

Assessment of Clarity of Bank of Russia Monetary Policy Communication by Neural Network Approach¹

Alina Evstigneeva, *Bank of Russia*

evstalina@gmail.com

Mark Sidorovskiy, *Bank of Russia*

mark.sidorovskiy@gmail.com

Inflation targeting requires clear and transparent central bank's communication. Analysts and market participants understand it as a broad list of information disclosed by the central bank. The general public understands it rather as the ability of a central bank to speak and explain its decisions in a plain language. In recent decades, monetary authorities in many countries have made significant progress in this direction. However, there has been no research on the quality of communication for the Bank of Russia. This paper aims to create a tool for automated evaluation of the readability of the Bank of Russia's monetary policy communication, taking into account the available experience of linguistic and textual analysis, including machine learning methods, as well as to provide recommendations for its improvement. This can contribute to improving the effectiveness of the Bank of Russia communication on monetary policy, which is vital for its credibility, anchoring inflation expectations, and predictability of the regulator's decisions.

Keywords:

*monetary policy,
communication, text analysis,
neural networks, NLP*

JEL Codes: E52, E58, D83

Citation: Evstigneeva, A.

and Sidorovskiy, M. (2021). Assessment of Clarity of Bank of Russia Monetary Policy Communication by Neural Network Approach. *Russian Journal of Money and Finance*, 80(3), pp. 3–33.

doi: 10.31477/rjmf.202103.03

1. Introduction

As the Bank of England analyst Jonathan Fullwood (Fullwood, 2016) has observed, today's central bank texts are unnecessarily complex. Understanding most articles in high quality media and most fictional texts produced by great writers

¹ The authors express their sincere gratitude to Yulia Shchadilova, Vadim Grishchenko, Maria Romanova, and Aydar Imaev for help with data processing and for informative discussions at different stages of the work. We also thank two anonymous reviewers for their comprehensive comments, which have greatly improved the paper.

requires no more than 10 years of education. Even understanding a speech from Nobel Laureate in physics Richard Feynman explaining Heisenberg's uncertainty principle would require only eight years of education. However, most central bank texts require 14 years of education.

Haldane et al. (2020) argue that central bank communications should be understandable for not only financial market participants or financial journalists, but also the broad public. First, this would allow central banks to be held accountable to the public and give the latter confidence that the 'independent central bank is fulfilling its social contract'. Second, understanding breeds credibility, which helps to better manage inflation expectations. Central bank credibility has a significant impact on economic growth and improved social well-being (Algan and Cahuc, 2014). Freedman and Laxton (2009) show that the higher the credibility, the closer inflation expectations are to the target, other things being equal. They also note that good communication supports the independence of central banks. A publicly understandable regulator confirms its political mandate and can pursue a more independent policy.

One of the pillars of inflation targeting regimes is openness, which is crucial for effective management of inflation expectations. The wide spread of those regimes has contributed to the fact that over the past three decades, central banks have transitioned from 'esoteric art' (Brunner, 1981), 'mumbling with great incoherence'² and following the 'Never apologise, never explain'³ rule to detailed explanations of the logic of their actions.

Good communication allows several things:

- 1) Provision of economic agents with guidance on further central bank actions, reducing the unpredictability of decisions (monetary policy 'surprises');
- 2) Influence on the formation of inflation expectations;
- 3) Increase of trust in the central bank and its policies.

As for Bank of Russia monetary policy 'surprises', the topic is covered in detail in the report of the International Monetary Fund (2018, Chapter 3) and in the review by Isakov et al. (2018). Both studies conclude that the predictability of Bank of Russia decisions is not high enough and that there is significant potential for improvement.

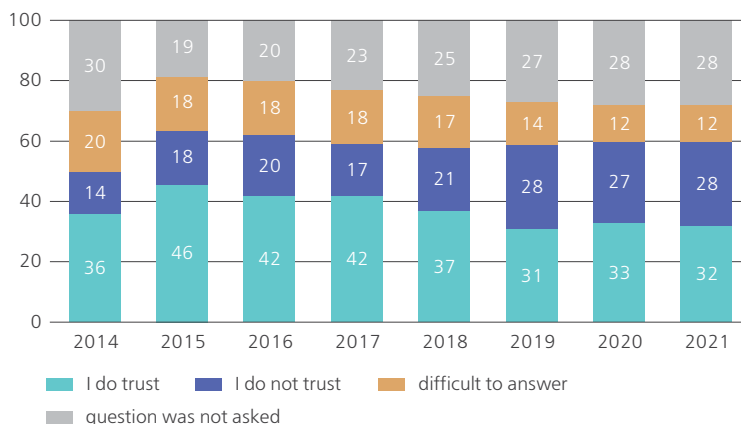
As regards inflation expectations, it can be stated that at present they are two to three times higher than the inflation target.⁴ Although inflation expectations exceed the inflation target in many countries, a several-fold excess is the exception rather than the rule.

Finally, public opinion polls show that the level of public trust in the Bank of Russia generally remains low (see Figure 1).

² Alan Greenspan, head of the US Federal Reserve System (FRS) from 1987 to 2006, noted shortly after his appointment: 'Since I've become a central banker, I've learned to mumble with great incoherence. If I seem unduly clear to you, you must have misunderstood what I said.'

³ This statement is known as the 'personal motto' of Sir Montagu Norman, the governor of the Bank of England from 1920 to 1944.

⁴ See http://cbr.ru/eng/analytics/dkp/inflation_expectations/

Figure 1. Bank of Russia credibility, % of total number of respondents

Note: Respondents were asked 'Generally speaking, do you trust the Central Bank of the Russian Federation or do you not?' The question was not asked of those respondents who had not previously heard of the Central Bank.

Source: InFOM (Public Opinion Foundation, Russia) (2021)

The literature tends to highlight two key characteristics of effective communication: transparency and readability. In this, transparency of communication refers to the degree to which the central bank discloses information about its decisions and the reasoning behind them (see, for example, Eijffinger and Geraats, 2006; Svensson, 1997; Laxton and Freedman, 2009; Dincer and Eichengreen, 2008, 2014). Readability is defined as the array of all the parameters of the text that help the reader to quickly understand the information contained in it (for more details, see Section 2, as well as the papers by Fracasso et al., 2003; Bulíř et al., 2008; Blinder, 2004; etc.). Some researchers, in particular Fracasso et al. (2003), also talk about the importance of adapting the style of communication to a specific audience. Striking examples of targeted communication, aimed in this case at the younger generation, are the comics on the subject of the US Federal Reserve's monetary policy⁵ or the Bank of Jamaica's reggae hit about inflation.⁶

Readability and transparency are important to all central bank audiences. If the central bank communicates clearly, unambiguously and fully reveals the logic of its decision-making, then it is better understood by both professional audiences (financial sector, investors, and analysts) and the wider public (real sector, population) (see Haldane and McMahon, 2018). In addition, as Robert Gunning, the creator of one of the world's most popular readability indexes, noted, the larger the target audience, the more simply the text should be written (Gunning, 1952). The audience of a national central bank is the entire population of the country, as well as foreign investors, analysts, and experts.

⁵ See <https://www.newyorkfed.org/medialibrary/media/outreach-and-education/comic-books/NewYorkFed-TheStoryOfMonetaryPolicy-WebColor.pdf>

⁶ See <https://www.youtube.com/watch?v=wm9L9lY1fiE>

The purpose of this study is to create a tool for automated assessment of the readability of economic texts in Russian taking into account existing experience in linguistic and textual analysis and using modern machine learning methods. The study also aims at drawing conclusions on how easily the general public understands the Bank of Russia's monetary policy communication.

To achieve this goal, we set the following tasks.

- 1) Study the experience of foreign and Russian linguists in assessing the readability of texts.
- 2) Study existing experience in solving the problems of assessing the readability of central bank communications.
- 3) Create a linguistic model for assessing texts that on economic topics and train it on a sufficient corpus of texts. The creation of such a corpus of texts is a subtask.
- 4) Assess the readability of the communication of the Bank of Russia and develop proposals for improvement.

The contribution of this paper to the scientific literature consists in the application of achievements from various fields of knowledge (mathematical linguistics, data science, Russian philology, and econometrics) to create a practically applicable neural network model for assessing the quality of the Bank of Russia's texts on monetary policy issues. An important advantage of the model is its interpretability: it equally takes into account the content of the text and its expert-extracted linguistic characteristics, which are responsible for the readability and quality of the text. Until now, there have been no automated tools for assessing the readability of texts on monetary policy issues to determine their comprehensibility to the general public, which is of particular importance in the context of an inflation targeting regime, with its focus on openness. Another important contribution is the compilation of a specialised training corpus out of 10,000 texts marked according to level of complexity. The simultaneous application of different scientific fields can be useful for future researchers in solving similar applied problems.

