

VII International Forum on Teacher Education

Measuring Text Features in Expository Discourse of Russian Students

Marina Solnyshkina (a), Elzara Gafiyatova* (b), Ekaterina Martynova (c), Andrey Danilov (d)

(a), (b), (c), (d) Kazan Federal University, 420008, Kazan (Russia), 2 Tatarstan street, rg-777@yandex.ru

Abstract

The study is aimed at investigating the potential for original automated Russian text analyzer RuLingva to assess linguistic metrics of written recalls of students of Russian as a Foreign language (RFL). Initially developed by the authors to estimate Russian texts readability indices, the public version of RuLingva reports on 33 metrics related to text length, readability indices, parts of speech classification, noun case, verb tenses, vocabulary frequency, lexical diversity, abstractness rating, etc. We hypothesize that the abovementioned metrics can be used to discriminate and score written expository discourse of RFL students. The corpus compiled for the study comprises written recalls of 407-word expository texts produced by 71 B2 students of Russian. Each subject's recall was scaled against the original reading text on the following metrics: text length, average sentence length, average word length, type token ratio, word frequency, abstractness rating, local and global noun overlap, local and global argument overlap. Prior to reading the expository text, we also assessed the subjects' general knowledge with WISC test and Russian language proficiency with Quick Russian placement Test. T tests of significance indicated a strong positive correlation (>0.05) of both general knowledge and Russian proficiency tests with the abovementioned metrics automatically assessed with RuLingva. The findings enable to narrow the range of text features predicting RFL writing quality and ways of estimating language proficiency. RuLingva as the first Automated Writing Assessment tool for the Russian Language has a potential to be successfully used in formative assessment motivating students to review their writing and contributing to both literal and inferential comprehension.

Keywords: Russian text analyzer RuLingva, written recalls, text metrics, Automated Writing Assessment.

© 2021 Marina Solnyshkina, Elzara Gafiyatova, Ekaterina Martynova, Andrey Danilov

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published by Kazan federal university and peer-reviewed under responsibility of IFTE-2021 (VII International Forum on Teacher Education)

* Corresponding author. E-mail: rg-777@yandex.ru

Introduction

Many studies have been conducted and contributed to our understanding of how readers comprehend foreign language texts and text parameters that influence readers' ability (or inability) to understand texts (Landauer, Laham, & Foltz, 2003). Comprehension is typically evaluated either with a recall or a test which are both time- and effort-consuming techniques (McNamara, Kintsch, Songer, & Kintsch, 1996). To identify and discriminate easier from more difficult input, researchers usually employ propositional analysis and contrast informativeness of discourse in the input (reading or listening text) and the texts produced by readers (written or oral recalls) (Crossley, Kyle, & McNamara, 2016). The first automated tools able to (1) assess a number of linguistic parameters of a piece of writing; (2) compute and contrast differences of an input text with those of written recalls were designed and developed in the late 20th century. The idea behind these tools was to reduce examiner's workload and time spent on evaluating writings (Cotos, 2015). Latest developments in Natural Language Processing (NLP) have enabled much higher quality of automatic analysis of students' writing. Modern algorithms for the work of artificial intelligence are based on methods of analyzing correlations and statistical relationships of the primary, i.e. input, given, and secondary, i.e. new, texts. Despite the recent advance of automated writing evaluation technologies and the increasing pursuits in applying these technologies in EFL, few studies have aimed at the effects of using similar technologies and tools to assess Russian language writing.

This study presents an innovative automated tool RuLingva (<https://rulingva.kpfu.ru/>) employed as an assessment tool to compute metrics of students' written recalls. The latter was later complemented with contrastive and propositional analysis of the texts performed by human raters.

Purpose and objectives of the study

We hypothesize that metrics related to text length, readability indices, parts of speech (POS) classification, noun case discrimination, verb tenses, vocabulary frequency, lexical diversity, abstractness rating can be used to discriminate and score expository discourse of students of Russian as a foreign language (RFL). The combination of assessments, i.e. written recalls and cloze tests, allow researchers to better evaluate students' comprehension and discriminate between testees with similar recalls or test scores. Two measures have the benefit of providing a more detailed assessment of students' comprehension also enabling a tester to validate test specifics and focus on comprehension of different parts in the text. The study was conducted to find an interaction of text comprehension on the one hand and General knowledge and Russian proficiency on the other for adult students of Russian as a Foreign language.

