Computing Descriptive Metrics and Propositions in Reading Texts and Recalls

Mariia Andreeva[1,2][0000-0002-5760-093], Marina Solnyshkina[2][0000-0003-1885-3039], Valery Solovyev[2,3][0000-0003-4692-2564], Artem Zaikin[4][0000-0002-5596-3176], Olga Bukach[2][0000-0001-8675-2243]

[1] Kazan State Medical University, 420012, Butlerova, 49, Kazan, Russia
[2] Kazan Federal University, Department of theory and practice of language teaching, Research laboratory 'Intellectual technologies for text management', 420008, Kremlyovskaya, 18, Kazan, Russia
[3] Linguistic research and education center, 420008, Kremlyovskaya, 18, Kazan, Russia
[4] Kazan Federal University, Laboratory of modern geoinformation and geophysical technologies, Department of mathematical statistics and information technologies, 420008, Kremlyovskaya, 18, Kazan, Russia
mesoln@yandex.ru

Abstract. This study addresses the problem of computational techniques to perform a multi-factor text analysis aimed at assessing text metrics and the amount of information in two contrasting texts. Assessing recalls in general and estimating the scope of information reproduced in recalls in particular are equally challenging. We introduce a computational linguistic tool that measures 28 linguistic parameters enabling conventional level of language assessment. The results are indicative of the tool distinguishing two versions of the low (OT51) versus high (MR51) cohesion of the texts but not the recalls. The results also showed that ubiquitously used descriptive metrics and readability indices inappropriately distinguish between reading texts and their recalls. Overall, the research advances our understanding of the relationship between conventional (quantitative, lexical) and semantic metrics providing foundations for more effective algorithms of assessing and profiling academic texts types.

Keywords: Qualitative Analysis, Descriptive Metrics, Cohesion, Multi-factor Text Analysis.

1 Introduction

Educators view instruments performing automated multi-factor text analysis of students' work as tools allowing to simplify a primary task of language assessment. A good text analyzer is expected to measure a wide range of conventional (descriptive, morphological, and lexical) and text-level (semantic) features. The latter include measurements which assess e.g. text cohesion, narrativity, informativeness, etc. Researchers all over the world aim at designing tools of this kind to enable users to se-

lect appropriate texts for different categories of readers on the one hand and assess students' work on the other. Another challenging area of research is an educational text pattern appropriate for certain categories of readers which can be presented as a set of conventional and text-level features.

In this paper we describe a two-stage comparative analysis of original high- and low cohesive versions of an educational text and their recalls. The research was designed in the following stages:

Stage 1. Psycholinguistic experiment in which two groups of students read one of the two versions of an educational text, an original, retrieved from the textbook on Social science, or a manipulated, high-cohesion version of the same text.

Stage 2. Conventional metrics analysis of the reading texts and texts of recalls conducted with the help of the online automated text analyzer TAR.

Stage 3. Text-level (semantic) analysis based on the assessment of text propositions.

2 Related work

Researchers distinguish various features influencing text comprehension. Among the validated parameters are readability indices, vocabulary knowledge, words frequency, abstractness and lexical diversity [1].

Text readability is predominately estimated with two metrics, i.e. average word length and average sentence length [2]. These are the key elements of the assumed notion of readability expected to change across grades from elementary to high school. Flash-Kincaid Grade level measured with the two parameters is expected to inform users of the appropriateness of the target audience for a reading text.

Text lexical parameters which we analyze in this article include word frequency, abstractness and lexical diversity.

A reader's vocabulary knowledge and lexical coverage relate to the amount of exposure the reader has received on words and as such they are viewed as good predictors of reading comprehension [3–8]. Typically, while assessing text difficulty researchers resort to one of the two techniques: (1) they either assess the number of 'difficult' words which a particular reader is unlikely to know or (2) compute ta text for word frequency and divide it into groups of words based on their frequency. In foreign languages studies, words are classified based on language proficiency levels (A1-C2, CEFR) [9–11], while vocabulary of texts for native speakers is assessed with the help of frequency lists [12, 13]. Another parameter close to the above, i.e. word frequency is regarded as highly indicative of word difficulty and correlates with numerous contributing factors [14]. Klare (1968) argues that frequency of words affects both the ease of reading and its comprehension. Frequency of words is strongly associated with two types of 'difficulty': the so-called 'perceived difficulty' and 'actual difficulty'. The first refers to the formal, external view of a word, while the second is related to users' ability to define or select the correct definition of the word among distracters [15]. High-frequency words are proved to be more easily perceived [16] and readily produced by readers [17]. High-frequency words are both perceived and

produced more quickly and more efficiently than low-frequency words [18–22], resulting in more efficient comprehension of the text [23].