Further, the paper is structured as follows. In Section 2, we summarize the previous studies, considering the construction the classical readability indices of the mid-20th century and more advanced linguistic research on text quality determination based on different characteristics (lexical, syntactic, morphological, phonetic, discursive). We also review the previous research in text quality determination via classification of texts' readability through machine learning and natural language processing (NLP). In addition, we consider the experience of Soviet and Russian linguists in identifying the Russian text characteristics and corresponding complications. Finally, we examine existing research on the text quality of central bank communication. The main part of this paper (Section 3) models the readability of economic texts in Russian. In Section 4, we apply the trained model to Bank of Russia texts on monetary policy issues and provide some linguistic recommendations to improve their readability. In Section 5, we

outline areas for future research and develop recommendations for the practical application of the resulting model. Section 6 concludes.

2. Literature review

2.1. Linguistic prerequisites for readability and classical indices

The earliest papers on determining the readability of texts date back to the middle of the 20th century. The increasing flow of information and the search for the optimal communicative language for the target audience (schoolchildren, students, immigrants) determined the motivations of the first researchers of the written language quality (Dale and Chall, 1948; Flesch, 1948).

Dale and Chall (1948) give the following classic definition of readability: Readability is the array of all the parameters of a text that help readers quickly understand the relevant information contained in it. DuBay (2004) details the history of the research on readability. In this work, we will not repeat it, but focus on the text parameters that Dale and Chall (1948) talk about.

All classic methods for assessing readability and complexity are based on statistical metrics of the text. Conditionally upon the assessed parameters of the text, these metrics can be divided into six groups:

- 1) Syntactic (structure of sentences);
- 2) Lexical (characteristics of individual words);
- 3) Morphological (weights of individual parts of speech or the structure of the words);
- 4) Phonetic (text's euphony);
- 5) Semantic (meanings of the text's constituent words);
- 6) Discursive (internal connections of the text and the parameters of functional styles of language).

The first readability indices were based on just two basic assumptions that short sentences and short words are easier to read than long ones. The famous Flesch index (Flesch, 1948), which is still used today, includes only one syntactic (average sentence length) and one lexical (average word length) characteristic of the text. Calibration data were obtained by empirical examination of a test set of texts, mapped for complexity by the author. Accordingly, if the same study is conducted on a different dataset, the coefficients in the formula may be different. In 1975, Kincaid et al. (1975) refine this methodology at the request of the US Navy, who wanted a convenient tool for assessing the simplicity of instructions for sailors. The output of the Flesch-Kincaid test is the number of years of study in the American educational system required to understand a text successfully.

In the subsequent years, many studies aimed at expanding the list of the linguistic characteristics used to assess the readability of a text, and to date, there are more than 200 formulas for such assessment (Solnyshkina and Kiselnikov, 2015).

Some modern models use hundreds of text characteristics. For example, Coh-Metrix (Graesser et al., 2014) uses 200, and the François model (2015) uses 406 characteristics. Models with the maximum number of linguistic characteristics is a popular approach to assessing the readability and quality of text today. This has largely become possible due to automatic information processing, which has significantly reduced the costs. The second approach is the search for new, more complex linguistic characteristics, primarily in the field of the discursive level of text.

The most important characteristics to determine readability are **lexical** (Chall and Dale, 1995). This group includes several dozen metrics. Their choice is based on the idea that if the reader knows all the words in the text and uses them often, understanding the text should not present any difficulty. Abbreviations, foreign, or even simply long words make reading more difficult. Collins-Thompson and Callan (2005) additionally use the frequency of N-grams, that is, sequences of syllables, words, or letters. Research on the lexical premises of textual complexity accounts for a significant proportion of readability research, from Lively and Pressey (1923) to the most recent by Imperial and Ong (2021).

Syntactic characteristics are a second popular area of research. Golub and Kidder (1974) significantly expanded the list of syntactic metrics, adding to the gold-standard 'average length of sentences' such characteristics as the proportion of complex sentences, the average length of the simple clauses in a complex sentence, as well as the weight of sentences burdened with gerund and participial phrases, and some others. Heilman et al. (2007) show the importance of the tense forms of verbs to the complexity of the text, and vor der Brück et al. (2008) points at the importance of distance that separates the predicate from the related parts of the sentence. Schwarm and Ostendorf (2005) propose an even more complex system involving the use of a graph of the related groups of words within a sentence (parse tree).

Morphological characteristics also affect text complexity, although they have attracted much less attention from researchers, which may be due to the greater morphological simplicity of the English language in comparison with many others, including Russian. In papers devoted to the readability of texts in non-English languages, interest in morphological characteristics is higher; for example, an approach to assessing readability using morphological characteristics for texts in German is discussed in Hancke et al. (2012). The main metrics for assessing morphological complexity are based on the prevalence of different parts-of-speech in the text. Such textual characteristic as the frequent usage of numbers may be of particular interest for the purposes of this research (see Curto et al., 2015).

Ivanov (2013) proposes a new approach to assessing readability, introducing **phonetic** readability indicators for the Russian language. Describing the research idea, the author quote Aristotle: 'What is written should be readable and easy to speak, which is the same thing.' Based on the stylistic rules of the Russian language, as well as the results of analysis of a corpus of tongue twisters, the author identifies criteria that worsen the euphony of individual words and sentences, and creates

a measure for assessing euphony, the euphonic index. The higher the index value, the more euphonious is the text.

The groups of features listed (lexical, syntactic, morphological, and phonetic) constitute what is called the superficial level of textual assessment. Many researchers (e.g. Bruce et al., 1981) criticise classical readability indices for their formal approach, use of the minimum number of metrics, and complete disregard for the actual content of the text and the complexity of the thoughts contained in it. There is now a vast layer of literature devoted to the study of textual characteristics associated with content.

For example, **semantic** characteristics allow for deeper analysis and inferences about complexity based on the meanings of words. Lorge (1939) shows that prepositional phrases in English affect readability. Graesser et al. (2014) note the importance of such properties as abstractness or, conversely, the ease of visualisation of the content of the text. A text full of abstract concepts is perceived as more difficult than one that deals with more concrete things. For this reason, for example, the shorter word ‘ion’ is more complex than the longer word ‘strawberry’ (which casts doubt on the classical criteria of lexical systems of readability). In Krioni et al. (2008), it was shown that this criterion can be used to automatically assess the readability of texts in Russian.

Additionally, researchers highlight the **discursive** characteristics of text. These are associated with meaning, style, and the reader’s understanding. This approach was proposed by the French linguist Benveniste (1966). Discursive characteristics are of a psycholinguistic nature and are the most difficult to extract. Some of these characteristics cannot be reliably extracted at all, such as, for example, the coherence or integrity of a text. That said, there is no consensus among researchers on how useful those characteristics are. For example, Feng et al. (2010) suggest that ‘discursive features do not appear to be very useful in constructing effective readability metrics’. At the same time, researchers’ interest in them remains high. The most popular metrics in this category are primarily related to topic modelling and Latent Semantic Analysis (LSA), which allows the identification and comparison of the internal connections in a text (Deerwester et al., 1990).

The main statistical parameters of a text used in the literature are given in Appendix 1 (see online version of the paper).

Despite the emergence of increasingly complex metrics, new formulas for assessing readability have barely been able to outperform classical ones (François, 2015). However, this has changed with the development of natural language processing (NLP) methods.

2.2. Machine learning and natural language processing methods for assessing the readability of texts

Modern readability studies most often use methods of machine learning, statistical assessments and automatic word processing of texts, in particular,

methods of random forest, support vector machines (SVM), and neural networks. They all can process huge amounts of data. Libraries that researchers may have at their disposal include hundreds of thousands or even millions of texts, making the model interpretations of the texts close to human perception. Consequently, modern methods predict the readability much more accurately than classical indices.

Si and Callan (2001) were among the first to use statistical language models (LM) to assess readability, with predictive power significantly higher than the classic Flesch-Kincaid index (the accuracy of classification of text levels was 70.5% versus 21.3%). Collins-Thompson and Callan (2005) use LM and a multinomial naïve Bayes classifier to achieve a score accuracy of 79% on a corpus of texts composed of web pages (i.e. the system correctly classifies 79% of texts). Moreover, their model includes only four predictors: words outside of a list of the simplest, unique words count in every 100 words, frequency of words in the English language, as well as the Flesch-Kincaid index itself. The fact that this approach performs very well on a corpus of relatively short web page texts is especially important, since classic readability indices have traditionally not worked well with short texts.