Literature review

The two popular approaches to the problem of students' writing are applicable to different secondary, i.e. produced by students, text genre: the term 'automated writing assessment' (AWA) implies assessment of written recalls performed by students, while 'automated writing evaluation' (AWE) is used in the practice of essay assessment. While AWA is based on comparing a written reproduction performed by a student with the input or original reading, i.e. primary text, AWE is programmed to compare each writing with a large database of essays of the same genre produced to answer a specific prompt (Balfour, 2013; Chapelle, Cotos, & Lee, 2015).

In both AWE and AWA, the tools compute text metrics including word count, tokens (all words in the text) and types (unique words), different parts of speech counts, grammar categories counts, readability indices, syntax parameters (number of clauses, number of words before the main verb, etc.), vocabulary range (academic words count, different professional contexts terms, obsolete and historic words, neologisms, etc.). Few sophisticated tools also measure text cohesion (local, global, referential), narrativity and abstractness (McNamara, Louwse, McCarthy, & Graesser, 2010). The overall score in AWE is suggested based on statistical modeling, while AWA algorithm provides two lists of metrics: one for the input and one for the secondary (generated by students) text (Petrova & Solnyshkina, 2021). The overall score in AWE can be provided with suggestions for students' writing. The most popular instruments designed for ELT students include Criterion, Write & Improve (Write and improve), WriteToLearn, MyAccess!, Writing Power, Writing Roadmap (Cotos, 2015). The tools benefit students, teachers and test designers saving assessment time and generating feedbacks on writing. In AWA, the secondary text is provided with an overall score based on the (1) contrasting analysis of the computed metrics of primary and secondary texts and (2) manual propositional analysis of both texts.

Methodology

General knowledge is known to play an important role in text comprehension (McNamara et al., 2010). The fact that expository text comprehension depends on readers' prior knowledge has been confirmed in a number of research with adults (McNamara et al., 1996). As B1 RFL students experience a transition from narrative to expository texts their knowledge and Russian language proficiency may become critical in comprehension of expository texts. As it was mentioned earlier the overarching goal of the current research is to study the factors that lead to comprehension difficulties among B2 RFL students. Following the goal, we examine the roles of language proficiency and general knowledge among B2 RFL students when exposed to expository texts.

We separately examined the effects of general knowledge and language proficiency on text comprehension. We hypothesized to find significant effects of both language proficiency and general knowledge.

Participants

Students were recruited by sending letters of invitation in which we described the aims and algorithm of the study and requested students to participate. The testing session was conducted on three Wednesday before students' regular classes. Students' participation was voluntarily and they received no bonus or payments for taking part in the study. Participants were tested in groups of 20–25 in the university classroom. The test session lasted approximately 90 minutes. The order of the tasks performed by the students was as follows: (1) General knowledge test, (2) Quick Russian test, (3) text reading (two times), (4) writing text recall, (5) multiple-choice text-based test. All the stages of the experiment were performed on University computers. Five Ph D students invigilated and administered all testing. Participants were 71 first year students majoring Russian as a Foreign Language. Students ranged in age from 17 to 25 years old. Females composed 76 % of the sample ($n = 55$), and males composed 24 % ($n = 18$). All students are not native Russian speakers, 71 are Turkmen, one participant is a Chinese citizen. Participation in the experiment was anonymous and at the beginning of the experiment each participants received a unique code.

WISC

We used General knowledge subtest of Wechsler test (WISC) (adapted by Grigoriev, Zhuravlev, Zhuravleva, Lapteva, & Noss, 2016) to measure participants' level of intellectual development as well as the ability to acquire, retain and retrieve information. General knowledge is also known to predict exam performance and exam results (Furnham, Monsen, & Ahmetoglu, 2009). The participants were given a series of general knowledge questions, such as: 1. What colour is the national flag of Russia? ... 5. What temperature does the water boil at? ... 9. What is rubber made of? 15. When do Christmas celebrated in Russia? 20. What is epistemology? (Grigoriev et al., 2016).

Quick Russian Placement Test

Russian language proficiency of the participants was tested with Quick Russian placement Test (Russian Proficiency Test, n.d.). It is a contemporary test of the Russian language which comprises 50 questions assessing Russian language competences in Language Use (vocabulary, grammar) and Reading Comprehension.

The test provides researchers and teachers with a reliable way to place non-native Russian language students into the correct level Russian class. The test is computer-based with four multiple choice questions.

