



**ФУНКЦИОНАЛЬНАЯ СЕМАНТИКА,
ГРАММАТИКА И СТИЛИСТИКА**
FUNCTIONAL SEMANTICS, GRAMMAR AND STYLISTICS

DOI: 10.22363/2313-2299-2024-15-3-872-894

EDN: IICRRK

UDC 811.161.1'42:371.671.11


Research article / Научная статья

**Towards a Taxonomy of Textbooks as a Genre:
the Case of Russian Textbooks**

Marina I. Solnyshkina¹  , Gulnoza N. Shoeva¹ , Ksenia O. Kosova² 

¹Kazan (Volga Region) Federal University, *Kazan, Russian Federation*

²RUDN University, *Moscow, Russian Federation*

 mesoln@yandex.ru

Abstract. The project is presented in the paper initially is launched to design a functional recognition or classification model of a modern Russian school textbook as a genre. In this study we test and confirm the hypothesis that detection of domain (subject area) and complexity level of a textbook can be reduced to a limited number of quantitative linguistic parameters provided with accurately identified and verified value ranges. We outlined our approach to genre analysis as multi-dimensional, compiled a corpus of over 1 mln. tokens, measured values of 15 linguistic parameters in 19 textbooks of two different subject areas and complexity levels, revealed 7 complexity predictors, 7 subject area predictors, and one — frequency — a metaparameter able to discriminate textbooks of History and Social Studies from texts of other genres. Our findings highlight the significance of the following parameters for textbooks across the selected subject areas: incidence of nouns, verb tenses (present, past and future), local and global argument overlap, type-token ratio. Complexity classification model is ascertained to be a function of sentence length, word length, incidence of nouns in genitive case and verbs, Abstractness score, verb/noun ratio, and adjective/noun ratio. The outcomes of this analysis will be used to interpret quantitative linguistic descriptions and classify texts.

Keywords: text profiling, multi-dimensional analysis, rulingva, parametrisation indices, classification models

Financing. Acknowledgements:

The research was supported by the RSF grant 24-28-01355 “Genre-discourse characteristics of the text as a function of lexical range”. This paper has been supported by the RUDN University Scientific Projects Grant System, project No 050738-0-000.

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

© Solnyshkina M.I., Shoeva G.N., Kosova K.O., 2024



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Conflicts of interest: The authors declare no conflicts of interest.

Authors' contribution: The authors contributed equally to this article.

Article history:

Received: 01.06.2024

Accepted: 15.06.2024

For citation:


Solnyshkina, M.I., Shoeva, G.N. & Kosova, K.O. (2024). Towards a Taxonomy of Textbooks as a Genre: the Case of Russian Textbooks. *RUDN Journal of Language Studies, Semiotics and Semantics*, 15(3), 872–894. <https://doi.org/10.22363/2313-2299-2024-15-3-872-894>

На пути к систематике учебников как жанра: на примере российских учебников

М.И. Солнышкина¹  , Г.Н. Шоева¹ , К.О. Косова² 

¹Казанский (Приволжский) федеральный университет, Казань, Российская Федерация

²Российский университет дружбы народов, Москва, Российская Федерация

 mesoln@yandex.ru

Аннотация. Проект, представленный в статье, первоначально ставил цель разработки функциональной модели распознавания или классификации современного российского школьного учебника как жанра. В этом исследовании мы проверяем и подтверждаем гипотезу о том, что определение предметной области и уровня сложности учебника может быть сведено к ограниченному числу количественных лингвистических параметров, имеющих точно заданные и проверенные диапазоны значений. Мы охарактеризовали наш подход к жанровому анализу как многомерный, собрав корпус из более чем 1 млн текстов. Лексемы, измерявшие значения 15 лингвистических параметров в 19 учебниках двух разных предметных областей и уровней сложности, выявили 7 предикторов сложности, 7 предикторов предметной области и один частотный — метапараметр, способный отличать учебники истории и обществознания от текстов других жанров. Наши результаты подчеркивают важность следующих параметров для учебников по выбранным предметным областям: частота встречаемости существительных, времена глаголов (настоящее, прошедшее и будущее), совпадение локальных и глобальных аргументов, соотношение типов лексем. Установлено, что модель классификации сложности зависит от длины предложения, длины слова, частоты встречаемости существительных в родительном падеже и глаголов, показателя абстрактности, соотношения глаголов и существительных, а также соотношения прилагательных и существительных нарицательных. Результаты этого анализа будут использованы для интерпретации количественных лингвистических описаний и классификации текстов.

Ключевые слова: профилирование текста, многомерный анализ, ru lingua, индексы параметризации, классификационные модели

Финансирование. Благодарности:

Исследование выполнено при поддержке гранта РНФ 24-28-01355 «Жанрово-дискурсивные характеристики текста в зависимости от лексического диапазона». Статья подготовлена в рамках проекта № 050738-0-000 Системы грантовой поддержки научных проектов РУДН. Работа выполнена за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета (ПРИОРИТЕТ-2030).

Заявление о конфликте интересов: Авторы заявляют об отсутствии конфликта интересов.

Вклад авторов: Вклад авторов равнозначен на всех этапах исследования.

История статьи:

Дата поступления: 01.06.2024

Дата приема в печать: 15.06.2024

Для цитирования:

Solnyshkina M.I., Shoeva G.N., Kosova K.O. Towards a Taxonomy of Textbooks as a Genre: the Case of Russian Textbooks // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2024. Т. 15. № 3. С. 872–894. <https://doi.org/10.22363/2313-2299-2024-15-3-872-894>

Introduction

In various fields of research, including information retrieval, corpus and computational linguistics, the demand for automatic classification of large amounts of texts and text taxonomies is recognized as increasing. Algorithms, aimed at designing classification text models, involve numerous procedures such as revealing relations between elements or structures of a text, attributing texts or groups of texts to specific ranges of variables and, finally, linguistic profiling [1].

The text model itself is nowadays viewed as an Abstract category able to (1) reproduce “relevant”, i.e. essential for a particular research, properties of the original text and (2) function as an analogue of the original. Functional recognition or classification models do not only correlate form and meaning but identify whether a text belongs to a certain category. Provided with a certain ‘input’, i.e. a segment of a text or a whole text, a researcher ‘applies the model’ conducting a comprehensive comparative and contrastive analysis for a set of linguistic parameters with the aim to identify idiosyncrasies and similarities of a text with a model. Based on the statistical analysis of the latter ‘the input text’ is recognized as a member of the class or an outcast [1]. The concept of a recognition or classification model also lies beyond the multi-dimensional method developed by D. Biber [2] who postulated his aim as identifying “clusters of co-occurring linguistic features and describing linguistic variation” [3]. Fostered by corpus linguistics and statistics, these methods enable to determine genre or discourse specific parameters, compare and contrast groups of texts with respect to different language dimensions [4; 5].

Classical genre classification models group texts based on systematic similarities including lexical, semantic, syntactic or discourse parameters. For example, B. Paltridge, in his seminal paper on genre analysis and identification of its textual boundaries emphasized the importance of semantic and syntactic features that play a key role in determining text genre [5]. D. Biber, on the other hand, while aiming at a text/register classification, applied multi-feature/multi-dimension (MF/MD) method and focused mostly on morphological parameters [2].

The past decade has witnessed a surge of interest in genre classification research based on multi-dimensional algorithm: it turned into a well-studied and developed area with applications in numerous fields [1; 4; 6; 7]. In the modern paradigm, linguists’ interest in designing classification models and their application increased

immensely due to the growing need of data processing automation in numerous areas including security and forensics, authorship profiling, NLP, intelligence, marketing etc. Nevertheless, Russian in this matter remains an under-resourced language with limited datasets at linguists' disposal and validated classification models to refer to [8–10].