Schwarm and Ostendorf (2005) pioneered the use of SVM, one of the simplest machine learning methods used to solve classification and regression problems, to assess readability. The authors' model uses as predictors the average length of sentences, the average syllable length of words, the Flesch-Kincaid index, the degree of use of rarely used words, and several characteristics in the representation of graphs (average length of connected groups of words within sentences, average sizes of groups of words around nouns, average sizes of groups of words around verbs, average number of subordinate clauses). The accuracy of the model ranges from 60% to 87%, depending on the level of education for which the classified texts were designed, and the share of classification errors was several times (in some cases by an order of magnitude) smaller than when using the classic Flesch-Kincaid index.

SVM has been used successfully in many other studies. François (2015) proves the power of the classical linguistic characteristics of a text in French to predict its readability. Falkenjack et al. (2013) analyse in detail the morphological differences between Swedish and English, taken into account in their model for evaluating texts in Swedish, and also provide an analysis of individual linguistic characteristics. Mohammadi and Khasteh (2019) test a method to assess the readability of arrays of texts in English and Farsi.

Somewhat more rarely than SVM, decision trees and random forests are used to classify texts; the essence of the latter method is the use of a large number of decision trees, which together have good predictive power. In Kauchak et al. (2014) and Santucci et al. (2020), random forests perform better than other models.

Today, the most modern class of models for solving various problems in working with text is neural networks. Thanks to giant datasets, millions of network

parameters, and massive computing power, they far outperform other machine learning algorithms when properly trained. Convolutional neural networks (CNN) and recurrent neural networks (RNN), primarily long short-term memory (LSTM) and gated recurrent units (GRU), are the most widely used for text classification (Luan and Lin, 2019).

There are two approaches to using neural networks for NLP. The first is to use the network to classify texts. In this case, the model takes raw texts as input and assigns them to classes. With this approach, researchers allow the network to decide for itself the criteria used to classify texts. The second approach is to use combined data: vectorised text and linguistic characteristics. In the first case, the researcher has no control over the operation of the network; in the second case, the model preserves the linguistic foundation and the ability to change the proportions of the network taking into account the text itself and its linguistic characteristics.

The first approach is used by Kim (2014), Deutsch et al. (2020), and Martinc et al. (2021), who investigate, among other things, the possibility of increasing the efficiency of neural network models for texts and find that adding linguistic characteristics to the models does not significantly improve their quality. The number of possible network parameters (millions) far exceeds the number of linguistic characteristics extracted statistically (hundreds).

The second approach is successfully applied, for example, by Meng et al. (2020). The authors propose the pre-processing of a text using the ReadNet system of linguistic characteristics, which allows the neural network to ‘decode’ the text and handle the classification task a little faster.

There are no significant differences in classification accuracy between these two approaches. However, the second approach makes it possible to make the network application more manageable thanks to the expert selection of linguistic prerequisites.

2.3. Readability of texts in Russian

There are not many papers devoted to the readability of texts in Russian. The Flesch index, developed in 1948, was adapted for the Russian language only after almost 60 years (Oborneva, 2006). The formula proposed is currently the most popular tool for assessing the complexity of texts in Russian.

However, the Soviet linguistics did not ignore the issue of text readability or its statistical characteristics. The foundation for future research in this area was laid in Lesskis (1962, 1964). The author notes that an indicator like the average length of sentences is associated with the nature of the text (for example, longer sentences are characteristic of narratives, short sentences are characteristic of segments with active action), while the longest sentences are characteristic of scholarly papers. Although Lesskis does not directly conclude

that scientific texts are difficult to read and understand, he describes in detail their linguistic differences from fiction.⁷

Several Soviet linguists developed their own readability formulas. In particular, the formula proposed by Mikk (1970) deserves mentioning. It assesses the readability of texts in Estonian, but the patterns discovered by the author are important for subsequent research. The author offers two predictors for such characteristics as ‘the difficulty of word connections’, the number of conscious grammatical connections between words and the total strength of semantic connections, and in addition, indicates that the complexity of a text can be estimated by the time the user spends reading it. Finally, the author states that the complexity of a text increases with increasing abstractness, while the repetition of unfamiliar words in the text contributes to the understanding of the text. He encourages authors to use shorter sentences, reduce the number of unfamiliar words, and prefer concrete words over abstract ones.

Tuldava (1975) proposes his own readability formula and expresses the idea of using an additional predictor to assess the difficulty of a text: the number of polysemous words.

Matskovskiy (1976) develops a formula for assessing readability (he calls it an assessment of the difficulty of a text) for the Russian language, which generally corresponds with the prevailing world practice at the time. As predictors, he chooses the average length of sentences (in words), as well as the percentage of words in the text consisting of more than three syllables. Matskovskiy verified the formula by asking readers to evaluate non-fiction texts at seven levels from simple to difficult. However, despite the successful testing of the methodology, the formula has not gained widespread acceptance (Oborneva, 2006).

Piotrowski et al. (1977), in their magnum opus *Mathematical Linguistics*, set the standard for the statistical processing of texts in Russian. This made it possible to formalise previously existing approaches. However, studies of the quality of a text in Russian based on statistical methods were hampered by the lack of a generally accepted linguistic understanding of a text quality. Solnyshkina and Kiselnikov (2015) directly indicate that Russian linguistics was unable to agree on what, in fact, is considered a readable text in Russian, and how the concepts of ‘complexity’, ‘difficulty’, ‘intelligibility’, and ‘readability’ of a text differ. Every linguist understood this issue in his or her own way. Razumovsky (1999; cited by:

⁷ In particular, scientific texts are characterised by high verbal noun content, the dominance of nouns in the genitive case (38.9% versus 21% for fiction), and the frequency of isolated constructions (18% versus 16.4%). The large percentage of verbal nouns somewhat compensates for the small use of verbs in it, Lesskis explains. The greatest contrast between the scientific and fictional text is manifested in syntax. Namely, scientific texts are dominated by the ‘molecular’ type of predicativity with the compositional type of connection, that is, predicative phrases that, in principle, can be decomposed into several phrases. For example, the phrase ‘These phenomena are based on the idea of the existence...’ can be transformed into the phrase ‘These phenomena are based on the idea that there exists...’ without destroying predicativity. In general, this observation is consistent with foreign studies regarding the average parse tree length as a significant factor in predicting readability.

Solnyshkina and Kiselnikov, 2015) also noted the lack of consensus in Russian linguistics on what to consider as criteria for the text complexity.

For an effective assessment of the readability of texts in Russian, a number of interdisciplinary studies are required, concerning the identification of the characteristics of a text that complicate its perception, statistical language models (LM), and methods of automatic text processing. Due to the interest in machine learning methods and big data, there have been more such studies in recent years.

Krioni et al. (2008) presented an automated algorithm for assessing the complexity of texts with deep linguistic elaboration. Particular emphasis is placed in the work on definitions included in the text: if users of the text can easily read them, or if they have difficulties. Nevdakh (2008) proposes a method for the automated determination of the text complexity using discriminant analysis based on the average paragraph length in words, the percentage of words with 11 letters or more, and the percentage of words with 13 letters or more. These textual characteristics turned out to be the most important among the 49 different characteristics. To reduce the attribute space, cluster and factor analyses, the method of correlation of galaxies and Wrocław taxonomy, and multidimensional scaling were used. This methodology can be used in other studies where it is necessary to select the most significant features of a text.

As far as we know, today there are two models for assessing the readability of texts in Russian based on linguistically grounded characteristics and machine learning methods: Karpov et al. (2014) and Reynolds (2016). Both models are designed to classify educational texts, primarily for those studying Russian as a foreign language.

Karpov et al. (2014) investigate the effect on readability of various predictors characterising the correlation of the text analysed with the active vocabulary. In total, the model uses 25 characteristics. Reynolds (2016) takes a similar approach, but his paper is based on a broader set of inputs. The classification of texts is carried out using the random forest method and 179 linguistic characteristics of text, of which 32 (including 14 morphological) are critically important. The model's F-score efficiency is 67%. Reynolds (2016) considers morphological as well as parse tree variations as the most important characteristics. This is an interesting result, which speaks of the significant difference between Russian and English, where morphology has little to do with readability. However, the author notes that he does not use semantic characteristics, LM, or fine tuning of syntax. In his assessment, these characteristics have the potential to improve models for assessing the readability of the Russian language.

Also noteworthy is the work of Rybka et al. (2015), which proposes a protocol for identifying the syntactic structure of a text using neural network algorithms. The model proposed by the authors partly compensates for the lack of systems developed for the Russian language for constructing parse trees similar to the existing English-language systems.