Yesterday my brother _____ a parcel at the post office.

A. Received /B. sent /C. Gave /D. signed

Oleg is a bad student, he often has to see _____.

A. the headmaster / B. to the headmaster /C. with the headmaster D./ for the headmaster

During lunch someone pushed me and I _____ soup.

A. spilled / B. poured C./ cast D./ dropped

Testing Reading Comprehension.

Text selection

The participants read an expository text twice. We did not limit their reading time, the average time spent reading the text was 4 minutes, after which participants proceeded to writing recalls. The reading text is from the book “Collection of tests in Russian as a foreign language” (Satretdinova, Glukhova, Matyushkova, & Kosmacheva, 2012) and consists of 8 paragraphs. Based on the research in the area (Arias, 2007) we selected the reading text based on its (1) relevance and (2) readability. The selection criteria imply that the text (1) is related to participants’ life and as such is viewed as meaningful and interesting (Kitao, 1997) (2) its complexity corresponds participants’ reading skills. The reading text offered to participants focuses on the two great roles culture plays in society: to preserve traditions and nature. Both roles are viewed by the authors to be relevant for students majoring Russian and Russian literature.

The extract of the reading text presented below (See Fig.1) was translated into English by the authors, the complete versions of the Russian text and its translation into English are uploaded at Laboratory site. Text readability alongside with a number of other qualitative parameters was assessed with the help of Rulingva (<https://rulingva.kpfu.ru/>) (see Table 1).

THE MEMORY OF CULTURE

I. Today a lot of scientists do everything they can to save air, seas, rivers, and forests from pollution. They want to preserve our planet's fauna and to save birds. Humanity spends a huge amount of money to preserve nature. The science that deals with nature conservation is called ecology. And it is already being taught at universities.

II. But ecology ought to deal not only with the nature conservation issues. After all, there is not only natural environment people live in, but also the environment, created by culture. If nature is necessary for us to keep living our biological life, then the cultural environment is necessary for his spiritual life. That's why an issue of the preservation of the cultural environment is no less significant than the preservation of nature. However, unfortunately, an issue of the ecology of culture has not yet been studied. Various aspects of culture and the culture of the past are studied though, but the entire cultural environment's significance for a person is not studied.

III. A person is being brought up in the surrounding cultural environment imperceptibly for himself. History and the past educate people. The past opens not only a window on the world for a person, but also doors. To live in a place where poets and writers of the great Russian literature lived, in a place where great critics and philosophers lived, to go to museums and exhibitions means gradually becoming spiritually richer. Streets, squares, and some separated houses tell us about those people who have been here before.

Fig. 1. The expository text read by participants of the experiment (fragment)

Text readability parameters measured with Rulingva report the text to correspond cognitive and linguistic abilities of Russian natives who had six years of formal schooling (Flesh-Kincaid Grade level (SIS)=6.42), its TTR is within a range of an academic text pattern (Churunina, Solnyshkina, Gafiyatova, & Zaikin, 2020) and its local and global overlaps are quite high for a Russian academic text (Gizatulina, Ismaeva, Solnyshkina, Martynova, & Yarmakeev, 2020).

Table 1. Text readability parameters measured with Rulingva.

#	Parameters	value
1	Word count	420
2	Syllable count	1051
3	Sentence count	38
4	Average sentence length (in words)	11,05
5	Average word length (in syllables)	2,5
13	Flesh-Kincaid Grade level (SIS)	6,42
15	Abstractness	2,66
16	Local noun overlap	0,14
17	Global noun overlap	0,07
18	Local argument overlap	0,59
19	Global argument overlap	0,25
21	TTR	0,67

Comprehension was assessed based on a combination of metrics of the written recalls and Cloze tests. Two measures are beneficial as they provide a better evaluation of the comprehension. As it was expected, there was a significant, moderate correlation between Russian language proficiency and comprehension. Comprehension was assessed (1) with the help of written recalls and (2) based on the cloze test (see Table 1 below).

After reading the text each participants received the following directive: “Write down everything you can remember about what you have just read. Provide as many details as possible”. Every participant was provided with a laptop. When the experiment was over we collected written (typed) recalls to assess their comprehension of the texts.

Pathfinder network

The authors also designed a mental map of the text of 9 macro-propositions or its Pathfinder network presented in Fig. 2. (Britton & Gulgoz, 1991) which was later used to develop a 22 multiple-choice questions test.

