

# The Structure of Cross-Linguistic Differences: Meaning and Context of 'Readability' and its Russian Equivalent 'Chitabelnost'

Marina I. Solnyshkina<sup>1</sup>, Elena V. Harkova<sup>1</sup>, Mariia B. Kazachkova<sup>2</sup>

<sup>1</sup> Kazan Federal University

<sup>2</sup> MGIMO University

Correspondence concerning this article should be addressed to Elena V. Harkova, Department of Theory and Practice of Teaching Foreign Languages, Institute of Philology and Intercultural Communication, Kazan Federal University, 2 Tatarstan str., Kazan, 420021, Russian Federation. E-mail: halenka@rambler.ru

The article presents the results of an original study aimed at finding (1) frequency fluctuations of the term 'readability' in American discourse and its Russian equivalent 'chitabelnost' in Russian discourse over the period from 1920s to the present; and (2) semantic similarities and differences between the English term 'readability' and its Russian equivalent 'chitabelnost' over the same period of time. A contrastive analysis of the words testified to inconsiderable differences in the semantic structures of the terms in the period under study: the term 'readability' has been used with the following meanings: (1) 'the quality of being legible or decipherable' and (2) 'the quality of being easy or enjoyable to read'. The Russian equivalent 'chitabelnost' has two contemporary meanings similar to the aforementioned English meanings as well as the obsolete 'library book checkouts'. With the help of the Google NgramViewer, we identified the 1980s frequency peak of both terms when the modern notion of the concepts was formed. The research into the topical context of readability as 'the quality of being easy or enjoyable to read' demonstrated empiricist tendencies in American studies focused on two types of parameters, i.e. the 'objective' parameters of texts, i.e. sentence length, word counts, number of high/low frequency words, ratio of high/low frequency words to total words, sentence complexity, etc. and 'individual' variables affecting a potential reader, such as 'word familiarity', cognitive and linguistic abilities, cultural and topic knowledge, etc. The Russian school's view, until the 1970s, had traditionally been more holistic and 'biased' towards an individuals' factors. The results of the study have the potential to contribute to cross-linguistic research in the area of text readability assessment, semantics, and scientific literature searches.

**Keywords:** readability, ngrams, topical context, semantic analysis, text parameters, sentence length, word counts, readability formulas

## Introduction

Communication as 'the imparting or exchanging of information by speaking, writing, or using some other medium' (Online Oxford English Dictionary, 1996)<sup>1</sup> implies either generating or comprehending a text, which may be handwritten, printed, electronic, or oral. Successful communication in its turn largely depends on whether the amount, content, and structure of the quanta of the information sent by its generator in the text and received by the addressee are similar or, in an ideal situation, the same. Thus, for the information of any text to be elicited, processed, and stored in a recipient's mind, it is important that the recipient has sufficient cognitive and linguistic abilities.

In the modern world, matching a text (both oral or written/electronic) to the target audience is a problem relevant in a number of spheres: education, PR, the military, government, law, advertising, business, publishing, medicine, and social relations, as these are areas where communication is the foundation of success. The

<sup>1</sup> Online Oxford English Dictionary (OED) (3rd ed.) (1996). Retrieved from <https://en.oxforddictionaries.com/>

research shows that companies suffer damages and take financial hits if the texts they expose their customers to are hard for an average reader to comprehend (Klare, 2000). On the other hand, editors and educators claim that if a text is too easy, i.e. below the audience's reading expertise, readers lose interest and stop reading it (Randall, 2013). The success of reading depends on the degree of reading comprehension at a certain reading speed and maintaining interest in the content of the text. In an ideal situation, if a text is beyond its target audience's reading level, it needs to be altered or leveled to match the reader. Modern text leveling procedures imply measuring two parameters: (1) the level of cognitive and linguistic abilities of the target audience and (2) text readability (Reading A-Z, 2018)<sup>2</sup>. According to Fernbach (1990), the latter is nowadays interpreted as "the ease with which a text can be read quickly, understood, and memorized". As for the concept of 'text readability' itself, it dates back to medieval times, when word counts were used to estimate the difficulty of the Talmud (see Taylor & Wahlstrom, 1986). In the 1880s, Professor L.A. Sherman conducted a research on the length of the average English sentence in different centuries and concluded that shorter sentences and more concrete terms in a text make the text easier for a reader. Sherman was also the first to argue that readability can be evaluated based on statistical analysis (Sherman, 1893). At about the same time, in 1889, Russian writer Nikolai A. Rubakin published a list of 1500 words 'known by all Russians' that he derived from over 10000 texts written by common people. Rubakin argued that reading comprehension is hampered by unfamiliar words and long sentences (see Choldin, 1979). Unfortunately Rubakin's ideas were soon forgotten in Russia, but text 'readability' studies have since been actively conducted in the USA, UK, and Germany.

Nowadays researchers worldwide are working to address the problem of 'text – reader correspondence' in two ways: (a) from the point of the reader and their subjective characteristics: age, education, background knowledge, memory span, etc.; and (b) from the point of view of the text and its objective parameters. Text objective parameters are generally classified into two types of categories: 'extra-textual' parameters, which include illustration support and graphic features, such as font, spacing, and indentation, and 'inter-textual' parameters, which comprise the following: total word count, number of different words, ratio of different words to total words, number of high frequency words, ratio of high frequency words to total words, number of low frequency words, ratio of low frequency words to total words, sentence length, sentence complexity, etc. (Reading A-Z, 2018<sup>3</sup>; Ivanov, 2013; Hiebert, 2012).

In this paper we aim at two research questions: (1) How different or similar were the frequencies of the word 'readability' in American discourse and its Russian equivalent 'chitabelnost' in Russian discourse over the period from the 1920s to the present? (2) How different or similar were readability constructs and meanings of the English-Russian equivalents 'readability' and 'chitabelnost' over the period from the 1920s to the present? As the Russian word 'chitabelnost' contains a borrowed suffix '-beln-' (Sologub, 2002), we may anticipate, and that is the first *hypothesis* of the study, that the word began functioning in Russian discourse later than its equivalent in English and its frequency has been considerably lower over the period under study. The implied *hypothesis* behind the second research question is that changes in the meaning of the term become evident once we contrast topical contexts in the specialized discourse. The focus is on the qualitative analysis of the topical context in the discourse, thus revealing diachronic changes in the semantic range of the words and structures of the corresponding concepts. The ultimate goal of the study is to define possible conceptual differences in the terms 'readability' and its Russian equivalent 'chitabelnost' over a period of time between the 1920s and 2019. To this end we synthesize ten historic meanings of the equivalents studied in each decade from the 1920s to 2019.

## Method

### Background Information

The 'linguistic diversity' of the word 'readability', naming a scientific concept, entails making a distinction between the language-specific semantic values or meanings of the word and the encyclopaedic concept(s) attributed to the word (see Willems, 2012). Acknowledging the two entities, Wierzbicka argues that 'meanings' are supposed to be conveyed in dictionaries while knowledge inherent to concepts is to be included in encyclopaedias (Wierzbicka, 1996). Thus, we resorted to unabridged monolingual dictionaries for the meaning

<sup>2</sup> Reading A-Z. (2018). Printable teacher materials. Retrieved from <http://www.readinga-z.com>

<sup>3</sup> Ibid.

of 'readability' in the English language<sup>4</sup>,<sup>5</sup> and its equivalent 'chitabelnost' in the Russian language<sup>6</sup> to derive all the registered meanings of the words corresponding to separate entries in dictionaries' definitions. Based on the assumptions that "a semantic value is intuitively available to the speaker" and "the context can provide enough evidence on its own to reveal separate senses of a word" (Willems, 2012, p.668) we also tested the registered 'senses' of the words 'readability' and 'chitabelnost' "in naturally occurring utterances" (Willems, 2012, p.665), i.e. contexts. The idea of using the context to derive senses goes back to J.R. Firth's dictum "you shall know a word by the company it keeps" (Firth, 1957, p.11) and in our case the goal was to verify the semantic range of the word used in the discourse of a particular period.

For the purposes of the research, contexts such as "linguistic environments <...> in which a particular word occurs" (Dash, 2008, p.22) were classified into local and topical (Ravin & Leacock, 2000). We predominantly used the local context, i.e. 1–5 words immediately before and after the words under study to verify, extend or narrow the meanings registered in dictionaries. The topical contexts as "the wider circle beyond the sentence level" (Dash, 2008, p.22) were essential in synthesizing the structure of the concepts.

For our future discussions it is also important that, as with any scientific concept, readability is designed and developed primarily for research purposes and, thus, related to other constructs in theoretical schemes. However, the area of implementing the findings on similarities and differences in constructing the concept of 'readability' is quite wide as, obviously, the way a piece of information is presented influences how and how well a reader comprehends it. Reading is a complex task in which text writers are supposed to determine the best means to promote comprehension. We also emphasize that, as with any scientific construct composed of a number of constituents (Kerlinger, 1973), we expect readability to become observable through indicators or manifestations of what researchers have agreed. Based on this, the strategy employed in the exploration of the concept under study was an in-depth analysis of the existing knowledge in the area that entails reviewing the published research on readability in American and Russian discourses in the period between the 1920s and 2019. We were mostly exploring the compiled topical contexts from the Google books corpus, which proved to be sufficient for extracting topical contexts and help discriminate between the historic paradigms of the concepts (1920s – 2019). It was this type of context that opened the major avenue of the research.

