and number of adjectives per sentence. Empirical evidence indicates that the proposed approach outperforms existing methods of text complexity assessment. The research results can be implemented in the system of scientific and educational content expertise for Russian school textbooks. They can also be of some use in the development of educational resources and further research in the field of text complexity.

**Key words:** *discourse complexology, cognitive complexity, text complexity, educational text, statistical analysis*



**For citation:**

Kupriyanov, Roman, Olga Bukach & Oksana Aleksandrova. 2023. Cognitive complexity measures for educational texts: Empirical validation of linguistic parameters. *Russian Journal of Linguistics* 27 (3). 641–662. <https://doi.org/10.22363/2687-0088-35817>


## **Оценка когнитивной сложности учебных текстов: эмпирическая валидация лингвистических параметров**

**Р.В. КУПРИЯНОВ<sup>1</sup>  , О.В. БУКАЧ<sup>2</sup> , О.И. АЛЕКСАНДРОВА<sup>3</sup> **

<sup>1</sup>*Казанский национальный исследовательский технологический университет,  
Казань, Россия*

<sup>2</sup>*Казанский (Приволжский) федеральный университет, Казань, Россия*

<sup>3</sup>*Российский университет дружбы народов*

 [kroman1@mail.ru](mailto:kroman1@mail.ru)

### **Аннотация**

В статье представлено исследование, проведенное в рамках дискурсивной комплексологии – интегрального научного направления, объединяющего лингвистов, когнитологов, психологов и программистов, которые занимаются проблемами сложности дискурса. Проблема когнитивной сложности текстов является одной из центральных в дискурсивной комплексологии. В работе показаны результаты исследования по выявлению и эмпирической валидации перечня предикторов сложности учебных текстов. Цель данного исследования – выявить дискриминантные лингвистические параметры, достаточные для установления уровня когнитивной сложности учебных текстов. Мы рассматриваем когнитивную сложность текста как конструкт, в основе которого лежит объем представленной информации и успешность взаимодействия читателя с текстом. В основе данного подхода – идея о том, что когнитивная сложность текста заметно возрастает в средних и старших классах общеобразовательной школы. Набор исследовательских данных включает восемь учебников по биологии общим размером 219 319 токенов. Метрики языковых особенностей текста оценивались с помощью автоматического анализатора RuLingva ([rulingva.kpfu.ru](http://rulingva.kpfu.ru)). Лингвистический и статистический анализ подтвердил гипотезу о том, что синтаксические и лексические параметры текста достаточно различны, чтобы позволить классифицировать различные уровни когнитивной сложности учебных текстов, используемых в средней и старшей школе. Параметры, манифестирующие различия в когнитивной сложности, включают лексическое разнообразие (TTR), локальную связность, индекс абстрактности, количество многосложных слов и индекс Флеша-Кинкейда, количество существительных и количество прилагательных в предложении. Эмпирические данные показывают, что предлагаемый подход является более эффективным по сравнению с другими существующими методами оценки сложности текста. Результаты исследования могут быть внедрены в систему экспертизы научно-образовательного содержания российских школьных учебников. Они также могут быть полезны при разработке образовательных ресурсов и дальнейших исследованиях в области сложности текста.

**Ключевые слова:** *дискурсивная комплексология, когнитивная сложность, сложность текста, учебный текст, статистический анализ*

**Для цитирования:**

Kupriyanov R.V., Bukach O.V., Aleksandrova O.I. Cognitive complexity measures for educational texts: Empirical validation of linguistic parameters. *Russian Journal of Linguistics*. 2023. V. 27. № 3. P. 641–662. <https://doi.org/10.22363/2687-0088-35817>

## 1. Introduction

The term ‘complexity’ appears in different research areas. As the modern world is becoming increasingly complex, unpredictable and non-linear, it is hardly surprising that in recent years the phenomenon of ‘complexity’ has attracted the attention of numerous scholars. One of the most famous works on this topic is the work “On Complexity” by the French philosopher Edgar Morin (2021), where the author presented the conceptual apparatus of his theory of complexity viewed by many as “a bold challenge to the fragmentary and reductionistic spirit that continues to dominate the scientific research” (Rodrigues et al. 2014: 1). According to Morin, it is complexity that underlies the majority of natural and social phenomena. “Complexity asserts itself first of all as an impossibility to simplify; it arises where complex unity produces its emergences” (Morin 2005: 386).

Complexity studies as a wide and dense research field provide numerous descriptions of ‘cognitive complexity’, investigated in computer science, psychology, pedagogy and linguistics (Andrews 2002, Weir 2008, Wang 2012, McComb 2016, Wijendra 2021, Lavazza 2022, Sharoff 2022, Solnyshkina 2022, Solovyev et al. 2022, Silva 2023). In computer science, researchers scrutinize cognitive complexity of a program code viewed as a risk factor which can cause problems with program debugging, technical support and program modernization. Special metrics have been developed to assess the level of program code cognitive complexity, as well as algorithms to reduce it (Bolbakov 2016, Gladkikh 2017, Wijendra 2021, Lavazza 2022). In psychology, the term ‘cognitive complexity’ is used as a characteristic of a person, or rather a psychological characteristic of a person’s cognitive sphere. Cognitive complexity in this sense reflects “the degree of categorical dismemberment (differentiation) of an individual’s consciousness, which contributes to the selective sorting of impressions of reality that mediates the individual’s activity” (Petrovsky 1998: 164).

Cognitive psychology views ‘cognitive complexity’ as a construct, a coherent whole comprised of psychological characteristics of independent components which are interrelated and connected. Cognitive complexity of a situation is viewed as a function of the number of its elements and their connections (Kholodnaya 2004). Cognitive complexity “is interrelated with a subject's real behavior, his flexibility and adaptability; it relies on the degree of an individual’s freedom to make decisions in a specific content area” (Petrenko 2010: 83).

For over a decade, cognitive complexity has also been an object of research in pedagogy and education (Kudzh 2018). Research in the area focuses on complexity of educational resources, comparative complexity of a subject domain, comparative difficulty of mastering a topic, etc. As students progress, they are expected to be

provided with resources of appropriate complexity, i.e. as they evolve, so do educational materials. The latter are the only foundations of the so-called “zone of proximal development” as “the distance between the actual development level, determined by independent problem solving and the level of potential development, determined through problem solving under adult guidance or in collaboration with more capable peer” (Vygotsky 1978: 86).

Linguistics lacks a unified approach to cognitive complexity of a ‘linguistic whole’ thus referring to the notion either as complexity, difficulty, or ‘accessibility’ (see Fulcher 1997, Solnyshkina 2015). The common doctrines regarding cognitive complexity of a text comprise three traditions and are viewed as a function of: (1) text informativeness (Hansen 1990, Valgina 2003, Nevdakh 2008, Zhu 2020); (2) linguistic features of a text (Ushakov 1980, Solnyshkina 2020, Gatiyatullina 2023); (3) reader’s (in)ability to process a text or accessibility of a text to the reader (Tsetlin 1980, Just 1987, Fulcher 1997, Valgina 2003, Koda 2005, Das 2020). Interdependence and interrelatedness of the phenomena is obvious: texts addressed to different categories of readers differ in their informativeness: a 200-word text for elementary school readers is cognitively less complex than a summary of the same length addressed to college students. Higher informativeness is likely to be manifested in more complex syntax and vocabulary of lower frequency thus being less accessible for less qualified readers. Hence, cognitive complexity of a text is viewed by the authors as a text ‘accessibility’, i.e., degree of cognitive efforts a reader employs to decode and comprehend it.

We aim to identify a list of linguistic features of informational texts exhibiting various degrees of cognitive complexity. We hypothesize that different degrees of cognitive complexity of informational texts are manifested in syntactic, lexical and morphological parameters. In other words, cognitive complexity ranks corresponding to different grade levels reveal themselves on the syntactic, lexical and morphological levels. We also hypothesize that grade differences are stable and as such, once identified for certain grade levels, they could be used to predict the cognitive complexity of a text.