Word frequency in Russian texts is assessed with the help of the data in S. Sharoff and O. Lyashevskaya Dictionary (2009) [24] which provides frequency ratings of 60 000 words of different types of discourse [25].

Lexical diversity or Type Token Ratio is the number of types divided by tokens in the text [26]. Type token ratio is low in case many words are repeated within one text or corpus. A high TTR suggests that a text or a corpus uses more diverse vocabulary. P. Baker claims that the larger is the corpus, the more likely it to have lower TTR due to the use of high frequency grammatical forms (e.g. articles and particles) [27]. M. Shermis et. al (2013) use TTR to assess students' essays. The authors argue that 'higher ratios indicate that more concepts are introduced in a given syntactic role, whereas lower ratios mean fewer concepts' [28].

The notion of cohesion was introduced by Halliday (1976) and is viewed as a device for connecting different parts of a text. It is achieved through lexical means, coreference, ellipsis and conjunctions [29]. Over the last 20 years, research into text comprehension and memorization has grown rapidly [30-33].

Based on experimental data of science and narrative texts comprehension researchers provide evidence that both college students and children face difficulties while reading low cohesion texts [34–37]. Texts with manually increased referential cohesion (primarily, by content word overlap) were reported by numerous researchers to be beneficial for comprehension and reasoning processes as compared to texts with low referential cohesion [38–40].

In this article based on the semantic roles defined as semantic frames [41] we also perform a propositional analysis of texts to compare 'newness and givenness of information' in the texts of recalls. The semantic roles are used in the paper as labels applied to the arguments of verbs to identify their roles in the events denoted by verbs. Sets of semantic roles vary in different studies and range from specific to general: Agent, Experiencer, Instrument, Object, Source, Goal, Location, Time, and Path [42].

3 Materials, tools and methods

The research data collected for the study comprises two reading texts (OT51 and MT51) described below and 65 texts of students' recall. The total size of the texts studied is over 7000 words. We computed the two reading texts with an automated tool, Text Evaluator of Russian texts (TAR) [43]) designed and developed by the authors (see below for more detailed description) to evaluate the following metrics:

1. Descriptive metrics: 1.1. words count (WC), 1.2. syllables count (SylC), 1.3. sentences count (SC), 1.4. average sentence length (ASL), 1.5. average word length (AWL)).
2. Flesch-Kincaid Grade Level (FKGL) [Kincaid, 1975], assessed with 2.1. Oborneva formula [44] and 2.2. SIS formula [45].

3. Morphological features (part-of-speech count: 3.1. adjectives (Adj), 3.2. adverbs (Adv), 3.3. pronouns (Pron), 3.4. nouns (N), and 3.5. verbs (V), 3.6. noun cases, 3.7. verb tenses.
4. Lexical features: 4.1. Type and token ration (TTR), 4.2. Word frequency (Freq.,), 4.3. Abstractness / Concreteness (A/C).

The processed text data are downloadable in spreadsheets (see Table 1).

Table 1. Text descriptive metrics (fragment)

	Metric	Value
1	Word Count	82
2	Syllables Count	250
3	Sentence Count	6
4	Average sentence length	13.666666666666666
5	Average word length	3.05
6	Adjectives Count	10
7	Adverbs Count	2
8	Pronouns Count	11
9	Nouns Count	43
10	Verbs Count	11

TAR, a Python-based tool estimates values of 28 features in Russian texts. Uploaded texts are processed with MyStem.3.0, a PoS tagger, developed by Yandex [46] which removes stop words and tags content words with the corresponding morphological categories: noun cases and gender, verb tense forms, etc. The tool stores texts in XML cue files as hierarchical structures in which each word is PoS-tagged (Fig. 1).