### Research Design

Synthesizing a concept based on the context in a text is a common approach practiced in modern corpus linguistics (see Zakharov et al., 2014). To define the trends in changes in the frequency of the words 'readability' and 'chitabelnost' between 1920 and 2019 we applied distributional semantic models (Firth, 1957), which allowed us to induce the meanings of words from texts. The algorithm of diachronic studies of words and concepts with Google Books Ngram, i.e. a database of 67 billion words in Russian Viewer and 361 billion words in English, has been successfully implemented by numerous researchers in Russia and abroad (see Solovyev, 2013, Zakharov et al., 2014, Hai-Jew, 2014).

The research was conducted in three stages:

- Stage 1, a lexicographic analysis, was aimed at defining the meanings of the two contrasting words ('readability' and 'chitabelnost') based on the data registered in dictionaries.
- Stage 2 is a frequency-based contrastive discourse analysis of the two words 'readability' and 'chitabelnost' using the words' recurrence in the following decades: 1925 – 1929, 1930 – 1939, 1940 – 1949, 1950 – 1959, 1960 – 1969, 1970 – 1979, 1980 – 1989, 1990 – 1999, 2000 – 2009, and 2010 – 2019.
- Stage 3, a conceptual or Topical Context Analysis, was conducted to synthesize and contrast 10 historical concepts of 'readability/chitabelnost' over the ten decades, i.e. from 1925 to 2019.

### Procedure

To contrast (1) the semantic volumes of the English 'readability' and its Russian equivalent 'chitabelnost' in different periods of time (the 1920s to 2019) and (2) the frequency of fluctuations in the terms in corresponding

<sup>4</sup> Fellbaum, C. (Ed.) (1998). WordNet: An electronic lexical database. MIT Press.

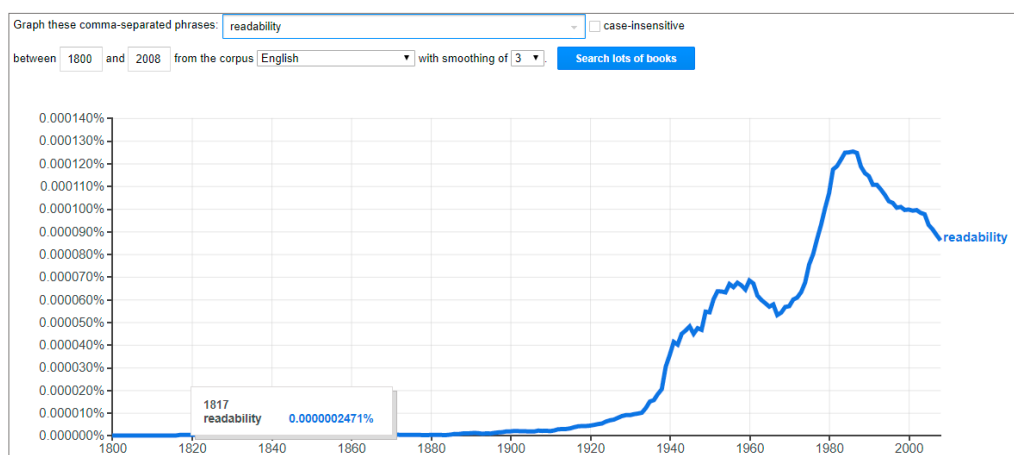
<sup>5</sup> Merriam Webster's collegiate dictionary. (1993). <https://www.merriam-webster.com/dictionary/readability>.

<sup>6</sup> Evgen'eva, A. P. (Ed.) (1999). MAS - Slovar' russkogoyazyka [The dictionary of Russian language]. <http://feb-web.ru/feb/mas/mas-abc/default.asp>.

discourses, we implemented the Ngram Viewer<sup>7</sup>, an online tool based on a ‘bag of words approach’.<sup>8</sup> We computed graphs of relative frequencies of the words ‘readability’ and ‘chitabelnost’, which demonstrate how often they were used in the Google Books corpora (Figures 1, 2 below).

The words ‘readability’ and ‘chitabelnost’ are viewed in the study as ‘1-grams’, i.e. strings of characters uninterrupted by a space. The x-axes in Figures 1 and 2 show the periods of the study, i.e. from 1800 to 2019.

*Figure 1*  
Ngram of Publications with the Word ‘Readability’, 1800 – 2019



The y-axes show the relative frequency of the word (RF), i.e. the percentage of the specified ngram of all ngrams in the corpus of books of that particular year. For the word ‘readability’ in year 1817 it is 0.0000002471 % (see Figure 1 above).

The graph in Figure 1 shows that the frequency of the word increased dramatically in the 1930s and in the 1970s as those were the periods when the term was actively used in many more printed sources than in previous decades.

Figure 2 below provides insights into the fluctuating periods of popularity and decline of the corresponding Russian term, ‘chitabelnost’, over time in Russian discourse: the term became popular in the 1920s, the 1960s – 80s and then, in the 2000s it experienced an unprecedented rise of recurrence.

*Figure 2*  
Ngram of Publications with the Word ‘Chitabelnost’ (Readability) in Russian Discourse, 1800 – 2019



<sup>7</sup> Google Ngram Viewer (2013). Professional web tool for calculating and tracing relative and absolute frequencies of a word or a phrase. <https://books.google.com/ngrams>

<sup>8</sup> Ong, M. (2014). Bag of Words (BoW) -Natural language processing [Blog]. <https://ongspxm.github.io/blog/2014/12/bag-of-words-natural-language-processing/>

We share the opinion that the term 'frequency counts' provides an indicator of trends in the corresponding research area (Hai-Jew, 2014) and the conceptual part of the study was based on the premise that the rate of knowledge growth in the area, as well as paradigm changes, are reflected in publications (Macias-Chapula, 1999).

### Limitations

Although the limitations of the Google Books corpus and the Ngram viewer as indicators "of the 'true' popularity of various words and phrases" have been discussed more than once in literature (Pechenick, Danforth, & Dodds, 2015), we advocate for the reliability of the tool based on the assertions presented below. The main criticism against the Google Books corpus revolves around it being a more lexicon-like than text-like dataset and designed as "a reflection of a library in which only one of each book is available" (Pechenick, Danforth, & Dodds, 2015, p.2/24). Another argument raised by Google Books corpus opponents is that it is imbalanced because "scientific texts have become an increasingly substantive portion of the corpus throughout the 1900s" (Pechenick, Danforth, & Dodds, 2015, p. 1/24. However, as we aim at the meanings and scientific constructs as well as the frequency of publications with the words under study in certain years and at certain periods (not the popularity of the words, which depends on the number of printed copies), this skewness to scientific discourse makes the Google Books corpus more scientifically representative and thus preferable for our research.

There were two problems we faced working with the Ngram Viewer. The first one was with optical character segmentation when the system recognized character sets of Russian 'реш' as 'чи' which resulted in referring to sources with the word 'rentabelnost' (Rus. profitability) not 'chitabelnost' (for optical character recognition errors in English texts, see Zhang, 2015). Thus, we had to manually check every snippet of the word 'chitabelnost' in the Russian version of the Google Books corpus.

The second problem was a lack of available resources published in a certain year. E.g. Ngram Viewer graph of readability (see Figure 1) starts in 1817, but the resource does not provide a reference to texts with the word 'readability' published until 1916. In the Russian corpus, the word 'chitabelnost' was used for the first time in 1922 with a relative frequency of 0.0000001530 % (see Figure 4. below). This was the reason we chose the 1920s as the starting point of the 'anthology'. As for the latest sources registered in both versions of the Google Books corpora in Russian and English, and ultimately in Ngram Viewer, they were released in 2019.<sup>9</sup> Thus, although the Google Books corpora designers argue that they provide graphs for the period of 1800-2009, in fact the data elicited covers the period of 1916-2019 for the word 'readability' and 1922-2019 for the word 'chitablnost'.

### Materials

As the Ngram Viewer typically provides texts of publications with an ngram under study, we expected to get access to the first registered resource in the English Google Books corpus, i.e. a book published in 1817 as that is the year when the graph starts (see Figure 1 above). Unfortunately, in our case, the earliest publication available in the Google Books corpus is the article "Development and Validation of a New Instrument to Assess the Readability of Spanish Prose" in Volume 65 of "The Modern Language Journal" published in 1916 by Patricia Vari-Cartier. The author refers to the two variables used to calculate readability, i.e. the average sentence length and the number of syllables: "Each new graph would require a different set of parameters (minimum and maximum sentence and syllable count) and *readability* designations" (Vari-Cartier, 1916, p. 145). Those were studies that were greatly influenced by the works of Sherman mentioned earlier in this paper (Sherman, 1893).

