

Topic Modeling for Text Structure Assessment: The case of Russian Academic Texts

Valery Solovyev¹, Marina Solnyshkina¹, Elena Tutubalina²

¹ Kazan Federal University, Kazan, Russia

² Ivannikov Institute for System Programming of the RAS, Moscow, Russia

ABSTRACT

Background: Automatic assessment of text complexity levels is viewed as an important task, primarily in education. The existing methods of computing text complexity employ simple surface text properties neglecting complexity of text content and structure. The current paradigm of complexity studies can no longer keep up with the challenges of automatic evaluation of text structure.

Purpose: The aim of the paper is twofold: (1) it introduces a new notion, i.e. complexity of a text topical structure which we define as a quantifiable measure and combination of four parameters, i.e. number of topics, topic coherence, topic distribution, and topic weight. We hypothesize that these parameters are dependent variables of text complexity and aligned with the grade level; (2) the paper is also aimed at justifying applicability of the recently developed methods of topic modeling to measuring complexity of a text topical structure.

Method: To test this hypothesis, we use Russian Academic Corpus comprising school textbooks, texts of Russian as a foreign language and fiction texts recommended for reading in different grades, and employ it in three versions: (i) Full Texts Corpus, (ii) Corpus of Segments, (iii) Corpus of Paragraphs. The software tools we implement include LDA (Latent Dirichlet Allocation), OnlineLDA and Additive Regularization Of Topic Models with Word2vec-based metric and Normalized Pairwise Mutual Information.

Results: Our findings include the following: the optimal number of topics in educational texts varies around 20; topic coherence and topic distribution are identified to be functions of grade level complexity; text complexity is suggested to be estimated with structural organization parameters and viewed as a new algorithm complementing the classical approach of text complexity assessment based on linguistic features.

Conclusion: The results reported and discussed in the article strongly suggest that the theoretical framework and the analytic algorithms used in the study might be fruitfully applied in education and provide a basis for assessing complexity of academic texts.

KEYWORDS

text structure, topic modeling, school textbooks, text complexity, Russian language

Citation: Solovyev V., Solnyshkina M., & Tutubalina E. (2023). Topic Modeling for Text Structure Assessment: The case of Russian Academic Texts. *Journal of Language and Education*, 9(3), 143-158. <https://doi.org/10.17323/jle.2023.16604>

Correspondence:
Valery Solovyev,
maki.solovyev@mail.ru

Received: January 04, 2023

Accepted: September 15, 2023

Published: September 30, 2023

INTRODUCTION

Approaches to Determining Text Complexity

Numerous attempts have been made to explain what text complexity is and recently it has become a feature of importance (Si and Callan 2001; McNamara et al., 2014; Gatiyatullina et al., 2023). The reason for this is obvious: a text is supposed to correspond to the proficien-

cy of the target audience in all possible areas including education, publishing, legislation, science, medicine, etc. In the modern world where educators are committed to providing high quality personalized teaching and distance education, validated assessment of text complexity has become of particular importance for textbooks writers.

A broad definition of text complexity as “a measure of how easy or difficult a text is” (Bailin and Grafstein, 2016) though



universally accepted but of no practical relevance as it provides no algorithm of its assessment. Similarly, analysis and assessment of the topical structure complexity seems potentially useful for identifying structural patterns which could distinguish between texts of different grade levels. Text topical structure as “a way of indicating the relationship between the progression of sentence topics and the topical depth which indicates the semantic hierarchy» (Chuang 1993, p.2) is especially in demand in education. The reason for this is that features of topical structure reveal qualitative differences among texts.

As a research problem, text complexity has been studied for over a century and, as a notion, it has given rise to numerous definitions (for complexity of Russian texts see Ivanov et al., 2018; Osborneva, 2006; Solnyshkina et al., 2014; Solovyev et al., 2019). However, all its descriptive definitions, as the one above, are of little practical use when it comes to measure its level. Many definitions of text complexity tend to lack operability since they do not provide order of the procedures to be applied to determine how comprehensible a text is.

Similarly with the text-reader multi-criteria approach: although psycholinguistics seeks to develop sophisticated methods of matching texts with specific categories of readers (Crossley et al., 2014), conducting experiments of the kind is not only time- and effort-consuming, applicability of their results is always questionable due to differences in cognitive and linguistic abilities of the respondents (readers) involved. Therefore, automation of text complexity assessment based on computing text metrics, not involving readers, is viewed by the authors as a research niche.

However, feasibility of the task to a great extent depends on (1) availability of a representative corpus annotated by experts for certain categories of readers and (2) adequacy of the applied methods. The history of the latter started with the first complexity formulas proposed in (Flesch, 1948) which relied on two simple variables of a text: the average word and sentence length. Over the years the formulas, known as readability formulas, gained popularity and, due to their simplicity and ease of calculation, have been ubiquitously used and even installed with Microsoft Word. However, a natural question asked by a number of researchers is whether these formulas fully reflect all aspects of text complexity or only some of them. An exemplary experiment of contrasting two versions of the same text presented in (Thorndyke, 1977) illustrates limitations of Flesch-Kincaid readability formulas: complexity levels estimated with readability formulas of a text and its jumbled version are the same, though it is obvious that the jumbled text is much more difficult to read and comprehend.

Back in the 1970s – 1990s, a number of formulas were developed and validated. They had more than those in Flesch-Kincaid variables and were expected to provide a much higher accuracy. Unfortunately, they also proved limited and were criticized in (Crossley et al., 2008; McNamara et al., 1996) for their ineffectiveness to take into account other parameters such as text informational capacity and discourse structure. While discriminating complexity levels in Wikipedia texts, standard readability formulas do not offer accuracy achieved with machine learning methods¹ (Eremeev & Vorontsov, 2020; Martinc et al., 2021).

In the 2000s – 2010s, hundreds of features were introduced and validated as text complexity predictors. E.g., Coh-Metrix (cohmetrix.com) computes 108 parameters, ReaderBench (Toma et al., 2021) calculates over 200 parameters, and TAALES (Kyle et al., 2018) calculates over 400 parameters. The majority of the parameters are interdependent thus forming the so-called clusters. For example, the average word length measured in letters/syllables is co-dependent with the average number of long words, i.e. words of two or more syllables. In fact, all readability formulas estimate the same, i.e. formal text complexity, and duplicate measurements using slightly different parameters. In (McNamara et al., 2014), all the parameters they estimate are streamed into five groups: narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion. The streaming is based on substantially different aspects of text complexity. However, despite a significant increase in the number and groups of parameters measured, there is always doubt if the list is exhaustive. Apparently, text complexity is a multifaceted phenomenon that encompasses a range of different aspects which is difficult to exhaust.