In a broader perspective, our objective is to contribute to classification models of Russian textbooks as a genre, i.e. identify, validate, and accumulate textbook typological parameters. Sharing W. Raible's viewpoint on text genres as major constructs "of the communicative economy of a society" which have their pragmatic settings, we regard a textbook to be a genre that comes "in series, with one text being the model for another, not without typical (and inescapable) changes during diachrony" [6]. Based on the above, we *hypothesize* that textbooks of one historical period, but different (1) subject areas and (2) grades or complexity levels though bearing certain differences in values of lexical and syntactic parameters, constitute one genre, and as such demonstrate similarities on numerous language levels¹. The

¹ Corpus Materials. Textbooks Arsent'ev, N.M., Danilov, A.A., Levandovskij, A.A. & Tokareva, A.Ja. (2016). *Istorija Rossii. 9 klass. Uchebnik dlja obshheobrazovatel'nyh organizacij. V 2 ch.* Moscow: Prosveshchenie. (In Russ.). Bogolyubov, L.N. (2010). *Obshhestvoznanie. 8 klass: ucheb. dlja obshheobrazovat. uchrezhdenij*, in Bogolyubov, L.N. & Gorodetskaya, N.I. (eds.). Moscow: Prosveshchenie. (In Russ.). Bogolyubov, L.N. & Ivanova, L.F. (2013). *Obshchestvoznanie. 5 klass : ucheb. dlya obshcheobrazovat. uchrezhdenii s pril. na elektron. Nositele.* Moscow: Prosveshchenie. (In Russ.). Bogoljubov, L.N., Matveev, A.I. & Zhil'cova, E.I. (2014). *Obshhestvoznanie 9 klass: ucheb, dlja obshheobrazovat. organizacij*, Bogoljubova, L.N. (Ed.). Moscow: Prosveshchenie. (In Russ.). Danilov, A.A. & Kosulina, L.G. (2015). *Istorija Rossii, XIX vek. 8 klass: ucheb. dlja obshheobrazovat. Organizacij.* Moscow: Prosveshchenie. (In Russ.). Danilov, D.D., Sizova, E.V., Turchina, M.E. (2015). *Obshchestvoznanie. 5 kl.: ucheb. dlya organizatsii, osushchestvlyayushchikh obrazovatel'nyu deyatel'nost'.* Moscow: Balass. (In Russ.). Judovskaja, A.Ja., Baranov, P.A. & Vanjushkina, L.M. (2023). *Istorija. Vseobshhaja istorija. Istorija Novogo vremeni. XVIII vek. 8-j klass. Uchebnik*, Iskenderov, A.A. (Ed.). Moscow: Prosveshchenie. (In Russ.). Judovskaja, A. Ja. & Baranov, P.A. (2019). *Vseobshhaja istorija. Istorija Novogo vremeni. 9 klass. Uchebnik*, Iskenderov, A.A. (Ed.). Moscow: Prosveshchenie. (In Russ.). Kravchenko, A.I. (2012). *Obshchestvoznanie: uchebnik dlya 5 klassa obshheobrazovatel'nykh uchrezh denii.* Moscow: Russkoe slovo. (In Russ.). Kravchenko, A.I. (2010). *Obshhestvoznanie: Uchebnik dlja 8 klassa obshheobrazovatel'nyh uchrezhdenij.* Moscow: Russkoe slovo. (In Russ.). Kotova, O.A. & Liskova, T.E. (2019). *Obshhestvoznanie. 8 klass. Uchebnik.* Moscow: Prosveshchenie. (In Russ.). Lyashenko, L.M., Volobuev, O.V. & Simonova, E.V. (2016). *Istorija Rossii: XIX — nachalo XX v. 9 kl. Uchebnik.* Moscow: Drofa. (In Russ.). Nikitin, A.F. & Nikitina, T.I. (2014). *Obshhestvoznanie. 8 klass. Uchebnik.* Moscow: Drofa. (In Russ.). Nikitin, A.F. & Nikitina, T.I. (2014). *Obshhestvoznanie. 9 klass. Uchebnik.* Moscow: Drofa. (In Russ.). Nikitin, A.F. & Nikitina, T.I. (2013). *Obshchestvoznanie. 5 kl.: ucheb. dlya obshcheobrazovat. Uchrezhdenii.* Moscow: Drofa. (In Russ.). Nikishin, V.O. & Strelkov, A.V. (2023). *Vseobshchaya istoriya. Istorija drevnego mira 5 klass. Uchebnik. Obnovlennyyi. FGOS.* Moscow: Russkoe slovo. (In Russ.). Saplina, E.V., Nemirovskii, A.A., Solomatina, E.I. & Tyrin, S.V. (2021). *Vseobshchaya istoriya. Istorija Drevnego mira. 5 klass.* Moscow: Drofa. (In Russ.). Vigin, A., Goder, G.I. & Svetsitskaya, I.S. (2020). *Vseobshchaya istoriya. Istorija Drevnego mira. 5 klass: ucheb. dlya obshcheobrazovat organizatsii.* Moscow: Prosveshchenie. (In Russ.).

latter is caused by the ability of academic discourses of different domains to interpret the world in particular ways, each drawing on different lexical, grammatical and rhetorical resources to create specialized knowledge [7].

In our previous study we solely focused on identifying differences which distinguish textbooks of different subjects [8–10] or literary texts and textbooks [11], to the neglect of similarities of a textbook as a genre. The present study *aims* at identifying a range of descriptive parameters ample to define a textbook as a genre with the following research questions posed for investigation:

1. What are the value ranges of linguistic parameters specific for modern Russian textbooks of History and Social Studies?
2. What is a set of domain-attributed text parameters for school textbooks of History and Social Studies?
3. What is a set of grade or complexity-attributed text parameters for school textbooks of Grades 5, 8 and 9?
4. What is a set of parameters common for school textbooks of different grades and subject areas?

Literature Review

Constructing a genre model of a textbook involves operating with the notions of a genre, text type, text, and a textbook, therefore we start with the distinction between text genres and text types. From these preliminaries we can further derive the notion of a recognition model of a textbook. As the genre under study, i.e. a textbook, employs a communicative event, in which language plays both a significant and an indispensable role, we restrict ourselves to focusing on language not pragmatic parameters, which may be crucial, for example, in communicative settings of a football match, visiting a photo gallery, figure skating training, etc.

J. Swales formulates a definition of a *text genre* based on its communicative purpose, arguing that “the principal criterial feature that turns a collection of communicative events into a genre is some shared set of communicative purposes” [12]. Applying this concept, we refer to *genres* as “classes of communicative events” [12], taxonomy of which is based on “intents of communicators” [9]. *Text types*, on the other hand, being identified by their structure and the way information is presented (i.e. descriptive, explanatory, narrative, etc.), are similar in their linguistic characteristics, regardless of their situational or register features [13].

Textbooks tend to accumulate different text types comprising descriptive, explanatory and narrative parts [1]. Engineering textbooks, for example, may contain instructions which deliver procedural information [14], while textbooks on Humanities are constituted predominantly of argumentative and narrative elements [9]. Confirming the idea of certain linguistic eclecticism in academic texts [15], emphasizes that all of them, including research genres, differ in their themes and communicative settings, and as such they facilitate text types convergence within the same genre.

Based on the concept that texts are “socially meaningful traces of a communicative event” [16], the modern research paradigm defines *textbooks* as secondary texts adapted for academic purposes. In proper educational settings they serve to make scientific ideas more comprehensible for learners thus realizing their informative function [17]. The latter predetermines that before becoming a textbook, a scientific text undergoes numerous procedures of simplification and clarification of complex concepts [18] as it is designed to be a means for systematic studies of a subject, and as such is expected to be structured and accumulate information, arguments, explanations, examples and exercises to develop and practice students’ skills [18].

The generic character of a *functional genre model of a textbook* implies that it comprises numerous parameters and filters texts based on at least two classification criteria: (1) subject area and (2) complexity level. It is obvious that textbooks of different domains and grades embody a certain amount of idiosyncrasies, on the one hand, but, on the other, all representatives of the genre resemble each other in certain essential respects. While idiosyncrasies jeopardize a text’s membership in the genre group, its common features with other texts of the same genre indicate its belonging.

Materials, methods, analysis

In this study we employ an integrated approach comprising automated linguistic analysis as well as statistical methods for data processing and interpretation.