Thus, based on the resources available, we focused on the time frame from 1925 to 2019. For an in-depth analysis of the concept, we sampled 10 topical contexts of the words 'readability' and 'chitabelnost' from 10 different texts of each decade from 1925 to 2019, thus compiling 10 sub-corpora of topical contexts for each decade of the period under study. The topical contexts were selected with the purpose of making the constituents of the concepts under study visible, thus the length of the topical contexts varied from nine sentences to a paragraph of 21 sentences. A compiled sub-corpus of 10 topical contexts contains on average 1912 tokens with the smallest being 1281 tokens (Russian Sub-Corpus, 1920s) and the biggest having 2357 tokens (Russian Sub-Corpus, 1990s). The data were extracted with the help of the Ngram Viewer application,<sup>10</sup> manually filed and marked as English

<sup>9</sup> Readability. (2019). <https://clck.ru/JAtwq>

<sup>10</sup> Google Ngram Viewer (2013).

Sub-Corpus, 1920s; Russian Sub-Corpus, 1920s; English Sub-Corpus 1930s; Russian Sub-Corpus, 1930s, etc. The topical contexts were elicited from query pages or Google Books pages. We resorted to Google Books pages only in cases when the topical context on the query page was not sufficient to reconstruct the structure of the concept. We compiled the corpus from the first editions of the books exclusively and excluded any re-editions on the assertion that they may present a scientific paradigm of another decade.

Table 1  
*Size of Topical Context Corpus*

Period	American Sub-Corpus, tokens	Russian Sub-Corpus, tokens
1925 - 1929	2264	1281
1930 - 1939	2216	unavailable
1940 - 1949	1991	unavailable
1950 - 1959	1936	unavailable
1960 - 1969	1755	unavailable
1970 - 1979	1276	2096
1980 - 1989	1886	2182
1990 - 1999	1709	2357
2000 - 2009	1913	2411
2010 - 2019	2189	2163
<b>Total</b>	<b>19135</b>	<b>12490</b>

## Results and Discussion

### Stage 1. Dictionary Meanings of ‘Readability’ and ‘Chitabelnost’

Modern English dictionaries register ‘readability’ either as a polysemous word defined as “(1) the quality of written language that makes it easy to read and understand; (2) a quality of writing (print or handwriting) that can be easily read”<sup>11</sup> or refer potential readers to the adjective ‘readable’<sup>12</sup>. The latter is defined as “able to be read easily: such as a: legible; b: interesting to read”<sup>13</sup>.

Dictionaries from earlier periods do not register the noun ‘readability’ but the adjective ‘readable’: “RE’ADABLE, *adjective*. That may be read; fit to be read”,<sup>14</sup> “RE’ADABLE The state of being readable; readableness”<sup>15</sup>. No substantive changes were made to the definition of ‘readable’ since that time, although the 1989 edition of Webster’s Dictionary separates the meanings of ‘legible’ and ‘readable’, which in modern dictionaries are very often defined as synonyms.

*illegible, unreadable*

This distinction has been noted by many commentators, dating back as far as Utter 1916. Several commentators, including Evans 1957, and Shaw 1975, have also noted that unreadable can sometimes mean “indecipherable,” Fowler 1926, Krapp 1927, Partridge 1942, and Phythian 1979 will not allow this sense of unreadable, but it is treated as standard in dictionaries. It was first recorded in 1830. According to our written evidence, its use in current English is extremely rare. In fact, the closest thing we have to recent evidence of its use is a single citation from the magazine *Infoworld*, in which its meaning is not so much “impossible to decipher” as it is “impossible to see clearly enough for reading”: . permitting them to function in reduced light conditions where LCDs [liquid crystal displays] are unreadable —Nancy Groth, *Infoworld*, 27 Jan. 1986.

<sup>11</sup> Fellbaum, C. (Ed.) (1998). *WordNet: An electronic lexical database*. MIT Press.

<sup>12</sup> Merriam Webster’s collegiate dictionary (1993). <https://www.merriam-webster.com/dictionary/readability>.

<sup>13</sup> *Ibid.*

<sup>14</sup> Webster’s Dictionary (1828). <http://webstersdictionary1828.com/Dictionary/readable>

<sup>15</sup> Online Webster’s Dictionary (1913). <https://www.websters1913.com/words/readable>

*readable*

1. interesting or easy to read
2. legible<sup>16</sup>

To verify the dictionary meanings we used either local or topical contexts of all the periods of study. Over the period 1925–2019, in full correspondence with the meanings registered in American dictionaries, the word readability was used as (1) “the property of being easy or engaging to read”: “Our editorial policy is based on a desire to promote readability without sacrificing either scholarly accuracy or a sense of the ease and irregularity of informal correspondence” (Blom & Blom, 1983, p. xix); (2) the quality of being legible or decipherable, the property of print that affects the ease with which a text can be read’: “Readability of technical training materials presented on microfiche versus offset copy” (Baldwin, & Bailey, 1971, p. 37).

In the Russian lexicography of 1925–2019, the word ‘chitabelnost’ was registered for the first time in Volume IV of Explanatory Dictionary of the Russian Language, also called Ushakov’s Dictionary, in 1940, although the lexicographer did not provide the definition but referred readers to the derivative adjective, ‘chitabelnyi’: “CHITABELNYI (coll). Easy, pleasant to read”<sup>17</sup>.

‘Chitabelnyi’, although mistakenly viewed by some linguists as neologism in the Russian language, was for the first time registered in Russian discourse as early as 1910, and the case is very well documented in the Google Books corpus:

What is this? Are they bad poems? No, not that bad, but only, as they say, in the editorial jargon “readable”: you can read them without resentment for the time spent, but they are not kept in one’s memory, and could be written by someone else, not Sasha Chernyi. A. V. Amfiteatrov, “On Sasha the Black”, 1910<sup>18</sup>. Modern Russian dictionaries also define ‘chitabelnost’ as the noun formed from the adjective ‘chitabelnyi’ (readable)<sup>19</sup>.

Chitabelnost, coll. Feature attributed to the adjective ‘chitabelnyi’, suitability for reading I myself am not young, I am already fifty-six years old, and I still read it (although the material there is much less interesting than before – it used to be 100% “readable” for me, and now it is 70 in the best case) “Knowledge is Power” 1987<sup>20</sup>.

The modern entry of the word ‘chitabelnyi’ in Wiktionary (2018), the free dictionary, registers two meanings which are also marked as colloquial: “1. coll. suitable for reading, worth reading; 2. coll. easy, pleasant to read; readable”<sup>21</sup>.

The word is also registered in the electronic Dictionary of Russian ARGO with the meaning ‘something you can read (about books, etc.)’<sup>22</sup>. Thus, Russian dictionaries do not register the meaning ‘legible’, even though we find contexts where it is realized from the late 1970s: “Factors hampering the readability of the “narrow” (“viczo”) font, are the following: density, insufficiency of the intra-letter clearance (Trudy Tiflisskogo gosudarstvennogo universiteta, Proceedings of the Tbilisi State University, 1977) “readability test: can I read the text on your business card in poor lighting conditions” (Rezak, 2008).

The context analysis proved that the first registered use of the word ‘chitabelnost’, which took place around 1920 (see Figure 2), denotes “library book checkouts”: “A common method applied by the majority of those studying a reader is based on counting digits demonstrating readability of this or that author in the library”.<sup>23</sup> “First

<sup>16</sup> Webster’s Dictionary of English Usage (1989). Merriam-Webster

<sup>17</sup> Ushakov, D. N. (1940). *Tolkovyj slovar’ russkogo yazyka* (IV th v.) [Explanatory dictionary of the Russian language]. Gosudarstvennoe izdatel’stvo inostrannyh i nacional’nyh slovarej.

<sup>18</sup> Wiktionary (2018). Chitabel’nost [readability]. <https://clck.ru/HSPSi>. Chitabel’nyj [readable]. Retrieved from: <https://clck.ru/HSPtT>.

<sup>19</sup> Slovar’ Akademik [Academician dictionary] (2017). <https://dic.academic.ru/dic.nsf/ushakov/1088802>.

<sup>20</sup> Wiktionary (2018).

<sup>21</sup> Ibid.

<sup>22</sup> Elistratov, V.S. (2000). Slovar’ russkogo ARGO [The Dictionary of Russian ARGO]. *Russkie slovari*. <https://clck.ru/HSPkc>.

<sup>23</sup> Golos Rabochoho Chitatel’ya [Voice of a Working Reader]. (1929). *Krasnaya Nov* [Red New], 5, 212. <https://clck.ru/HSPG3>.

and foremost, the writer's readability is worth mentioning. For instance, one of the mobile libraries reports 21 copies of Lavrenev books to have been read almost 87 times within a month period – that is from January the 10<sup>th</sup> until February the 14<sup>th</sup>"<sup>24</sup>.