The study dataset comprises textbooks used in Russian schools, the quality, sophistication and complexity of which have widely been discussed over decades (see Solnyshkina et al., 2020 for a review). Recent reductions in the Federal List of Textbooks recommended for use in mainstream schools² did not make the situation less challenging: although educators in Russia have fewer choices, they still need reliable tools either to assist in choosing or modifying available texts for the target audience. Moreover, the quality of school textbooks language is another important issue in Russian education. At the moment, the quality of textbook language is evaluated quite subjectively by experts, the existing expertise tests almost nothing but compliance with grammar, vocabulary and spelling norms. With the advent of online resources for the Russian language and modern natural language processing models, there is an opportunity to develop an algorithm to assist in addressing these challenges.

¹ Although, the authors clarify that the machine learning models are supposed to be trained on representative text samples and their complexity levels are to be graded by qualified experts (Eremeev & Vorontsov, 2020).

² [Publication.pravo.gov.ru/document/0001202307280015](https://publication.pravo.gov.ru/document/0001202307280015)

Complexity of Text Topical Structure

In this paper, we introduce a new notion, i.e. complexity of text topical structure, which, to the best of our knowledge, has not been previously recognized as a quantifiable measure estimated with a limited number of parameters. We refer to these parameters as ‘predictors of topical structure complexity’ which include the number of topics, topic coherence, topic distribution, and topic weight.

The types of topic development identified by researchers indicate that while conveying information, writers do not limit themselves to one topic but may divert to other topics, boundaries of which are sometimes hard to identify (Watson, 2016). Besides, as topics are developed differently in various text types and genres, e.g. instructional and expository texts may develop different types of topic progression following which may present additional difficulty for a reader (Ninio & Snow, 1999). Moreover, if topics are implicit or intertwined, the text is also hard to comprehend. The latter makes us conclude that structural organization of topics is directly related to text complexity.

Topic coherence is viewed in the article as semantic similarity of the words forming a topic (McNamara et al., 2014; McNamara et al., 1996; Balyan et al., 2018), whereas text coherence is referred to as “sense relations between single units (sentences or propositions) of a text” (www.glottopedia.org/index.php/Coherence) and manifests itself in repetitions or synonyms, as well as cohesion devices. As for comprehensive definition of ‘complexity of a text topical structure’, as a notion and an aspect of text complexity, it is, in our view, to exhibit and be based on precise mathematical formalism and specific features designating it. For the purpose indicated, we suggest implementing topic modeling apparatus developed by distinguished scholars in recent decades (Boyd-Graber et al., 2017; Mulunda et al., 2018; Rehurek & Sojka, 2010).

As it was mentioned earlier, similar numerical experiments require collections of texts annotated for complexity. For this purpose, researchers tend to use either collections of school textbooks, foreign language tests or collect corpora for specific research goals. As a rule of thumb in discourse complexology, suitability and ‘complexity’ of the dataset is to be evaluated by experts (Solovyev et al., 2022).

Present Study

Following the tradition developed in the area, for the current study we compiled four subcorpora: (1) a subcorpus of textbooks on Social Science, (2) a subcorpus of textbooks on Biology used in secondary and high schools of the Russian Federation, (3) a subcorpus of literary fiction read by students in secondary and high schools of the Russian Federation; (4) a subcorpus of texts used in tests for learners of Russian as a foreign language (A2-C1, CEFR).

In comparison with our previous conference papers (Sakhovskiy et al., 2020a, 2020b), we significantly extended experimental exploration of text complexity and focused on text complexity correlation with semantic and statistical properties of its topics. More specifically, in this research we have conducted new experiments, moving from one collection of educational texts on Social Science to four collections, including collections of school textbooks on Biology, fiction texts, and texts of Russian as a foreign language. We also extended evaluation of topic models and investigated effectiveness of previously suggested topic-based complexity features using texts of different domains. Finally, we investigated the correlation between text complexity and topic variety measured by the entropy of a topic distribution.

For the current analysis, we employ Latent Dirichlet Allocation (LDA) (Blei et al., 2003), and additive regularization of topic models (ARTM) (Vorontsov & Potapenko, 2015). The revealed topics are evaluated based on a number of several standard quality measures. Importantly, we show correlation of the revealed topics with the textbooks grade levels using Spearman’s rank correlation.

Our objectives in this paper include (1) introducing a new concept of topical structure complexity as a quantifiable measure assessed with a limited number of parameters, and (2) testing the hypothesis that topic modeling methods enable to estimate complexity of a text topical structure. More specifically, we study the following questions:

- (1) What is the reference range of topics in a collection of thematically related academic texts? We hypothesize that in order to achieve its goals and ensure readability, a text is supposed to have a certain optimum number of topics. A text typically contains more than one topic but topics excess hampers comprehension.
- (2) Is topic coherence related to text complexity? Does topic coherence increase or decrease across grades?
- (3) What topic parameters are co-dependent with the grade level?
- (4) What type of topic progression is characteristic of textbooks? How do types of topic progression change across grades? Can these types and changes be quantified?

LITERATURE REVIEW

It’s worth mentioning that the majority of research in the field of computer linguistics is often published in conference proceedings to establish precedence swiftly.

Text structure is viewed as an essential aspect of its complexity and has been explored as a part of reading comprehension and writing processes (McNamara et al., 2019;

Williams, 2005). In the majority of investigations, researchers focus on one or more specific types of text structures. (Kendeou & Broek, 2007; Diakidoy et al., 2003) present experimentally validated influence of text structure in refutation and non-refutation texts types. (Williams, 2005) reports on differences of respondents comprehending compare vs contrast text structures. In (Roehling et al., 2017), the authors identify text structure as one of the aspects of its complexity. A number of works explore ways of improving “text coherence and text information structure” for students with learning disabilities (cf. Arfe and Mason, 2018).

The impact of text structure on understanding and storing information has been a focus of a number of psychological and discourse studies. E.g. intensive studies on a sentence level have been conducted to test Chomsky’s theory. (McBride & Cutting, 2015) offer a thorough overview of classical works on the influence of text structure on understanding individual sentences or short texts.

One of the main theories accepted by the majority of researchers working in the area, is W. Kintsch’s theory of macro-propositions emphasizing a hierarchical organization of texts and revealing the structure and mechanisms of text comprehension (Kintsch, 1998). In (Kintsch & Vipond, 1979), W. Kintsch’s theory was applied to contrast complexity of speeches delivered by US presidential candidates, D.D. Eisenhower and A.E. Stevenson. The authors argue that D.D. Eisenhower’s speeches, if estimated with standard readability formulas, are more complex than those of A.E. Stevenson’s, but much simpler to comprehend which ultimately explains his victory in the election. The modern paradigm of text complexity studies also implements theory of macro-propositions to demonstrate that structural organization is manifested in features reflecting referential and deep cohesion of texts (McNamara et al., 2014; McNamara et al., 2010; Balyan et al., 2018). The general conclusion of these studies is that texts of higher coherence are easier to comprehend. However, these studies were conducted at sentence and paragraph levels, not the topical level of a corpus of texts, thus leaving a niche for our research.