The *research design* comprised three stages:

(a) *Data collection*:

We employed two collections of textbooks on Social Studies and History: 8 textbooks for Grade 5 with the total size of 317,318 tokens compiled for the current study and 11 textbooks for grades 8 and 9 with the total size of 716,882 tokens, which were also used in our previous study [11]. Thus, the total size of the corpus exceeds 1 mln tokens.

To ensure representativeness and complexity balance of the selected texts, we compiled each of the two subcorpora with about the same number of textbooks on the grade levels selected (see Table 1). However, as textbooks on History are typically much longer, the corpus is highly skewed to History.

Table 1

Size of Textbooks Corpus

| Subject area | Grade 5 (textbooks) | Subcorpus (tokens) | Grades 8–9 (textbooks) | Subcorpus (tokens) | Corpus size |
|--------------------------|---------------------|--------------------|------------------------|--------------------|----------------|
| History textbooks | 4 | 226529 | 5 | 138463 | 644672 |
| | | | | 279680 | |
| Social Studies textbooks | 4 | 90789 | 6 | 180549 | 389528 |
| | | | | 118190 | |
| Total | | 317318 | | 716882 | 1034200 |

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

All the selected textbooks belong to the “Ministry List”, i.e. Federal Register of the textbooks recommended for use in all state schools of the Russian Federation, and as such they are viewed a priori as high quality texts approved by psycho-pedagogical examination committee (fpu.edu.ru/).

b) Text parameters measurement:

We started this stage with parsing the textbooks into documents of about 1000 tokens, keeping all sentences whole and not compromising their integrity. Segmentation was finalized with deleting the final passage of the textbook which contained fewer than 1000 tokens and thus resulted in a slight decrease of the Corpus (see Table 1). Later, all texts were annotated and processed using RuLingva, a text analyzer developed for Russian academic texts.

At the moment RuLingva offers 49 indices for Russian text parametrization including quantitative descriptive, morphological, lexical, and discourse (cohesion) indices ascribing corresponding values to each text parameter and parameterization categories to each word in the document.

To confirm the theoretical assumptions about the variability of textbooks of different disciplines, we processed the texts on numerous morphological, lexical and syntactic parameters following the research design developed in [11].

Based on our previous research, of the 49 parameters available for extraction in RuLingva [8; 10; 11], we limited the list for analyses to the following 15: Average number of words per sentence, Average number of syllables per word, incidence of Nouns, Genitive case (Noun), incidence of Verbs, Present tense (Verb), Future tense (Verb), Past tense (Verb) (Norm), Frequency (by Sharoff), Abstractness score, Type-Token Ratio (average), Local argument overlap, Global argument overlap, Verb/Noun ratio, Adjective/Noun ratio. As the previous research indicates, these parameters play crucial roles in defining text complexity and genre specific characteristics.

Below we provide brief descriptions of these parameters to introduce readers into the system of measured by RuLingva indices.

- *Average number of words per sentence* is related to text readability or syntactic complexity: longer sentences suggest more elaborate and formal texts, such as academic or literary works. Shorter sentences are common in journalistic or children’s literature, prioritizing clarity and brevity [19].
- *Average number of syllables per word* is also used as an index or variable to predict certain text characteristics. By analysing word length researchers identify important cross-genre patterns, predict texts structure, which, in turn, affects text readability and complexity [20].
- *Nouns* incidence in a text provides insights into its informational density and the nature of its content. Texts with a high incidence of nouns tend to be more descriptive and information-heavy, e.g. academic texts and technical documents.

- *Genitive case* of nouns which typically marks possession (e.g., ‘the book’s cover’), is also indicative of descriptive, detail-oriented language. Exploring variability of genitive constructions, C. Lyons (1986) emphasizes their role in numerous syntactic and semantic relationships [21]. By highlighting possessive and descriptive nuances, which are particularly important in genres requiring precision and detail, *genitive case* constructions are also known to contribute to annotating genres.
- *Verb incidence* is an important parameter in determining syntactic structure and content of a text, adding to identification of genre-specific features and (dis) similarities between genres. The ratio of verbs to other parts of speech varies from genre to genre, affecting the overall style and text sentiment [22; 23].
- *Verb tenses* (Present, Future and Past) reflect time frames, purposes and functions of different genres enabling a more accurate text interpretation. D. Biber et al. (1999) in their seminal work on corpus linguistics point out that the use of different verb tenses helps create a sense of time and place in a text, which is essential for its genre identification [24].
- Cumulative text *frequency* refers to the sum of all the tokens frequencies in a text and is viewed useful in comparing complexities of two or more texts. RuLingva assesses cumulative text ‘*frequency*’ based on Lyashevskaya and Sharoff’s “Frequency Dictionary of Russian» which provides a comprehensive list of the most frequent words in Russian [25].
- Text *Abstractness score* reflects its cognitive or informative complexity. More Abstract texts require readers to use complex cognitive strategies to build mental models and interpret content. The latter is due to readers’ ability to integrate different levels of information, which makes the reading process more complex [26].
- *Type-token ratio* or *TTR* is widely used to measure lexical ‘richness’, ‘diversity’ or ‘variety’ of a text and revealed to be higher in discourses offering a broad range of unique concepts and events [27; 28]. TTR is proved to be sensitive to text length: shorter texts tend to have higher TTRs, while in longer texts, due to numerous repetition of functional words, TTR is lower. TTR is one of the reasons researchers are expected to parse texts under study into 1000 tokens as the modern paradigm requires [29].
- *Local argument overlap* measures repetition of nouns and pronouns in adjacent sentences while *Global argument overlap* is a metric of all repeated nouns and pronouns in a text indicating cohesion and thematic focus. High *argument overlap* implies a tightly focused subject matter, while low metrics of argument are typical of narrative texts, where the story progresses with varying agents [20–32].
- *Verb/noun ratio* is indicative of a text’s dynamism. Characterizing this parameter, Biber (1988) and Conrad (2000) highlight its utility

in distinguishing narrative against descriptive text types as high ratio suggests action-oriented content typical of narratives and procedural texts, while a lower ratio is indicative of a descriptive or expository style, common in academic and scientific writing [1; 19].

- *Adjective/Noun ratio* aids to genre classification by processing ‘emotive and descriptive quality of sentences’ [21]. Research shows that adjectives’ frequency and context vary considerably across academic disciplines. Some adjectives are particularly dominant in specific areas of academic texts, suggesting differences in style and vocabulary across academic disciplines [33].

c) *Comparative ↔ contrastive analysis*.

On this stage of research we addressed the research questions with the view to identify

(1) a set of domain-attributed text parameters for school textbooks of History and Social Studies;

(2) a set of complexity predictors or grade-attributed text parameters for school textbooks of Grades 5, 8 and 9;

(3) a set of parameters common for school textbooks of different grades and subject areas.

Using STATISTICA software, we processed RuLingva output data identifying typical ranges of core parameters values (Table 2) and assessed statistically significant (dis) similarities between the subcorpora indices (Table 3).

We also implemented a cross-sectional comparison and juxtapositioned A. History textbooks vs Social Studies textbooks and B. Grade 5 textbooks vs Grade 8–9 textbooks. We focused on the role of each parameter across subject areas discourses and grade levels. To identify the relationship between each pair of values, we employed Mann-Whitney correlation analysis. Table 3 below shows correlations of the parameters separately in grade levels and subject areas investigated.

Results

We derived several statistical measures to comprehensively analyse linguistic features of the textbooks across the selected grades and domains with the aim to identify value ranges of parameters specific for modern Russian textbooks of History and Social Studies.

Description and analysis of these measures is provided in the following order: Mean sentence length, Mean word length (In syllables), Nouns, Genitive case (Noun), Verbs, Present tense (Verb), Future tense (Verb), Past tense (Verb), Frequency (by Sharoff), Abstract index, TTR, Local argument overlap, Global argument overlap, Verb/Noun ratio, Adjective/Noun ratio (Table 2).