The study was conducted to answer two Research Questions:

RQ 1: What linguistic parameters discriminate texts cognitive complexity?

RQ 2: What are the ranges of values of cognitive complexity predictors inherent in informational texts for Grades 5–11?

## **2. Text comprehension and complexity**

Text comprehension depends on a range of factors. Cognitive psychology describes this process as developing inner understanding, based on the knowledge previously gained (Matlin 1998, Polya 2015). Reading comprehension begins and finishes with non-verbal representation, also referred to as a mental model (Johnson-Laird 1983). It is also noted that comprehension is individual and closely connected with the background knowledge a reader acquired: no matter what people are trying to comprehend, they draw on their background knowledge (Matlin 1998,

Wang 2012, Polya 2015). It is for this reason separating the study of text comprehension and the study of a reader's specific features would seem unfeasible.

Weir's model of reading comprehension reflects both lower-level processes (e.g., decoding) and higher-order processes (e.g., comprehension) and comprises seven aspects (Weir & Khalifa 2008): 1) word recognition, the ability to "match the form of a word in a written text with a mental representation of the orthographic forms of the language" (Weir 1993: 6); 2) lexical access, the "retrieval of a lexical entry from the lexicon, containing stored information about a word's form and its meaning"; the form addresses orthographic and phonological mental representations of a lexical item and possibly information on its morphology (Field 2004); 3) syntactic parsing which involves grouping "words into phrases, and into larger units at the clause and sentence level to understand the message of the text" (Weir & Khalifa 2008: 6); 4) establishing propositional (core) meaning at the clause or sentence level, "a literal interpretation of what is on the page, the reader has to add external knowledge to it to turn it into a message that relates to the context in which it occurred" (Weir 2003: 6); 5) inferencing as "a creative process whereby the brain adds information which is not stated in a text in order to impose coherence" (Weir 2003: 6); 6) building a mental model "entails an ability to identify main ideas, to relate them to previous ideas, distinguish between major and minor propositions and to impose a hierarchical structure on the information in the text" (Field 2004: 241), it is when "the propositions representing the meaning of a text are linked together, usually by argument overlap, to form a hierarchical text base" (Kintsch & van Dijk 1978: 374); 7) creating a text (or discourse) implies not only recognizing the hierarchical structure of the whole text but also determining which items of information are central to the meaning of the text and which are secondary propositions, that is the ability to recognize significance of different parts of the text to the writer or reader (Weir 2003: 6).

In cognitive linguistics, text comprehension is irrevocably linked with the analysis of a fundamental problem known as cognitive complexity of texts and its relationship to syntactic and semantic complexity in natural languages. According to McCarthy, text comprehension is a function of several factors (McCarthy 2019). First, there is complexity of the very idea that the author conveys through the text, i.e. the so-called text informational complexity. Hence, the amount of information, i.e. text information intensity is often seen as a text complexity predictor (Hansen 1990, Valgina 2003, Nevdakh 2008, Zhu 2020). Second, there are linguistic means selected by the author to express his/her ideas (Valueva 2017). Consequently, text complexity is viewed as a text characteristic dependent on the text linguistic parameters. This type of complexity is referred to as linguistic complexity (Solnyshkina 2020). Text linguistic parameters associated with text complexity include but are not limited to the following: average sentence length, number of polysyllabic words, genitive case mean, word frequency, narrativity, number of abstract words, lexical density, etc. (Ushakov 1980, Solnyshkina 2020, Gatiyatullina 2020, 2023). Mathematical methods introduced into complexity

studies confirm statistical significance of numerous text parameters with *linguistic* complexity thus enabling to categorize them as complexity predictors (Krioni 2008, Nevdakh 2008, Sheehan 2010, Fitzgerald 2015, Valueva 2017, Solovyev 2021, Kupriyanov 2022, Shardlow 2022). The third factor is reader's ability to process information. The idea behind it is that text comprehension involves cognitive activity, intensity of which is related to the type of information to be processed (Dehaene 2007). Consequently, a reader is expected to meet certain cognitive requirements, and readers' characteristics affect the understanding of the information embedded in the text. The most significant among the readers' characteristics are general knowledge, knowledge of the subject domain, verbal intelligence, cognitive abilities of the reader including working memory and motivation, etc. (Tsetlin 1980, Just 1987, Valgina 2003, Koda 2005, Das 2020).

### **3. Cognitive complexity of informational texts and cognitive abilities of schoolchildren**

The research shows that the phenomenon of cognitive complexity manifests itself in the interaction of a person with the outside world and therefore is associated with the human factor (Kudzh 2018). Hence, evaluating cognitive complexity of a text implies considering two types of factors, i.e. text parameters which are objective and stable, and reader's characteristics which are individual and variable. Thus, the same text may be quite easy to comprehend for one category of readers and demanding for another group of readers. This occurs due to the fact that the concept of complexity is de facto formed on the basis of psychological or cognitive factors (Kudzh 2018).

As a construct and psychological characteristics of a person's cognitive sphere (Petrovsky 1998), cognitive complexity of a person also implies numerous abilities, including the ability to identify characteristics of an object or the ability to evaluate characteristics and reveal links between and/or among them. In other words, cognitive complexity encapsulates cognitive differentiation and integration of individual consciousness (Kalinkina 2021). The concept of cognitive complexity, first propounded by James Bieri as early as in 1955, is to do with the organization of constructs and their similarity. Cognitive complexity has also been defined as "an aspect of a person's cognitive functioning which at one end is defined by the use of many constructs with many relationships to one another (complexity) and at the other end by the use of few constructs with limited relationships to one another (simplicity)" (Pervin 1984: 507). Cognitive complexity describes an individual's ability to perceive things in the world around them. It also describes the number of cognitive processes required to solve a problem or complete a task. Individuals with more complex cognition can see shades of nuance and meaning. A person of high cognitive complexity is able to perceive nuances and see subtle differences between the objects he/she perceives, understand connections between events and phenomena.

Cognitive complexity is also argued to be essential for understanding a complex and uncertain environment (Da'as 2020). One of the methods to evaluate a person's cognitive complexity involves measuring the number of classification bases that an individual uses consciously or unconsciously while differentiating between objects in any content area. The basic condition necessary for an individual to acquire a high level of his/her cognitive abilities is through verbal intelligence which enables a person to classify objects on various grounds. Inability to name different types of objects causes difficulties while performing mental operations with them.

In the most general terms, cognitive characteristics of people are traditionally ranked based on their age. While selecting books for readers, educators use their age as the basic classifying principle matching texts for primary or middle school, high school or University level. Aging is strongly associated with cognitive complexity growth, since life experience allows people to perform cognitive operations they were unable to perform in childhood. Therefore, cognitive complexity of informational texts increases alongside with the readers' age they are addressed to. This is especially evident with books assigned to schoolchildren when their cognitive abilities undergo intensive development. As regards informational texts specifically, their classification is based on the level of readers' education, for example, "Grade 6 textbook", "Books for College Students", etc. Readers' age characteristics and their level of education are generally viewed as interdependent.

Cognitive complexity of a person in childhood and adolescence is closely related to the psychophysiology of developing his/her cognitive sphere, as cognitive abilities develop gradually alongside with the extension of memory, attention, etc. The research shows that the very style of thinking changes as a person matures (Perry 1981). Perry (1981) argues that learning changes students' ideas about the nature of knowledge. Another area of development is reasoning, the nature of which modifies dramatically as students learn to organize and evaluate knowledge. Perry (1981) grouped these changes into four cognitive stages: dualism, multiplicity, relativism, and commitment. The first stage, that is, dualism, is characterized by a dichotomous structure in which information is divided into two categories: correct and incorrect. In the second stage, the dualistic structure is discarded and replaced by uncertainty. Knowledge becomes subjective: there are conflicting decisions, therefore, one needs to trust their "inner voice", but not some external authority. The third stage is relativism. Knowledge is relative to the context in which a decision is made, hence it is necessary to learn how to evaluate decisions based on the situation. In the fourth stage, there is an integration of knowledge received from others with personal experience and reflective analysis. People rely not only on their knowledge, but also on values; they use moral and ethical positions to make decisions.