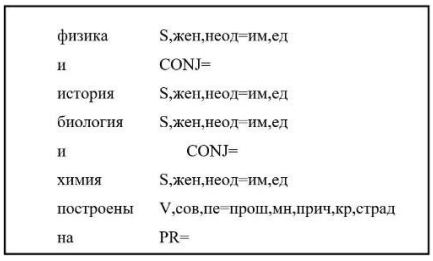

Fig. 1. PoS tagging in TAR

TAR is also supplied with a Stemmer and Lemmatizer, thus providing users with lists of stems and lemmas of the uploaded texts (Fig. 2).

ты наверн соглас с так вывод искусств обязательн треб от человек умен твор создава прекрасн произведен котор поража бы друг сво новизн и необычн легенд рассказыва что жил когда-т мастер нестор котор постро без един гвозд удивительн красот деревя церков преображен на остров киж в онежск озер в начал век когд росс утвержда на берег балтийск мор и станов морск держав мастер возвел праздничн жизнерадостн ни на что не похож двадцатидвухглав

Fig. 2. TAR Stemming (fragment)

Text readability indices are estimated based on modified for the Russian language FKGL formulas: 1. FKGL (O) designed by I.V. Oborneva (2006) for fiction texts: FKGL (O) = 206.836 – (1.52 x ASL) – (65.14 x AWL). 2. FKGL (SIS) for academic texts: FKGL (SIS) $0.36 \times ASL + 5.76 \times ASW - 11.97$ [45].

TAR also computes TTR, word frequency and text abstractness (see Part 4). The number of propositions and subpropositions were manually assessed.

4 Research design

Stage 1. Psycholinguistic experiment: reading and recalling.
The respondents (n 177), 5th graders, Russian natives aged 11–12, were offered to read one of the two versions of a text: (1) an original unabridged 200-word text from "Social Science 5" (Text OT51) or (2) a manually manipulated version of OT51 (further referred to as MT51) which contains words and ideas that overlap across the adjacent sentences and the entire text, with explicit threads that connect parts of the texts for the reader.

The assessment of text comprehension was preceded by Wechsler General Knowledge Subtest for children (WISC GK) [47] and streaming the respondents into two sections. Of the 118 respondents participating in WISC GK we selected 65 with the average GK index, i.e. 13–16. 34 students read and recalled the original, unchanged version of the text retrieved from the textbook, OT51, 31 subjects read and recalled MT51, the modified version of the original text. The subjects' reading comprehension was tested by recalls.

Stage 1a. Manipulations of text OT51 included the following: sentence splitting, adding a topic sentence, adding temporal markers and demonstrative pronouns (a–d in Fig. 3).

We increased the number of sentences from 11 (OT51), to 18 (MT51)) (see Table 2). We also increased text cohesion by the following (see Fig. 4):

(a) adding the topic sentence: '*This text focuses on works of art and culture'* (MT51.1) and paragraphing which involved splitting three original paragraphs of OT51 into six paragraphs in MT51;

(b) adding a temporal marker to connect two adjacent sentences '*In the early 18th century, when Russia was becoming a sea power, the Master erected a festive, cheer-*

ful church' (OT51.3) → '*This happened in the early 18th century, when Russia was becoming a sea power', 'It was at this time that Master Nestor built a festive, cheerful church*' (OT51.6, OT51.7);

(c) adding demonstrative pronouns to specify the referents: '*works of art* (OT51.1) → *these works of art* (OT51.3), *church* (OT51.3) → *this church'* (OT51.8));

(d) introductory '*for example'* to exemplify the arguments;

(e) content words overlaps (CWO) and anaphoric replacements: '*works*' (MT51.2/3), '*Master (Nestor)'* (MT51.5/7/13), '*church*' (MT51.8/9/10/12).

The flow chart of the Text manipulations is presented in Figure 3 below.

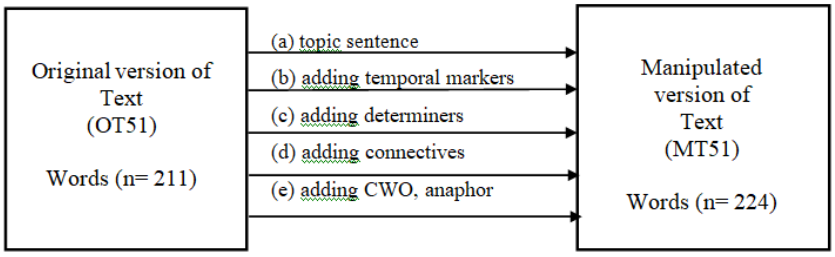

Fig. 3. Manipulations of OT51

Texts OT51 and MT51 are presented in Table 2.