2.4. Clarity of central bank texts on monetary policy issues

Central banks have become concerned about the clarity of their communication, largely due to the transition to inflation targeting regimes. Research shows that in these monetary policy regimes, good communication is very important to hitting the targets (Blinder et al., 2008; Bulíř et al., 2013).

Fracasso et al. (2003) analyse which qualities of inflation reporting (an important tool for explanations about monetary policy) ensure the predictability of monetary policy decisions for the general public. The authors evaluate reports with the help of several readers invited to participate in the project. The estimates obtained show that there is a trade-off between the length of a text and its readability, and that such qualities of reports as the persuasiveness and amount of information provided negatively correlate with their comprehensibility to non-economists. The authors recommend avoiding long reports, since in long texts information that is important to readers may go unnoticed.

Dincer and Eichengreen (2008, 2014) develop a transparency index for central bank communication. The index goes through a list of monetary policy issues about which information can be disclosed by the central bank, and can take values from 0 to 15 ('0' means very closed communication, '15' means maximally transparent communication). The questions in the index are lumped into five groups that reflect political, economic, procedural, and operational openness, as well as openness in decision-making (see Appendix 5 for more details). Having calculated index values for central banks in more than 100 countries, the authors show that transparency of central bank policies reduces the volatility of inflation.

Jansen (2011a, 2011b) and Bulíř et al. (2014) propose the use of the classical Flesch-Kincaid index to assess the clarity of communication and demonstrate the existence of an inverse relationship between the readability of central bank texts and the volatility of financial markets.

Following Jansen (2011b), Hernández-Murillo and Shell (2014) analyse statements by the Federal Open Market Committee (FOMC) of the US Federal Reserve using the Flesch-Kincaid index. The authors show that, as FOMC statements become more detailed (and this may indicate more transparency in decision-making), they also become less intelligible. By the end of Alan Greenspan's chairmanship in January 2006, the FOMC statements had an average length of 210 words, requiring 14 years of education to understand. In subsequent years, in the context of the global financial crisis, the size and complexity of the texts increased dramatically: by January 2009, during Ben Bernanke's chairmanship, the average length of statements reached 400 words, and under Janet Yellen, it exceeded 800 words, so 16 years and 18–19 years of education were required, respectively, to understand these texts. Thus, with the transition to unconventional monetary policy, the FOMC statements became little understood even by university graduates.

Bulíř et al. (2013) uses the Flesch-Kincaid index to conclude that low readability of central bank materials is associated with increased volatility of inflation.

Bruno (2017) uses NLP techniques (in particular, latent semantic analysis), together with the classic automated readability index (ARI) and the modification of the Flesch-Kincaid Index for the Italian language, to assess the readability of Bank of Italy financial stability reports, and comes to the conclusion that these documents are understandable only to people with higher education.

Omotosho (2019) uses text mining techniques to assess the communication of the Bank of Ghana. In this work, the method of sentiment analysis (or opinion mining) is also used, which allows the drawing of conclusions about the dominance of optimism or pessimism in communication, which correlates with general economic situation.

The communication of central banks under inflation targeting regimes is aimed at minimising monetary policy surprises. As Benchimol et al. (2020) point out, the impact of surprises spreads through two channels: the financial asset channel (investors are forced to revise their portfolios after unexpected decisions by the central bank) and the behavioural channel (increasing uncertainty complicates the task of forecasting the future). In the latter case investors' uncertainty about the future creates the risk that negative sentiments will be more prevalent in the markets. The authors note that better understanding of the risks associated with monetary policy surprises has led to greater transparency in central banks, but quantifying it using the method proposed by Dincer and Eichengreen (2008) does not take into account aspects of communications such as volume, readability, or the sentiment that they transmit. Benchimol et al. (2020) propose combining the readability and transparency criteria when assessing the clarity of communication. We follow this approach in the empirical part of our research.

As for research on the clarity or readability of Bank of Russia communication, to the best of our knowledge there are none so far.

3. Development of a readability metric for economic texts in Russian

3.1. Linguistic characteristics of text

The first step in our paper is to develop a linguistically based set of characteristics that affect readability. A summarised list of the characteristics we formulated on the basis of domestic and foreign research is presented in Table 1 (see Appendix 1 for a detailed list, and more detail in Appendix 2). The list includes 44 textual characteristics which cover all levels: syntactic, lexical, morphological, phonetic, semantic, and discursive. We tried to take into account the main conclusions and recommendations of previous researchers, including the importance of the

morphological features for the Russian language, the genitive case in scientific texts, the importance of euphony, the coherence of the internal structure, and others.

Table 1. Textual characteristics (summary)

Category	Number of features
Syntactic	12
Lexical	16
Morphological	7
Phonetic	1
Semantic	5
Discursive and others	3
Total	44

Source: compiled by the authors

We also added some new textual characteristics that may be useful in working with economic texts, such as the weight of specific punctuation (colon, semicolon, parentheses), the average number of verbs in the passive voice and reflexive form, the weight of constructions with consecutive nouns ('noun strings') and derivative prepositions, the weight of markers of uncertainty, and the weight of markers of 'wateriness' (it is assumed that this metric will be higher for non-economic texts as they are less concentrated; the metric accounts for words that do not carry the main semantic load and serve, for example, as links between individual elements within the text or deviations from the main topic). In addition, we introduce an adapted ARI for economic texts in Russian (see below).

We use the following dictionaries in our paper: OpenCorpora,⁸ Tikhonov (2002), Lyashevskaya and Sharov (2009), Andryushina and Kozlova (2013), and Klobukova et al. (2014).

In the group of **syntactic** features of the text, we include the classical criteria for the syntactic complexity of text, from the average length of sentences in words to the average length of the parse tree. We also measure the proportion of compound and complex sentences, participles and verbal adverbs that impede the perception of text. As already noted, we single out sentences with passive and reflexive verbs into a special group. In a way similar to Karpov et al. (2014), we add a metric that captures the proportion of extra-long sentences (we consider sentences of more than 25 words to be extra-long). Since the model classifies economic texts, we also keep track of the share of the numbers and complex punctuation that are typical in such publications.

The group of **lexical** features of text is represented by three main subgroups. The first subgroup contains metrics that reflect the proportion of words that are not included in lists of lexical minimums. For such lists, we took the A2 and B1 lists from the dictionaries of Andryushina and Kozlova (2013) and Klobukova et al. (2014): together they give about 3,000 of the most common words, which corresponds to the popular method of Chall and Dale (1995). In addition, we

⁸ See <http://opencorpora.org/dict.php>

introduced an inverse metric that calculates the proportion of words from the bottom 30% in the frequency dictionary, as well as words not in the dictionary at all. Presumably, this metric penalises the system for using too much terminology.

The second subgroup is various metrics of lexical diversity, including the ratio of the number of unique lexical units (type) and the total number of word forms (token) in the text (Type-Token Ratio, TTR). The best texts are those in which the TTR approaches one, that is, the author is able to express a thought in different words and avoids repetitions, tautologies, and pleonasms. Since the TTR depends on the length of a text (it decreases with increasing length), which is a significant drawback, modern researchers use various modifications of it, which we have also taken into consideration.

The third subgroup is responsible for identifying excessive word length (based on previous studies, we took 6, 8, 10, and 12 letters as thresholds). In addition to the metrics from these three subgroups, we use a metric that reflects the proportion of verbal nouns (an important sign of ‘officialese’), as well as the classic indicator of the average word length (in characters).

The **morphological** level of a text is assessed using the Pymorphy2 library (Korobov, 2015), which works on one of the largest marked-up corpora for the Russian language, OpenCorpora. We used the part-of-speech tagging (POS tagging) standard in these studies. The metrics selected reflect the proportion of nouns, verbs, adjectives, adverbs, and pronouns. Additionally, we record the proportion of the genitive nouns characteristic of the scientific style and the average length of meaningful words in morphemes (using Tikhonov’s 2002 dictionary).

The **phonetic** level is represented by one characteristic accounting for euphony. The formula for calculating this metric is based on the euphony index (Ivanov, 2013), while we exclude from the formula the part that penalises word length and sentence length in order to avoid correlation with our other metrics. Since a third of the Bank of Russia’s main communication is presented in an oral format (statements by the Governor following meetings of the Board of Directors), we considered it necessary to evaluate euphony as well.

Semantic characteristics allow us to get an idea of the semantic content of text. We include five such characteristics, two of which are mentioned above (markers of ‘officialese’ and ‘wateriness’). Based on Mikk (1981), we estimate the proportion of abstract nouns. Using the Stanza library (Qi et al., 2020), an entity density indicator was calculated, that is, the proportion of fixed concepts in the text (e.g. ‘inflation targeting’, ‘GDP’, ‘USA’, etc.). The fifth metric from this group is the weight of markers of uncertainty, or the proportion of lexical markers that account for the author’s degree of confidence (for a variant of the calculation method, see Rubin et al., 2006). Since this metric may be of great importance in monetary policy communication, its separation and application are of interest for future research.