Theoretical, formal and reflective thinking typically begins to develop in the cognition of middle school students and is viewed as their age-related feature. The cognitive system of an early adolescent still has a relatively small number of loosely

coupled constructs thus reflecting his/her low cognitive differentiation and low values of cognitive integration. It will be fully formed at the next stage of development (youth) in high school (Konogorskaya 2014). The transition from childhood to adulthood is usually divided into two stages: adolescence and youth (early and late). As regards the research described in the given article, it focused on middle school students of 11 to 15 years of age and high school students of 15 years of age and above. The research shows that the ability to reflect and introspect one's inner world develops in youth. The brainwork of a high school student, unlike that of a middle school student, is more affective and personalized, it is at this age that people form their worldview, search for their place in society and identify their life goals. Cognitive abilities in adolescents reach a maximum, they form and develop skills of abstract thinking, from 13 to 16 years of age the ability to memorize increases dramatically, theorizing becomes their age-related feature. High school students are able to independently perform planning, putting forward and testing hypotheses, which indicates the ability for scientific thinking (Klyueva 2003). Academic success in high school largely depends on the extent to which a person managed to develop conceptual thinking (Konogorskaya 2014).

Levels of individual cognitive complexity, though they vary from child to child, are expected to be matched with the reading texts children are exposed to. Texts that are not suitable for readers' cognitive abilities, i.e. either too easy or too difficult, may cause lack or loss of students' motivation, their boredom or frustration. The latter may result in inability to develop high reading skills (McCarthy 2019). Therefore, selection of educational texts assigned to readers of a certain age is viewed as a global problem and a practical task in numerous fields. The modern interdisciplinary paradigm of discourse complexology identifies appropriateness of a text to readers' cognitive complexity, i.e. text cognitive complexity, based on a number of text variables. The latter comprise significant characteristics of the object(s) described or events narrated, dimension of the semantic space presented in the text and links between its elements.

#### **4. Dataset and research methodology**

The dataset comprises Russian textbooks for middle (Grades 5 and 6) and high school (Grades 10 and 11). To reduce the statistical noise which the author's individual style and subject domain of academic disciplines may produce, we selected eight biology textbooks by the same authors for middle and high school and grouped them into "Level II" (Grades 5, 6) and "Level III" (Grades 10 and 11). The research was based on the two premises that are widely accepted in the modern interdisciplinary paradigm of discourse complexology mentioned above: a) high school students, on average, possess higher levels of cognitive complexity compared to middle school students; b) textbooks for different academic levels differ in cognitive complexity. A textbook designed for a higher level has a higher cognitive complexity.



The Research was designed in 3 stages. The first, preparatory, stage included selecting and preprocessing the dataset, i.e. 8 biology textbooks. To ensure consistency of the genre and content, we deleted text meta descriptions, prefaces, authors' introductory words, contents, illustrations, inscriptions, phrases like "Figure 1", notes, self-control questions, laboratory tasks, chapter titles, subheadings, footers and afterword. Then we divided the materials into 220 texts with about 1000 tokens in each. The range of the text size varied between 959 and 1143 tokens, all texts comprised full sentences. The total size of the research corpus comprised 219,319 tokens (Table 1).

Table 1. Corpus Size

Academic Level	Textbook code	Grade	Number of tokens	Number of texts
Level II	V-5	5	23 919	24
	S-5	5	13 784	14
	V-6	6	22 994	23
	S-6	6	15 689	16
Total			76 658	77
Level III	V-10	10	43 871	44
	S-10	10	24 871	25
	V-11	11	33 969	34
	S-11	11	39 950	40
Total			142 661	143

In the second stage, we measured metrics of text parameters with the automatic text profiler RuLingva<sup>1</sup>. Indices of educational texts comprise relative indices, i.e. assessed on two relative scales, 1000 tokens and per sentence, and composite indices. Composite indices include more than one variable. All parameters analyzed were divided into four groups:

1. Syntactic indices: average word length, average sentence length, average number of nouns, verbs, adjectives per sentence.
2. Descriptive indices: readability (Flesch-Kincaid index, FK(SIS)), abstractness, local argument overlap, global argument overlap, lexical diversity (or Type token ratio, TTR per 1000 tokens), ratio of verbs to nouns, ratio of adjectives to nouns or descriptiveness, rate of nouns in the genitive case.
3. Morphological indices: number of nouns in a certain case per 1000 tokens.
4. Phonological indices: number of monosyllabic, disyllabic, trisyllabic and four-syllabic words per 1000 tokens.

The third stage was analytical. In this stage, we employed Mann-Whitney test to compare and contrast indices of educational texts used on Level II and Level III. Statistical analysis of the data was carried out with the software STATISTICA.

<sup>1</sup> [rulingva.kpfu.ru/](http://rulingva.kpfu.ru/) (ENA, 20.08.2023).

## 5. Results

The results of statistical analysis demonstrate that educational texts of different cognitive complexity have statistically significant differences on the syntactical, morphological and phonological levels. As Table 2 below indicates, most of the indices measured demonstrated statistically significant differences of Level II (Grades 5 and 6) and Level III (Grades 10 and 11) texts. The exceptions are Number of verbs per sentence and Global argument overlap.

Table 2. **Biology textbook indices**

I	Indices	Level II (N = 77)		Level III (N = 143)		Mann-Whitney U	p-value
		Mean	SD	Mean	SD		
	II	III	IV	V	VI	VII	VIII
Syntactic indices							
1.	Mean sentence length (in tokens)	11,66	1,34	12,81	1,54	3032	< .01*
2.	Mean word length (in syllables)	6,10	0,24	6,69	0,22	392	< .01*
3.	Mean nouns per sentence	4,68	0,59	5,39	0,69	2258	< .01*
4.	Mean verbs per sentence	1,54	0,27	1,53	0,23	5336	0,71
5.	Mean adjectives per sentence	1,63	0,30	2,04	0,33	2084	< .01*
Descriptive indices							
6.	FK (SIS)	7,27	0,89	9,30	0,74	451	< .01*
7.	Abstractness	2,57	0,10	2,66	0,10	2888	< .01*
8.	Local argument overlap	0,61	0,20	0,55	0,16	4480	0,023*
9.	Global argument overlap	0,23	0,08	0,21	0,07	4750	0,093
10.	TTR per 1000 tokens	0,45	0,04	0,47	0,04	4067	< .01*
11.	Ratio of verbs to nouns	0,33	0,05	0,28	0,04	2483	< .01*
12.	Descriptiveness (adjective-to-noun)	0,35	0,05	0,38	0,05	3486	< .01*
13.	Rate of nouns in the genitive case (to the total number of nouns in the text)	0,34	0,04	0,40	0,03	1487	< .01*
Morphological indices							
14.	Nominative case (NOUNS) per 1000 tokens	113,98	16,55	101,63	12,84	2927	< .01*
15.	Genitive case (NOUNS) per 1000 tokens	138,62	20,47	169,03	18,53	1468	< .01*
16.	Dative case (NOUNS) per 1000 tokens	14,07	5,16	15,91	6,03	4536	0,03*
17.	Accusative case (NOUNS) per 1000 tokens	69,13	13,07	64,17	11,20	4201	< .01*
18.	Instrumental case (NOUNS) per 1000 tokens	28,78	7,19	26,28	6,40	4251	< .01*
19.	Prepositional case (NOUNS) per 1000 tokens	35,87	9,01	39,52	8,76	4194	< .01*
Phonological indices							
20.	Number of one-syllable words per 1000 tokens	195,50	20,55	169,92	18,03	1865	< .01*
21.	Number of two-syllable words per 1000 tokens	258,88	27,93	201,19	21,68	543	< .01*
22.	Number of three-syllable words per 1000 tokens	215,83	20,38	203,95	20,47	3720	< .01*
23.	Number of four-syllable words per 1000 tokens	165,01	21,86	181,31	23,86	3286	< .01*

\*  $p < .05$  — statistically significant differences

The most significant differences observed across Levels II and III include the following:

1. The number of nouns per sentence increases by 15% (from 4.68 on Level II to 5.39 on Level III).
2. The average number of adjectives per sentence grows by 25% (from 1.63 to 2.04).
3. Readability (FK (SIS)) rises by 28% (from 7.27 to 9.30).
4. Ratio of verbs to nouns decreases by 15% (from 0.33 to 0.28), which means that the nominativity of the text increases.
5. The number of nouns in the genitive case increases by 22% (from 138.62 to 169.03), while the proportion of nouns in the genitive case to nouns in other cases increases by 18% (from 0.34 to 0.40).
6. The number of disyllabic words decreases by 22% (from 258.88 to 201.19).