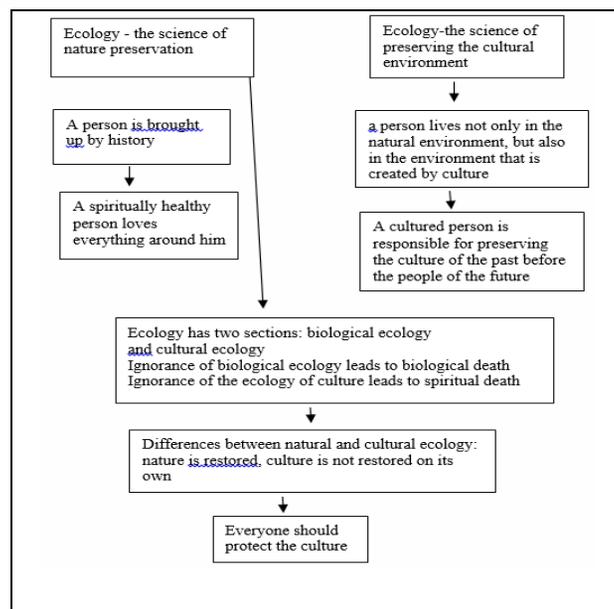


Fig. 2. Its Pathfinder network of the text **The Memory of Culture**

Cloze Test

After writing recalls, the participants answered 22 multiple-choice questions based on the Pathfinder network designed by the authors. Participants did not have an access to the text while answering the questions.

Table 2. Cloze Test

Write down your code number _____

Specify your gender _____

Write down your name _____

Choose the correct variant that matches the information in the text

1. Nature's ecology is the science of

A. nature conservation

B. animal world

C. pollution of nature

D. A B C

2. People around the world are investing in

A. teaching ecology at universities

B. development of ecology as a science

C. protecting the nature of planet Earth

D. maintaining peace on our planet

3. Humanity seeks to save

A. B C D

B. rivers, seas

B. animals of the planet

D. forests and air

4. The ecology of culture deals with

A. preservation of the natural environment

B. biological life issues

B. the problems of the past of the country

D. the study of the cultural environment

5. Nature is necessary for man to

A. spiritual development

B. biological life

C. money enrichment

D. A and B

6. Spiritual life is impossible without

A. nature conservation

B. biological life

C. natural environment

D. cultural environment

7. A person is brought up by

A. poets and philosophers

B. ecology and nature

C. cultural environment and history

D. works of literature

8. A person is spiritually enriched if

A. visits museums and exhibitions

B. A, B, D

C. walks through the old streets and squares

D. lives where the poets lived

9. Knowledge of the past helps a person

A. to become spiritually richer

B. understand art

C. to love Russian literature

D. to solve the problems of the world around us

10. To conservation the culture of the past, it is necessary to

A. open your soul to people

B. be responsible for the past

C. respect your predecessors

D. take care of people and nature

11. The cultured person is responsible for

A. the memory of the people of the past

B. the care of the future generation

C. the preservation of the culture of the past

D. A, B, C

12. To enter into the past means in paragraph № IV

- A. to remember the culture of the past
B. to learn respect
C. B and C
D. be responsible to children
13. If a person is spiritually healthy, he
A. loves his planet and the whole world
B. loves his family, his childhood and past
C. loves his home, his school, his city,
D. loves his country, his culture and language
14. Ignoring the laws of ecology leads to
A. destruction and / or annihilation of monuments
B. destruction of the natural environment of man
C. spiritual death and destruction of culture
D. moral and biological devastation
15. There are two sections in ecology: biology and
A. philosophy
B. culture
C. literature
D. history
16. Ignorance of the ecology of culture can destroy
A. human spiritually
B. human biologically
C. cultural environment
D. biological environment
17. The main difference between natural and cultural ecology:
A. nature educates, culture destroys
B. nature is destroyed, culture is restored
C. nature is restored, culture is destroyed
D. nature destroys, culture educates
18. It recovers over time:
A. art
B. monuments
C. culture
D. nature
19. Nature is capable of
A. B and C
B. rebirth
C. clean up
D. persist
20. Define the main idea of the Paragraph VIII
A. Everyone's task is to join a public organization
B. People's duty to society is to preserve culture
C. People should create cultural monuments
D. The goal of humanity is to protect beauty
21. According to the author of the text, culture-
A. is studied
B. monotonous
C. unrecoverable
D. protected
22. The protection of culture should be handled by
A. all people
B. the government
C. public organizations
D. scientist

Results

Each participant's recall was computed with RuLinva and later we also performed the Spearman Rank Order Correlations of the recalls (See Table 3).