## Stage 2. Absolute and Relative Frequencies of the Words 'Readability' and 'Chitabelnost'

Ngram Viewer also provides an opportunity to calculate the absolute frequency of an n-gram (AF) in some particular year in Ngram Viewer with the following formula:

$$AF = RF \times 0.01 \times T \text{ (Google Ngram Viewer, 2013),}$$

where AF is Absolute Frequency,  
RF is Relative Frequency,  
T is the total number of tokens in the corpus of that particular year.

The total numbers or the raw data for the English corpus are available online in the file 'total\_counts' (Risi, 2016). As the total counts for year 1922 is 1413237707 and the relative frequency is 0.0000050374 (see Figure 3) we compute Absolute Frequency of 'readability' (AF<sup>R</sup>) as follows:

$$AF^R = 0.0000050374 \times 0.01 \times 1413237707 = 711.90 \approx 712.$$

In other words, the data point of year 1922 on the graph is caused by 712 appearances of 'readability' in texts published in 1922.

But smoothing makes frequencies look more stable and with the standard setting of 3 for smoothing<sup>25</sup>, 712 is in fact the average number of appearances for seven years: three years before 1922 (1919, 1920, 1921), year 1922 and three year after (1923, 1924, 1925).

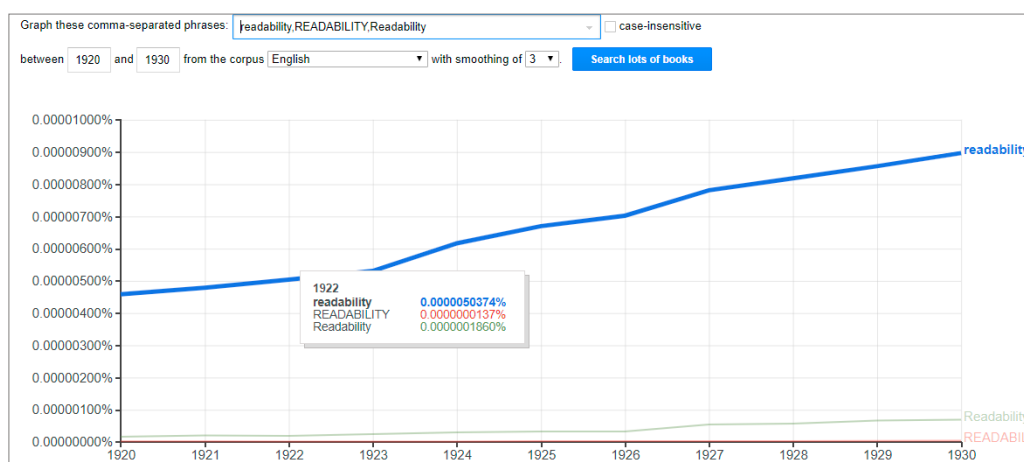
The same is true for the Russian version of the Ngram Viewer.

The relative frequency of the word 'readability' in American discourse in 1922 was 0.0000050374% (Fig.3 above) i.e. 38.37 times higher than that of the Russian term – 0.0000001530% (see Figure 4. below).

For the purposes of our research we set smoothing to 0, which gave the yearly, not average, values of the relative frequencies of 'chitabelnost' (RF<sup>C</sup>) and 'readability' (RF<sup>R</sup>) (see Figure 5, Figure 6 below). The procedure also made the peaks on the graph look higher and the pits lower.

Figure 3

Relative Frequency of Readability in the English Discourse in 1922, Standard Smoothing 3



<sup>24</sup> Shteiman, Z. (1928). Literatyrnie epizody [Literary episodes]. <https://books.google.ru/books?id=3uoGAQAIAAJ>

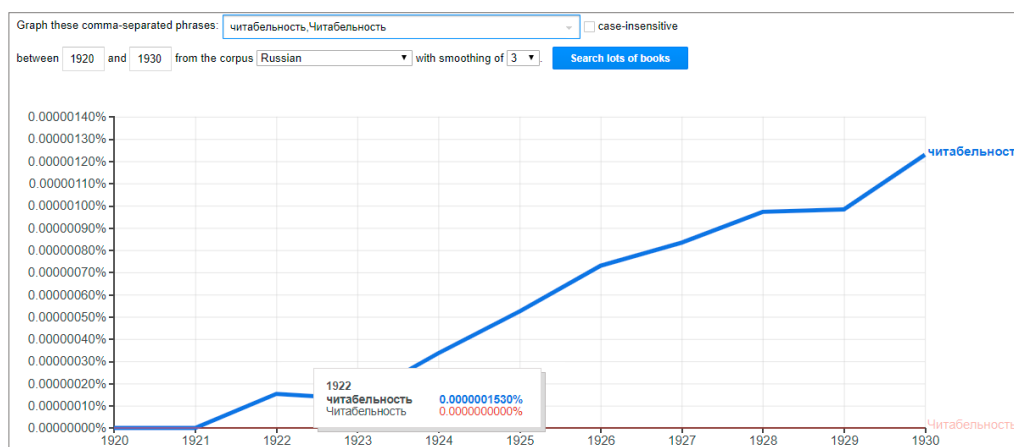
<sup>25</sup> Google Ngram Viewer.



## MEANING AND CONTEXT OF ‘READABILITY’ AND ‘CHITABELNOST’

Figure 4

Relative Frequency of ‘Chitabelnost’ (Readability) in Russian Discourse in 1922, standard Smoothing 3



Another finding of the procedure is the absence of any recorded Russian texts before 1925. In 1925,  $RF^C$  was as low as 0.0000009182 (Figure 6 below). The corresponding value on the ‘readability’ graph –  $RF^R$  – reached the level of 0.0000065810 and that is 6.8 times higher than  $RF^C$ . With 1113107246 tokens registered in 1925 (Risi, 2016), we calculate  $AF^R$  in the year 1925 as follows:

$$AF^R \text{ in } 1925 = 0.0000065819 \times 0.01 \times 1113107246 = 699.99 \approx 700.$$

As Google Viewer does not provide the raw data for the Russian corpus, we put the actual figures of  $RF^R$  and  $RF^C$  in ratio terms and simplify the ratio as follows:

$$0.0000065819 : 0.0000009182 = 7.1 : 1.$$

As absolute frequencies are to be in the same ratio, we calculate  $AF^C$ :

$$RF^R : RF^C = AF^R : AF^C; \quad 7.1 : 1 \approx 700 : 102.9 \approx 103.$$

Thus, in 1925 we may expect to find 700 records of the word ‘readability’ in American discourse and about 103 records (6.8 times fewer) of the word ‘chitabelnost’ in Russian discourse.

Figure 5

Relative Frequency of the Word ‘Readability’ in English Discourse 1925 – 2000s, Smoothing 0

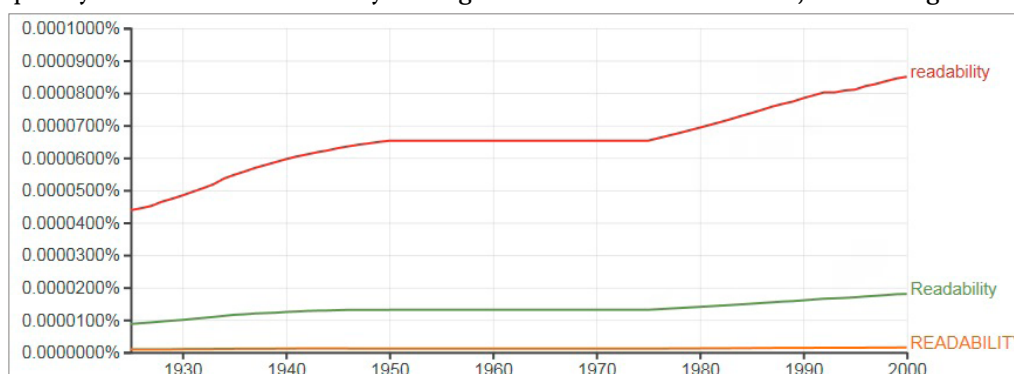
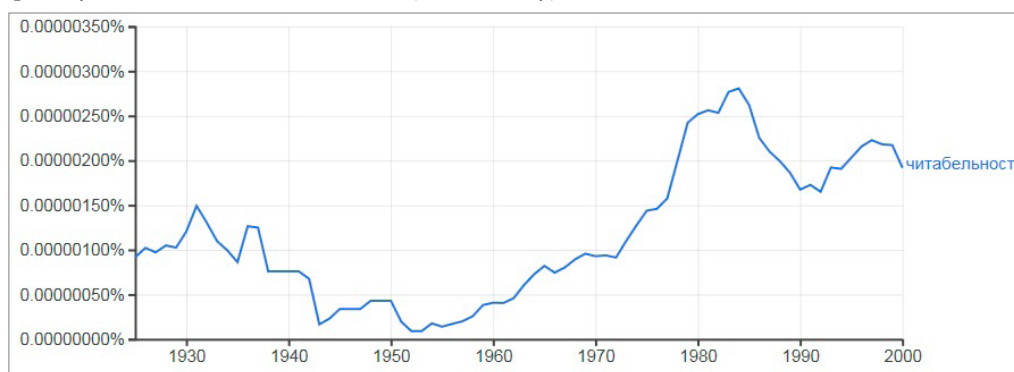


Figure 6

Relative Frequency of the Word 'Chitabelnost' (Readability) in Russian Discourse in 1925, Smoothing 0



### Stage 3. Topical Context Analysis, 1920s–2019

From the late 19<sup>th</sup> century to the present, Russian and American research schools experienced a number of explosions of interest in the area: in the 1950s and 1980s in the USA and in the late 1920s, 1980s, and 2000s in Russia (see Figures 1, 2 above), representatives of these two schools developed original theories and provided a number of tools to measure text readability. In different years, readability studies focused on various aspects of the effects of textual variables on readers' comprehension.