Below we interpret and explain the meanings and effects of the trends observed in light of the theory of cognitive complexity. The most obvious growth trajectory is that of readability (FK(SIS)): from 7.27 on Level II to 9.30 at Level III (Fig. 1). As readability metrics is based on two indices, i.e. sentence and word length, its growth implies that Level III readers are expected to possess higher cognitive complexity. Successful comprehension of Level III texts is cognitively more difficult and requires more effort.

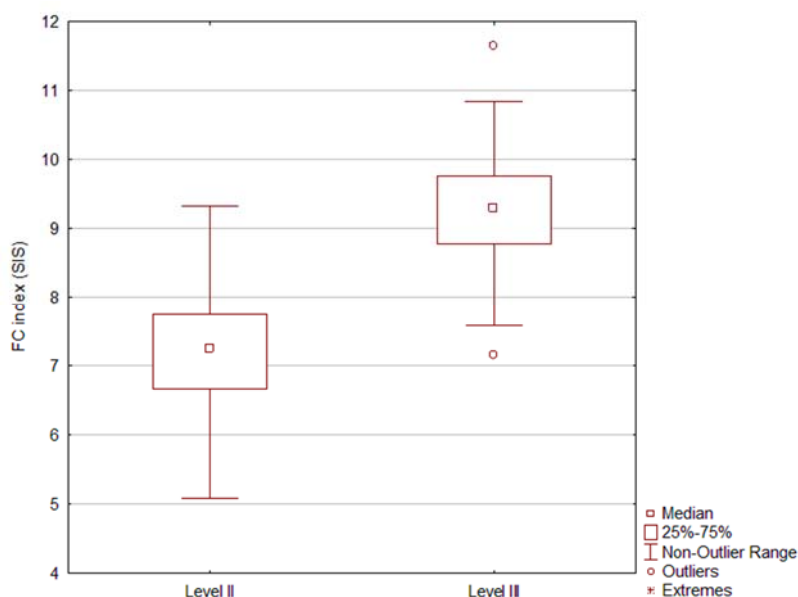


Figure 1. Flesch-Kincaid (SIS) growth across Levels II – III

Statistically significant are also dynamics of Local argument overlap and TTR values: the values of local argument overlap decreases from 0.61 to 0.55, TTR per 100 tokens (Lexical diversity) rises across levels from 0.45 to 0.47 (cf. Fig. 2 below).

The increase in Lexical diversity of educational texts and the decrease in cohesion manifested in both Local and Global argument overlap (cf. Table 2) in Level III texts are accompanied by Abstractness growth (from 2.57 on Level II to 2.66 on Level III). The latter indicates that Lexical diversity increases to a certain extent due to appearance of abstract words, which is typical of scientific terminology (Fig. 3).

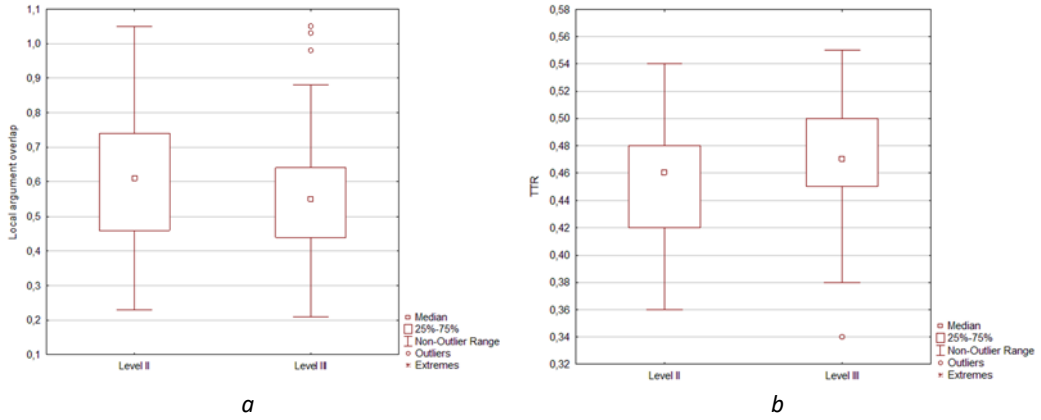


Figure 2. a) Local argument overlap; b) TTR

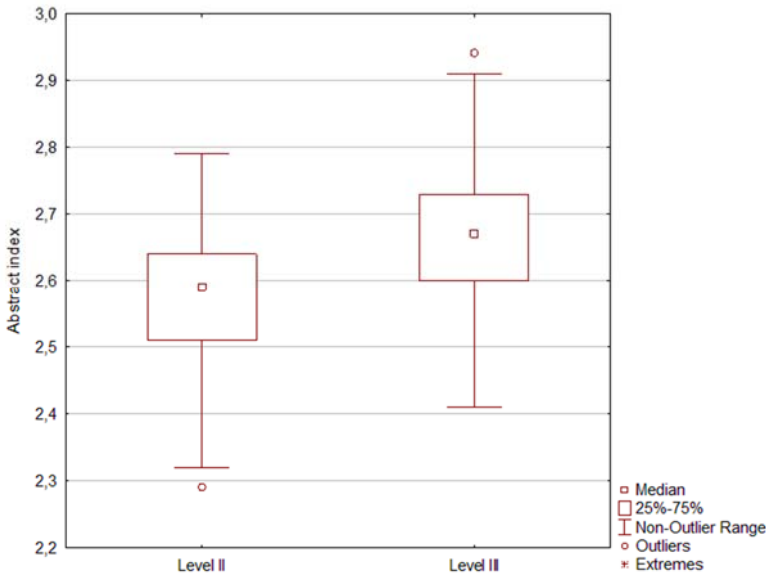


Figure 3. Abstractness

Lexical diversity growth coincides with the increase of two other metrics, i.e. the number of nouns (from 4.68 to 5.39) and number of adjectives per sentence (from 1.63 to 2.04), while the metric of number of verbs across levels remains intact. Thus, on Level III we observe the growth of nominativity and

descriptiveness (Fig. 4) which in fact manifests higher informativeness and cognitive complexity of texts.

A detailed analysis of noun morphology indicates that the noun per sentence increase occurs mainly due to the sharp increase of genitive case (Fig. 5).