Table 3. Spearman Rank Order Correlations of the recalls

		GKT	Rus	GK	Frequency (Sharoff)	FKGL (SIS)	Abstractness	Local noun overlap	Global noun overlap	Local argument overlap	Global argument overlap	TTR
1.	Rus	0,57*	1,00	0,68*	0,04	0,14	0,17	0,36*	0,32*	0,41*	0,24*	-0,54*
2.	General Knowledge test	0,49*	0,68*	1,00	-0,22	0,42*	0,09	0,31	0,19	0,36*	0,05	-0,41*
3.	Cloze tests	1,00	0,57*	0,49*	-0,01	0,19	0,07	0,33*	0,21	0,34*	0,14	-0,34*
4.	Word count	0,47*	0,71*	0,63*	0,05	0,06	0,23	0,48*	0,30*	0,52*	0,23	-0,82*
5.	Sentence count	0,41*	0,60*	0,52*	0,10	-0,13	0,07	0,41*	0,18	0,47*	0,08	-0,71*
6.	Average number of words in a sentence	0,24*	0,47*	0,47*	-0,02	0,28*	0,29*	0,33*	0,37*	0,33*	0,38*	-0,51*
7.	Adjectives	0,45*	0,65*	0,65*	-0,12	0,15	0,23	0,48*	0,27*	0,47*	0,17	-0,63*
8.	Adverbs	0,36*	0,63*	0,54*	0,03	-0,01	0,14	0,33*	0,27*	0,40*	0,23*	-0,61*
9.	Pronouns	0,32*	0,39*	0,50*	0,07	0,13	-0,15	0,16	0,08	0,24*	-0,06	-0,44*
10.	Nouns	0,42*	0,66*	0,61*	0,12	0,06	0,29*	0,49*	0,30*	0,54*	0,24*	-0,84*
11.	Verbs	0,41*	0,66*	0,54*	0,02	0,06	0,07	0,43*	0,21	0,46*	0,12	-0,78*
12.	Frequency (Sharoff)	-0,01	0,04	-0,22	1,00	-0,42*	-0,05	0,26*	0,40*	0,36*	0,32*	-0,25*
13.	FKGL (SIS)	0,19	0,14	0,42*	-0,42*	1,00	-0,09	-0,08	-0,18	-0,14	-0,20	-0,06
14.	Abstractness	0,07	0,17	0,09	-0,05	-0,09	1,00	0,26*	0,26*	0,24*	0,25*	-0,23
15.	Local noun overlap	0,33*	0,36*	0,31	0,26*	-0,08	0,26*	1,00	0,64*	0,76*	0,46*	-0,59*
16.	Global noun overlap	0,21	0,32*	0,19	0,40*	-0,18	0,26*	0,64*	1,00	0,64*	0,73*	-0,45*
17.	Local argument overlap	0,34*	0,41*	0,36*	0,36*	-0,14	0,24*	0,76*	0,64*	1,00	0,65*	-0,57*
18.	Global argument	0,14	0,24*	0,05	0,32*	-0,20	0,25*	0,46*	0,73*	0,65*	1,00	-0,33*

	overlap											
19.	TTR	- 0,34*	- 0,54*	- 0,41*	-0,25*	-0,06	-0,23	-0,59*	-0,45*	-0,57*	-0,33*	1,00

Note. * — $p < 0.05$ — statistically significant differences

The research shows a number of statistically significant relationships between the metrics of parameters (1-19). It is significant that participants' general knowledge correlates with recalls metrics, i.e. word counts (0.47) and sentence counts (0.49). The latter is easy to explain: the wider a person's outlook and level of general knowledge, the more informative and longer texts he can generate. In addition, general knowledge has positive correlations with text cohesion, specifically with Local noun overlap (0.33) and Local argument overlap (0.34). There is a possibility that it is also a function of a subjects' general knowledge which enables him/her to create coherent texts with the main idea traceable throughout the text.

Interestingly, the level of lexical diversity (TTR) is negatively correlated with the General Knowledge test and Russian language proficiency. This seemingly contradictory fact can be explained by the fact that in this experiment, students with a low level of Russian proficiency wrote very short texts with few repetitive words, which resulted in a high value of TTR.