Another aspect to emphasize is that highly cohesive texts are not necessarily coherent, i.e., cohesive ties per se do not constitute quality well-structured texts, but embody numerous deliberate repetitions, which may make texts boring and unattractive. Well-written coherent texts, on the opposite, are not repetitive, but can usually quite “uncontroversially be divided into successively smaller segments down to the level of the clause, yielding a hierarchical structure” (Hobbs, 1990, p. 111).

Higher thematic coherence manifested in semantic proximity of the most significant terms of topics is likely to contribute to better comprehension of a text. However, if the topics are too close, a reader faces an interference effect as retroactive and proactive inhibition (Loftus, 1983). The latter

is proved to hamper text comprehension. All the above establishes a new area to research in the area and highlights the importance of topic level of text complexity as a focus of new studies. Within topic modeling as one of the most frequently implemented approaches aimed at defining the content of text collections with the help of automated tools, researchers accumulated a number of methods (Boyd-Graber et al., 2017) with a potential to be applied in a wide range of spheres (Mulunda et al., 2018). In this study we offer unsupervised learning methods enabling automatic extraction of those text features that affect text topic complexity.

METHOD

Background

The modern research paradigm requires that parameters selected to measure complexity of a text topical structure are to be tested on a representative corpus able to provide reproducibility of results. As it was mentioned earlier, to achieve the stated objectives, i.e. to test the algorithm proposed and apply topic modeling to assess the nature of a text structure and its complexity, we use Corpus of school textbooks. We view school textbooks as a suitable type of texts for the research as they are sequenced from lower to higher grade levels based on their assigned complexity indices. K. Berendes and S. Vajjala test (2018) this assumption and refer to it as ‘systematic complexification’. They also contrasted textbooks from different German publishers and validated the systematic complexification assumption with varying degrees of consistency (Berendes & Vajjala, 2018). Collections of school textbooks have also been widely used to teach and test various models assessing text complexity in different languages (cf. Al-Tamim et al., 2014; Chen et al., 2013; Chen & Daowadung, 2015; Si & Callan, 2001; Santucci et al. 2020; Gazzola et al., 2022). Specifically, the tradition in the area is that a typical size of training collections of books is approximately a million tokens. Researchers use a corpora limited either by the number of subjects or grades. E.g. in Chen et al. (2013), these are textbooks on Mandarin, Social Studies, and Life Science for Grade 6 only. Si & Callan (2001) use a collection of Mathematics textbooks. Tanaka-Ishii et al. (2010) qualify a corpus of school textbooks as desirable and emphasize numerous challenges in compiling one. Another possible source of texts with assigned complexity levels is a corpus of foreign language tests described by Laposhina et al. (2018).

In this paper, we report on the algorithm to identify, extract and match topics exemplified in 4 sets of text collections, i.e. textbooks on Social Studies, Biology used in the 5th - 11th grades of secondary and high schools of the Russian Federation, fiction texts selected to read by schoolchildren in secondary and high schools, and Russian texts used to teach Russian as a foreign language. It is noteworthy that

the research is conducted on a battery of textbooks written by one author or a collective of authors for all grades. This fact eliminates any influence of authors' style or pedagogical concepts implemented in textbooks of different grades and enables to focus on text complexity only. Contrasting textbooks of the allegedly same complexity level written by different authors provides the possibility to identify the impact of authors' style.

In our experiments, we fit each topic model on the whole text collection. We have D documents and each document is described by frequencies of W words from the vocabulary elicited from the texts. The documents we have are of three types: full textbooks, segments, or paragraphs. The whole collection is viewed as an W by D matrix and the goal of topic modeling is decomposition of a large matrix into smaller matrices: a document-topic matrix and the word-topic matrix. Thus, we do not operate on the level of separate documents but instead fit a topic model on the full collection in one step. As for our experiments on correlation analysis, we first elicit the topic distribution of full textbooks using the observed word frequencies and then conduct our experiments on these distributions. For details refer to (Sakhovskiy et al., 2020b) where the authors provide mathematical foundations of topic modeling calculations.

Topic Models Quality: Metrics of Assessment

To assess the quality of a topic model, we utilized the following metrics: word2vec-based metric (Nikolenko, 2016) and normalized pairwise mutual information (NPMI) (Bouma, 2009). For our experiments, we employed 300-dimensional Rusvectors (Kutuzov & Kuzmenko, 2016) skip-gram models trained on the Russian National Corpus (RNC) (ruscorpora.ru/) and the Taiga corpus (Shavrina & Shapovalova, 2017).

We use standard approaches based on the distributive hypothesis which implies that semantics of a word is identified based on its contexts. Contexts are set by the adjacent words frequency vectors, and the metrics are calculated as distances between vectors. The normalized pairwise mutual information metric between two words indicates how likely the words are to occur together in a corpus.

A list of topic words is viewed as coherent if these words frequently occur in the same documents. To quantify the degree of topic coherence, we utilized the NPMI. In this work, we calculated the frequencies using RNC. The larger the NPMI measure, the more often the words of the topic occur together in the texts, i.e. the topic is more coherent.

The NPMI and word2vec-based metric are known to correlate well with human estimates of topic interpretability (Nikolenko, 2016; Newman et al., 2010a; Newman et al., 2010b), but they characterize topics from slightly different points of view. In Topic modeling studies, interpretability

refers to subjective evaluation of experts to what extent a topic corresponds to any specific topic.

The word2vec-based Q-metric characterizes topics by how semantically close their words are, regardless of their relative location in the texts. The smaller the Q-metric, the closer is the semantics of the words, i.e. the topic is more coherent. NPMI is a more complex measure. It takes into account 'words joint occurrence' and reflects both the semantic proximity of words and their syntactic properties. This measure is a function of two factors. On the one hand, the closer semantically are the words, i.e. the more coherent is the topic, and the higher is the NPMI. On the other hand, high NPMI values can also be due to the fact that topic keywords form stable combinations, thus reflecting not so much text coherence but its stereotyped style and occasionally the writer's desire to facilitate text perception. If the collection of texts is stylistically homogeneous, and designed for similar categories of readers, then the second factor is leveled. However, in our case, i.e. with the collection of textbooks for different grades, the second factor is not leveled. Findings in (Nikolenko, 2016) indicate that the word2vec-based Q-metric reflects topic coherence better than NPMI.

Besides we introduce a new parameter, i.e. topic weight, which is defined as the average frequency of topic words in an auxiliary corpus.

To increase the stability and robustness of topic evaluation (cf. Lau and Baldwin, 2016), we computed all metrics using 5, 10, 15, 20 top frequency words of each topic and took the mean over the four values as a topic estimate. The model quality is the average topic quality over all its topics.