Table 2

The range of linguistic parameters values

| Parameter | Social Studies, Grade 5 (N* = 88) | Social Studies, Grades 8–9 (N = 298) | History, Grade 5 (N = 226) | History, Grades 8–9 (N = 1089) |
|---------------------------------------|--------------------------------------|--|-------------------------------|-----------------------------------|
| Mean sentence length | 11.29±1.97 | 14.98±2.18 | 10.06±1.00 | 14.89±1.66 |
| Mean word length (In syllables) | 2.45±0.18 | 2.77±0.17 | 2.46±0.12 | 2.61±0.10 |
| Nouns | 350.92±40.76 | 397.50±32.94 | 412.65±36.62 | 410.14±32.83 |
| Genitive case (Noun) | 95.85±27.82 | 140.74±30.89 | 149.34±23.53 | 145.98±21.24 |
| Verbs | 158.23±25.88 | 125.14±16.86 | 124.50±20.99 | 130.61±17.89 |
| Present tense (Verb) | 62.53±15.57 | 64.92±12.77 | 22.77±11.73 | 15.10±9.68 |
| Future tense (Verb) | 4.78±3.51 | 3.21±2.64 | 2.00±2.20 | 1.24±2.12 |
| Past tense (Verb) | 48.17±25.76 | 33.13±12.41 | 101.85±21.83 | 96.6±17.33 |
| Frequency (by Sharoff) | 397.54±144.84 | 265.51±58.59 | 235.99±66.37 | 203.28±34.20 |
| Abstract index | 2.66±0.10 | 2.78±0.10 | 2.61±0.10 | 2.77±0.08 |
| TTR | 0.48±0.03 | 0.48±0.04 | 0.51±0.03 | 0.53±0.03 |
| Local argument overlap | 0.53±0.16 | 0.80±0.28 | 0.30±0.12 | 0.36±0.17 |
| Global argument overlap | 0.19±0.07 | 0.29±0.10 | 0.11±0.05 | 0.16±0.07 |
| Nominative ratio (Verb/ Noun) | 0.47±0.13 | 0.32±0.07 | 0.36±0.07 | 0.32±0.06 |
| Descriptive ratio (Adjective/Noun) | 0.31±0.06 | 0.37±0.05 | 0.30±0.05 | 0.40±0.06 |

N* — number of 1000 tokens fragments processed.

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Value ranges of text core indices

Below we offer boxplots visualizing the value range of each selected parameter and briefly describe text core indices. The dotted boxes represent the most typical values of the selected indices.

Mean sentence length

Fig. 1. visualizes the range and growth of sentence length in Social Studies and History textbooks from Grade 5 to Grades 8–9. The dotted line indicates the typical core of indices from minimum 9 words in History 5 to maximum of over 16 words in Social studies 8–9, on average

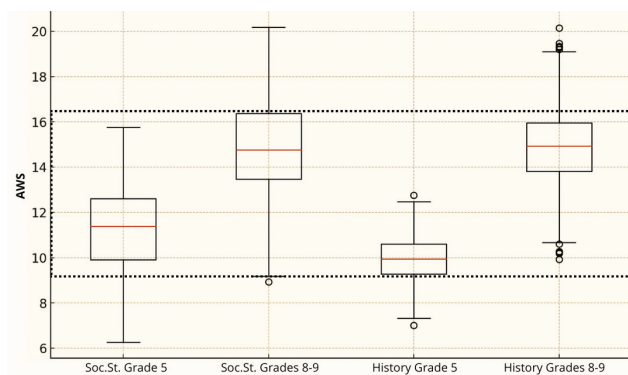


Fig. 1. Range of sentence length

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Mean word length (In syllables)

The average number of syllables per word is an indicator of vocabulary complexity, as words with more syllables are usually more complex and require a higher level of language skills [21]. Fig. 2 shows that in contrast to Grade 5 texts, high school texts contain longer words, thus reflecting their complex and specialized nature. History texts contain proper names of historical personalities, cities, states, locations, while specificity of Social Studies texts is reinforced by presence of terms which are longer than average common words. Such outliers are not anomalies in academic texts, they rather reflect specificity of the educational materials addressed to high school students.

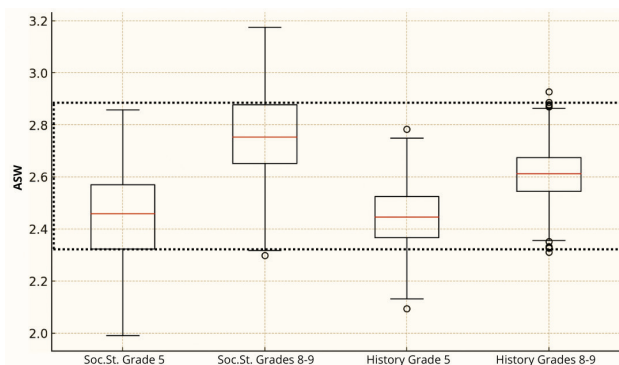


Fig. 2. Range of word length

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Nouns

Fig. 3. shows that the average number of nouns varies by grade level and subject. The average number of nouns in Grade 5 Social Studies textbooks is approximately 350 per 1000 tokens. Textbooks for this age group contain a relatively small number of nouns, which corresponds to the more simplified structure of the text intended for younger students. And in texts for Grade 8–9 learners, the average number of nouns increases to about 398. The higher average and wide range of meanings indicate the presence of more specialized vocabulary and complex grammatical structures. The average number of nouns in History textbooks for Grade 5 is about 413. This shows that even on Grade 5 level, history texts contain a significant number of nouns. The presence of top outliers in the graph (>500 words <330), especially in Grades 8–9 History texts, indicates high volatility of the parameter in the selected texts and, consequently, need for further investigation and development of available data

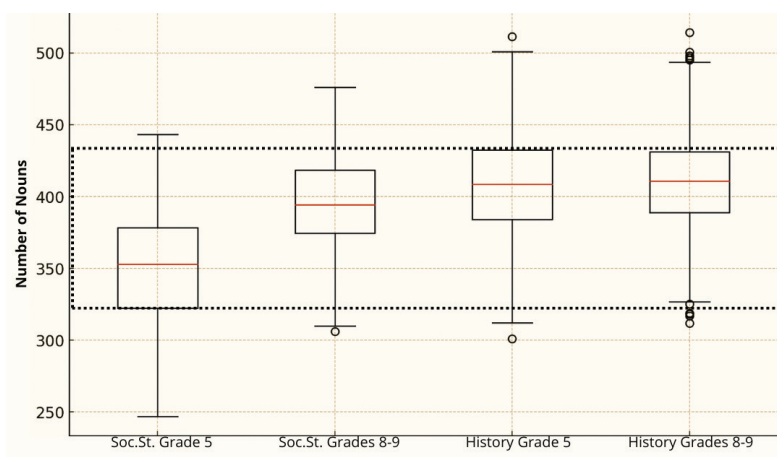


Fig. 3. Incidence of nouns

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Genitive case (Noun)

Genitive case of nouns typically denotes and signifies belonging, a part of a whole, or clarification and is used in grammatical constructions Noun + Noun^{gen}. Its incidence is a validated morphological complexity predictor [12] and its growth may indicate readers' difficulties in comprehension. History textbooks for both grades and Social Studies textbooks for Grades 8–9 demonstrate nearly identical, rather high, values of genitive case incidences: 149 and 142 per 1000 tokens respectively.