Table 2. OT51 Sentence Splitting (Fragment)

S-ce #	OT51	S-ce #	MT51
OT51.2	The legend says that once upon a time there lived Master Nestor, who built an amazingly beautiful wooden church of Transfiguration on Kizhi island in Onega lake without any nail at all[1].	MT51.4	There is such a legend.
		MT51.5	Once upon a time there lived Master Nestor, who built an amazingly beautiful wooden church of Transfiguration on Kizhi island in Onega lake without any nail at all.
OT51.3	In the early 18th century, when Russia was settling in the Baltic sea and becoming a sea power, the Master built a 22-domed festive, cheerful church, which was different from any other.	MT51.6	This happened in the early 18th century, when Russia was becoming a sea power.
		MT51.7	It was at this time that Master Nestor built a festive, cheerful church.
		MT51.8	This 22-domed church was different from any other.

4.1 The parameters assessed and computed for OT51 and MT51

Stage 2. Conventional metrics analysis of the reading texts and texts of recalls conducted with the help of the online automated text analyzer TAR.

Elaboration of OT51 comprised extension of its length (211 words → 224 words), which manifests itself in more nouns (62 → 71) and verbs (43 → 46). The number of

[1] The original Russian text was translated into English by the authors of the article. In translation we mostly aimed at word for word translation to demonstrate the performed syntactic and lexical manipulations.

sentences rose from 11 in OT51 to 18 in MT51, which caused a dramatic decrease in a number of words per sentence: from 17,42 in OT51 to 12.33 in MT51. The descriptive metrics of texts OT51 and MT51 are presented in Table 3.

Table 3. Quantitative metrics of Original and Manipulated Texts

	Metrics	OT51 Text	MT51 Text
Descript.	Words count (WC)	211	224
	Syllables count (SylC)	506	528
	Sentence count (SC)	12	18
	Average sentence length (ASL)	17.42	12.33
	Average word length (AWL)	2.42	2.38
	FKGL (SIS)	8.25	6.17
	FKGL (O)	13.46	10.56
Morphol.	Adjectives count (Adj)	30	26
	Adverbs count (Adv)	5	5
	Pronouns count (Pron)	20	23
	Nouns count (N)	62	71
	Verbs count (V)	42	46
Lex.	Word frequency (Freq)	170.10	117.8
	Abstractness (A/C)	-1.48	-1.41
	TTR	0.78	0.76

The table shows significant increase in descriptive metrics, such as word count (211 (OT51) – 224 (MT51)) and syllables count respectively (506 (OT51) – 528 (MT51)). As the number of sentences increased from 12 in OT51 to 18 in MT51, ASL and AWL decreased. Parts of speech count within the morphological metrics is also increased in MT51.

4.2 Experiment

To collect data and find differences in quantitative metrics of reading texts and recalls we designed a three-stage experiment (see Fig. 4):

1. We conducted the Russian version of Wechsler Intelligence Scale test for Children (WISC) aimed at selecting a representative sampling of respondents with average test results. WISC comprises 27 close-ended questions and estimates participants' general rather than topic-specific or theoretical knowledge as well as strengths and weaknesses associated with working memory, processing speed, and long-term memory [48]. The results of the statistical analysis of WISK aimed at selecting respondents with similar General knowledge index are presented in Fig. 6. Of 177 native Russians, 11-year old 5th graders, we selected 64 with the average GK of 13–17 (see Fig. 5) and streamed them into two sections: 33 subjects in Section OT51 and 31 subjects in Section MT51.

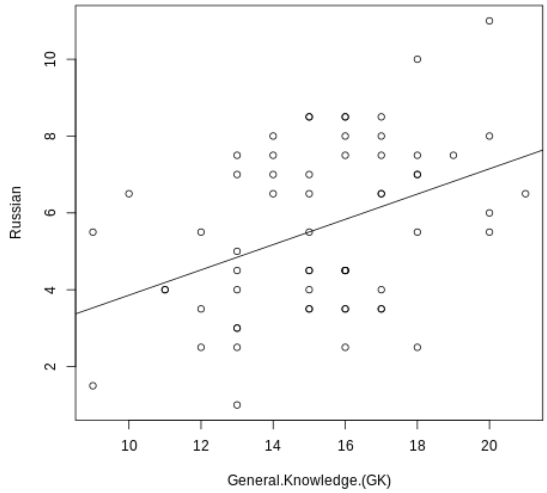

Fig. 4. WISC performance results

2. Each subject read and recalled a text (either OT51 or MT51) to an individual expert.
3. The recalls were recorded and later transcribed (see Fig. 5).