Data and Preprocessing

When constructing topic models, we used the following corpora: (i) Corpus of Full Texts, (ii) Corpus of Segments, (iii) Corpus of Paragraphs. Corpora (ii) and (iii) were obtained by splitting full book corpus documents into smaller documents. The corpus of segments was compiled by the algorithm designed to get a maximum possible number texts so that the trained model was of excellent quality. The algorithm included numerous interdependent steps: sentences were sequentially added to the segment until the end of the book was reached, or until situations when adding a sentence resulted in an excessive size of the segment. And in that case, the sentence was added to a new 'empty' segment. The maximum segment size was set at 1000 tokens. We also removed punctuation, rare words (i.e. words registered in fewer than 3 documents), stop-words and auxiliary parts of speech. The Stop-word list was adopted from github.com/stopwords-iso/stopwords-ru. Ultimately the segments' length appeared to be shorter than 1000 words. At the final stage of preprocessing we also used UDPipe library for lemmatization and POS tagging. Table 1 below shows the corpora statistics received after preprocessing the dataset.

Table 1

Number of documents and document lengths after preprocessing the collection

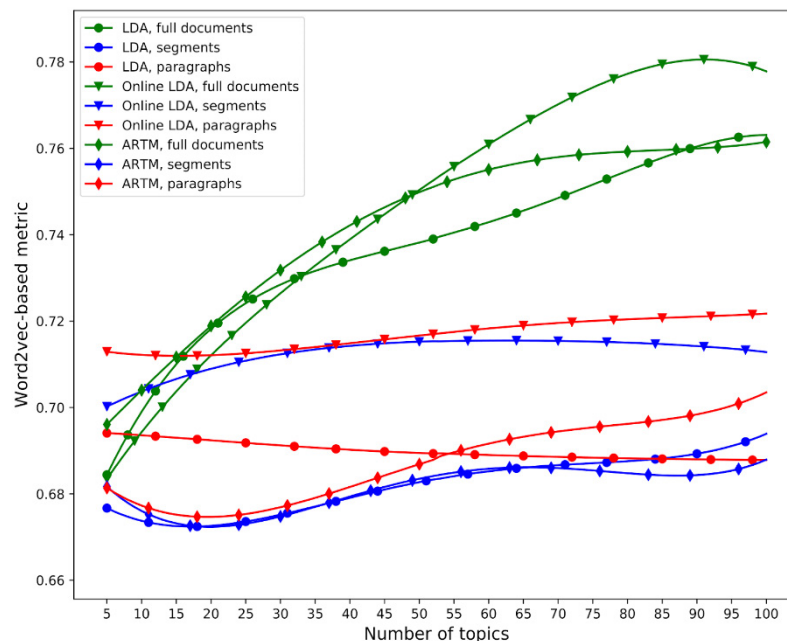
Collections	Number of full texts	Average document length (tokens)		
		full texts	segments	paragraphs
Social Studies	16	29214	576	26
Biology	25	22523	607	25
Russian as a Foreign Language	199	133	127	20
Fiction	111	19809	390	28

The Table above shows that the average length of full texts in “Russian as a Foreign Language” subcorpus is small and does not differ much from the average length of a segment. Thus, splitting this collection into segments did not significantly increase the number of documents, and was not used in further work.

The Topic models we used included the LDA model from Mallet library (McCallum, 2002), OnlineLDA (Hoffman et al. 2010). and ARTM model from BigARTM library (Vorontsov, 2015). While constructing LDA models, we performed 1000 iterations in the training corpus. To tune hyperparameters of ARTM model, we introduced regularizers of selection and decorrelation of topics, regularizers of thinning distributions ($d|t$) and $p(w|t)$, the effect of which was gradually increased. The models were trained with 280 training iterations for the corpus of full books and 175 iterations for smaller documents. 25 training iterations were performed in each document. We constructed models with the range of 5 to 100 topics and 1 step.

Figure 1

Quality Assessments of Topic Models Trained on Social Studies Texts: Word2vec Metric with RNC³



³ In Fig. 1, metrics and quality of the model are inversely dependent: the smaller is the metric, the higher is the quality of the model.

The received estimates of the topic models were used to define the number of topics and the type of the document split for further experiments. The resulting graphs demonstrate the main findings: the best quality values are observed for the LDA and ARTM segment models, and the best quality value of these models is achieved with a number of topics close to 20. The diagram also shows calculation results for the OnlineLDA model, but as they turned out to be the worst of the three models considered, the OnlineLDA was excluded from further analysis.

For further experiments of correlation tests of texts on Social Studies, we used the segment model. For all other collections we implemented paragraph-trained models. The number of topics assigned are 40 for fiction texts and 20 for all other collections of texts.

Correlation Analysis of Topic Properties and Text Complexity

The topics obtained as a result of topic modeling were later used to define the topic properties able to identify the type of relations between a text and a certain level of complexity. To this end we implemented distributions of topics in Corpus of Full Texts in the collections of Social Studies, Biology, Fiction, and Russian as a Foreign Language. Each document in this collection had an assigned complexity either by Common European Framework of Reference, i.e. A1-C2 (www.coe.int/en/web/common-european-framework-reference-languages), or a grade level. The relationship between complexity levels and topic distribution metrics were assessed with Spearman’s correlation tests.

With the topic defined as 20 most probable words in a text, we conducted a number of experiments aimed at assessing the relationship between text complexity and the following text parameters: the average NPMI of topics and average distance between word vectors in the topic in the word-2vec space. The research was performed on corpus Taiga and two subcorpora of RNC, i.e. Subcorpus “Sovremennyye pis’mennyye teksty” of the Main Corpus and the Spoken Subcorpus. We applied two approaches to assess texts. The first implies (1) selecting a subset of topics which exceeded

a certain established threshold of a topic probability in the text and (2) identifying a new text parameter as a certain subset of topics in a text. For our experiments, we use a threshold equal to $1/|T|$, where $|T|$ is the number of topics in the topic model. For example, the probability threshold for a topic in a book is 0.05 for a topic model with 20 topics and 0.025 for a topic model with 40 topics. The disadvantage of this approach is information loss of the topics below the established threshold.

Another approach requires assessment of each topic contribution to the book topics and is conducted based on the probability values of each topic in a certain book. In our work, this idea is realized in the cosine distance between the vector of topics probabilities in the document and the vector of assessments of individual topics. We also compared both approaches presented above to assess a document and implemented the average topic score as a parameter for a subset of topics selected based on the probability threshold.

Table 2 shows the values of Spearman’s correlation coefficients for the complexity levels of texts and the topics weight. The results obtained allow us to make the following main observations. Firstly, the cosine distance between the vectors of estimates and topics probabilities in the text enables higher absolute values of correlation coefficients. The results obtained are of little statistical significance if the probability threshold is employed as a method of selecting topics. Second, for all the collections of books, except for the collection of Russian as a foreign language, there is a negative correlation between the cosine distance and text complexity. Third, the phenomena are observed for both models. Thus, with text complexity increasing, the vector of topics probabilities in texts acquires similarities with the vector of topics weights.

Spearman’s correlation indices for text complexity and NPMI are provided in Table 3 below.