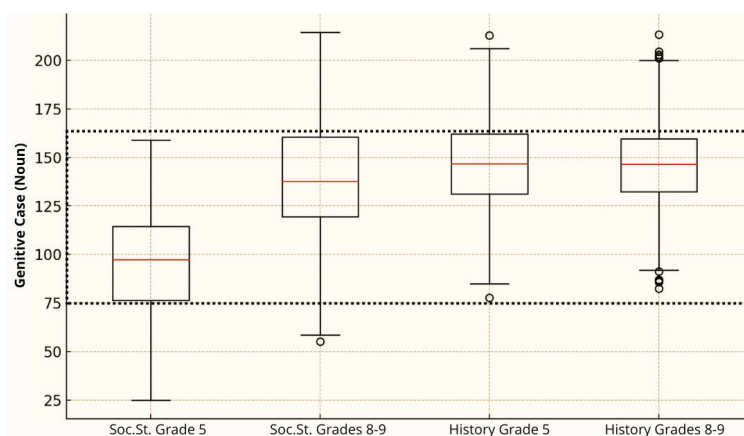


Fig. 4. Incidence of genitive case (nouns)

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Verbs

The graph below compares the incidence of verbs in Social Studies and History texts. The average number of verbs in Grade 5 Social Studies textbooks is approximately 158, but in Grades 8–9 textbooks the average number of verbs drops to 125 indicating that the texts lose narrativity, acquire descriptiveness and analyticism. The latter is likely to be related to changes in incidence of nouns and adjectives. History textbooks for younger students have relatively stable verb incidence index, which favors a more dynamic and narrative style of presenting historical events and facts. In Grade 8–9, the average number of verbs slightly increases to 130. Outliers on the graph, especially in History Grades 8–9, may be caused by topics or quotations from other discourses where the number of verbs fluctuates. For example, some chapters may comprise more narrative or dynamic elements thus resulting in increased verb use.

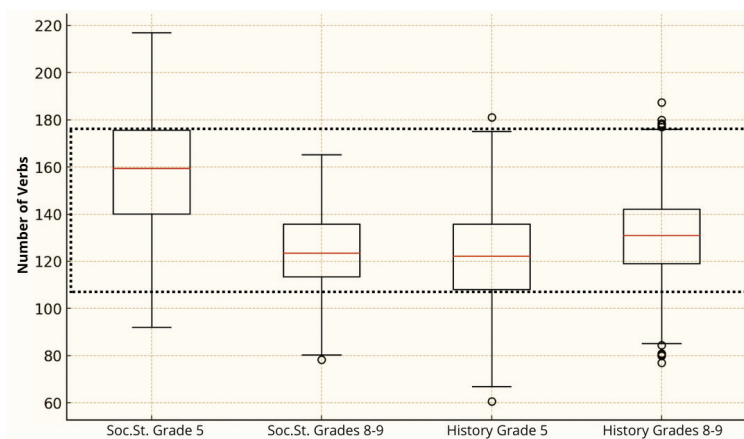


Fig. 5. Incidence of verbs

Source: compiled Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Frequency (by Sharoff)

Research shows that lexical frequency in textbooks decreases from junior to senior grades, reflecting changes in academic styles. Senior grade textbooks place more emphasis on specialized and academic vocabulary with lower frequency indices [9].

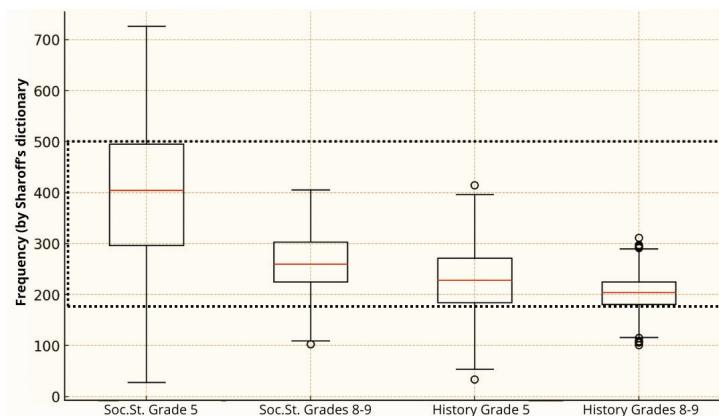


Fig. 6. Frequency

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Abstract index

Abstractness is a validated complexity predictor with a tendency to increase across grades in instructional texts [8; 11]. The current study (Fig. 7) confirmed the previous research results [8; 11]: texts of both subjects indicate a noticeable growth of the parameter value across grades. We also observe that Abstractness indices in both Social Studies and History demonstrate relatively small standard deviation, indicating homogeneous use of Abstract concepts. Another significant finding is that Abstractness in History 5 is lower than in Social Studies 5.

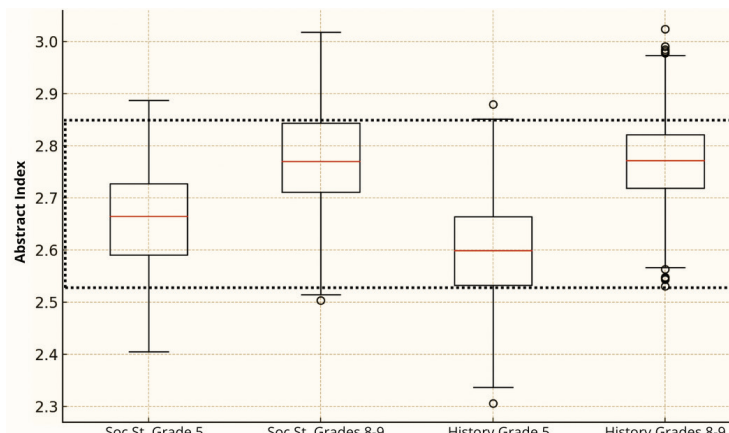


Fig. 7. Abstractness

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Type Token ratio (TTR)

History books TTR increases gradually across grades from 5 to 8–9, reflecting more diverse vocabulary in high school where students are expected to be exposed to more Abstract notions. TTR in Social Studies textbooks of both levels is 0.48, which is impressively lower than in History, i.e. 0.52 and 0.54 respectively, with the range of values being slightly wider in Grade 8–9.

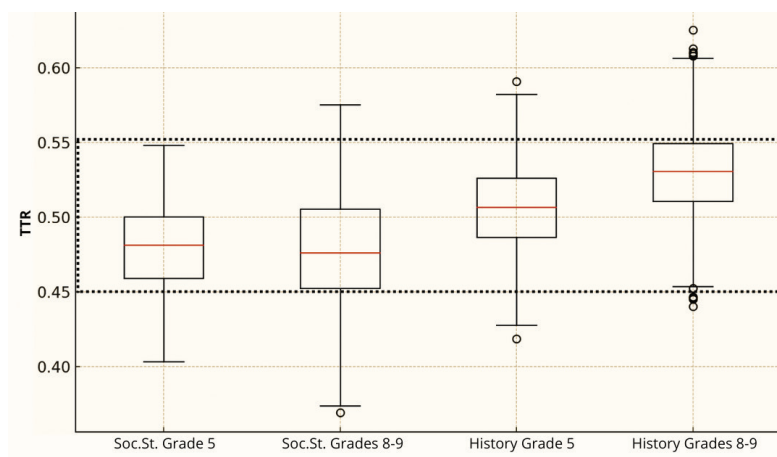


Fig. 8. Type Token Ratio

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

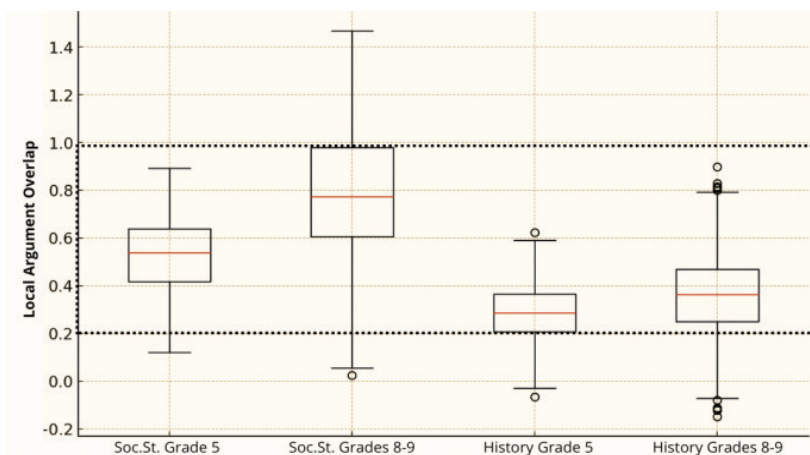
Local / Global argument overlap

The average of Local argument overlap in Grade 5 Social Studies is 0.53 with a range from 0.2 to 1.0. This indicates a moderate degree of argument, i.e. key nominal, elements repetition, which is intended to help readers comprehend information. Noteworthy, that the average level of local argument overlap in Grades 8–9 Social Studies is much higher at 0.80, with a range of values from 0.2 to 1.4. The latter reflects the so called ‘complexity trade-off’, when a higher level of one complexity predictor, i.e. cohesion, due to the repetition of nominal elements, simplifies the text. It is viewed as a kind of balance to increased cognitive complexity, realized in longer sentences, higher Abstractness or TTR in higher grades.