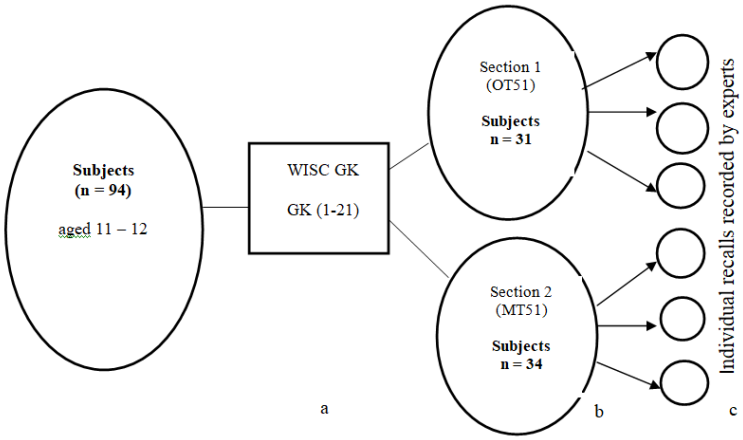

Fig. 5. Design of the experiment

The transcribed recalls of OT51 and MT51 are comprised in the Corpus of Transcripts with the total size of 7282 tokens (Table 4).

Table 4. Corpus of Transcripts

Tokens	OT51	MT51
Original texts	211	224
Recalls	3107	3740
Total	3318	3964

The Corpus of over 30 samplings of recalls of each text is viewed as representative to conclude on the results of the experiment [49].

5 Analysis

Contrastive analysis of the reading texts (OT51, MT51) and participants' recalls was conducted based on two groups of parameters: (a) descriptive metrics and lexical features computed with TAR and (b) propositional analysis conducted manually.

Initially we computed the recalls with the help of TAR to measure the following metrics: word count (W), syllable count (SylC), sentence count (S), average sentence length (ASL), average word length (AWL (in syllables), FKGL (SIS), FKGL (O), adjective count (ADJ), adverb count (ADV), pronoun count (Pron), noun count (N), verb count (V), word frequency (Freq), Abstractness / Concreteness (A/C), type-token ratio (TTR) (see Table 5).

On the next stage of the research we had to exclude the majority of the metrics estimated with TAR from the analysis based on the following observations: a) length of the recalls was fewer than 200 words, thus insufficient to measure text readability; b) due to differences in the length of pauses, sentence length was not always correctly estimated in transcripts either and therefore caused unreliable values of sentence length; c) numerous repetitions of the same word in recalls while respondents were hesitating or formulating thoughts makes computing the number of tokens in recalls useless.

Table 5. Descriptive Parameters of MT51 and OT51 Recalls

		MT51 Recalls		**OT51 Recalls**	
	Metrics / Subject	K5P22	K5B21	K185A09	K185A03
Descript.	WC	49	128	174	109
	SylC	110	299	364	211
	SC	7	12	13	7
	ASL	7	10.6	13.3	15.5
	AWL	2.24	2.34	2.09	1.94
	FKGL (SIS)	3.48	5.33	4.9	4.79
	FKGL (O)	6.77	9.37	8.67	8.46
Morphol.	Adj	4	12	24	10
	Adv	2	2	10	7
	Pron	4	13	14	11
	N	13	36	38	20

Lex.	V	14	31	28	16
	Freq	297.24	83.71	101.76	161.31
	A/C	-1.28	-1.59	-0.88	-1.28
	TTR	0.78	0.80	0.61	0.72

Stage 3. Text-level (semantic) analysis based on the assessment of text propositions. As TAR is not yet programmed to compute semantic metrics or idea units of a text, propositional modeling was performed manually. To assess the information scope reproduced by respondents in recalls we applied a quantitative structural approach and estimated the number of propositions in recalls to compare it with those in the original text. Thus, the information in each sentence was broken into main propositions (P) comprising the main idea and sub-propositions or arguments identifying their roles in the events denoted by main propositions. The taxonomy of the sub-propositions estimated includes the following: (1) Actant (Act) which comprises Agent (Ag)), Experiencer (Exp); Object (Obj)), Source (Sour), Goal (Gl) and Instrument (Instr); (2) Locative (Loc), Time (Temp), and Path (Pth) [50, 51]. E.g., the sentence 'In the early 18th century the Master erected a 22-domed festive, cheerful church' (OT51.3) receives the following code: 'Time, Act (Ag) P1 (Verb) Mod1, Mod2, Mod3 (Obj).