As in the previous series of experiments on interdependency between the topic weight and text complexity level, we observed negative correlations while using the cosine distance as a way to calculate a text metric. However, for

Table 2

Correlations between the Average Topic Weight (RNC Frequencies) and Text Complexity Level

Collection of text	ρ, q , where ρ is Spearman’s correlation coefficient, $\rho < 10^{-q}$			
	probability threshold		cosine distance	
	ARTM	LDA	ARTM	LDA
Social Studies	0.55, 1	0.51, 1	-0.84, 3	-0.72, 2
Biology	-0.13, 0	-0.18, 0	-0.54, 2	-0.60, 2
Fiction	-0.07, 0	0.14, 0	-0.36, 3	-0.34, 3
Russian as a foreign language	0.11, 0	0.18, 2	-0.00, 0	-0.20, 2

Note. In Table 2 and below, q index is preceded by a comma, and statistically significant if ≥ 2 .

NPMI metric, the absolute values of the correlation coefficient are smaller, and the revealed dependencies have a lower level of statistical significance.

We also identified few values inconsistencies when the negative correlation alternates with positive. A probable reason for this has to do with the influence of two opposing factors on NPMI metric noted above.

Table 4 presents results of correlation tests for text complexity level and the mean cosine distance between word2vec of the topic words. The results obtained are similar to the results of correlation tests for topic weight and text complexity level: we observe a negative correlation for both types of topic models when using the cosine distance for all text collections, except for the collection of texts of Russian as a foreign language.

Dependencies between Text Complexity and Topic Structure

The next series of correlation tests was aimed at testing the hypothesis of interdependency between text complexity and topic distribution structure in texts. We define topic distribution entropy as Shannon entropy of document-topic distribution $p(t|d)$ of a pre-trained topic model. Since Shannon entropy reaches its maximum on uniform probability distribution, its smaller values would indicate that a text is focused on a small number of topics. Vice versa, higher en-

trophy indicates a text focused on a wide range of topics with generally sparse and shallow coverage.

The results of the correlation tests between text complexity and topic distribution entropy are provided in Table 5 below.

Spearman's correlation tests indicate statistically significant positive correlation between topic distribution entropy of texts and their complexity level in Biology and Fiction collections. We also observe maximum of topic distribution entropy in cases when probabilities of all topics in texts are equal. That is true about polythematic texts with equally distributed coverage of topics. On the contrary, topic distribution entropy is minimal, if probability of one of the topics is close to 1, hence the text is monothematic.

Based on the results obtained, we conclude that more complex Biology and Fiction texts cover more topics of comparable sizes. As texts of Russian as a foreign language are much shorter and their complexity does not grow across texts, their results differ dramatically: a short text cannot cover numerous topics and embodies monothematicity.

In texts of Social Studies collection, as in Biology and Fiction texts, we observe a positive correlation for both types of models. However, statistical significance of the results is lower, which is either caused by a smaller size of the sample or by the fact that text complexity of this collection is largely

Table 3

Correlations between NPMI Average Number of Topics and Text Complexity Level

Collection of texts	ρ, q , where ρ is Spearman's correlation coefficient, $\rho < 10^{-q}$			
	probability threshold		cosine distance	
	ARTM	LDA	ARTM	LDA
Social Studies	0.41, 0	0.63, 1	-0.68, 2	-0.61, 1
Biology	-0.01, 0	0.33, 0	-0.32, 0	-0.46, 1
Fiction	-0.17, 1	-0.16, 1	-0.25, 2	-0.16, 1
Russian as a foreign language	0.13, 1	-0.07, 0	-0.13, 1	-0.09, 0

Table 4

Correlations between Mean Topic Word2vec (RNC Model) and Text Complexity

Collection of texts	ρ, q , where ρ is Spearman's correlation coefficient, $\rho < 10^{-q}$			
	probability threshold		cosine distance	
	ARTM	LDA	ARTM	LDA
Social Studies	-0.69, 2	-0.78, 3	-0.62, 1	-0.43, 0
Biology	-0.22, 0	-0.13, 0	-0.58, 2	-0.59, 2
Fiction	0.16, 1	0.04, 0	-0.39, 4	-0.28, 2
Russian as a foreign language	0.12, 1	-0.01, 0	0.09, 0	-0.17, 1

Table 5*Correlations between Text Complexity Level and Topic Distribution Entropy*

Collection of texts	ρ, q , where ρ is Spearman's correlation coefficient, $p < 10^{-q}$	
	ARTM	LDA
Social Studies	0.59, 1	0.38, 0
Biology	0.76, 4	0.59, 2
Fiction	0.42, 5	0.38, 4
Russian as a foreign language	-0.10, 0	0.21, 2

dependent on different parameters. The findings may also be caused by frequency and semantic properties of the most probable topics which correspond to the observations obtained in the previous experiments (see Tables 2-5 above). Thus, the foregoing suggests that, as a function of the subject area, properties of different types of topics affect text complexity diversely. Consequently, a more accurate definition of complexity requires taking into account numerous properties of various texts including semantics, frequency, text topic structure.

Topic structures of individual texts were scrutinized based on the topic distributions of ARTM models trained on Social Science texts and a segment model with 20 topics. The results of the experiment are presented in Fig. 2.

Graphs in Fig. 2 show that textbooks of higher grades demonstrate lower probability of the most probable topic. On the opposite, the probability of the most probable topics in texts of the 5th and 6th grades is exponentially higher than probability of any other topic. Therefore, text complexity growth results in texts becoming polythematic, while lower complexity texts predominantly maintain their monothematicity. This observation is consistent with the findings of the study on interdependency of entropy and complexity: text complexity growth leads to the increase of topic distribution entropy.

Topics Qualitative Analysis

Analysis of words representing topics is a mandatory step in the Topic Modeling algorithm. To this end, we selected several most interpreted topics of ARTM segment model with 20 topics, trained them on a collection of Social Studies texts. Topics interpretability was assessed with word2vec scores. Table S1 in Appendix shows examples of the selected topics and their corresponding word2vec scores. In addition, for each topic selected we identify 5 texts, in which the share of the topic is maximum.

Analysis of the above topics reveals, that most of the topics, represented by the words semantically close in the word2vec, have the highest weight in the texts of the 9th - 11th

grades. This observation is consistent with the results of the correlation tests presented in Table 4. However, there are cases inconsistent with this observation, for example, topics 1 and 4, which have a high weight in texts of the 6th and 7th grades, respectively. This observation indicates that text complexity estimate implies assessment of more than one topic metric, i.e. topic interpretability word2vec.