The rise of the frequency of the word 'readability' in English discourse in the early 1920s (see Figure 1) reflects 'the revolution' in the research area, when the first readability formulas based on linear regression model were published. Those were the years when "librarians and educators realized the need of providing appropriate reading material to people of various reading levels" (Vieth, 1988, p. vii). In those years, close attention was paid to the legibility and illustrations of reading materials as well as vocabulary and sentence length of reading texts (Kitson, 1921). Readability formulas were also developed in the 1920s: the first formula – by Lively and Pressey (1923), then – by Vogel and Washburne (1928). In those years, the concept of readability was viewed as a function of two metrics: word count and sentence length.

A noticeable rise of interest in the problem of text readability in Russian research writing was stimulated mostly due to numerous illiteracy eradication projects, when adults learning to read found children's textbooks offered to them too difficult. In those days, the problem of matching a text and a reader was mainly solved with the help of experts' judgment: "metrics of readability" of a text were aligned in the collection annotated by experts, with the level of text readability being assessed by an expert (see Karpov, 2014). We did not find recordings of any empirical research on the concept of readability in the Russian corpus of Google Books. Paradoxically the adjective 'chitabelnyy', with the meaning of 'readable, easy or enjoyable to read'<sup>26</sup>, was used by Lenin in his letter to Karpinsky, probably in 1917: "We decided to publish the attached manifesto instead of not readable theses" (Lenin, 1917, p.10).

*In the 1930s*, Thorndike published a number of profound studies on frequency of words in English discourse: his "Teacher's Word Book of 20,000 Words" (1932) and "Teacher's Word Book of 30,000 Words" (1944) were used as instruments of assessing text readability by generations of educators (Thorndike 1916, 1921, 1932, 1944, 1974). In the English tradition of the 1930s, reading was viewed as one of the aspects of mass communication and Miller developed pros and cons of illustrations in reading texts (see Miller, 1937).

*In the 1940s*, Johnson made a list of readability factors: "sentence length, difficulty of words, personal pronouns, prepositional phrases, monosyllables, and affixes have an advantage over some of the other factors influencing readability (Johnson, 1947). When evaluating the readability of educational material, scholars in the 1940s also addressed the impression produced by illustrations in children's reading materials, thus interpreting readability as a category related to the quality and number of pictures used in the text (Halbert, 1944; Strang, 1941). In the 1940s, Paterson and Tinker also introduced the term 'relative readability', which was in those times used in reference to the type of font and print that make a text legible or decipherable: "The relative readability of newsprint and book print" (Paterson & Tinker, 1946).

<sup>26</sup> Online Oxford English Dictionary (OED) (3rd ed.) (1996). <https://en.oxforddictionaries.com/>

*In the 1950s*, researchers continued to explore the effectiveness of the readability formulas developed in the 1920s: "This evaluative study of readability formulas is based on ratings of 52 books and 3 reading tests <...>. Suggestions are given for making adequate ratings of readability without excessive expenditure of time" (Klare, 1952, p.385). "Recall and prediction scores were correlated with Flesch and Dale-Chall readability scores. All the correlations were positive and high. Both readability formulas showed a higher correlation with learning than with prediction, but the difference was not significant" (Rubenstein & Aborn, 1958, p.28). The range of the metrics offered in those times was rather limited: "Authors of recent formulas have emphasized certain structural aspects of readability: (1) vocabulary level, (2) sentence length and structure, and (3) human-interest words" (Peterson, 1955, p.455).

*In the 1960s*, readability studies were accelerated by the two outstanding investigations: (1) the Readability Graph named after its designer, Fry, who claimed the Readability Graph to be suitable for estimation of text readability "for all ages, from infant to upper secondary" (Fry, 1968, p. 514); and (2) the SMOG Readability Formula developed in 1969 by a professor at Syracuse University, G. Harry McLaughlin. The formula estimated the age of a reader of a prosaic work based on the calculation of the square root of the proportion of polysyllabic words in the text. The research, known as the SMOG Readability Formula, triggered a number of studies published in English (see Figure 1) and strengthened the notion of readability as "the degree to which a given class of people find certain reading matter compelling and comprehensible" (McLaughlin, 1969).

Thus, the interest in text readability in American discourse was stable throughout the period from the 1930s to the 1960s, as those were the years when researchers widened the range of the parameters influencing text readability and validated readability formulas (see Dale, 1967; Flesch 1949, 1964; Gunning, 1952; Klare & Buck, 1954). Among the 'subjective' factors added to the notion in those years were readers' reading ability, cultural background and knowledge, and readers' motivation.

The Google Books corpus does not provide Russian texts exemplifying the development of the concept 'chitabelnost' from the 1930s to 1960s. We may also argue that American and Russian research paradigms of those times had limited connections: neither reflection on the Flesch readability index nor the SMOG Readability Formula or publications of McLaughlin (1969) are found in Russian discourse before the 1970s.

*In the 1970s*, Estonian researcher Y. Mikk conducted a series of well-acknowledged studies on the readability of textbooks translated from Russian into Estonian. He grounded three levels of noun abstractness and argued that readability is to be viewed as a function of two variables: the average length of a sentence in the text and noun abstractness. His readability formula, adapted for the Estonian language, had two variables:  $\text{Readability} = (0.131 \times \text{average sentence length in characters}) + (9.84 \times \text{average abstractness of repeated nouns}) - 4.59$  (Mikk, 1974).

Mikk's works (1974, 1981) contributed greatly to the development of the concept 'chitabelnost' in Russian discourse. Although in the early 1970s, the term 'chitabelnost' was still functioning with the meaning "the quality of being legible or decipherable" (see Aron, 1972), those were the years when the first empirical studies of readability were successfully conducted in Russia. In 1976, Mackovskij developed and introduced the first readability formula for the Russian language:  $\text{Readability} = (0.62 \times \text{average sentence length}) + (0.123 \times \% \text{ of } \geq 3\text{-syllable words}) + 0.051$  (Mackovskij, 1976).

Unfortunately after the publications of Matskovskiy (1976), Mikk (1981), and Tuldava (1975), interest in the subject gradually faded, the developed readability formulas were not validated on texts of different genre, and for another decade reports on readability experiments were not published. The list of quantitative parameters of Russian texts readability suggested by the Russian scholars of that period was much shorter than the corresponding English one and included the following: the average sentence length (word counts), percentage of words which have more than three syllables, percentage of words of 11 or more letters, noun abstractness (Mackovskij, 1976; Mikk, 1981; Tuldava, 1975).

American readability discourse during this decade developed rapidly, widening the range of text readability variables. The Coleman-Liau index of readability, which appeared in 1975, caused another rise of publications as seen on the graph in Figure 1. The index estimates the number of years of formal education a reader requires to comprehend a text on their first reading and is computed based on two variables, (1) L. i.e. the average number

of letters per 100 words and (2) S, i.e. the average number of sentences per 100 words. Thus, the readability of a text is measured based on the following formula:  $R = 0.0588 L - 0.296S - 15.8$  (Coleman & Liau, 1975).

In the middle of the 1970s, extensive studies were also conducted on the impact of syntax on readability of English texts. Empirical data from Dawkins argued that “because oversimplistic syntax violates basic principles of clear writing, it would be fair to conclude that it is actually difficult to process” (Dawkins, 1975, p.36), thus predicting the introduction of referential and deep cohesion as text readability variables in the 2000s. Although characterizing the state of affairs in the area the scholar also admitted that “Even in our area of syntax (an aspect of readability about which we do know something) we are uncertain about many analyses, lacking in empirical data, amazed by the complexity and variety of elements, clumsy in our methods, and doubtful of our oversimple results” (Dawkins, 1975, p.44).

In the discourse of the mid 1970s, the concept ‘chitabelnost’ was also actively discussed in the contexts of famous publications of Russian physiologists Zimniya and Leontyev (Zimnyaya, Dridzhe, & Leontyev, 1976). In 1976, Leontyev wrote: “All the variety of factors affecting the readability of printed publications can be reduced to four main groups. These factors are related to: 1) the content of this material; 2) the form or style of its presentation; 3) the organization of the material (the sequence of basic and secondary positions, division into paragraphs, concluding phrases, etc.) and 4) its external design (font, illustrations, cover etc.) (Leontyev, 1976). It was this research that was followed by numerous studies exploring the idea (see Figure 2 above).