The obtained 65 recall files (34 for OT51, 31 for MT51) were contrasted based on the number of propositions reproduced. Two shortest recalls (K5A14) had only 11 words forming seven total propositions. The longest recall (K5P21) was organized in a 214-word text with 114 total propositions. An average MT51 recall consisted of 58 total propositions, whereas the average OT51 recall comprised 43 total propositions.

Based on the data received we estimated the average number of each type of propositions reproduced by respondents for MT51 and OT51. To contrast difference in recall results we visualized the normalized recall results in boxplots in Fig. 6–8 separately for the total number of propositions, main propositions and sub-propositions.

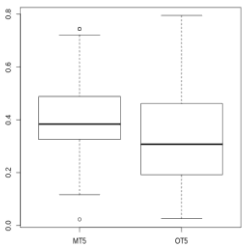

Fig. 6. MT51 and OT51 recall main propositions performance

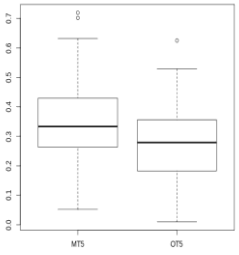

Fig. 7. MT51 and OT51 recall sub propositions performance

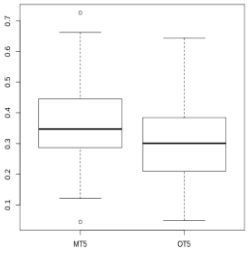

Fig. 8. MT51 and OT51 recall total propositions performance

We also conducted a set of statistical tests and implemented the binomial model to reveal differences in the total number of propositions, main propositions and sub-propositions separately for recalls of OT51 and MT51 [52]. The obtained p-values suggest better results in recalls of text MT51 (See Table 6 and 7) manipulated with the purpose to increase its cohesion.

Table 6. Comparison of P-values of Propositions MT51

	Recalls MP. MT51	Recalls SP.MT51	Recalls. TP. MT51
Recalls. MP. MT51	1	0,84	0,91
Recalls. SP. MT51	0,84	1	0,99
Recalls. TP. MT51	0,91	0,99	1

Table 7. Comparison of P-values of Propositions OT51

	Recalls MP. OT51	Recalls SP.OT51	Recalls. TP. OT51
Recalls. MP. OT51	1	0,88	0,94
Recalls. SP. OT51	0,88	1	0,99
Recalls. TP. OT51	0,94	0,99	1

Low cohesive text OT51 proves to present a higher degree of complexity for readers and therefore and caused worse performance.

6 Conclusion

To pursue a contrastive analysis of low (OT51) and high (MT51) cohesive version of the reading text with texts of recalls, we computed the following descriptive quantitative metrics: sentence count, word count, average sentence length, average word length in syllables, noun count, verb count, adjective count, adverbs count, pronouns count, propositions count. Experiments with descriptive and lexical text metrics measured with TAR demonstrated that they do not suffice to comprehensively describe the quality of recalls. The metrics measured with the tool prove insufficient to assess the quality and scope of the information in the recalls based on their small length and numerous repetitions elevating assessment results. The feature discriminating the scope of the information reproduced and as such the quality of recalls is proposition count. Judged by the lower number of propositions reproduced (mean OT51 – 39,7, MT51 – 52,7), the low cohesion original text (OT51) proved to be more difficult the for the subjects.

The study confirms that cohesion is a text parameter able to improve comprehension and as such increase the scope of the reproduced in recalls information. We also suggest implementing propositional analysis to compare original reading texts and recalls. The results of the current study are relevant in modern Russia to assess texts of classroom books for cohesion and the scope of the information recalled. Based on the analysis conducted, the research offers a guide to design a more sophisticate tool

which could discriminate propositions and semantic roles in texts. The study also provides useful information for researchers, educators and can be of assistance for textbook writers and test developers.