DISCUSSION

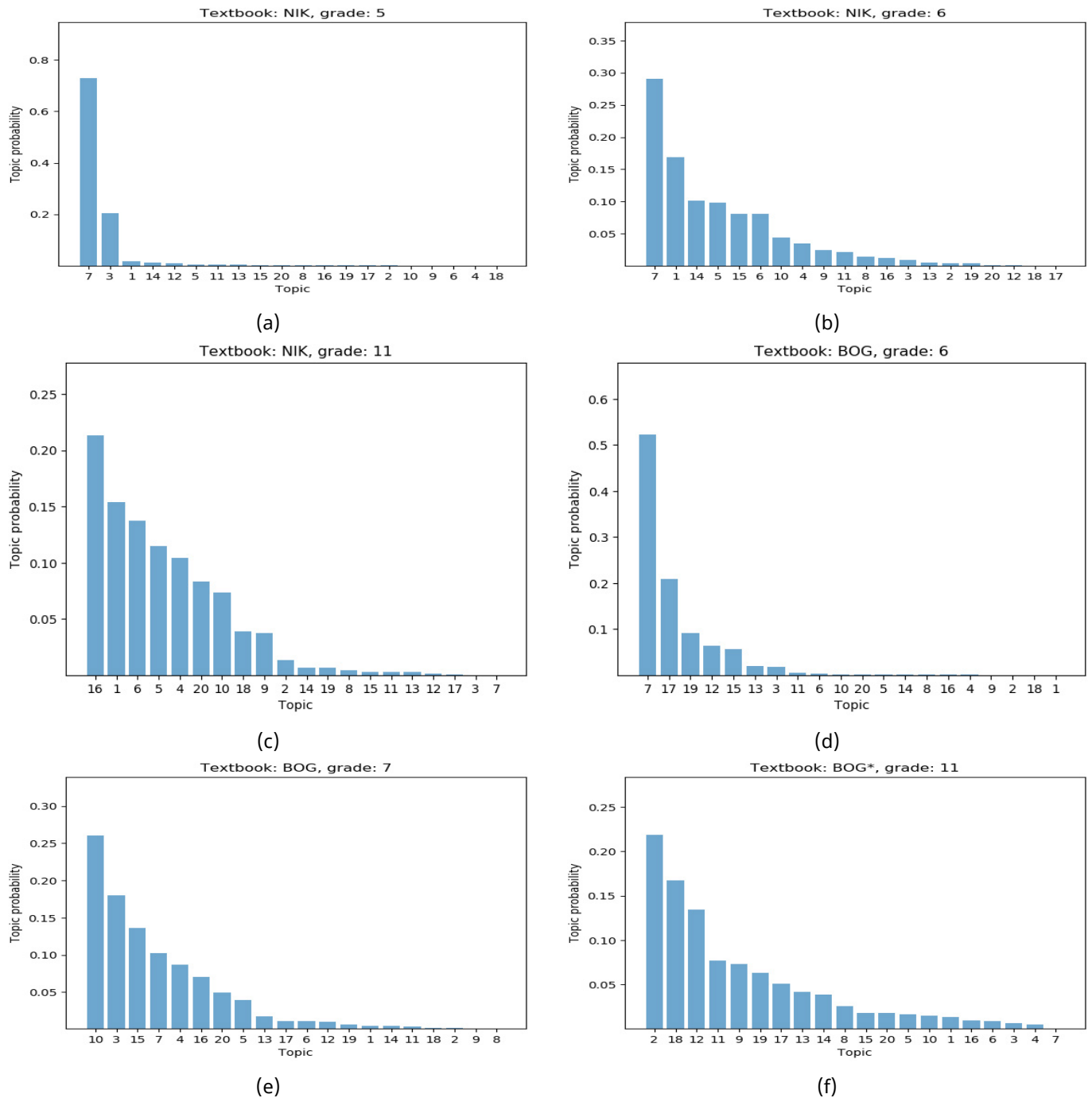
In this research, we investigate applicability of the two state-of-the-art topic models, i.e. Latent Dirichlet Allocation (LDA), and Additive regularization of topic models (ARTM), for assessment of text complexity in schoolbooks. We adopt three training strategies for representing books to train topic models: (i) full-length textbooks, (ii) segments with a maximum size of 1000 words, (iii) paragraphs. When assessing topic coherence, we used two metrics (word2vec, NPMI) and two corpora (RNC, Taiga).

We also validated the topical pattern of Russian school textbooks: typically, there are 15-20 topics in a collection of academic texts. This result does not depend either on the chosen method for assessing the quality of topics (word2vec or NPMI) or the text corpus (RNC or Taiga). Based on Social Studies textbooks, we presume that the selected 20 topics are well interpreted by experts.

ARTM demonstrated better evaluation results compared to LDA in terms of topic coherence metrics. Based on the trained topic models, we revealed correlation between book grades and properties of the highest weights topics.

Spearman rank correlation results demonstrate the following statistically significant dependencies: (i) higher-grade texts, i. e. more complex texts, are characterized by higher topics coherence; (ii) higher-grade texts contain more topics and the topics are equally distributed in the texts (distribution entropy of topics correlates with text complexity); (iii) higher grades texts contain fewer topic words of average frequency. Specifically, (i) implies that narrower, more specialized topics are taught in senior grades, i.e. there is a

Figure 2
Distribution of Topics in Social Science Texts



Note. (a) - (c) refer to textbooks by Nikitin A.F., (d) - (f) refer to textbooks by Bogolyubov L.N.
* marks advanced complexity levels of books

transition from general to more specific in introduction of educational material. This finding is also supported by (iii), which confirms that rarer, i.e. more specific terms are used in senior grades.

As for (ii), it implies that author’s attention in lower grades texts is focused on a limited, much lower number of topics than in the senior grades (Figure 2). Thus, throughout the school, textbooks acquire more topics, widening world view

from narrow to broad in senior grades. Although this statement may seem self-evident, the methods we have developed make it possible to obtain a quantitative assessment of this process for each subject and each series of textbooks of different grades.

The correlation coefficient of grade level received in the study is moderate, which indicates that we observe a lack of systematic complexification of textbooks across grades.

The latter is similar to the results obtained in (Berendes & Vajjala, 2018) for German school textbooks. We have also established correlation of two additional parameters, i.e. topic weight and distribution entropy of topics with text complexity. In addition to this we confirmed the previously obtained results (Sakhovskiy et al., 2020a, Sakhovskiy et al., 2020b) and replicated them on a larger representative balanced collection of texts.

CONCLUSION

The article offers a consistent description of experimental application of topic modeling algorithms to evaluating a text structure. The dataset used in the study was compiled of graded texts of four academic subcorpora, i.e. textbooks of Social studies and Biology for Russian schools, tests materials for Russian as a foreign language and texts for extra-curricular reading. We revealed and described patterns of structural change in educational texts as they become more complex from grade to grade. The study confirmed the hypothesis that topic properties change systematically across grade levels. We also offer a list of parameters discriminating various educational texts structures, and present the latter as a set of topics designed of keywords. We validated the list of introduced parameters, i.e. number of topics, topic coherence, topic distribution, and topic weight as predictors of (1) a topics change from specific to general; and (2) an increasing text complexity.