The topical context of Russian discourse of *the 1980s* demonstrates the focus of Russian researchers on both qualitative and quantitative text parameters: “Linearity and connectivity are therefore the most essential characteristics of any text, providing its ‘readability’. As a guarantee of the existence of the text, ‘readability’ thus acts as a function of the context” (Filologicheskie nauki [Philological Sciences], 1988, p.160). There were numerous studies on adapting the Flesch-Kincaid formula for the Russian language: “The average readability of text A measured with the Flesch formula is 27.7, text B - 36.9. Thus, the average readability of the textbook is effective, since according to Flesch, texts are readable if their readability is 15-20 or higher” (Mutt, 1984). On the other hand, it was admitted that “In the Russian literature the term ‘chitabelnost’ is not generally accepted” (Leonov & Elepov, 1986) and researchers used the terms ‘trudnost’ (Rus. text difficulty), ‘prostota’ (Rus. text ease), or ‘sloznost’ (Rus. text complexity) interchangeably.

The peak in the occurrence for the word ‘readability’ (Figure 1) was registered in the 1980s, with the highest number (-0.000125) in 1986. In 1984, Barr, Pearson, and Kamil published Volume 1 of “The Handbook of Reading Research” where they defined readability as a notion with three separate meanings: “1. Legibility, of either the handwriting or the typography. 2. Ease of reading, owing to the interest-value of the writing. 3. Ease of understanding, owing to the style of writing” (Pearson, 1984, p.681). The scholars argued that “Though the first and second meanings still occur, usage now clearly favors the third meaning, especially in the field of reading” (Pearson, 1984, p.681). Illustrations of that particular sense in the Google Books corpus are numerous: “his novels have few equals in readability, a sometimes deceptive readability” (Killam, 1984, p.192). “Queneau’s novels are as remarkable for the sheer readability of their stories as for their other qualities” (p.75) (Shorley, 1985). It was also in the 1980s when P. Fries (1986) equated readability and coherence stating that “mere counting of the language forms contained in a text will not lead to useful judgments of the readability or coherence of that text” (Fries, 1986, p.13).

*In the 1990s and early 2000s*, the peak from the 1980s was followed by a decrease similar to that observed in the late 1970s. In the 1990s, readability researchers addressed paragraph length as a function of text readability and proved that an average reader has a more positive attitude to texts with short paragraphs of fewer than 100 words than to longer paragraphs (Markel, Vaccaro, & Hewett, 1992). Another study quite extensively quoted in the 1990s was “The effect of syntactic simplification on reading EST texts as L1 and L2” (Ulijn & Strother, 1990). Their empirical study validated the hypothesis that syntactic complexity “does not significantly affect the level of reading comprehension for both expert and novice readers in a particular professional field” (Ulijn & Strother, 1990, p.54). In the 1990s, American discourse registered two main meanings of the word: 1) ‘the quality of being legible or decipherable’ as in “As illuminance contrast decreased, readability also decreased. However, the relationship between illuminance contrast and readability was direct” (Lomperski, 1995, p. iii); and 2) ‘the quality of being easy or enjoyable to read’ as in “He examines what each formula is good for and based on vocabulary grading tentatively recommends the Gunning and Fry Readability formula which he successfully applies to

<...>" (Chia, 1998, p.37). Researchers continued discussing readability formulas: "Formally determining grade level according to a standard, computer-based readability formula would heighten the awareness of the those responsible for writing the impartial analysis to the problem of public understanding" (Dubois & Floyd, 1998, p. 178) and Microsoft Word 97 was the first to display readability statistics: Flesch Reading Ease score, Flesch-Kincaid Grade Level score, and passive sentence occurrences. In 1998, Sides uses 'usability' as a synonym of 'readability' and made the revolutionary conclusion that "Correctly structured with a feel for rhythm the ebb-and-flow of a phrase, even the long sentence is readable" (Sides, 1999, p.11), thus extending the list of potential metrics affecting readability.

Russian discourse from the 1990s provides examples of 'chitabelnost' in institutional IT discourse where the term is used to modify the word program: "It is also worth noting such advantages of ADA language as modularity, structurality, readability, and documentability of the programs" (Gricenko, 1991, p.65). Topical contexts of the word 'chitabelnost' identify the meaning "the quality of being easy or enjoyable to read<sup>27</sup>", but no modifications to the concept in the 1990s were registered. A possible explanation to the situation could be the reluctance of Russian researchers to use a borrowed word and prefer 'trudnost' (Rus. text difficulty) and 'sloznost' (Rus. text complexity), although verification of this hypothesis is beyond the scope of this article.

*In the 2000s*, the highest peak of 'readability' frequency took place in 2005 when RFR reached 0.0000998954%, which means the Google Books corpus compiled about 125 texts with the word 'readability' ( $AFR = 0.0000998954 \times 0.01 \times 12519922882 = 125.06 \approx 125$ ). The 2005 vertex was mostly obtained due to "The Principles of Readability" published by DuBay (2004) in August 2004. It soon became one of the most cited articles in the area, which gave a profound boost to studies of readability and readability formulas. Researchers from that time also addressed the disadvantages of using readability formulas<sup>28</sup> and offered online tools to measure readability<sup>29</sup> based on multiple correlation analysis<sup>30</sup>.

The Russian research on 'chitabelnost' in those years was mostly focused on websites and texts being legible<sup>31</sup> ([http://www.bhv.ru/books/full\\_contents.php?id=419](http://www.bhv.ru/books/full_contents.php?id=419)). Another popular research object was text and technical drawings layout<sup>32,33</sup>.

It was also in the *early 2010s* when school textbooks began being assigned 'reading levels'<sup>34</sup> and automatic readability tools started computing readability levels, not only of public speeches (Kayam, 2018; Schumacher & Eskenazi, 2016) but students' writings as well (Peng, 2015). The theme of the decade was the readability of books<sup>35</sup> and webpages (Taylor, 2017; Boztas et al., 2017) with the goal for writers defined by Kirk as "to practice composing a sentence that requires only one reading to decipher the intended message"<sup>36</sup> and their the intended audience.

*In the 2010s*, a number of researchers in the area validated correlations between companies' documents readability and their business performance "indicating that companies with stronger CSR<sup>37</sup> performance are more likely to have CSR reports with higher readability" (Wang, Hsieh, & Sarkis, 2018, p.66, see also Kim, Wang, & Zhang, 2019; Bonsall & Miller, 2017). Another area of interest at that time was the readability of health information for patients, the representatives of which indicated that writers of medical information on diseases and possible treatment "overestimate the reading ability of the overall population which may "have its greatest impact among those with low literacy and limited access to health care" (Storino et al., 2016, p.831). However, the primary focus of the research *in the late 2010s* was on "the identification of the linguistic features that predict text readability judgments, and how these features perform when compared to traditional text readability formulas such as the Flesch-Kincaid grade level formula" (Crossley et al., 2017, p.340). The new

<sup>27</sup> Lexico. Oxford English Dictionary. <https://www.lexico.com/definition/readability>

<sup>28</sup> Readability formulas. (2004). <http://www.readabilityformulas.com/articles/advantages-and-disadvantages-of-readability-formulas.php>

<sup>29</sup> Free readability formulas. (2005). <http://www.readabilityformulas.com/articles/advantages-and-disadvantages-of-readability-formulas.php>

<sup>30</sup> Methods for measuring text readability. (2005). <http://www.standards-schmandards.com/2005/measuring-text-readability/>

<sup>31</sup> BHV. Web site svoimi rekami [BHV. Website. DIY]. [http://www.bhv.ru/books/full\\_contents.php?id=419](http://www.bhv.ru/books/full_contents.php?id=419).

<sup>32</sup> Google Books Corporate Social Responsibility. (2004). Readability. <https://clck.ru/BUwT>

<sup>33</sup> Google Books. (2004). Readability. <https://clck.ru/BUwp>

<sup>34</sup> Google Books. (2005). Readability. <https://clck.ru/BUxD>

<sup>35</sup> Google Books. (2015). Readability. <https://clck.ru/BUyF>

<sup>36</sup> Kirk, K. (2010). Writing for Readability. Association for Talent Development.

<sup>37</sup> CSR - Corporate Social Responsibility.

findings included evidence that “the traditional readability formulas are less predictive than models of text comprehension, processing, and familiarity derived from advanced natural language processing tools” (Crossley et al., 2017, p. 340).

Thus, to answer the research questions on similarities/differences of the words ‘readability’ and ‘chitabelnost’ in American and Russian discourse over the period from 1925 to 2019, we brought together the resources that did not only help to evaluate the development of the concept in two different academic environments, but are also of great scientific value on their own. We used the Google Books Ngram Viewer to observe trends in the usage of the terms ‘readability’ and ‘chitabelnost’ as well as the stages of concept formation. The major obstacle to this endeavor was the inaccessibility of some of the resources, which caused low compatibility in the datasets of the 1920s to 1960s.