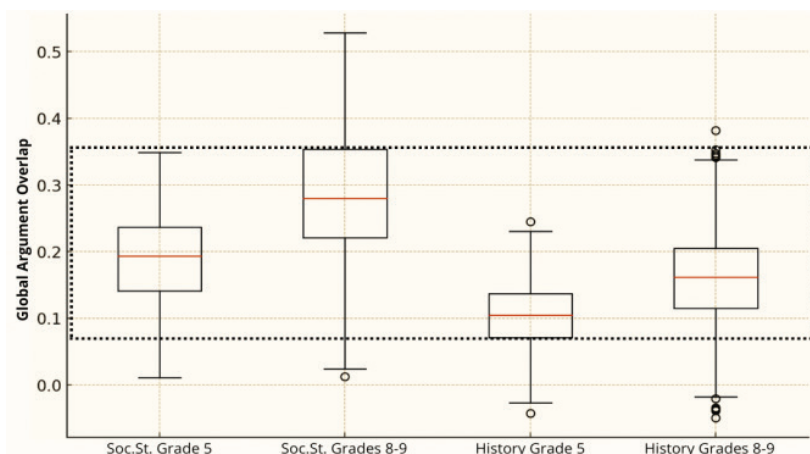
Another interesting observation is that both Local and Global argument overlaps in History textbooks are of lower values compared to Social Studies textbooks, which is another marker of Social Studies textbooks writers’ efforts to assist text comprehension by inclusion of more repetitions.

Nominative ratio (Verb/Noun)

The ratio decreases from Grade 5 to Grade 8–9 in both subjects which is predictable in academic texts, as higher values indicate higher frequency of verbs compared to nouns and vice versa.



a



b

Fig. 9. a) Local argument overlap; b) Global argument overlap
Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

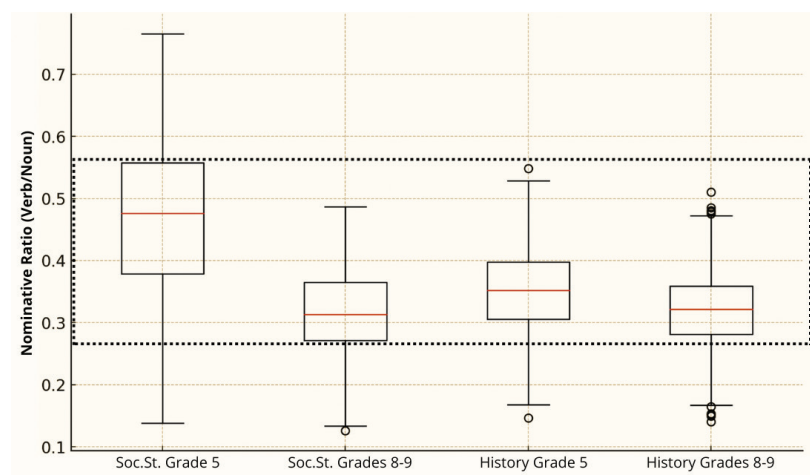


Fig. 10. Verb/Noun ratio
Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Descriptive ratio (Adjective/Noun)

As for the Descriptive ratio, it tends to grow across grades alongside with cognitive development of students and text complexity in texts of both subjects.

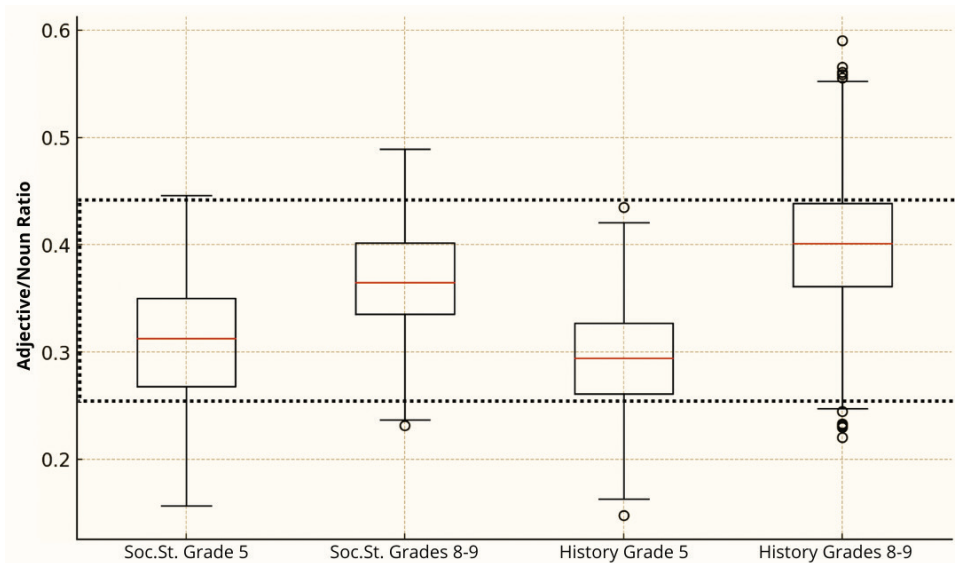


Fig. 11. Adjective/Noun ratio

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Verb tenses

Dynamics of verb tenses has subject peculiarities: unexpected higher narrativity in Social Studies Grade 5 is obvious in all three Figures, where the ratio of verbs in all tenses is higher in Grade 5 books than in Grades 8–9. We also observe anticipated lower levels of present and future tenses in History textbooks and their skewness to past tenses.

Thus, our experimental data, presented in 4.1, indicate that defining value ranges of the core textbook parameters are although feasible but rather challenging. The possible reasons to the above is a simple fact that a textbook categorization or assigning it to a category, involves at least two classification principles, i.e. complexity (or a grade level) and a subject area. A textbook is always denoted in these two parameters, i.e. domain and the year of schooling, e.g., Geography 5 or Mathematics 3. Due to the above, the most practical decision for textbook classification model is to comprise two steps. The first, i.e. application of the subject inherent parameters, must be followed by the (non) attribution of the text to a particular subject area and the second presupposes application of the complexity/grade-inherent parameters followed by the (non) attribution of the text to the complexity (grade) level.

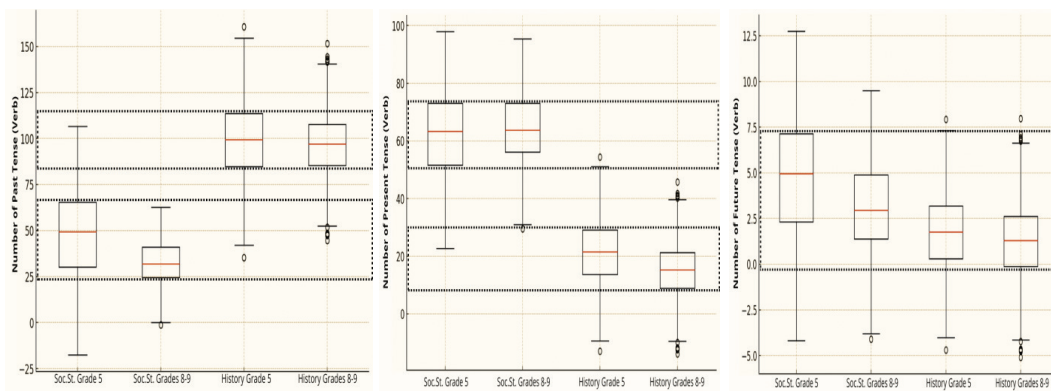


Fig. 12. Verb tenses

Source: compiled Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Domain- and Grade-attributed parameters

For the objective specified above, in the following part of the article we identify two sets of text parameters:

- 1) domain-attributed parameters for school textbooks of History and Social Studies only. They are supposed to be similar or common for different complexity levels and as such do not or poorly discriminate texts of different grades. Their main function is to “attribute” a text as a member of History/Social Studies subject area or not;
- 2) grade-attributed text parameters for school textbooks of Grades 5, 8 and 9, which are common for different subject areas and as such demonstrate similar tendencies in texts of the same complexity level. In our research they do not discriminate texts of History and Social Studies, but as complexity predictors may categorize grade level of a text only.