We conclude that topic models can be used to assess text structure dynamics. Due to the ease of computing values of these parameters with available software programs, they can be used along with traditional text complexity assessment tools. We also emphasize that the studied parameters characterize only one aspect of text complexity, i.e. structural organization of text topics.

In the algorithm suggested, complexity is assessed not by commonly used linguistic parameters (length of sentences,

number of long words, TTR, etc.), but by computational parameters related to textbook topics. The proposed approach offers new insights into the problems of text complexity and methods of presenting educational material in a textbook. Automatically obtained metrics of the introduced parameters, i.e. number of topics, topic coherence, distribution, and weight, enable to evaluate sustainability of strategy and presentation techniques across a text/textbook. The algorithm designed and developed for Russian texts can be further extrapolated to other languages and texts, provided the language is well-resourced and a representative corpus is available to compute word2vec and NPMI.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Higher Education of the Russian Federation, agreement No. 075-15-2022-294 dated 15 April 2022 and by the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030).

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS CONTRIBUTIONS

Valery Solovyev: Conceptualization, Methodology, Supervision, Project administration.

Marina Solnyshkina: Resources, Writing - Original Draft.

Elena Tutubalina: Resources, Writing - Original Draft.

REFERENCES

- Al Tamimi, A. K., Jaradat, M., Al-Jarrah, N. & Ghanem, S. (2014). AARI: Automatic Arabic readability index. *International Arab Journal of Information Technology*, 11(4), 370-378.
- Arfe, B., Mason, L. & Fajardo, I. (2018). Simplifying informational text structure for struggling readers. *Reading and Writing*, 31, 2191-2210. <https://doi.org/10.1007/s11145-017-9785-6>
- Bailin, A., & Grafstein, A. (2016). *Readability: Text and context*. Palgrave Macmillan.
- Balyan, R., McCarthy, K.S., & McNamara, D.S. (2018). Comparing machine learning classification approaches for predicting expository text difficulty. In *The Thirty-First International Flairs Conference (FLAIRS-31)* (pp. 421-426). AAAI press.
- Berendes, K., & Vajjala, S. (2018). Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track? *Journal of Educational Psychology*, 110(4), 518-543. <https://doi.org/10.1037/edu0000225>
- Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.

- Borda, M. (2011). *Fundamentals in information theory and coding*. Springer.
- Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCCL* (pp. 31-40). Gunter Narr Verlag Tubingen.
- Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of topic models. *Foundations and Trends® in Information Retrieval*, 11(2-3), 143-296. <http://dx.doi.org/10.1561/15000000030>
- Chen, Y.-H., & Daowadung, P. (2015). Assessing readability of Thai text using support vector machines. *Maejo International Journal of Science and Technology*, 9(3), 355-369.
- Chen, Y.-T., Chen, Y.-H., & Cheng, Y.-C. (2013). Assessing Chinese readability using term frequency and lexical chain. *International Journal of Computational Linguistics and Chinese Language Processing*, 18(2), 1-18.
- Chuang, Hsiao-yu (1993). Topical structure and writing quality: A study of students' expository writing. *Theses Digitization Project*, 686. California State University.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475-493.
- Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language*, 26(1), 92-113.
- Diakidoy, I.-A. N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology*, 28(3), 335-356. [https://doi.org/10.1016/S0361-476X\(02\)00039-5](https://doi.org/10.1016/S0361-476X(02)00039-5)
- Eremeev, M. A., & Vorontsov, K. V. (2020). Quantile-based approach to estimating cognitive text complexity. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"*, 19, 256-269. РГГУ.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Gatiyatullina, G.M., Solnyshkina, M.I., Kupriyanov, R.V., & Ziganshina, C.R. (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. [10.18413/2313-8912-2023-9-1-0-2](https://doi.org/10.18413/2313-8912-2023-9-1-0-2)
- Gazzola, M., Leal, S., Pedroni, B., Theoto Rocha, F., Pompéia, S., & Aluísio, S. (2022). Text complexity of open educational resources in Portuguese: mixing written and spoken registers in a multi-task approach. *Language Resources and Evaluation*, 56(2), 621-650.
- Hobbs, J. (1990). *Literature and cognition*. Stanford.
- Hoffman, M.D., Blei, D., & Bach, F. (2010). Online inference for latent Dirichlet allocation. In *Neural Information Processing Systems* (pp. 856-864). Curran Associates, Inc.
- Ivanov, V.V, Solnyshkina, M.I., & Solovyev, V.D. (2018). Efficiency of text readability features in Russian academic texts. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2018"* (pp. 267-283). RGGU.
- Kendeou P., & Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, 35(7), 1567-1577. <https://doi.org/10.3758/BF03193491>
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. Nilsson (Ed.), *Perspectives on memory research* (pp. 329-365). Psychology Press.
- Kutuzov, A., & Kuzmenko, E. (2017). Web vectors: A toolkit for building web interfaces for vector semantic models. In *International Conference on Analysis of Images, Social Networks and Texts* (pp. 155-161). Springer Cham.
- Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*, 50(3), 1030-1046.
- Laposhina, A. N., Veselovskaya, T. V., Lebedeva, M. U., & Kupreshchenko, O. F. (2018). Automated text readability assessment for Russian second language learners. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2018"* (pp. 396-406). RGGU.
- Lau, J. H. & Baldwin, T. (2016). The sensitivity of topic coherence evaluation to topic cardinality. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technology* (pp. 483-487). Association for Computational Linguistics.
- Loftus, G.R. (1983) The continuing persistence of the icon. *Behavioral and Brain Science*, 6(1), 28. <https://doi.org/10.1017/S0140525X00014461>