Acknowledgement

The research was financially supported by the Russian Science Foundation, grant 18-18-00436.

References

1. Solnyshkina, M. I., Harkova, E. V., Kazachkova, M. B.: The Structure of Cross-Linguistic Differences: Meaning and Context of 'Readability' and its Russian Equivalent 'Chitabelnost'. Journal of Language & Education, 6 (1), 103–119 (2020).
2. DuBay, W. H.: The Principles of Readability. Online Submission (2004).
3. Bernhardt, E., Kamil M.: Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. Applied Linguistics, 16(1), 15–34 (1995).
4. Laufer, B.: How much lexis is necessary for reading comprehension? In: H. Bejoint and P. Arnaud, editors, Vocabulary and applied linguistics, 126–132. Macmillan, Basingstoke & London (1992).
5. Nation, P.: Learning vocabulary in another language. Cambridge University Press, Cambridge (2001).
6. Nation, P.: How large a vocabulary is needed for reading and listening? The Canadian Modern Language Review, 63(1), 59–82 (2006).
7. Qian, D.: Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. The Canadian Modern Language Review, 56(2), 282–308 (1999).
8. Ulijn, J., Strother, J.: The effect of syntactic simplification on reading EST texts as L1 and L2. Journal of Research in Reading, 13(1), 38–54 (1990).
9. Carrell, P. L.: Second language reading: Reading ability or language proficiency? Applied linguistics, 12(2), 159–179 (1991).
10. Zareva, A.: Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. System, 33(4), 547–562 (2005).
11. Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., Dobbs, C. L.: Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. Reading Research Quarterly, 50 (3), 337–356 (2015).
12. Lyashevskaya, O. N. Sharov, S. A.: Chastotnyj slovar' sovremennogo russkogo yazyka (na materialah Nacional'nogo korpusa russkogo yazyka) [Frequency dictionary of modern Russian language (based on Russian national corpora)]. M., Azbukovnik (2009).
13. Feng, G., Hao, W., Davies, M.: The Graded Vocabulary of Contemporary English: 5,000 Upper-intermediate Words. Kindle Edition, 579 pages. Published November 1st 2016 by Transnational Academic Exchange Service (TrAES.info) (2016).
14. Ryder, R. J., Slater, W. H.: The relationship between word frequency and word knowledge. The Journal of Educational Research, 81(5), 312–317 (1988).
15. Leroy, G., Kauchak, D.: The effect of word familiarity on actual and perceived text difficulty. Journal of the American Medical Informatics Association, 21(e1), 1–4 (2014).

16. Bricker, P., Chapanis, A.: Do incorrectly perceived stimuli convey some information? Psychological Review, 60(3), 181–188 (1953).
17. Haseley, L.: The relationship between cuevalue of words and their frequency of prior occurrence. Unpublished master's thesis, Ohio university (1957).
18. Balota, D., Chumbley, J.: Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception & Performance, 10(3), 340–357 (1984).
19. Howes, D., Solomon, R.: Visual duration threshold as a function of word-probability. Journal of Experimental Psychology, 41(6), 401–410 (1951).
20. Jescheniak, J., Levelt, W.: Word ¨ frequency effects in speech production: Retrieval of syntactic information and of phonological form. Journal of Experimental Psychology: Learning, Memory, & Cognition, 20(4), 824–843 (1994).
21. Monsell, S., Doyle, M., Haggard, P. N.: Effects of frequency on visual word recognition tasks: Where are they? Journal of Experimental Psychology: General, 118(1), 43–71 (1989).
22. Rayner, K., Duffy, S. A.: Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition, 14(3), 191–201 (1986).
23. Klare, G. R.: The role of word frequency in readability. Elementary English, 45(1), 12–22 (1968).
24. Lyashevskaya, O. N., Sharov, S. A.: Chastotnyj slovar' nacional'nogo korpusa russkogo yazyka: koncepciya i tekhnologiya sozdaniya. [Frequency dictionary of Russian national corpora: concept and technology of compiling)], http://www.dialog-21.ru/digests/dialog2008/materials/html/53. htm, last accessed 2020/10/30.
25. Word frequency data, https://www.wordfrequency.info/, last accessed 2020/10/29.
26. Templin, M.: Certain language skills in children: Their development and interrelationships. Minneapolis: The University of Minnesota Press (1957).
27. Baker, P.: Using corpora in discourse analysis. A&C Black (2006).
28. Shermis, M. D., Burstein, J.: Handbook of automated essay evaluation: Current applications and new directions. Routledge (2013).
29. Maybury, M.: Advances in automatic text summarization. MIT press (1999).
30. Pratt, M. W., Luszcz, M. A., MacKenzie-Keating, S., Manning, A.: Thinking about stories: The story schema in metacognition. Journal of Verbal Learning & Verbal Behavior, 21, 493–505 (1982).
31. Shaughnessy, J. J.: Confidence-judgment accuracy as a predictor of test performance. Journal of Research in Personality, 13, 505–514 (1979).
32. Wearn, Y., Askwall, S.: On some sources of metacomprehension. Scandinavian Journal of Psychology, 22, 17–25 (1981).
33. Lefèvre, N., Lories, G.: Text cohesion and metacomprehension: Immediate and delayed judgments. Memory & cognition, 32(8), 1238–1254 (2004).
34. McNamara, D. S., Kintsch, W.: Learning from text: Effects of prior knowledge and text coherence. Discourse Processes, 22, 247–287 (1996).
35. Ozuru, Y., Briner, S., Best, R., McNamara, D. S.: Contributions of self-explanation to comprehension of high-and low-cohesion texts. Discourse Processes, 47(8), 641–667 (2010).
36. Ozuru, Y., Dempsey, K., McNamara, D. S.: Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. Learning and instruction, 19(3), 228–242 (2009).