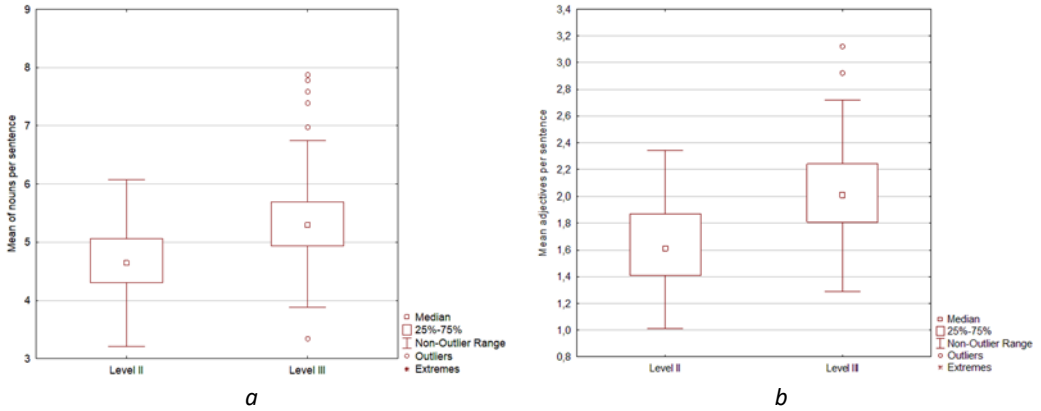


Figure 4. a) Mean nouns per sentence; b) Mean adjectives per sentence

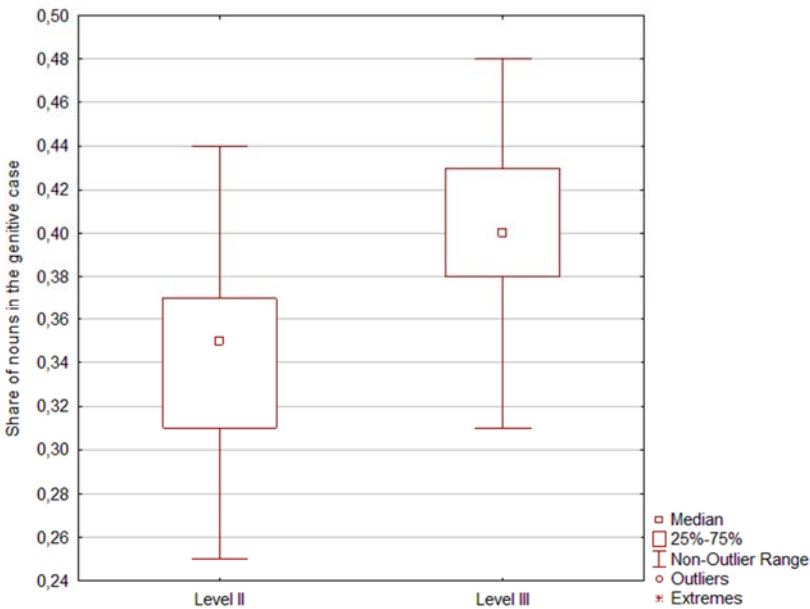


Figure 5. Genitive case nouns

Significantly, increase of nouns in the genitive case (from 138.62 to 169.03) is accompanied by the decrease in the number of nouns in the nominative case (from 113.98 to 101.63) (cf. Fig. 6 below).

Text phonological parameters reflect statistically significant interdependence between text academic levels and word length. Trajectories of mono-, di-, three- and four-syllabic words fluctuations in educational texts of different cognitive

complexity indicate that the most significant changes are manifested by disyllabic words (Fig. 7). In Level III texts, the number of disyllabic words decreases by 22% (from 258.88 to 201.19), while a similar tendency is not observed in Level II texts. The number of four-syllabic words is significantly higher in Level III texts in comparison with Level II texts: 181.31 and 165.01, respectively.

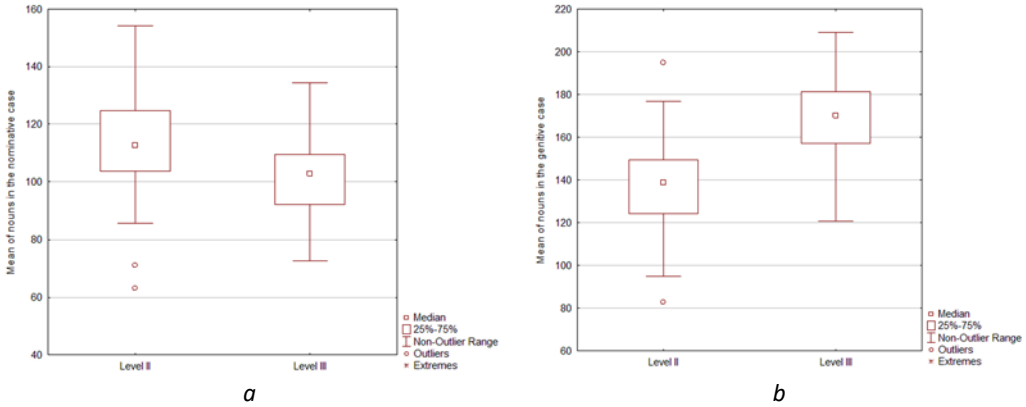


Figure 6. a) Nouns in the nominative case; b) Nouns in the genitive case

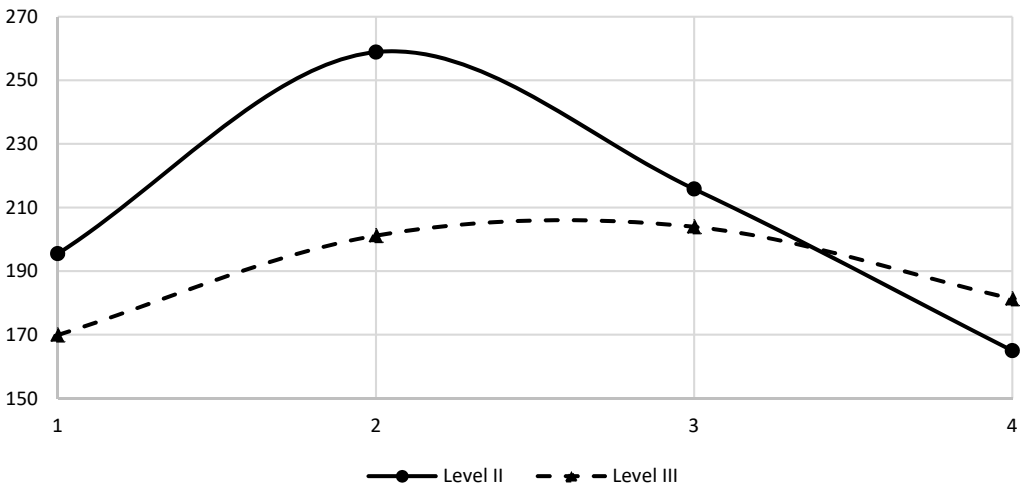


Figure 7. Numbers of mono-, di-, three- and four-syllabic words in educational texts of different levels \*

## 6. Discussion

The data received provides insights into specifics of cognitive complexity of educational texts and its manifestation on the phonological, morphological, lexical and syntactical levels.

Our main findings referring to the *phonological* level features fluctuations reveal that text complexity across academic levels grows mostly due to the dramatic decrease of disyllabic words and slight increase of four-syllabic words. These findings are new and as such need further investigation. To the best of our

knowledge, trajectories of metrics of phonological parameters of Russian educational texts across academic levels have been predominantly studied in terms of word length but not specified to the ratio of mono- or disyllabic words (Solnyshkina 2015). Matskovskiy readability formula based on publicistic texts (Matskovskiy 1976) incorporates the variable of three-syllabic words:  $X1 = 0,62 X2 + 0,123 X3 + 0,051$ , where  $X1$  – text complexity;  $X2$  – average sentence length (in words);  $X3$  – percentage of three-syllabic words. However, researchers argue that Matskovskiy formula was not validated and cannot be used as such to assess text complexity.

Our findings of the notable increase of nouns in genitive case as a *morphological* manifestation of cognitive complexity growth are in correspondence with the conclusions made by Gatiyatullina et al (2020) and drawn on the research of Social Science texts. We share the researchers' view on a noun in the genitive case as a dependent word denoting belonging, composition, participation and origin of an object (see Blake 2001), and its increase as such reflects the growing number of multi-element terms that complement and clarify meanings in educational texts.

As regards the findings on the share of different parts of speech per sentence, they do correspond to those of O. B. Sirotina (Sirotina 2009) and A. F. Zhuravlev (Zhuravlev 1988), but nevertheless are new with regard to educational texts and may be used as referential indices in assessment of textbook complexity.

Comparison of Level III and II educational texts metrics also revealed a significant increase in Lexical diversity and Abstractness, which implies that higher academic levels correspond to higher levels of students' verbal intelligence, general knowledge and scientific background. Increase of Lexical diversity and Abstractness of educational texts in high school (Level III) reflects age-related changes in human psyche, i.e. developing the ability for abstract thinking. Thus, we confirm Konogorskaya's (2014) conclusion that high school students are expected to exercise the ability for abstract thinking, which is viewed as the basic intellectual ability determining students' progress.