- Martinc, M., Pollak, S., & Robnik-Šikonja, M. (2021). Supervised and unsupervised neural approaches to text readability. *Computational Linguistics*, 47(1), 141-179.
- McBride, D. M., & Cutting, J. C. (2015). *Cognitive psychology: Theory, process, and methodology*. Sage.
- McCallum, A.K. (2002). *Mallet: A machine learning for language toolkit*. University of Massachusetts Amherst.
- McNamara, D.S., Graesser, A. C., McCarthy, P. M., & Cai, Zh. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1-43. https://doi.org/10.1207/s1532690xci1401_1
- McNamara, D. S., Louwrese, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47(4), 292-330. <https://doi.org/10.1080/01638530902959943>.
- McNamara, D. S., Roscoe, R., Allen, L., & Balyan, R., & McCarthy, K.S. (2019). Literacy: From the perspective of text and discourse theory. *Journal of Language and Education*, 5(3), 56-69. <https://doi.org/https://doi.org/10.17323/jle.2019.10196>.
- Mulunda, C.K., Wagacha, P.W., & Muchemi, L. (2018). Review of trends in topic modeling techniques, tools, inference algorithms and applications. In *2018 5th International Conference on Soft Computing & Machine Intelligence (ISCFMI)* (pp. 28-37). IEEE.
- Newman, D., Jey, H. L., Karl, G., & Timothy, B. (2010a). *Automatic evaluation of topic coherence. Human language technologies*. (pp. 100-108). Association for Computational Linguistics.
- Newman, D., Noh, Y., Talley, E., Karimi, S., & Baldwin, T. (2010b). Evaluating topic models for digital libraries. *Proceedings of the 10th annual joint conference on Digital libraries* (pp. 215-224). Association for Computing Machinery.
- Nikolenko, S.I. (2016). Topic quality metrics based on distributed word representations. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 1029-1032). Association for Computing Machinery.
- Ninio, A., & Snow, C. (1999). The development of pragmatics: Learning to use language appropriately. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of language acquisition* (pp. 347-383). Academic Press.
- Oborneva, I. V. (2006). *Automated assessment of the complexity of educational texts based on statistical parameters* [Unpublished doctoral dissertation,]. Institute of Contents and Methods of Training RAO.
- Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45-50). University of Malta.
- Roehling, J. V., Hebert, M., Ron Nelson, J., & Bohaty, J. J. (2017). Text structure strategies for improving expository reading comprehension. *The Reading Teacher*, 71(1), 71- 82. <https://doi.org/10.1002/trtr.1590>
- Sakhovskiy A., Solovyev V., & Solnyshkina M. (2020a). Topic modeling for assessment of text complexity in Russian textbooks. *Proceedings of Ivannikov ISPRAS Open Conference* (pp. 102-108). IEEE.
- Sakhovskiy, A., Tutubalina, E., Solovyev, V., & Solnyshkina, M. (2020b). *Topic modeling as a method of educational text structuring. DeSE* (pp. 399-405). IEEE.
- Santucci, V., Santarelli, F., Forti, L., & Spina, S. (2020). Automatic classification of text complexity. *Applied Sciences*, 10(20), 7285.
- Shavrina, T., & Shapovalova, O. (2017). To the methodology of corpus construction for machine learning: "Taiga" syntax tree corpus and parser. In *Proceedings of the International Conference "Corpus Linguistics-2017"* (pp. 78-84). St Petersburg State University.
- Si, I., & Callan, J. (2001). A statistical model for scientific readability. In *Proceedings of the tenth international conference on Information and knowledge management* (pp. 574-576). Association for Computing Machinery.
- Solnyshkina, M.I., Harkova, E.V., Kazachkova, M.B. (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. <https://doi.org/10.17323/jle.2020.7176>
- Solnyshkina, M.I., Harkova, E.V., & Kiselnikov, A.S. (2014). Unified (Russian) state exam in English: Reading comprehension tasks. *English Language Teaching*, 7(12), 1-11. <https://doi.org/10.5539/ELT.V7N12P1>
- Solovyev, V., Solnyshkina, M., Ivanov, V., & Batyrshin, I. (2019). Prediction of reading difficulty in Russian academic texts. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4553-4563. <https://doi.org/10.3233/JIFS-179007>
- Solovyev, V. D., Solnyshkina, M. I., & McNamara, D. S. (2022). Computational linguistics and discourse complexology: Paradigms and research methods. *Russian Journal of Linguistics*, 26(2), 275-316.

- Tanaka-Ishii, K., Tezuka, S., & Terada H. (2010). Sorting texts by readability. *Computational Linguistics*, 36(2), 203-227. <https://doi.org/10.1162/coli.09-036-R2-08-050>
- Thorndyke, P.W. (1977). Cognitive structures in comprehension and memory in narrative discourse. *Cognitive Psychology*, 14, 560-589. [https://doi.org/10.1016/0010-0285\(77\)90005-6](https://doi.org/10.1016/0010-0285(77)90005-6)
- Toma, I., Marica, A. M., Dascalu, M., & Trausan-Matu, S. (2021). Readerbench—automated feedback generation for essays in Romanian. *University Politehnica of Bucharest Scientific Bulletin Series C-Electrical Engineering and Computer Science*, 83(2), 21-34.
- Vorontsov, K., Frei, O., Apishev, M., Romov, P., & Dudarenko, M. (2015). Bigartm: Open source library for regularized multimodal topic modeling of large collections. In *International Conference on Analysis of Images, Social Networks and Texts* (pp. 370-381). Springer Cham.
- Vorontsov, K., & Potapenko, A. (2015). Additive regularization of topic models. *Machine Learning*, 101(1-3), 303-323.
- Watson T.R. (2016). *Discourse topics*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Williams J.P. (2005). Instruction in reading comprehension for primary-grade students: A focus on text structure. *The Journal of Special Education*, 39(1), 6-18.

APPENDIX

Examples of the selected topics and their corresponding word2vec scores

Table S1

ARTM Segment Model Topics and Texts with Maximum Topic Weights. A Lower Word2vec Q_t Score Corresponds to a Better Topic

#	Q_t	Topic	Most probable words of the topic	Books with the highest weight	Weights of topics in texts
1	0.61	Political system	federation power Russian statelaw state organ Constitution of RF federal	Nik-6	0.17
				Nik-10-11	0.16
				Nik-11	0.15
				Bog-9	0.11
				Nik-8-9	0.08
4	0.64	Law and order	crime law court criminal administrative punish- ment offense law responsibility authority	Nik-10-11	0.23
				Nik-7	0.17
				Bog-9	0.14
				Nik-9	0.12
				Nik-11	0.10
13	0.65	Science	science knowledge scientific education human cognition scientist research activity truth	Bog-10*	0.17
				Nik-10	0.10
				Bog-10	0.08
				Bog-8	0.05
				Nik-8-9	0.05
18	0.65	Development of society	society country development economic life social production economy modern social	Bog-11*	0.17
				Nik-10	0.09
				Bog-10	0.09
				Bog-8	0.09
				Bog-10*	0.08
16	0.65	Economy	economy money commodity state country price market income market economic	Nik-11	0.21
				Nik-9	0.18
				Bog-8	0.18
				Bog-7	0.07
				Nik-8-9	0.05
8	0.66	Religion	religion religious person society philosopher life history century state god	Nik-10	0.12
				Nik-8-9	0.09
				Bog-10*	0.08
				Nik-7	0.05
				Nik-8	0.04
14	0.66	National identity	people person country national state Russia cul- ture Russian conflict language	Nik-6	0.10
				Nik-10	0.07
				Nik-8-9	0.07
				Nik-8	0.06
				Bog-8	0.05

#	Q^t	Topic	Most probable words of the topic	Books with the highest weight	Weights of topics in texts
19	0.67	Personality	person activity society personality social life consciousness need social spiritual	Bog-10*	0.25
				Bog-10	0.12
				Nik-10	0.12
				Bog-9	0.09
				Bog-11*	0.06
11	0.68	Culture	culture art spiritual mass society human value artistic cultural work	Bog-10	0.10
				Bog-11*	0.08
				Nik-8	0.05
				Nik-10	0.04
				Bog-10*	0.04