Applying Mann — Whitney U statistical technique, we computed a cross-sectional comparison between textbooks of the same discourse, i.e. Humanities, but different subject areas: History and Social Studies, on the one hand, and textbooks of different complexity levels, i.e. Grade 5 and Grades 8–9, on the other (Table 3).

The list of domain-attributed classification parameters comprises the following: Average number of words per sentence, Average number of syllables per word, Genitive case (Noun), incidence of Verbs, Abstractness score, Local argument overlap, Global argument overlap, (Table 3). The rest with only one exception, i.e. Frequency, are grade-attributed text parameters including incidence of Nouns, Present tense (Verb), Future tense (Verb), Past tense (Verb) (Norm), Type-Token Ratio (average), Verb/Noun ratio, Adjective/Noun ratio.

Frequency constitutes a separate parameter discriminating neither complexity nor subject areas in textbooks, and as such demonstrates a narrow range of quantitative diversity characteristic of a metaparameter able to potentially differentiate textbooks from texts of other genres.

Domain- and grade-attributed textbook parameters

| Variable | Complexity (Grade level) | | | | Subject area | | | |
|--------------------------------------|--------------------------|---------|-------------------------|---------|----------------|---------|-------------------------|---------|
| | Grade 8–9 | Grade 5 | p-value, Mann-Whitney U | p-value | Social Studies | History | p-value, Mann-Whitney U | p-value |
| Average number of words per sentence | 14.59 | 10.53 | 0.50 | < .01* | 13.21 | 12.52 | 39.50 | 0.68 |
| Average number of syllables per word | 2.70 | 2.45 | 4.00 | < .01* | 2.65 | 2.54 | 24.00 | 0.09 |
| Nouns | 405.29 | 383.81 | 29.00 | 0.23 | 382.24 | 411.80 | 19.00 | < .01* |
| Genitive case (Noun) | 144.08 | 110.80 | 9.00 | < .01* | 124.18 | 136.61 | 30.00 | 0.23 |
| Verbs | 127.78 | 153.58 | 0.00 | < .01* | 138.93 | 138.33 | 38.00 | 0.59 |
| Present tense (Verb) | 37.77 | 41.99 | 38.00 | 0.65 | 57.91 | 19.15 | 9.00 | < .01* |
| Future tense (Verb) | 42.92 | 41.99 | 44.00 | 0.96 | 63.57 | 19.15 | 0.00 | < .01* |
| Past tense (Verb) | 60.28 | 75.95 | 28.00 | 0.20 | 39.46 | 97.35 | 1.00 | < .01* |
| Frequency (by Sharoff) | 3545.63 | 3383.70 | 24.00 | 0.10 | 3439.13 | 3520.02 | 36.00 | 0.48 |
| Abstractness score | 2.75 | 2.62 | 0.00 | < .01* | 2.71 | 2.68 | 39.00 | 0.65 |
| Local argument overlap | 0.58 | 0.40 | 29.00 | 0.23 | 0.67 | 0.32 | 2.50 | < .01* |
| Global argument overlap | 0.11 | 0.08 | 24.50 | 0.11 | 0.12 | 0.07 | 11.00 | < .01* |
| Type-Token Ratio | 0.51 | 0.50 | 29.50 | 0.24 | 0.49 | 0.53 | 5.50 | < .01* |
| Verb/Noun ratio | 0.32 | 0.41 | 8.50 | < .01* | 0.37 | 0.34 | 36.00 | 0.48 |
| Adjective/Noun ratio | 0.38 | 0.31 | 0.00 | < .01* | 0.34 | 0.35 | 35.50 | 0.46 |

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

Source: compiled by Marina I. Solnyshkina, Gulnoza N. Shoeva, Ksenia O. Kosova.

Our experiments indicate that textbook classification decisions can be made with reasonable accuracy on the basis of 15 parameters. Sentence length, word length, incidence of genitive case, incidence of verbs, Abstractness score, verb/noun ratio, and adjective/noun ratio as complexity predictors classify textbooks of different grade levels but not subject domains. Incidence of nouns, verb tenses (present, past and future), local and global argument overlap as well as type-token ratio discriminate textbooks of different subject areas but not complexity (grade) level. ‘Frequency’ behaves differently from all other parameters correlating neither with complexity nor subject area differences between groups of texts and establishing itself as a metaparameter able to discriminate between genres.

Discussion and conclusion

The study of genre profiling as an important area of computer and corpus linguistics based on predominantly quantitative analysis aims at revealing relations between elements of a text and designing a functional recognition model of a genre. Thus, the primary focus of the current research is a multi-dimensional profile of a modern Russian school textbook. Similar to our previous contrastive study of adventure stories and textbooks we identified a list of genre inherent quantitative lexical and grammatical features, in this paper we contrast textbooks of different disciplines and grades with the objective to define typical features of a modern Russian textbook as a genre.

Cross-sectional comparison was conducted between (1) textbooks of the same discourse, i.e. Humanities, but different subject areas: History and Social Studies and (2) textbooks of different complexity levels, i.e. Grade 5 and Grades 8–9.

By using a set of 15 linguistic parameters automatically extracted with RuLingva we singled out and statistically verified three distributional patterns: (1) textbooks of History and Social Studies are similar to one another in sentence length, word length, incidence of genitive case, incidence of verbs, Abstractness score, verb/noun ratio, and adjective/noun ratio; (2) parameters similar in texts of different complexity levels include incidence of nouns, verb tenses (present, past and future), local and global argument overlap as well as type-token ratio; (3) Frequency, discriminating neither complexity nor subject areas in textbooks, demonstrates a narrow range of quantitative diversity characteristic of a metaparameter able to differentiate textbooks with texts of other genres.

Prospects of the study are in synthesizing textbook classification model implementing the identified ranges of parameters, patterns and distinctions that characterize textbooks as a genre. We also plan to verify our assumption that frequency index may serve as a good classifier between genres in our future contrastive studies of textbooks and political, folklore and fiction texts.

References / Библиографический список

1. Kessler, B., Nunberg, G. & Schuetze, H. (1997). *Automatic Detection of Text Genre*.
2. Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
3. Biber, D. & Conrad, S. (2013). Introduction: Multi-dimensional analysis and the study of register variation. In: S. Conrad & D. Biber (Eds.) *Variation in English: Multidimensional studies*. Routledge. pp. 3–12.
4. Biber, D. & Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: Cambridge University Press.
5. Paltridge, B. (1994). Genre Analysis and the Identification of Textual Boundaries. *Applied Linguistics*, 15, 288–299.
6. Raible, W. (2019). Variation in Language: How to Characterise Types of Texts and Communication Strategies between Orality and Scriptuality. Answers given by Koch / Oesterreicher and by Biber. *International Journal of Language and Linguistics*, 6(2). <https://doi.org/10.30845/ijll.v6n2p19>