## Conclusion

The study demonstrated language-specific features of the semantic values of the words ‘readability’ and its Russian analogue ‘chitabelnost’. The American word ‘readability’ has had two meanings verified in the contexts of the studied period: ‘the quality of being legible or decipherable’ and ‘the quality of being easy or enjoyable to read’. The word ‘chitabelnost’ has been used in Russian discourse with three meanings: two are similar to the English mentioned above, and the third, “a number of library book checkouts”, was used only in the late 1920s. Russian dictionaries published between the 1920s and the 2000s register only one meaning of the noun ‘chitabelnost’, defining it as a derivative of the adjective ‘chitabelnyi’ (Russian ‘readable’). Another difference between the word ‘chitabelnost’ and ‘readability’ is that it is marked in the dictionaries as colloquial, but its active functioning in Russian academic and scientific discourse over the period from the 1920s to the 2000s testifies to its belonging to the high register of communication.

The research also showed that debates over the concept of readability have involved a certain level of disagreement on the range of variables affecting it. All the existing approaches to the notion of ‘readability’ differ in the number of parameters they explore. American and Russian schools experienced peaks of activities in the 1980s when scientists advanced the concept to the next level of its development and added a number of new constituents to its structure. By the 2000s, the American concept of ‘readability’ had manifested in a wide range of scientific publications and was viewed as a function of lexical, syntactic, semantic, and pragmatic parameters of the text. As for the Russian school of text readability of that period, their records of achievements were mostly gained in the 1970s and 1980s. By the 2000s, the Russian school had accumulated extensive data on qualitative parameters of readers’ cognitive and linguistic abilities, but possessed limited resources in the quantitative domain: text readability was estimated based on four variables only, i.e. word count, word length, sentence length, and noun abstractness as functions.

The features summarized in the article as ‘predicting readability’ provide a stable ground for further research on the metrics and parameters of different readability levels both in English and Russian.

## Acknowledgements

This research was financially supported by the Russian Science Foundation, grant No 18-18-00436.

## Conflict of interests

The authors declare that they have no conflict of interest.

## References

Aron, E. I. (1972). *Metody issledovaniya i proyektirovaniya organizatsii truda na predpriyatii* [The methods of research and planning the work organization in the enterprise]. <https://books.google.ru/books?id=K7cwAQAAIAAJ> –

- Baldwin, T. S., & Bailey, L. J. (1971). Readability of technical training materials presented on microfiche versus offset copy. *Journal of Applied Psychology*, 55(1), 37-41. <http://psycnet.apa.org/record/1971-21933-001>
- Blom, M., & Blom, T. (1983). *Canada Home: Juliana Horatia Ewing's Fredericton Letters, 1867-1869*. University of British Columbia Press.
- Bonsall, S. B., & Miller, B. P. (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies*, 22(2), 608-643. <https://doi.org/10.1007/s11142-017-9388-0>
- Boztas, N., Omur, D., Ozbilgin, S., Altuntas, G., Piskin, E., Ozkardesler, S., & Hanci, V. (2017). Readability of internet-sourced patient education material related to "labour analgesia". *Medicine*, 96(45), 1-5. <https://clck.ru/JBxuL>
- Chia, E.N. (1998). Guide to readability in african languages. *Linguistic Society of America*, 74(3), 665. <https://muse.jhu.edu/article/451344/pdf>
- Choldin, M. T. (1979). Rubakin, Nikolai Aleksandrovic. In A. Kent, H. Lancour & J. E. Daily (Ed.), *Encyclopedia of library and information science* (pp. 178-179). CRC Press.
- Coleman, M., & Liau, T. L. (1975). A Computer Readability Formula Designed for Machine Scoring. *Journal of Applied Psychology*, 60(2), 283-284. <https://doi.org/10.1037/h0076540>
- Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., & Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, 54(5-6), 340-359. <https://doi.org/10.1080/0163853X.2017.1296264>
- Dale, E. (1967). *Can you give the public what it wants?* World Book Encyclopedia.
- Dash, N. S. (2008). Context and contextual word meaning. *SKASE Journal of Theoretical Linguistic*, 5, 21-31. [http://www.skase.sk/Volumes/JTL12/pdf\\_doc/2.pdf](http://www.skase.sk/Volumes/JTL12/pdf_doc/2.pdf)
- Dawkins, J. (1975). *Syntax and Readability*. International Reading Association. <https://files.eric.ed.gov/fulltext/ED103814.pdf>
- DuBay, D. H. (2004). The principles of readability. <https://ru.scribd.com/document/346200577/Dubay-principles-of-readability-pdf>.
- Dubois, P. & Floyd, F. (1998). Lawmaking by initiative: issues, options, and comparisons. Agathon.
- Fellbaum, C. (Ed.) (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
- Fernbach, N. (1990). La lisibilité dans la réaction juridique au Québec [Legibility in the legal reaction in Quebec]. Le Centre de promotion de la lisibilité, Centre Canadien d'information juridique.
- Filologicheskie nauki [Philological Sciences]. (1977). Trudy Tiflisskogo gosudarstvennogo universiteta [Proceedings of the Tbilisi State University]. Tbilisi University Press.
- Filologicheskie nauki [Philological Sciences]. (1988). *Nauchnye doklady vysshej shkoly* [Scientific Reports of the Higher School]. Vysshaya shkola.
- Firth, J. R. (1957). A synopsis of linguistic theory. In *Philological society: Studies in Linguistic Analysis* (special vol., pp. 1-32). Basil Blackwell.
- Flesch, R. (1949 and 1974). *The art of readable writing*. Harper.
- Flesch, R. (1964). *The ABC of style: A guide to plain English*. Harper.
- Fries, P. (1986). Language features, textual coherence and reading. *Word*, 37, 13-29. <https://doi.org/10.1080/00437956.1986.11435764>
- Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading*, 11, 513-516. <http://www.jstor.org/stable/40013635>
- Gray, S. (2019). Readability of Patient-reported Outcome Measures for Persons with Aphasia. <https://clck.ru/JAtzz>
- Gricenko, V.I. (Ed.) (1991). *Upravlyayushchie sistemy I mashiny* [Control systems and machines]. Nauk. Dumka.
- Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
- Hai-Jew, S. (2014). *Querying google books ngram viewer's big data text corpuses to complement research*. <https://www.researchgate.net>
- Halbert, M. G. (1944). The teaching value of illustrated books. *American school board journal*, 108(5), 43-44.
- Hiebert, E. H. (2012). *The common core state standards and text complexity*. Text Project & University of California.
- Ivanov, V. V. (2013). K voprosu ob ispolzovanii lingvisticheskikh kharakteristik slozhnosti teksta pri issledovanii okulomotornoy aktivnosti pri chtenii u podrostkov [Back to the linguistic characteristics of the complexity of the text in the study of oculomotor activity in reading in adolescents]. *New research*, 2, 42-50. <https://cyberleninka.ru/article/>
- Johnson, K. G. (1947). *Factors influencing the readability of science*. University of Wisconsin.
- Karpov, N. V. (2014). *Identifikatsiya urovnya slozhnosti teksta i ego adaptatsiya* [Identification of the complexity level and its adaptation]. <http://www.slideshare.net/karprv/ss-31225145#14356960593761&fbinitialized>