Having confirmed the general idea that short texts are easier to comprehend and better to remember (Kotova 2021), we also identified the specifics of general interdependence of sentence length on the one side and academic levels on the other. Our findings on sentence length specify metrics range characteristic for Level II and III texts thus serving as a framework for recommendations addressed to test and educational material developers. The revealed increase of cognitive complexity of educational texts across academic levels implies that high school students are expected to be able to apply more cognitive effort to comprehend longer sentences and complex syntactic constructions while processing a text (Dobrynina 2019). The data obtained are also in good agreement with Valueva's (2017) findings on cognitive complexity of literary texts.

## 7. Conclusions

Linguistic and statistical analysis confirmed the hypothesis on linguistic differences between educational texts of different cognitive complexity. These differences manifest themselves in statistically significant dynamics of numerous parameters from Level II (middle school) to Level III (high school). Lexical diversity (TTR), text abstractness, sentence and word lengths, Flesh-Kincaid index, nominativity (nouns per sentence) and descriptiveness (number of adjectives per sentence) increase notably, while local cohesion (Local argument overlap) demonstrates a tendency to decrease. Thus, the research revealed that cognitive complexity of educational texts manifests itself in longer sentences, higher abstractness and lexical diversity, growth of polysyllabic words and lower local cohesion. All these set higher cognitive demands for the activities of high school students who are expected to have a high capacity of working memory and sustained attention involving concentration.

The research possesses certain strengths as well as a number of limitations. This study examines the influence of linguistic parameters on the cognitive complexity of texts. These parameters are only some of the factors that determine the cognitive complexity of a text. Additional research is needed to assess the influence of other factors, such as propositional density (P-density) and terminological density of the text and characteristics of its readers (students' background knowledge and level of their cognitive development). It is also important to note that the scope of our research is to a certain extent limited by the existing functionality of RuLingva text profile; more specifically, at present the said profiler does not provide its users with the possibility to automatically single out scientific terms from the analyzed texts.

Nevertheless, it seems reasonable to conclude that the data obtained may serve as a framework for a text complexity profiler able to identify markers of cognitive complexity on the syntactical, lexical, morphological and phonological levels. The research results can also be useful for textbook writers and test developers, as well as researchers in text and discourse complexity.

## Acknowledgements

This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030).

We would like to thank Konstantin R. Sheriyazdanov, student at Kazan Federal University, for his assistance in compiling the corpus of academic texts while conducting the research. This paper has been supported by the RUDN University Scientific Projects Grant System, project No 050738–0–000.

## REFERENCES

- Andrews, Glenda & Graeme S Halford. 2002. A cognitive complexity metric applied to cognitive development. *Cognitive Psychology* 45 (2). 153–219. [https://doi.org/10.1016/S0010-0285\(02\)00002-6](https://doi.org/10.1016/S0010-0285(02)00002-6)



- Blake, J. Barry. 2001. *Case* (2nd ed., Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139164894>
- Bolbakov, Roman G. 2016. Slozhnost' informatsionnykh konstruksiy (Complexity of informative structures). *Obrazovatel'nye resursy i tehnologii* (In Russ.)
- Da'as, Rim'a, Chen Schechter & Mowafaq Qadach. 2020. School Leaders' Cognitive Complexity: Impact on the Big 5 Model and Teachers' Organizational Citizenship Behavior. *Journal of School Leadership* 30 (5). 398–423. <https://doi.org/10.1177/1052684619896535>
- Das, Syaamantak, Kumar S. Das Mandal & Anupam Basu. 2020. Cognitive complexity analysis of learning-related texts: a case study on school textbooks. In Vittorini, P., Di Mascio, T., Tarantino, L., Temperini, M., Gennari, R., De la Prieta, F. (eds.), *Methodologies and Intelligent Systems for Technology Enhanced Learning*, 10th International Conference. MIS4TEL 2020. *Advances in Intelligent Systems and Computing*, vol 1241. Springer, Cham. [https://doi.org/10.1007/978-3-030-52538-5\\_9](https://doi.org/10.1007/978-3-030-52538-5_9)
- Dehaene, Stanislas. 2007. *Les Neurones de la Lecture*. Paris: Éditions Odile Jacob.
- Dobrynina, Oxana L. 2019. Akademicheskoe pis'mo dlya publikatsionnykh tselei: stilisticheskie pogreshnosti. (Academic writing for publication purposes: Stylistic faults). *Vysshee obrazovanie v Rossii* 10. 38–49. (In Russ.)
- Field, John. 2004. Psycholinguistics: The Key Concepts. London, Routledge (in preparation). *Cognitive Validity*. Taylor L. ed. In *Examining Speaking: (Studies in Language Testing)*. Cambridge, Cambridge University Press / Cambridge ESOL.
- Fitzgerald, Jill, Jeff Elmore, Heather Koons, Elfrieda H. Hiebert, Kimberly Bowen, Eleanor Sanford-Moore & Jackson A. Stenner. 2015. Important text characteristics for early-grades text complexity. *Journal of Educational Psychology* 107 (1). 4–29. <https://doi.org/10.1037/a0037289>
- Fulcher, Glenn. 1997. Text difficulty and accessibility: Reading formulae and expert judgement. *System* 25 (4). 497–513. [https://doi.org/10.1016/S0346-251X\(97\)00048-1](https://doi.org/10.1016/S0346-251X(97)00048-1)
- Gatiyatullina, Gallya, Marina Solnyshkina, Valery Solovyev, Andrey Danilov, Ekaterina Martynova & Iskander E. Yarmakeev. 2020. Computing Russian Morphological distribution patterns using RusAC Online Server. *Proceedings – International Conference on Developments in eSystems Engineering*. 393–398.
- Gatiyatullina, Gallya, Marina I. Solnyshkina, Valery Solovyev, Andrey Danilov, Ekaterina Martynova & Iskander Yarmakeev. 2020. *Computing Russian Morphological distribution patterns using RusAC Online Server Proceedings – International Conference on Developments in eSystems Engineering, DeSE*. 2020-December. 393–398.
- Gatiyatullina, Galiya M., Marina I. Solnyshkina, Roman V. Kupriyanov & Chulpan R. Ziganshina. 2023. Lexical density as a complexity predictor: The case of Science and Social Studies textbooks. *Research Result. Theoretical and Applied Linguistics* 9 (1). 11–26. <https://doi.org/10.18413/2313-8912-2023-9-1-0-2>
- Gladkikh, Anatoliy A., Sergey M. Namestnikov & Nikita A. Pchelin. 2017. Effectivnoe perestanovochnoe dekodirovanie dvoichnykh blokovykh izbytochnykh kodov (*Efficient Permutation Decoding of Binary Block Redundant Codes*). *Avtomatizatsiya processov upravleniya* 1 (47). 67–74. (In Russ.)
- Hansen, Kathleen A. 1990. Information richness and newspaper pulitzer prizes. *Journalism Quarterly* 67 (4). 930–935. <https://doi.org/10.1177/107769909006700447>
- Johnson-Laird, Philip N. 1983. *Mental Models*. London et al.: Cambridge University Press.
- Just, Marcel A. & Patricia A. Carpenter 1987. *The Psychology of Reading and Language Comprehension*. MA, US: Allyn & Bacon.
- Kalinkina, Evgeniya M., Tatyana A. Poyarova & Aida V. Yablokova. 2021. Uchet dinamiki kognitivnoi slozhnosti u podrostkov pri postroenii obrazovatel'nogo protsessa (Taking