7. Hyland, K. (2004). *Disciplinary Discourses: Social Interactions in Academic Writing*. Michigan: University of Michigan Press.
8. Kupriyanov, R.V., Solnyshkina, M.I. & Lekhnitskaya, P.A. (2023). Parametric Taxonomy of Educational Texts. *Science Journal of Volgograd State University. Linguistics*, 22(6), 80–94. <https://doi.org/10.15688/jvolsu2.2023.6.6> (In Russ.).
Куприянов Р.В., Солнышкина М.И., Лехницкая П.А. Параметрическая таксономия учебных текстов // Вестник Волгоградского государственного университета. Серия 2. Языкознание. 2023. Т. 22. № 6. С. 80–94. <https://doi.org/10.15688/jvolsu2.2023.6.6>
9. Gatiyatullina, G., Solnyshkina, M., Solovyev, V., Danilov, A., Martynova, E. & Yarmakeev, I. (2020). Computing Russian Morphological distribution patterns using RusAC Online Server Proceedings. *International Conference on Developments in eSystems Engineering, DeSE 9450753*. pp. 393–398. <https://doi.org/10.1109/DeSE51703.2020.9450753>
10. Paraschiv, A., Dascalu, M. & Solnyshkina, M.I. (2023). Classification of Russian textbooks by grade level and topic using ReaderBench. Research Result. *Theoretical and Applied Linguistics*, 9(1), 50–63. <https://doi.org/110.18413/2313-8912-2023-9-1-0-4>
11. Solnyshkina, M.I., Kupriyanov, R.V. & Shoeva, G.N. (2024). Linguistic profiling of text genres: adventure stories vs. textbooks. Research Result. *Theoretical and Applied Linguistics*, 10(1), 15–132. <https://doi.org/10.18413/2313-8912-2024-10-1-0-7>
12. Swales, J.M. (1990). *Genre Analysis English in Academic and Research Settings*. Cambridge: Cambridge University Press.
13. Lüdeling, A. & Kytö, M. (2009). *Corpus Linguistics. An International Handbook* (HSK 29.1 und 29.2). Berlin, New York: Mouton de Gruyter. <https://doi.org/10.1515/zrs-2012-0019>
14. Biber, D. & Conrad, S. (2019). *Register, Genre, and Style*. Cambridge: Cambridge University Press.
15. Alekseeva, L.M., Annushkin, V.I. & Bazhenova, E.A. (2003). *Stilisticheskii entsiklopedicheskii slovar' russkogo yazyka*. Moscow: Nauka: Flinta. (In Russ.).
Алексеева Л.М., Аннушкин В.И., Баженова Е.А. Стилистический энциклопедический словарь русского языка. М.: Наука: Флинта, 2003.
16. Kuznetsova, J. (2015). *Linguistic profiles: going from form to meaning via statistics*. New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110361858>
17. Yakhibbaeva, L.M. (2008). Uchebnyi tekst kak osobyi vid vtorichnogo teksta i sostavlyayushchaya uchebnogo diskursa. *Vestnik Bashkirskogo universiteta*, 13(4), 1029–1031.
18. Vedyakova, N.A. (2016). Uchebnyi tekst — nauchnyi tekst? *Lingua mobilis*, 1(54), 19–26.
19. Plavén-Sigray, P. & Matheson, G.J., Schiffler, B.Ch. & Thompson, W.H. (2017). *Research: The readability of scientific texts is decreasing over time eLife*. URL: <https://elifesciences.org/articles/27725#cite-this-article> (accessed: 12.01.2024). <https://doi.org/10.7554/eLife.27725>
20. Ember, M. & Ember, C. (1999). Cross-Language Predictors of Consonant-Vowel Syllables. *American Anthropologist*, 101(4), 730–742.
21. Lyons, C. (1986). The syntax of English genitive constructions. *Journal of Linguistics*, 22(1), 123–143. <https://doi.org/10.1017/S0022226700010586>
22. Tang, J. (2024). Variation in metadiscourse verb patterns in English academic papers from intra- and interdisciplinary analysis. *Applied Mathematics and Nonlinear Sciences*, 9(1).
23. Hodošček, B. (2011). Word Class Ratios and Genres in Written Japanese: Revisiting the Modifier Verb Ratio. *Acta Linguistica Asiatica*, 1(2).
24. Biber, D., Conrad, S. & Reppen, R. (1999). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
25. Sharoff, S., Umanskaya, E. & Wilson, J. (2013). *A Frequency Dictionary of Russian: Core vocabulary for learners*. Routledge.
26. Van Dijk, T.A. (1988). *News as Discourse*. Lawrence Erlbaum Associates, Inc.

27. Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
28. Treffers-Daller, J., Parslow, P. & Williams Sh. (2018). Back to Basics: How Measures of Lexical Diversity Can Help Discriminate between CEFR Levels. *Applied Linguistics*, 39(3), 302–327. <https://doi.org/10.1093/applin/amw009>
29. Yang, J.S., Rosvold, C. & Bernstein, R.N. (2022). Measurement of Lexical Diversity in Children’s Spoken Language: Computational and Conceptual Considerations. *Front. Psychol*, 13, 905789. <https://doi.org/10.3389/fpsyg.2022.905789>
30. Foltz, P.W., Kintsch, W. & Landauer, T.K. (1998). The Measurement of Textual Coherence with Latent Semantic Analysis. *Discourse Processes*, 25 (2–3), 285–307. <https://doi.org/10.1080/01638539809545029>
31. McNamara, D.S., Louwse, M.M., McCarthy, P.M. & Graesser, A.C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47, 292–330.
32. Crossley, S.A. & McNamara, D.S. (2011). Text cohesion and judgments of essay quality: Models of quality and coherence. *Cognition and Instruction*, 29(6), 569–589.
33. Marza, L.E. (2011). A comprehensive corpus-based study of the use of evaluative adjectives in promotional hotel websites. *Odisea*, 12, 97–123.

Information about the authors:

Marina I. Solnyshkina, Ds.Dc. (Philology), Head and Chief Researcher, Text Analytics Research Laboratory, Professor of the Department of Theory and Practice of Teaching Foreign Languages, Institute of Philology and Intercultural Communication, Kazan Federal University (18, Kremlevskaya str., Kazan, Russian Federation, 420008); *Research interests*: applied linguistics, quantitative linguistics, psycholinguistics, semantics; *e-mail*: mesoln@yandex.ru
ORCID: 0000-0003-1885-3039; Scopus Author ID: 56429529500; ResearcherID: E-3863-2015; SPIN-code: 6480-1830; AuthorID: 440177.

Gulnoza N. Shoeva, PhD student of the Department of Theory and Practice of Teaching Foreign Languages, Text Analytics Research Laboratory, Institute of Philology and Intercultural Communication, Kazan Federal University (18, Kremlevskaya str., Kazan, Russian Federation, 420008); *Research interests*: applied linguistics, quantitative linguistics, psycholinguistics, semantics; *e-mail*: gnshoeva@yandex.ru
ORCID: 0009-0005-0438-0404.

Ksenia O. Kosova, PhD student of the Department of Foreign Languages, Faculty of Philology, RUDN University (6, Miklukho-Maklaya str., Moscow, Russian Federation, 117198); *Research interests*: applied linguistics, quantitative linguistics, psycholinguistics, semantics; *e-mail*: kosova-ko@rudn.ru
ORCID: 0009-0007-5606-9604; SPIN-code: 2675-2106; AuthorID: 1200147.

Сведения об авторах:

Солнышкина Марина Ивановна, доктор филологических наук, профессор, профессор кафедры теории и практики преподавания иностранных языков, руководитель НИЛ «Текстовая аналитика», Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет (420008, Российская Федерация, г. Казань, ул. Кремлевская, 18); *сфера научных интересов*: прикладная лингвистика, квантитативная лингвистика, семантика; *e-mail*: mesoln@yandex.ru
ORCID: 0000-0003-1885-3039; Scopus Author ID: 56429529500; ResearcherID: E-3863-2015; SPIN-код: 6480-1830; AuthorID: 440177.

Шоева Гулноза Нармахмадовна, аспирант кафедры теории и практики преподавания иностранных языков, младший научный сотрудник, НИЛ «Текстовая аналитика», Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет (420008, Российская Федерация, г. Казань, ул. Кремлевская, 18); *сфера научных интересов*: прикладная лингвистика, квантитативная лингвистика, семантика; *e-mail*: gnshoeva@yandex.ru
ORCID: 0009-0005-0438-0404.

Косова Ксения Олеговна, аспирант кафедры иностранных языков филологического факультета, Российский университет дружбы народов (117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6); *сфера научных интересов*: прикладная лингвистика, квантитативная лингвистика, семантика; *e-mail*: kosova-ko@rudn.ru
ORCID: 0009-0007-5606-9604; SPIN-код: 2675-2106; AuthorID: 1200147.