The level of Russian language proficiency has statically significant strong positive correlation with the following parameters: General Knowledge (0.57), word count (0.71), syllable count (0.70) and sentence counts (0.60) of the recalls and the Cloze test (0.68). In addition, Russian language proficiency positively correlates with the parameters of text cohesion: Local noun overlap (0.36), Global noun overlap (0.32), Local argument overlap (0.41), Global argument overlap (0.24). No statistically significant relationship was found between the Russian language proficiency and indices of abstractness or readability (Flash – Kinkaid Grade level (SIS)).

The research also shows that the backbone parameter in evaluating recalls is Local argument overlap, since it has statistically significant correlations with all text metrics, except for the Flash-Kincaid Readability Index. In other words, changing Local argument overlap leads to a change in almost all metrics of all parameters of the analyzed text.

Effects of Reader Abilities and General Knowledge

On two screening measures, RFL students demonstrated the average Russian language proficiency to be 28.5 ($SD = 3.7$) and General Knowledge performance as 8.39 ($SD = 3.98$) thus indicating variability of the group. We performed a median split on the individual difference scores resulting in high (Rus = 40-28) and

low (Rus=27-15) groups for General knowledge and Russian proficiency measures. The mean scores for the high and low groups on the two individual difference measures are presented in Table 4 below.

Based on General Knowledge test results, all students were streamed into 3 groups. The average value for the entire sample (n = 71) in General Knowledge Test was 8.39 ± 3.98 . The group with high scores in General Knowledge test comprises subjects with the score 12.4 (n = 11), the group with low scores included subjects scoring less than 4.4 (n = 10). The respondents with mean scores in General Knowledge Test did not participate in further analysis.

Table 4. Mean Scores for Participants Assigned to the Low and High Groups

Individual difference test	Ability group	
	High	Low
General Knowledge	Over 12.4	Less 4.4
Russian Proficiency	40-28	27-15

We conducted Mann-Whitney Test on performance on free recalls and cloze tests. The high and low ability groups demonstrated statistically significant differences in the following parameters: word count, sentence count, average number of words in a sentence, local noun overlap, global noun overlap, Local argument overlap, TTR (see Table 5).

Table 5. Mann-Whitney Test for Participants Assigned to the Low and High Groups

	High group (n=11)		Low group (n=10)		Mann-Whitney U Test, p-value
	Mean	SD	Mean	SD	
General knowledge test	17,60	1,82	11,00	4,08	0,022991*
Word count	157,18	74,21	55,00	34,52	0,000376*
Syllable count	383,09	189,83	133,30	78,86	0,000934*
Sentence count	13,64	7,72	5,50	4,30	0,003888*
Average sentence length(words)	12,81	4,20	17,20	27,14	0,113103
Average sentence length(syllables)	2,41	0,25	2,49	0,25	0,418053
FKGL (SIS)	6,50	1,98	8,55	9,01	0,751335
Frequency (Sharoff)	273,88	53,11	260,68	76,75	0,503515
Abstractness	2,79	0,14	2,69	0,23	0,359965
Local noun overlap	0,31	0,15	0,10	0,15	0,018325*
Global noun overlap	0,21	0,09	0,12	0,16	0,029040*
Local argument overlap	0,68	0,27	0,16	0,24	0,001946*
Global argument overlap	0,50	0,18	0,35	0,27	0,121336
TTR	0,78	0,06	0,87	0,09	0,002462*

Note. * — $p < 0.05$ — statistically significant differences

The results show (see table above) that High group students' recalls are significantly better than Low group students recalls: their recalls are almost three times longer (in words) and two times longer (in sentences)

than Low group recalls. Interestingly, the average number of words in a sentence and the average number of syllables in a word in the Low group are higher than those in the High group, although not much, nevertheless, these are statistically significant differences. In other words, a higher level of Russian language proficiency does not imply an increase in the length of sentences and the use of longer polysyllabic words, on the contrary, the number of words in a sentence is reduced, but at the same time, as indicated above, the length of the text increases. Apparently, a good command of the language allows a person to perform a more accurate selection of words and structure of the text to convey the meaning, which results in shorter sentences. To some extent, this assumption is confirmed by the higher average values of frequency and abstractness of the words used by the High group participants, but these differences are not statistically significant. Longer sentences in the recalls of the Low group participants cause higher Flash-Kincaid readability index thus implying these recalls to be hard to read. Meanwhile, the texts created by High group subjects demonstrate a higher level of coherence computed with Local/Global noun overlap and Local argument overlap. The latter testifies to these texts to be easier to read and understand.