- Kayam, O. (2018). The readability and simplicity of Donald Trump's language. *Political Studies Review*, 16(1), 73–88. <https://doi.org/10.1177/1478929917706844>
- Kerlinger, F. N. (1973). *Foundations of behavioral research* (2nd ed.). Holt Rinehart and Winston.
- Killam, G. D. (Ed.) (1984). *The writing of East and Central Africa*. Heinemann Educational Books Ltd.
- Kim, C., Wang, K., & Zhang, L. (2019). Readability of 10-K reports and stock price crash risk. *Contemporary accounting research*, 36(2), 1184–1216. <https://doi.org/10.1111/1911-3846.12452>
- Kitson, H. D. (1921). *The mind of the buyer*. Macmillan.
- Klare, G. R. (1952). Measures of the readability of written communication: an evaluation. *Journal of Educational Psychology*, 43(7), 385–399. <https://doi.org/10.1037/h0058972>
- Klare, G. R. (2000). Readable computer documentation. *ACM journal of computer documentation*, 24(3), 148–168. <https://doi.org/10.1145/344599.344645>
- Klare, G.R., & Buck, B. (1954). *Know your reader, the scientific approach to readability*. Hermitage House.
- Lenin, V. I. (1917). *Polnoye sobraniye sochineniy* [Complete Works]. <http://uaio.ru/vil/49.htm>
- Leonov, V. P., & Elepov, B.S. (1986). *Referirovanie I annotirovanie nauchno-tekhnicheskoy literatury* [Precis-writing and annotating of the scientific and the technical literature]. Nauka.
- Leontyev, A. A. (1976). *Smyslovoe vospriyatie rechevogoso obshcheniya (v usloviyah massovoj kommunikacii)* [Semantic perception of speech communication (in the conditions of mass communication)]. Nauka.
- Lively, B. A., & Pressey, S. L. (1923). A method for measuring the 'vocabulary burden' of textbooks. *Educational administration and supervision*, 9, 389–398. <https://doi.org/10.1080/00220671.1931.10880190>
- Lomperski, T. J. (1995). *Enhancing interior building sign readability for older adults: Lighting color and sign color contrast*. University of Wisconsin.
- Macias-Chapula, C.A. (Ed.) (1999). *Seventh conference of the international society for scientometrics and informetrics*. Universidad de Colima
- Mackovskij, M. S. (1976). *Problemy chitabel'nosti pechatnogo materiala. Smyslovoe vospriyatie rechevogo soobshcheniya v usloviyah massovoj kommunikacii* [Problems of readability of printed material. Semantic perception of speech messages in conditions of mass communication]. Nauka.
- Markel, V., Vaccaro, M., & Hewett T. (1992). Effects of paragraph length on attitudes toward technical writing. *Technical Communication*, 39(3), 454–456. <http://www.jstor.org/stable/43090117>
- McLaughlin, H. (1969). SMOG grading a new readability formula. *Journal of reading*, 22, 639–646. [http://webpages.charter.net/ghal/SMOG\\_Readability\\_Formula\\_G\\_Harry\\_McLaughlin\\_\(1969\).pdf](http://webpages.charter.net/ghal/SMOG_Readability_Formula_G_Harry_McLaughlin_(1969).pdf)
- Mikk, J. A. (1974). Metodika razrabotki formul chitabel'nosti [Methods for developing readability formulas]. *Sovetskaja Pedagogikai Skola*, 9, 78–163.
- Mikk, Y. A. (1981). *Optimizaciya slozhnosti uchebnogo teksta: V pomoshch' avtoram i redaktoram* [Optimizing the complexity of the text: To help authors and editors]. Prosveshchenie.
- Miller, W. A. (1937). Reading with and without pictures. *Elementary School Journal*, 38, 676–682. <https://doi.org/10.1086/462248>
- Mutt, O. (1984). *Aktual'nye voprosy otbora uchebnogo materiala dlya vuzovskogo kursa inostrannogo yazyka* [Actual questions of selection of the educational material for the university course of a foreign language]. The Tartu State University.
- Paterson, D. G., & Tinker, M. A. (1946). The relative readability of newsprint and book print. *Journal of Applied Psychology*, 30(5), 454–459. <https://doi.org/10.1037/h0054012>
- Pearson, P. D. (Ed.) (1984). *Handbook of reading research*. Lawrence Erlbaum Associates.
- Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books Corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLOS ONE*, 10(10), e0137041. <https://doi.org/10.1371/journal.pone.0137041>
- Peng, C. (2015). Textbook readability and student performance in online introductory corporate finance classes. *The Journal of Educators Online-JEO*, 13(2), 35–49. <https://doi.org/10.9743/JEO.2015.2.6>
- Peterson, E. M. (1955). Aspects of readability in the social studies. *Journal of Applied Psychology* 39(2), 141–142. <https://doi.org/10.1037/h0039753>
- Randal, R. (Ed.) (2013). *Novel and short story writer's market*. Writer's digest books.
- Ravin, Y., & Leacock C. (2000). *Polysemy: Theoretical and computational approaches*. Oxford University Press.
- Rezak, D. (2008). *Svyazi reshayut vse. Biznes-skazka o Carevne-lyagushke* [Communications solve everything. Business tale about the Frog Princess]. MIF.
- Risi, S. (2016). *Google Ngrams: From relative frequencies to absolute counts*. [http://stanford.edu/~risi/tutorials/absolute\\_ngram\\_counts.html](http://stanford.edu/~risi/tutorials/absolute_ngram_counts.html)
- Rubenstein, H., & Aborn, M. (1958). Learning, prediction, and readability. *Journal of Applied Psychology*, 42(1),



- 28-32. <https://doi.org/10.1037/h0039808>.
- Schumacher, E., & Eskenazi, M. (2016). *A readability analysis of campaign speeches from the 2016 US Presidential campaign*. Carnegie Mellon University.
- Sherman, L.A. (1893). *Analytics of literature: A manual for the objective study of English prose and poetry*. Ginn and Co.
- Shorley, Ch. (1985). *Queneau's fiction: An introductory study*. Cambridge University Press.
- Sides, C.H. (1999). *How to write and present technical information*. Cambridge University Press.
- Sologub, O. P. (2002). *Usvoenie inoyazychnyh strukturnyh ehlementov v russkom yazyke* [The assimilation of foreign language structures in the Russian language]. In Proceedings of the 2002 the 3rd Scientific Conference (pp. 130-134). Nauka.
- Solovyev, V.D. (2013). A frequency-based approach to language dynamics. <https://events.spbu.ru/eventsContent/files/corpling/corpora2013/Solovjov.pdf>.
- Storino, A., Castillo-Angeles, M., Watkins, A. A., Vargas, C., Mancias, J. D., Bullock, A., Moser, A. J., Demirjian, A., & Kent, T. S. (2016). Assessing the accuracy and readability of online health information for patients with pancreatic cancer. *JAMA surgery*, 151(9), 831-837. <https://doi.org/10.1001/jamasurg.2016.0730>
- Strang, A. (1941). *A study of gains and losses in concepts as indicated by pupils' reading scores after the addition of illustrations to reading material* [Unpublished doctoral dissertation]. Temple University.
- Taylor, M. C., & Wahlstrom, M. W. (1986). Readability as applied to the assessment instrument. *International Journal for Basic Education*, 10(3), 155-170. <http://en.copian.ca/library/research/report4/rep31-35/rep34.pdf>
- Taylor, Z.W. (2017). A failure to communicate: are university websites readable? *International Journal of Education and Multidisciplinary Studies* 8(2), 149-163. doi: <http://dx.doi.org/10.21013/jems.v8.n2.p1>.
- Thorndike, E. L. (1916). An improved scale for measuring ability in reading. *Teachers college record*, 17, 40-67.
- Thorndike, E. L. (1921). *The teacher's word book*. Columbia University.
- Thorndike, E. L. (1932). *A teacher's word book of 20,000 words*. Columbia University.
- Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. Columbia University.
- Thorndike, R. L. (1973-1974). Reading as reasoning. *Reading research quarterly*, 9, 135-137. Columbia University.
- Tuldava, Yu. A. (1975). Ob izmerenii trudnosti tekstov [About the measurement of the difficulty of texts]. *Uchenye zapiski Tartuskogo universiteta*, 345(IV), 102-120. [http://dspace.ut.ee/bitstream/handle/10062/25443/acta\\_827\\_1988.pdf](http://dspace.ut.ee/bitstream/handle/10062/25443/acta_827_1988.pdf).
- Ulijn, J. M., & Strother, J. B. (1990). The effect of syntactic simplification on reading EST texts as L1 and L2. *Journal of research in reading*, 13(1), 38-54. <https://eric.ed.gov/?id=EJ405025>
- Vari-Cartier, P. (1916). Development and validation of a new instrument to assess the readability of Spanish prose. *The Modern Language Journal*, 65, 141-148. <https://clck.ru/HSPNr>
- Vieth, N. A. (1988). *Mathematics in composition: A defense of Flesch's readability formula* [Unpublished MA thesis, Iowa State University]. <https://clck.ru/HSPQD>
- Vogel, M., & Washburne, C. (1928). An objective method of determining grade placement of children's reading material. *Elementary school journal*, 28, 373-381. <https://doi.org/10.1086/456072>.
- Wang, Z., Hsieh, T. S., & Sarkis, J. (2018). CSR performance and the readability of CSR reports: Too good to be true? *Corporate Social Responsibility and Environmental Management*, 25(1), 66-79. <https://doi.org/10.1002/csr.1440>
- Wierzbicka, A. (1996). *Semantics: Primes and universal*. Oxford University Press.
- Willems, K. (2012). Intuition, introspection and observation in linguistic inquiry. *Language Sciences*, 34, 665-681. <https://doi.org/10.1016/j.langsci.2012.04.008>
- Zakharov, V., & Masevich, A. (2014). Diachronic investigations on the base of Russian corpus of Google Books Ngram Viewer. <https://www.academia.edu/8592136>
- Zhang, S. (2015). *The Pitfalls of using Google Ngram to study language*. <https://www.wired.com>.
- Zimnyaya, I. A., Dridzhe, T. M., & Leontyev, A. A. (1976). *Smyslovoye vospriyatiye rechevogo soobshcheniya: V usloviyakh massovoy kommunikatsii* [Semantic perception of the voice message: In the context of mass communication]. <https://books.google.ru/books?id=SI8ZAAAAMAAJ>.