Finally, we take into account **discursive** characteristics, of which there are only three: the average distance between sentences (allowing the assessment of the

connectivity of sentences among themselves, one popular metric for connectivity in text), the weight of discourse markers (non-semantic linguistic units), and an adapted ARI for economic texts in Russian.

To create the adapted ARI, we took the classic ARI formula with linear dependence of readability on the average length of words in characters and the average length of sentences in words. We estimated the coefficients of the respective regression (with the help of the stats package in the R program), we adapted its coefficients, taking into account the goals of our research. To do this, we formed a corpus of 10,000 texts, which we divided into 10 levels of readability. The entry levels, with the highest readability (10 to 7), are represented by home reading texts. The sixth level is information style texts (notes, news, articles, interviews, and other texts on general political topics). The following levels are represented by fictional, educational, and academic literature of increasing complexity. Finally, the first level (the worst readability) corresponds to laws, court orders, and monographs on financial and economic disciplines (for more details, see Section 3.2, as well as Appendix 4).⁹

After estimating a linear regression, we obtained the following coefficients for the adapted ARI:

$$\text{Level of textural clarity} = 19.025 - 2.184 \left(\frac{\text{character}}{\text{words}} \right) - 0.048 \left(\frac{\text{words}}{\text{sentences}} \right). \quad (1)$$

At this stage, we do not implement metrics based on the log likelihood function that give a probabilistic idea of the sentiment of a text, which have proven themselves useful for the English language (Pitler and Nenkova, 2008), due to the lack of the necessary dictionaries for the Russian language. Perhaps such dictionaries are absent because individual units of the lexical corpus of the Russian language themselves are much less emotionally coloured than units of the English language. Context plays an important role in the Russian language, and there are more ambiguous words in it. At the same time, the use of specialised libraries of sentiment analysis (such as Dostoevsky) would not help meet the objectives of this research, since the classification of texts on the basis of emotional colouring does not say anything about their quality: a negatively coloured text may be written as clearly and intelligibly as a positive one.

Table 2. Regression coefficients of the adapted ARI for economic texts in Russian

Coefficient	Coefficients	Standard error	t-value
Characters/words	-2.184***	0.021	-102.65
Words/sentences	-0.048***	0.002	-26.08
	19.025***	0.113	168.77
Significance: *** – 0.001; ** – 0.01; * – 0.05			
R ² : 0.63; adjusted R ² : 0.63			

Source: authors' calculations

⁹ For examples of texts in Russian, see https://econs.online/upload/medialibrary/66b/txt_exam.pdf

3.2. Corpus of texts

For the research tasks, it was necessary to form a corpus of pre-classified texts in Russian, since the already existing text corpora (for example, SinTagRus) do not include sufficient number of texts on economic topics.

One of the subtasks of this paper was the formation of a new dataset with data sufficient for the application of machine learning algorithms.

We based the expert classification of the supervised learning sample on the principle used in classical readability indices. The levels of textual complexity are determined depending on the number of years of education required to understand them. In this system, we took one important nuance into account. Since the models have to work with economic texts, we include such texts in the corpus starting at the sixth level of readability, with increasing complexity from easy-to-understand articles in the style of ‘personal finance advice’ to monographs on finance. Thus, at levels 10 to 7, 100% of the texts are fiction, at the sixth level, 10% of the texts are represented by the simplest notes from social and political media on topics related to economics (for example, news about an increase in pensions). At the fifth level, 25% of the texts are popular business literature and ‘success stories’, at the fourth level, 40% of the texts are articles in the genre of ‘personal finance’, at the third level, 50% of the texts are texts from business media, and 50% are textbooks on economic disciplines. At the second level, all 1,000 texts are articles from professional economic and financial publications. At the first level, all 1,000 texts are laws, governing financial activities, or monographs on economics and finance.

The characteristics of the corpus are set out in Appendix 4. In total, there are 10,000 texts in the corpus with lengths of about 300 words each, 1,000 texts at each level.

To test the models, we rated each text according to the 44 linguistic characteristics described in Section 3.1. As a result, we obtained a matrix with 46 columns and 10,000 rows. The first column is the expertly determined level of complexity of the text, the second is the text itself, and the remaining 44 are the results of the functions described in Appendix 2. Each model worked with all 10,000 texts.

3.3. Readability modelling

We use two models as basic ones. The first is the Flesch index adapted for the Russian language (Oborneva, 2006):

$$Oborneva\ Index = 206.835 - 1.52 \left(\frac{total\ words}{total\ sentences} \right) - 65.14 \left(\frac{total\ syllables}{total\ words} \right). \quad (2)$$

Since this formula gives an index value ranging from 0 to 100, and in our classifier there are ten levels, we divide the Oborneva index by ten and round to integers, thus bringing the output of the Oborneva regression model closer to the

output of our classifier for comparing the results. We were unable to find codes of classifying models for assessing readability in Russian in open sources, which, along with the popularity of Osborne's formula, predetermined the choice of this model as the basic one.

The second basic model is the classic Flesch index evaluated on our corpus. Using this model separates the effect of the corpus from the effect of the model.

$$\text{Flesch Index} = 17.9773 - 0.0551 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 4.7351 \left(\frac{\text{total syllables}}{\text{total words}} \right). \quad (3)$$

For testing, we chose models that have successfully proven themselves in the work of previous researchers: the naïve Bayes classifier, k-nearest neighbours, linear discriminant analysis, random forest, support vector machine, XGBoost gradient boosting algorithm, neural network of LSTM type with textual input data, and Transformer neural network with textual input data and linguistic characteristics. For training all the models, except for the last two, a dataframe with dimensions of $8,500 \times 44$ was provided, that is, 44 linguistic characteristics extracted from each of the 8,500 texts. For the convolutional neural networks (CNN), only vectorised text was used as inputs. For the Transformer neural network, both types of data were used as inputs: vectorised text and 44 expertly selected linguistic characteristics.

Table 3 shows the results of the models' operation on the test dataset (1,500 texts, about 150 texts per complexity level). The most successful (F1-score = 94.8%) is the Transformer neural network model, which is consistent with the results obtained by Meng (2020) for the English language.

Before proceeding to the description of the successful model, let us dwell on the significance of the linguistic characteristics selected. The random forest model allows a comparison of their contribution to the overall readability score. The top 5 characteristics with the highest predictive power are: the proportion of nouns in the text (POSN), the proportion of verbal nouns (VN), the proportion of words longer than eight letters (Lex8), the average word length (AvWord), and the ARI adapted for economic texts. Three of these five characteristics are lexical, which is consistent with the observations of previous studies. The importance of morphological features such as the richness of the text with nouns and, especially, verbal nouns allows the models to capture the scientific style. In our dataset, samples in the scientific style are placed in the most difficult levels.

The Transformer deep neural network model is an advanced version of the neural network model with attention mechanism. This model, proposed by Vaswani et al. (2017), has gained a lot of popularity in industrial models for textual analysis. As the basis for our model, we take a basic assembly from the TensorFlow library and improve it in such a way that the model directly takes into account the text and 44 linguistic characteristics preliminarily extracted from it in a ratio of 60% to 40%. Thus, we retain expert control over the neural network and can influence

its decision making through linguistic data. The information received by the neural network from the vectorised text is almost identical in weight to the information obtained from the data array with linguistic characteristics.

Table 3. Outputs the models

Model	Precision	Recall	F1-score
Oborneva index (basic model 1)	0.1149	0.0755	0.0534
Flesch index (basic model 2)	0.1372	0.0951	0.0854
Naïve Bayes (44)	0.5175	0.5120	0.4919
K-nearest neighbours (44)	0.5923	0.5800	0.5775
Linear discriminant analysis (44)	0.6169	0.6060	0.6096
Random forest (44)	0.6456	0.6433	0.6415
SVM (44)	0.6546	0.6583	0.6564
XGBoost (44)	0.6561	0.6592	0.6576
CNN (text-only input)	0.7270	0.7126	0.7173
Transformer neural network with text and linguistic characteristics (44)	0.9506	0.9480	0.9484

Source: authors' calculations

Figure 2. Error matrix of classification of the test dataset by the Transformer neural network

1	143	0	1	0	0	0	0	0	0	0
2	0	153	2	1	2	1	0	0	0	0
3	1	2	150	1	1	1	0	0	0	0
4	0	1	1	141	6	2	0	0	0	0
5	0	1	2	5	141	2	0	2	0	0
6	0	1	3	3	4	140	2	0	3	6
7	0	0	0	0	0	0	136	3	0	1
8	0	0	0	0	0	0	7	145	1	1
9	0	0	0	0	0	0	11	1	129	2
10	0	0	0	0	0	0	2	2	1	133
	1	2	3	4	5	6	7	8	9	10

Note: Columns correspond to the actual levels of texts' complexity; rows correspond to the model assessed levels of texts' complexity.