Cloze Test Analysis

The mean value for the entire sample of Cloze tests is 13.14 ± 3.86 . The high comprehension group included students who scored over 17.0 ($n = 7$), the low comprehension group included students who scored less than 9.3 ($n = 7$). The respondents with the mean score of General knowledge did not participate in further analysis.

Table 6. *Mann-Whitney Test for Participants Assigned to the Low and High Comprehension*

	High comprehension (n=7)		Low comprehension (n=7)		Mann-Whitney U Test, p-value
	Mean	SD	Mean	SD	
Rus	34,857	4,5251	21,571	6,373	0,006012*
General knowledge test	11,286	4,5356	4,143	2,795	0,015194*
Word count	158,143	44,7416	73,143	38,989	0,012717*
Syllable count	356,571	112,8433	171,571	87,405	0,021451*
Sentence count	14,429	3,1547	5,714	3,498	0,006012*
Average sentence length(words)	10,951	1,8489	22,240	31,747	1,000000
Average sentence length(syllables)	2,231	0,1212	2,413	0,308	0,159865
FKGL (SIS)	4,831	1,1900	9,934	10,683	0,110224
Frequency (Sharoff)	306,066	74,4799	263,099	118,501	0,443289
Abstractness	2,889	0,1212	2,849	0,293	0,898327
Local noun overlap	0,416	0,1712	0,170	0,290	0,040914*
Global noun overlap	0,244	0,0810	0,179	0,197	0,159865
Local argument overlap	0,850	0,1757	0,427	0,498	0,063920

Global argument overlap	0,523	0,1211	0,504	0,371	0,482203
TTR	0,747	0,0519	0,830	0,127	0,306686

Note. * — $p < 0.05$ — statistically significant differences

The table indicates that respondents with "high comprehension" has statistically significant differences when contrasted with the group of "low comprehension" in a number of parameters, in particular, they have a better command of the Russian language and have a higher level of General knowledge. The respondents of the "high comprehension" group created longer recalls compared to the other group; their recalls are more cohesive (Local noun overlap). There are no statistically significant differences in other parameters, which is probably caused by the paucity of the identified groups. In the future, increasing the sample will allow us to draw more informative conclusions when studying this phenomenon. Thus, we can conclude that text comprehension depends on the readers' language proficiency and general knowledge.

Discussion

The identification of the reproduced versus lost information within a written recall with the help of software is still a research niche for the Russian language. Despite theoretical and practical advances for the English language (Mccarthy, Dufty, Hempelmann, & Graesser, 2012), computational methods for evaluating the amount of given information in written Russian recalls, to the best of our knowledge, have not previously been implemented. The current study offers a new computational instrument to analyze metrics of expository recalls against those of the primary (i.e. given) text. Our findings suggest metrics computed with RuLingva outperform the existing Russian text analyzers.

Features measured with RuLingva, i.e. text length, readability indices, morphological parameters, vocabulary frequency, lexical diversity and abstractness rating, objectively indicate differences in RFL students' explanatory discourse and are recommended by the authors to be used by professionals in assessment of written expression. Automation of contrasting metrics of the abovementioned features of the reading text and a written recall of a certain student allows a rater to better assess students' comprehension and provides an opportunity to focus on other features of students' writing.

The findings indicate that comprehension was enhanced by increased knowledge: high knowledge readers showed better comprehension than low knowledge readers and the input texts were comprehended better by readers with high general knowledge. Interactions between readers' general knowledge and Russian proficiency levels on the one hand and written recalls characteristics on the other indicate that the readers with higher levels of proficiency showed large effects of general knowledge for generating more cohesive and longer texts.

Conclusion

We examined comprehension of RFL readers' as a function of their abilities (language proficiency and world knowledge). The overarching purpose of this study was to contribute to our understanding on the use of RuLingva, an automated tool computing over 33 metrics for written Russian texts. RFL students read a 400 word text, the comprehension of which was assessed with the help of its written recalls and a cloze test designed and developed by the authors. Overall, the study confirms that the quality of written recalls of RFL students is at least partially attributable to the readers' prior knowledge. The results also suggest that RuLingva has a potential to be used in automated assessment of Russian writing. This study also suggests future research to view RuLingva functions in a new area of identifying and classifying writing patterns at different stages of writing skills development. Our findings has potential utility to be used both in research and RFL teaching and assessment.