- into account the dynamics of cognitive complexity in teenagers in educational process design). *Problemy sovremennogo pedagogicheskogo obrazovaniya* 72 (4). 328–331. (In Russ.)
- Kholodnaya, Marina A. 2004. *Kognitivnye stili. O prirode individual'nogo uma* (Cognitive styles. On the nature of individual brainwork). Saint Petersburg. (In Russ.)
- Kintsch, Walter & Teun A. van Dijk. 1978. Toward a model of text comprehension and production. *Psychological Review* 85. 363–394.
- Kluyeva, Nadezhda V. 2003. *Osobennosti vospitaniya na raznykh vozrastnykh etapakh* (Peculiar features of upbringing at different age). Moscow. (In Russ.)
- Koda, Keiko 2005. *Insights into Second Language Reading*. Cambridge: Cambridge University Press.
- Kongorskaya, Svetlana A. 2014. Vozrastnye osobennosti razvitiya prostranstvennogo myshleniya podrostkov i starshikh shkol'nikov: ikh vzaimosvyaz' s uchebnoy uspevaemost'yu (Age features of developing spatial thinking of adolescents and high school students: their relationship with academic performance). *Vestnik Buryatskogo gosudarstvennogo universiteta* 5. 59–65. (In Russ.)
- Kotova, Ekaterina O. 2021. Lingvoekologicheskaya otsenka udobochitaemosti rossiiskikh longridov (statisticheskii podkhod) (Linguoecological assessment of the readability of Russian longreads (Statistical approach). *Izvestiya Yuzhnogo federal'nogo universiteta. Philologicheskie nauki* 25 (2). 67–76. (In Russ.)
- Krioni, Nikolai K., Alexey D. Nikitin & Anastasia V. Filippova 2008. Avtomatizirovannaya sistema analiza parametrov slozhnosti uchebnogo teksta (Automated system for analyzing the complexity parameters of instructional text). *Tekhnologiya i organizatsiya obucheniya* 155–161. (In Russ.)
- Kudzh, Stanislav A. & Viktor Ya. Tsvetkov 2018. Faktory kognitivnoi slozhnosti (Cognitive complexity factors). *Informatsionnye tekhnologii v nauke, obrazovanii i upravlenii* 6 (10). 34–41. (In Russ.)
- Kupriyanov, Roman V., Marina I. Solnyshkina, Mihai Dascalu & Tatyana Soldatkina. 2022. Lexical and syntactic features of academic Russian texts: A discriminant analysis. *Research Result. Theoretical and Applied Linguistics* 8 (4). 105–122. <https://doi.org/10.18413/2313-8912-2022-8-4-0-8>
- Lavazza, Luigi, Abedallah Abualkishik, Geng Liu & Sandro Morasca. 2022. An empirical evaluation of the “Cognitive Complexity” measure as a predictor of code understandability. *Journal of Systems and Software* 197. 111561. <https://doi.org/10.1016/j.jss.2022.111561>.
- Matlin, Margaret W. 1998. *Cognition, 4th edn.*: Harcourt Brace College Pub. NY.
- Matskovskiy, Mikhail S. 1976. Problemy chitabel'nosti pechatnogo materiala (Readability issues of printed materials). *Smyslovoe vospriyatie rechevogo obscheniya v usloviyakh massovoi kommunikatsii*. 126–142. (In Russ.)
- McComb, A. Sara & Jane M. Kirkpatrick. 2016. Impact of pedagogical approaches on cognitive complexity and motivation to learn: Comparing nursing and engineering undergraduate students. *Nursing Outlook*. 64 (1). 37–48. <https://doi.org/10.1016/j.outlook.2015.10.006>.
- McCarthy, Kathryn Soo, Danielle Siobhan Mcnamara, Marina I. Solnyshkina, Fanuza Kh. Tarasova & Roman V. Kupriyanov. 2019. The Russian language test: Towards assessing text comprehension. *Science Journal of Volgograd State University. Linguistics* 18 (4). 231–247. <https://doi.org/10.15688/jvolsu2.2019.4.18>
- Morin, Edgar. 1992. *Method: Towards a Study of Humankind*. Vol. 1. New York; Berlin, Bern; Frankfurt/M.; Paris; Wien; Lang.
- Morin, Edgar. 2005. *Metod. Priroda prirody* (Method. The nature of nature). Progress-Traditsiya. Moscow. (In Russ.)

- Morin, Edgar. 2021. *O slozhnosti* (On complexity). Institut obschegumanitarnykh issledovani. Moscow. (In Russ.)
- Nevdakh, Marina M. 2008. Issledovanie informatsionnykh kharakteristik uchebnogo teksta metodami mnogomernogo statisticheskogo analiza (The study of the instructional text information characteristics by the methods of multivariate statistical analysis). *Prikladnaya informatika* 4. 117–130. (In Russ.)
- Perry, William G., Jr. 1981. “Cognitive and ethical growth: The making of meaning”. In Arthur W. Chickering & Associates (eds.), *The Modern American College*. 76–116. San Francisco: Jossey-Bass.
- Pervin, Lawrence A. 1984. *Current Controversies and Issues in Personality*. 2nd ed-n. John Wiley & Sons.
- Petrenko, Victor F. 2010. *Osnovy psikhosemantiki* (Basic psychsemantics). Moscow: Eksmo. (In Russ.)
- Polya, George. 2015. *How to Solve It: A New Aspect of Mathematical Method* (Princeton Science Library, 34). Princeton Science Li Edition.
- Rodrigues, Virginia T., Luis A. Gonçalves, Soares P. Paolinelli Maciel & Augusto P. Rodrigues de Paiva 2014. Ethical education of an engineer with responsibility for a sustainable world. *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering*. 1–7. <https://doi.org/10.1109/ETHICS.2014.6893426>
- Shardlow, Matthew, Richard Evans & Marcoz Zampieri. 2022. Predicting lexical complexity in English texts: the Complex 2.0 dataset. *Lang Resources & Evaluation* 56. 1153–1194. <https://doi.org/10.1007/s10579-022-09588-2>
- Sharoff, Serge. 2022. What neural networks know about linguistic complexity. *Russian Journal of Linguistics* 26 (2). 371–390. <https://doi.org/10.22363/2687-0088-30178>
- Sheehan, Kathleen M., Irene Kostin, Yoko Futagi & Michael Flor. 2010. Generating automated text complexity classifications that are aligned with targeted text complexity standards. *ETS Research Report Series*. 2010 (2). 1–44. <http://dx.doi.org/10.1002/j.2333-8504.2010.tb02235.x>
- Silva, Susana, Filomena Inácio, Daniel Rocha e Sousa, Nuno Gaspar, Vasiliki Folia & Karl Magnus Petersson. 2023. Formal language hierarchy reflects different levels of cognitive complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 49 (4). 642–660. <https://doi.org/10.1037/xlm0001182>
- Sirotinina, Olga B. 2009. *Razgovornaya rech v sisteme funktsionalnyh stiley sovremennogo russkogo literaturnogo yazyka: grammatika* (Spoken language within the system of functional styles of the Russian literary language: grammar) 3rd ed-n, Moscow: Librekom. (In Russ.)
- Solnyshkina, Marina I., Elena V. Harkova & Mariia B. Kazachkova. 2020. The structure of cross-linguistic differences: Meaning and context of ‘readability’ and its Russian equivalent ‘chitabelnost’. *Journal of Language and Education* 6 (1). 103–119.
- Solnyshkina, Marina I., Ekaterina V. Martynova & Mariya I. Andreeva. 2020. Propositional’noe modelirovanie dlya otsenki informativnosti teksta (Propositional modeling for evaluating the informative value of a text). *Uchenye zapiski natsional’nogo obschestva prikladnoi lingvistiki* 3 (31). 47–57. (In Russ.)
- Solnyshkina, Marina I. & Alexander S. Kisel’nikov. 2015. Slozhnost’ teksta: Ehtapy izucheniya v otechestvennom prikladnom yazykoznanii (Text complexity: Stages of study in domestic applied linguistics). *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya* 6. 86–99. (In Russ.)
- Solnyshkina, Marina I., Danielle S. McNamara & Radif R. Zamaletdinov. 2022. Natural language processing and discourse complexity studies. *Russian Journal of Linguistics* 26 (2). 317–341.