Source: authors' calculations

Initially, Transformer neural networks were intended for machine translation, since their architecture allows effective work with context. The internal mechanisms of this class of networks (the presence of an encoder and a decoder as groups of specialised layers) also make it possible to effectively solve the problems of the classification of texts. These models reveal the importance of all the other words in a sentence for each particular with the help of the attention mechanism. The output of this layer is fed into the feedforward neural network. An important advantage of Transformer networks is their faster learning compared with recurrent networks that use blocks of memory and back propagation of information.

Detailed results from the operation of the Transformer neural network according to the classification of our dataset are presented in Figure 2. The figure shows how the neural network classified all 1,500 test texts. It experienced difficulty only at the 7th level, which may be due to the thematic and stylistic proximity of the texts of this level to the texts of 8th and 9th levels; all three levels are composed of fiction. Since the most important levels for the purposes of our research are levels 3 to 6, we consider the slight difficulties that arose in the neural network with the classification of level 7 texts insignificant. Appendix 3 presents the architecture of the Transformer neural network. It has over 1.3 million learning parameters.

4. Assessment of the readability of Bank of Russia monetary policy communication and recommendations

We assessed the readability of Bank of Russia communications in 2013–2021 using the model obtained. The results are presented in Figures 3 to 7.

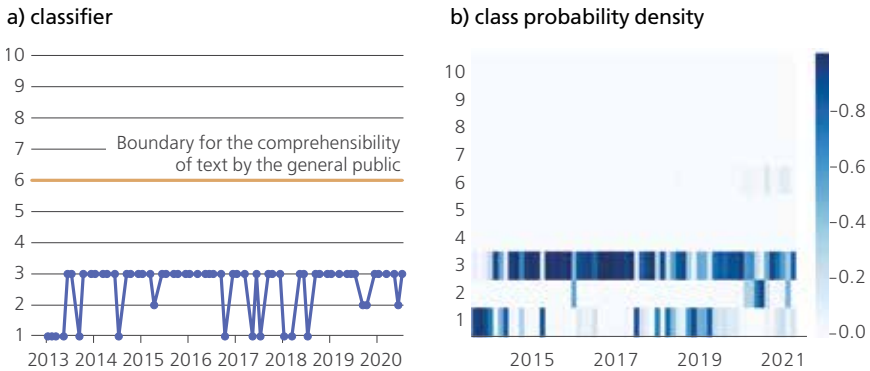
The graphs show that Bank of Russia monetary policy communications are classified by the model mainly as falling into the first, second, and third levels of readability, that is, they are comprehensible to people with higher education in economics or even with academic degree in economics.

It can be noted that during the period under review, the probability of text falling into the sixth (target) level (communication comprehensible to school graduates) increased. The graphs of the class probability density in Figures 3 to 7 indicate that around 2018, the Bank of Russia began to gradually simplify its communication, and now some parts of it are more comprehensible to the general public.

Nevertheless, for communication to reach a level of readability as a whole that is fully comprehensible to the general public, these individual manifestations of improvements to the quality of text are still not enough.

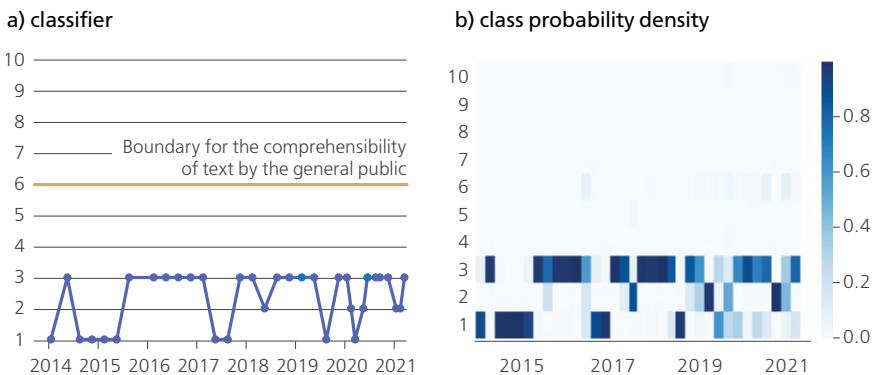
It is worth noting that the readability of press releases and statements by the Governor of the Bank of Russia on the key interest rate may be subject to current economic uncertainties. Most of the time, their readability was at the third level and dropped lower during the 2014–2015 crisis, the volatility in the financial markets in the first half of 2018, the VAT hikes in early 2019, and the pandemic in 2020. Glas and Müller (2021) found similar results when assessing the readability of statements by members of the Executive Board of the European Central Bank.

Decomposition of the results of the operation of the model according to the 44 linguistic characteristics allows us to offer a number of recommendations for improving communication. Thus, the model indicates a need for less use of terminology (a large number of terms increases the average length of words) and a reduction in the proportion of nouns and complex polynomial sentences. It is also necessary to avoid such stylistic flaws as ‘noun strings’ and the widespread use of verbal nouns; these characteristics significantly reduce the readability of Bank of Russia texts.

Figure 3. Press releases on the key rate

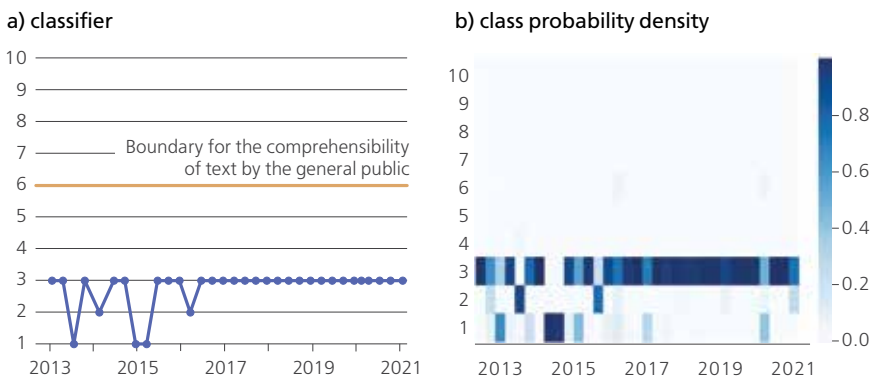
Note: Left scale – levels (classes) of texts' readability; Figure 3b, right scale – probability density of classes.

Source: authors' calculations

Figure 4. Statements of the Governor of the Bank of Russia on the key rate

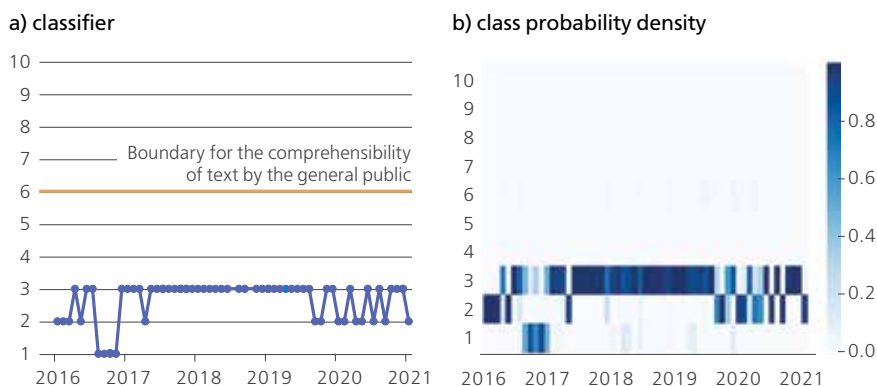
Note: Left scale – levels (classes) of texts' readability; Figure 4b, right scale – probability density of classes.