Acknowledgements

This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program.

References

- Arias, I. J. (2007). Selecting reading materials wisely. *LETRAS*, 1(41), 131–151.
- Balfour, S. (2013). Assessing writing in MOOCs: automated essay scoring and calibrated peer review. *Research and Practice in Assessment*, 8: 40–48.
- Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's Computational Model to Improve Instructional Text: Effects of Repairing Inference Calls on Recall and Cognitive Structures. *Journal of Educational Psychology*, 83(3), 329–345.
- Chapelle, C. A., Cotos, E., & Lee, J. Y. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32, 385–405.
- Churunina, A., Solnyshkina, M., Gafiyatova, E., & Zaikin, A. (2020). Lexical Features of Text Complexity: the case of Russian academic texts. *International Scientific Forum - Issues of Modern Linguistics and the Study of Foreign Languages in the Era of Artificial Intelligence (dedicated to World Science Day for Peace and Development)*, 88, 01009.

- Cotos, E. (2015). Automated Writing Analysis for writing pedagogy: From healthy tension to tangible prospects. *Writing and Pedagogy*, 6, 1-29.
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48(4), 1227–1237.
- Furnham, A., Monsen, J., & Ahmetoglu, G. (2009). Typical intellectual engagement, Big Five personality traits, approaches to learning and cognitive ability predictors of academic performance. *British Journal of Educational Psychology*, 79(4), 769–782. DOI: 10.1348/978185409X412147. PMID 19245744.
- Gizatulina, D., Ismaeva, F., Solnyshkina, M., Martynova, E., & Yarmakeev, I. (2020). Fluctuations of text complexity: the case of Basic State Examination in English. *SHS Web of Conferences*, 88, 02001.
- Grigoriev, A. A., Zhuravlev, I. V., Zhuravleva Y. V., Lapteva E. M., Noss I. N. (2016). Approbation of the general awareness test in Russia. *Psychology. HSE Journal*, 13(4), 667-677.
- Kitao, K. (1997). Selecting and developing teaching/learning materials. *The Internet TESL Journal*, 4(4). Retrieved from <http://iteslj.org/Articles/Kitao-Materials.html>
- Landauer, T. K., Laham, D., & Foltz, P. W. (2003). *Automatic scoring and automation of essays with the intelligent essay assessor. Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Mccarthy, Ph. M., Dufty, D., Hempelmann, Ch. F., & Graesser, A. C. (2012). Newness and Givenness of Information: Automated Identification in Written Discourse. In Ph.M. McCarthy & Ch. Boonthum-Denecke (Eds.), *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 457-478). IGI Global.
- 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.
- McNamara, D. S., Louwerson, M. M., McCarthy, Ph. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing Linguistic Features of Cohesion. *Discourse Processes*, 47(4), 292–330. Retrieved from <https://www.tandfonline.com/doi/abs/10.1080/01638530902959943>

Petrova, A. A., & Solnyshkina, M. I. (2021). Immediate recall as a secondary text: Referential parameters, pragmatics and propositions. *Russian Journal of Linguistics*, 25(1), 221–249. DOI: 10.22363/2687-0088-2021-25-1-221-249.

Russian Proficiency Test. (n.d.). *Transparent language*. Retrieved April 19, 2021, from <https://www.transparent.com/learn-russian/proficiency-test.html>

Satretdinova, A. H., Glukhova, N. V., Matyushkova, A. I., & Kosmacheva, O. Yu. (2012). *Collection of tests on Russian as a foreign language (basic level, I certification level II certification level)*. Astrakhan: Publishing house of FGBOU VPO AGMA.

Solnyshkina, M., Guryanov, I., Gafiyatova, E., & Varlamova, E. (2018). Readability Metrics: the Case of Russian Educational Texts. In F. Uslu (Ed.), *ADVED–4th International Conference on Advances in Education and Social Sciences: proceedings* (pp. 676-681). Ocerint Publishing.

Solnyshkina, M., Kazachkova, M., Gafiyatova, E., & Varlamova, E. (2018). Linking Words in Russian Social Studies Course Books: a Study on Text Complexity. *SOCIOINT, 5th International Conference on Education, Social Sciences and Humanities: proceedings* (pp. 764–771). Ocerint Publishing.