- Solovyev, Valery, Musa Islamova, Marina Solnyshkina, Roman Kupriyanov & Elzara Gafiyatova. 2021. Sentiment analysis for Russian academic texts: A lexicon-based approach. *CEUR Workshop Proceedings*. 89–97.
- Solovyev V., M. Solnyshkina & D. McNamara. 2022. Computational linguistics and discourse complexology: Paradigms and research methods. *Russian Journal of Linguistics* 26 (2). 275–316. <https://doi.org/10.22363/2687-0088-30161>
- Tsetlin, Valentina S. 1980. Didakticheskie trebovaniya k kriteriyam slozhnosti uchebnogo materiala (Didactic requirements to the complexity criteria of educational material). *Novye issledovaniya v pedagogicheskikh naukakh* 1 (35). 30–33. (In Russ.)
- Ushakov, Konstantin M. 1980. O kriteriyakh slozhnosti uchebnogo materiala shkol'nykh predmetov (On the criteria of complexity of teaching material of school subjects). *Novye issledovaniya v pedagogicheskikh naukakh* 2 (36). 33–35. (In Russ.)
- Valgina, Nina S. 2003. *Teoriya teksta* (Theory of text). Moscow: Logos. (In Russ.)
- Valueva, Ekaterina A., Nina Danilevskaya, Ekaterina Lapteva & Dmitriy Ushakov 2017. Kognitivnaya slozhnost' khudozhestvennykh tekstov dlya detei: kvantitativnye metody otsenki (Cognitive complexity of literary texts for children: Quantitative methods of evaluation). *Voprosy psikholingvistiki* 42–61. (In Russ.)
- Vygotsky, Lev S. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Wang, Yingxu, Robert Berwick & Xiangfeng Luo X. 2012. A formal measurement of the cognitive complexity of texts in cognitive linguistics. *2012 IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing, Kyoto, Japan*. 94–102. <https://doi.org/10.1109/ICCI-CC.2012.6311132>
- Weir, Cyril J. 1993. *Understanding and Developing Language Tests*. London, Prentice Hall.
- Weir, Cyril J. & Hanan Khalifa. 2008. A cognitive processing approach towards defining reading comprehension. *Research Notes, Cambridge ESOL* 31. 2–10.
- Wijendra, Dinuka & K. Priyantha Hewagamage. 2021. Analysis of cognitive complexity with cyclomatic complexity metric of software. *International Journal of Computer Applications* 174. 14–19. <https://doi.org/10.5120/ijca2021921066>.
- Zhu, L., He Li, Wu He & Chuang Hong. 2020. What influences online reviews' perceived information quality? Perspectives on information richness, emotional polarity and product type. *The Electronic Library* 38 (2). 273–296. <https://doi.org/10.1108/EL-09-2019-0208>
- Zhuravlev, Anatoly F. 1988. Opyt kvantitativno-tipologicheskogo issledovaniya raznovidnostey ustnoy rechi (An experience of quantitative and typological investigation of spoken registers). *Raznovidnosti gorodskoy ustnoy rechi, Moscow, Nauka*. 84–150. (In Russ.)

### Dictionary

- Kondakov, Igor M. 2000. *Psikhologiya. Illyustrirovannyi slovar'* (Psychology. Illustrated dictionary). Saint Petersburg, Moscow. (In Russ.)
- Petrovskiy, Artur V. (ed.). 1998. *Kratkiy psikhologicheskii slovar'* (Brief dictionary of psychology). Rostov-on-Don: Phenix. (In Russ.)
- Voronin, Alexander S. 2006. *Slovar terminov po obschei i sotsial'noy pedagogike* (Glossary of terms on general and social pedagogy). Yekaterinburg. (In Russ.)

## Dataset

### Textbook

#### code

- S-5 [Sivoglazov, Vyacheslav I. & Pleshakov, Andrey A. 2019. Biology, 5<sup>th</sup> grade. Moscow: Prosveshcheniye. (In Russ.)]
- S-6 [Sivoglazov, Vyacheslav I. & Pleshakov, Andrey A. 2019. Biology, 6<sup>th</sup> grade. Moscow: Prosveshcheniye. (In Russ.)]
- S-10 [Sivoglazov, Vyacheslav I. et.al. 2020. Biology. General biology. 10<sup>th</sup> grade. Textbook. Base level. Moscow: Prosveshcheniye. (In Russ.)]
- S-11 [Sivoglazov, Vyacheslav I. et.al. 2020. Biology. General biology. 11<sup>th</sup> grade. Textbook. Base level. Moscow: Prosveshcheniye. (In Russ.)]
- V-5 [Lovyagin, Sergei N. et al. 2012. Biology. 5<sup>th</sup> grade. Moscow. (In Russ.)]
- V-6 [Lovyagin, Sergei N. et al. 2011. Biology. 6<sup>th</sup> grade. Moscow. (In Russ.)]
- V-10 [Vakhrushev, Alexander A. et al. 2013. Biology. General patterns. 10<sup>th</sup> grade. Moscow. (In Russ.)]
- V-11 [Vakhrushev, Alexander A. et al. 2013. Biology. General patterns. 11<sup>th</sup> grade. Moscow. (In Russ.)]

### Article history:

Received: 31 August 2023

Accepted: 18 September 2023

### Bionotes:

**Roman V. KUPRIYANOV** is Doctor of Psychology and Associate Professor of the Department of Social Work, Pedagogy and Psychology at Kazan National Research Technological University; Chief Researcher of the “Text Analytics” Research Lab at the Institute of Philology and Intercultural Communication, Kazan Federal University (Kazan, Russia). His areas of research are psycholinguistics, pedagogy of higher education, social psychology and social work. He is the author of more than 120 research articles.

*e-mail:* kroman1@mail.ru

<https://orcid.org/0000-0001-9794-9607>

**Olga V. BUKACH** is Doctor of Philology and Associate Professor of the Department of Theory and Practice of Foreign Language Teaching at the Institute of Philology and Intercultural Communication, Kazan Federal University (Kazan, Russia). She is in charge of organizing and carrying out language testing procedures at the Institute, including but not limited to test construction, as well as using statistical methods for handling and analyzing the obtained data.

*e-mail:* olga.bukach1987@gmail.com

<https://orcid.org/0009-0009-8638-5119>

**Oksana I. ALEKSANDROVA** is Doctor of Philology and Associate Professor of the Department of General and Russian Linguistics at RUDN University (Moscow, Russia). Her areas of research are semantics, cognitive linguistics and discourse analysis. She is the author of more than 60 research articles.

*e-mail:* alexandrova-oi@rudn.ru

<https://orcid.org/0000-0002-7246-4109>

**Сведения об авторах:**

**Роман Владимирович КУПРИЯНОВ** – кандидат психологических наук, доцент кафедры социальной работы, педагогики и психологии Казанского национального исследовательского технологического университета (Казань, Россия); старший научный сотрудник НИЛ «Текстовая аналитика» Института филологии и межкультурной коммуникации Казанского федерального университета. Сфера его научных исследований включает психолингвистику, педагогику высшей школы, социальную психологию и социальную работу. Он является автором более 120 научных публикаций.

*e-mail:* kroman1@mail.ru

<https://orcid.org/0000-0001-9794-9607>

**Ольга Владиславовна БУКАЧ** – кандидат филологических наук, доцент кафедры теории и практики преподавания иностранных языков Института филологии и межкультурной коммуникации Казанского федерального университета (Казань, Россия). Она является руководителем проекта по внутреннему языковому тестированию в ИФМК КФУ; осуществляет отбор материалов для разрабатываемых тестов, занимается использованием статистических методов для обработки и анализа данных, полученных в результате проведения языковых тестирований.

*e-mail:* olga.bukach1987@gmail.com

<https://orcid.org/0009-0009-8638-5119>

**Оксана Ивановна АЛЕКСАНДРОВА** – кандидат филол. наук, доцент кафедры общего и русского языкознания филологического факультета РУДН, заместитель декана по научной работе. Область исследования – семантика, когнитивная лингвистика, дискурс-анализ. Является автором более 60 научных публикаций.

*e-mail:* alexandrova-oi@rudn.ru

<https://orcid.org/0000-0002-7246-4109>