Source: authors' calculations

Figure 5. Monetary Policy Report

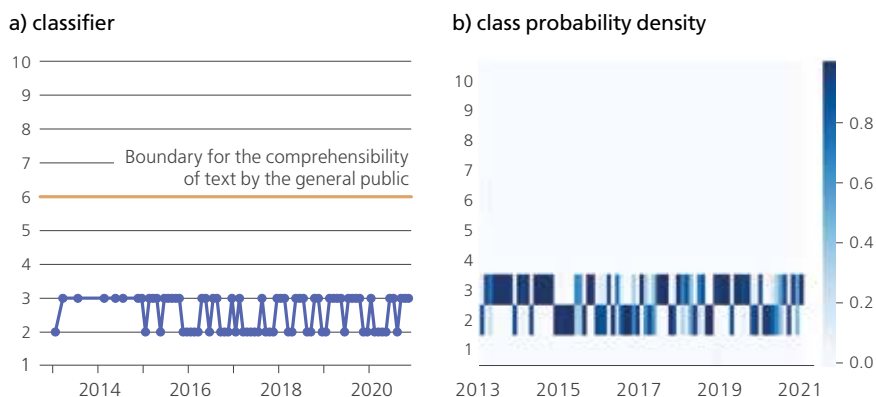
Note: Left scale – levels (classes) of texts' readability; Figure 5b, right scale – probability density of classes.

Source: authors' calculations

Figure 6. Analytical Commentary on Dynamics of Consumer Prices

Note: Left scale – levels (classes) of texts' readability; Figure 6b, right scale – probability density of classes.

Source: authors' calculations

Figure 7. Analytical Commentary on Inflation Expectations and Consumer Sentiment

Note: Left scale – levels (classes) of texts' readability; Figure 7b, right scale – probability density of classes.

Source: authors' calculations

5. Adaptation of the model for practical use and prospects for future research

We have created a model for the automated assessment of the readability of texts on monetary policy in Russian, but the following important points must be taken into account for its practical application.

- 1) The model's assignment of a text to one or another level is not very informative from the point of view of the prospects for improving the material. Therefore, along with this information, it is advisable to extract the key linguistic characteristics (out of the 44 selected) whose values are far from those of the sixth (target) level.

- 2) The clarity of communication goes far beyond readability.¹⁰ More information about the quality of communication can be obtained by adding other parameters to the readability model. For example, transparency, that is, the extent to which the central bank discloses information about monetary policy.

In this section, we will outline paths for the further improvement of the model in terms of adding characteristics for transparency.

From a theoretical perspective, transparency of the monetary policy is the absence of an information asymmetry between the central bank and economic agents (Blinder et al., 2008). In practice, this means that the central bank discloses, in a timely and easily accessible form, information about its strategy, about monetary policy decisions and its decision-making procedure, about its assessment of the economic situation, and much more. A large number of papers were released in the 2000s that prove the importance of transparency and develop tools for its assessment (e.g. Freedman and Laxton, 2009; Blinder et al., 2008; Eijffinger and Geraats, 2006).

In Russian studies (for example, Evdokimova et al., 2019; Kuznetsova and Merzlyakov, 2015), the index proposed by Dincer and Eichengreen (2008, 2014) is used to assess the transparency of the Bank of Russia (see Section 2.4. and Appendix 5). However, this index does not take into account the current monetary policy regime and, therefore, misses distinctions that are important for an inflation targeting regime. Al-Mashat et al. (2018) developed an index that focuses directly on inflation targeting countries (see Appendix 6 for the components of the index).

Assessments of the transparency of the Bank of Russia's communication on monetary policy using both indices are presented in Figure 8 (for the assessment of individual components of the indices, see Appendices 5 and 6). Since 2013–2014, the Dincer and Eichengreen (2014) index values have been in line with those of the central banks of inflation targeting emerging market countries.¹¹ At the same time, the Al-Mashat et al. (2018) index grew steadily, more accurately reflecting the changes that were taking place. The discrepancy in the results between the two indices is also characteristic of other central banks. For example, the transparency of the Czech National Bank by the first index was at the level of 14.5 out of 15 for 2011 to 2017, while by the second index, it was at the level of 11.75 to 12 out of 20.

Let us combine the data on transparency and readability by averaging the indicators:

$$Y = \frac{RDBL + \frac{TRSPR}{2}}{2}, \quad (4)$$

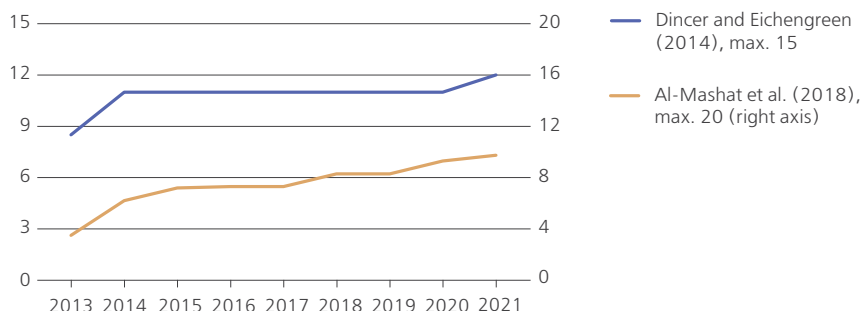
¹⁰ Bulíř et al. (2013), as well as Benchimol et al. (2020), point out that the clarity of central bank communication cannot be measured by a simple single-dependent-variable model; for an at least partly relevant definition of the level of clarity of communication, several parameters must be taken into account.

¹¹ For example, Mexico – 10, Thailand – 11, Brazil – 11.5, Poland – 13, Hungary – 14 (estimates for 2019; see Evdokimova et al., 2019).

where Y is the clarity metric, $RDBL$ is the readability metric (from one to ten), and $TRSPR$ is the transparency metric, which we divide by two to match the scales (the maximum value of the Al-Mashat et al. (2018) index is 20 points).

The results of the calculation are shown in Figure 9. It is clearly seen that the ultimate clarity of communication is gradually increasing. This growth is almost entirely determined by the expansion of the amount of information published as the level of readability has been almost constant, while the level of transparency reflects the lists of published documents.

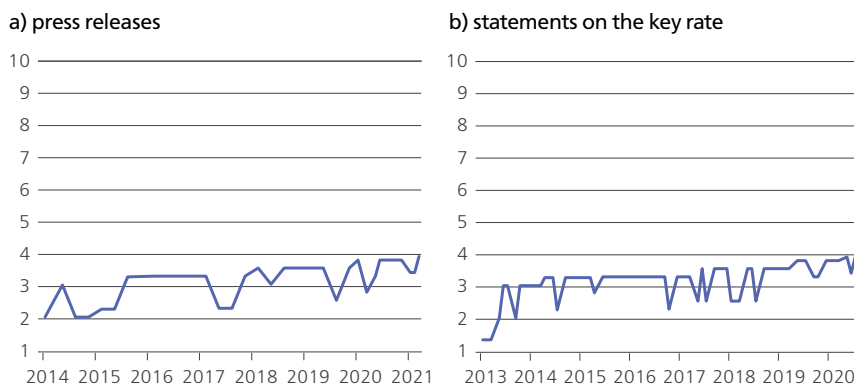
Figure 8. Indices for the transparency of Bank of Russia monetary policy communication



Note: The values of the indices for 2020 are calculated based on data for January – April.

Source: authors' calculations

Figure 9. Clarity of Bank of Russia monetary policy communication



Source: authors' calculations

6. Conclusion

In this paper, we propose a tool for the automated assessment of the readability of economic texts in Russian in view of existing experience with linguistic and textual analysis and the use of modern machine learning methods.

The most successful model (the Transformer Neural Network) was trained on a corpus of 10,000 texts specially compiled for our study, which were expertly ranked from level 1 (the most complex texts, for example, monographs on economics) to level 10 (elementary texts for the extra-curricular reading of young schoolchildren). The classification accuracy of the successful model is close to 95%, which corresponds to the best examples for the English language.

An important feature of our model is the combination of input data. The model takes as input a vectorised text and linguistic characteristics extracted from it (44 at all levels of the text: syntactic, lexical, morphological, phonetic, semantic, and discursive). The first three groups show the greatest predictive power, as in other studies. Due to the lack of the necessary dictionaries, we were unable to extract some textual characteristics. This could be an area for future study.

Having assessed the texts of the Bank of Russia on monetary policy (press release, statements, and reports, as well as two types of analytical commentaries on inflation and inflation expectations) using the successful model, we concluded that the current level of communication is difficult for the general public to understand. At the same time, it is important to note the improvement in the readability which began around 2018. The results of the model show that the Bank of Russia is introducing certain elements to improve the clarity of its communication with the general public, but so far this has not been enough to reach a qualitatively new level.

From the point of view of readability, the model indicates a need to use less terminology (which increases the average length of words) and to reduce the proportion of nouns and complex polynomial sentences. It is also necessary to avoid such stylistic flaws as ‘noun strings’ and the widespread use of verbal nouns. These textual characteristics significantly reduce the readability of Bank of Russia texts. In explaining its decisions, the Bank of Russia actively uses specialised terminology and scientific style, which makes it difficult for the general public to understand the texts and hinders the increase of public confidence.