37. Cain, K., Nash, H. M.: The influence of connectives on young readers' processing and comprehension of text. Journal of Educational Psychology, 103(2), 429 p. (2011).
38. Solnyshkina, M. I., Solov'ev, V. D., Andreeva, M. I., Danilov, A. V.: Vliyanie svyaznosti teksta na ego vospriyatie: eksperimental'nyj podhod. In: I.A. Boduen de Kurtene i mirovaya lingvistika [Influence of text cohesion on its comprehension: experimental approach], 212–216 (2019).
39. Bouquet, P., Warglien, M.: Mental models and local models semantics: the problem of information integration. In: Proceedings of the European Conference on Cognitive Science, 169–178. University of Siena Italy (1999).
40. Walsh, C. R., Johnson-Laird, P. N.: Co-reference and reasoning. Memory & Cognition, 32(1), 96–106 (2004).
41. Fillmore, C. J.: Frame semantics and the nature of language. In: Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech, 280, 20–32. New York Academy of Sciences, New York (1976).
42. Fillmore, Charles J.: Some problems for case grammar. In: R. J. O'Brien, editor, 22nd Annual Round Table. Linguistics: Developments of the Sixties – Viewpoints of the Seventies. Volume 24 of Monograph Series on Language and Linguistics, 35–56. Georgetown University Press, Washington, D.C. (1971).
43. TAR text analyser, http://tykau.pythonanywhere.com/, last accessed 2020/10/30.
44. Oborneva, I.V.: Avtomatizirovannaya otsenka slozhnosti uchebnykh tekstov na osnove statisticheskikh parametrov [Automated estimation of complexity of educational texts on the basis of statistical parameters]. Pedagogy Cand. Diss. Moscow (2006).
45. Solovyev, V., Ivanov, V., Solnyshkina, M.: Assessment of reading difficulty levels in Russian academic texts: Approaches and metrics. Journal of intelligent & fuzzy systems, 34(5), 3049–3058 (2018).
46. MyStem, https://yandex.ru/dev/mystem/, last accessed 2020/10/30.
47. Wechsler, D.: Wechsler intelligence scale for children-fifth edition. Bloomington, MN: Pearson (2014).
48. Filimonenko, Y. I., Timofeev, V. I.: Rukovodstvo k metodike issledovaniya intellekta u detei D. Vekslera (WISC) (1994).
49. Biber, D.: Representativeness in corpus design. Literary and linguistic computing, 8(4), 243–257 (1993).
50. Starodumova, E. A.: Sintaksis sovremennogo russkogo yazyka [Russian syntax]. Vladivistok: Far East University Publishing (2005).
51. Tesnière, L.: Éléments de syntaxe structural (1959).
52. Lehmann, E. L.: Testing Statistical Hypotheses. Wiley, 600 p. (1986).