As areas for future research, we recommend considering the questions of supplementing the model with other characteristics for the quality of communication, such as transparency, document length, and targeting a specific audience.

The communication clarity assessment tool developed in this paper can be used by the Bank of Russia for quick assessment of its main communication products. Improving the readability of the Bank of Russia communication on monetary policy will help increase public confidence in its policies, improve the predictability of its decisions, and reduce and anchor inflation expectations.

Appendices are available at
<http://rjmf.econs.online/en>;
doi.org/10.31477/rjmf.202103.03

References

- Algan, Y. and Cahuc, P.** (2014). Trust, Growth, and Well-Being: New Evidence and Policy Implications. In: P. Aghion and S. Durlauf, eds. *Handbook of Economic Growth*, Vol. 2, pp. 49–120.
- Andryushina, N. P. and Kozlova, T. V.** (2013). *Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Basoviy Uroven'. Obshchee vladenie* [Russian as Foreign Language Basic Dictionary. Basic Level. General Knowledge]. Saint-Petersburg: Zlatoust. [In Russian].
- Al-Mashat, R. A., Bulíř, A., Dinçer, N., Hlédik, T., Holub, T., Kostanyan, A., Laxton, D., Nurbekyan, A., Portillo, R. A. and Wang, H.** (2018). *An Index for Transparency for Inflation-Targeting Central Banks: Application to the Czech National Bank*. IMF Working Paper, N 210.
- Benchimol, J., Kazinnik, S. and Saadon, Y.** (2020). *Communication and Transparency Through Central Bank Texts*. Preprint presented at ASSA/AEA Annual Conference.
- Benveniste, E.** (1966). *Problèmes de Linguistique Générale*. Paris: Gallimard.
- Blinder, A. S.** (2004). *The Quiet Revolution: Central Banking Goes Modern*. Yale University Press. <https://doi.org/10.12987/yale/9780300100877.001.0001>
- Blinder, A., Ehrmann, M., Fratzscher, M., De Haan, J. and Jansen, D.** (2008). Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence. *Journal of Economic Literature*, 46(4), pp. 910–945.
- Bruce B., Rubin, A. and Starr, K.** (1981). Why Readability Formulas Fail. *IEEE Transactions on Professional Communication*, PC–24(1), pp. 50–52.
- Brunner, K.** (1981). Epilogue: Understanding the Great Depression. In: K. Brunner, ed. *The Great Depression Revisited. Rochester Studies in Economics and Policy Issues*, Vol 2. Dordrecht: Springer, pp. 316–358. https://doi.org/10.1007/978-94-009-8135-5_18
- Bruno, G.** (2017). *Central Bank Communications: Information Extraction and Semantic Analysis*. Paper presented at IFC-Bank Indonesia Satellite Seminar on 'Big Data' at the ISI Regional Statistics Conference, Bali, 21 March. BIS.
- Bulíř, A., Čihák, M. and Jansen, D.-J.** (2013). What Drives Clarity of Central Bank Communication About Inflation? *Open Economies Review*, 24, pp. 125–145. <https://doi.org/10.1007/s11079-012-9259-z>
- Bulíř, A., Čihák, M. and Jansen, D.-J.** (2014). *Does the Clarity of Inflation Reports Affect Volatility in Financial Markets?* IMF Working Paper, N 175.
- Bulíř, A., Čihák, M. and Šmídková, K.** (2008). *Writing Clearly: ECB's Monetary Policy Communication*. IMF Working Paper, N 252.
- Chall, J. S. and Dale, E.** (1995). *Readability Revisited: The New Dale-Chall Readability Formula*. Brookline Books.
- Collins-Thompson, K. and Callan, J.** (2005). Predicting Reading Difficulty with Statistical Language Models. *Journal of the American Society for Information Science and Technology*, 56(13), pp. 1448–1462. <https://doi.org/10.1002/asi.20243>
- Curto, P., Mamede, N. and Baptista, J.** (2015). Automatic Text Difficulty Classifier: Assisting the Selection of Adequate Reading Materials for European Portuguese Teaching. In: *Proceedings of the 7th International Conference on Computer Supported*

Education – Volume 1 (CSEDU 2015). Setubal, Portugal: SCITEPRESS – Science and Technology Publications, pp. 36–44. <https://doi.org/10.5220/0005428300360044>

Dale, E. and Chall, J. (1948). A Formula for Predicting Readability. *Educational Research Bulletin*, 27(1), pp. 11–28.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41(6), pp. 391–407.

Deutsch, T., Jasbi, M. and Stuart, S. (2020). Linguistic Features for Readability Assessment. In: *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*. Seattle, USA: Association for Computational Linguistics, pp. 1–17. <https://doi.org/10.18653/v1/2020.bea-1.1>

Dincer, N. and Eichengreen, B. (2008). Central Bank Transparency: Where, Why, and With What Effects? *Central Banks as Economic Institutions*. Edward Elgar Publishing, pp. 105–141.

Dincer, N. and Eichengreen, B. (2014). Central Bank Transparency and Independence: Updates and New Measures. *International Journal of Central Banking*, 10(1), pp. 189–259.

DuBay, W. (2004). *The Principles of Readability*. Costa Mesa, US: Impact Information.

Eijffinger, S. C. W. and Geraats, P. M. (2006). How Transparent Are Central Banks? *European Journal of Political Economy*, 22(1), pp. 1–21.

Evdokimova, T., Zhirnov, G. and Klaver, I. (2019). The Impact of Inflation Anchor Strength and Monetary Policy Transparency on Inflation During the Period of Emerging Markets Volatility in Summer 2018. *Russian Journal of Money and Finance*, 78(3), pp. 71–88. <https://doi.org/10.31477/rjmf.201903.71>

Falkenjack, J., Mühlenbock, K. H. and Jönsson, A. (2013). Features Indicating Readability in Swedish Text. In: *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)*, pp. 27–40.

Feng, L., Jansche, M., Huenerfauth, M. and Elhadad, N. (2010). A Comparison of Features for Automatic Readability Assessment. In: *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Beijing: Coling 2010 Organizing Committee, pp. 276–284.

Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology*, 32(3), pp. 221–233. <https://doi.org/10.1037/h0057532>

Fracasso, A., Genberg, H. and Wyplosz, C. (2003). *How Do Central Banks Write? An Evaluation of Inflation Targeting Central Banks*. Geneva Reports on the World Economy, Special Report 2. Geneva: ICMB; London: CEPR.

François, T. (2015). When Readability Meets Computational Linguistics: A New Paradigm in Readability. *Revue Française de Linguistique Appliquée*, XX(2), pp. 79–97. <https://doi.org/10.3917/rfla.202.0079>

Freedman, C. and Laxton, D. (2009). *Why Inflation Targeting?* IMF Working Paper, N 86.

Fullwood, J. (2016). A Cat, a Hat and a Simple Measure of Gobbledygook: How Readable is Your Writing? *Bank Underground*, 4 October. <https://bankunderground.co.uk/2016/10/04/a-cat-a-hat-and-a-simple-measure-of-gobbledygook-how-readable-is-your-writing/> [accessed on 26 February 2021].

- Glas, A. and Müller, L. S.** (2021). *Talking in a Language that Everyone Can Understand? Transparency of Speeches by the ECB Executive Board*. FAU Discussion Papers in Economics, N 1.
- Golub, L. and Kidder, C.** (1974). Syntactic Density and the Computer. *Elementary English*, 51, pp. 1128–1131.
- Graesser, A., McNamara, D., Cai, Z., Conley, M., Li, H. and Pennebaker, J.** (2014). Coh-Matrix Measures Text Characteristics at Multiple Levels of Language and Discourse. *Elementary School Journal*, 115(2), pp. 210–229. <https://doi.org/10.1086/678293>
- Gunning, R.** (1952). *The Technique of Clear Writing*. New York: McGraw-Hill.
- Haldane, A., Macaulay, A. and McMahon, M.** (2020). *The 3 E's of Central Bank Communication with the Public*. Bank of England Staff Working Papers, N 847.
- Haldane, A. and McMahon, M.** (2018). *Central Bank Communications and the General Public*. AEA Papers and Proceedings, 108, pp. 578–583. <https://doi.org/10.1257/pandp.20181082>
- Hancke, J., Sowmya, V. and Meurers, D.** (2012). Readability Classification for German Using Lexical, Syntactic, and Morphological Features. In: *24th International Conference on Computational Linguistics – Proceedings of COLING 2012: Technical Papers*. Mumbai, India: The COLING 2012 Organizing Committee, pp. 1063–1080.
- Heilman, M., Collins-Thompson, K., Callan, J. and Eskenazi, M.** (2007). Combining Lexical and Grammatical Features to Improve Readability Measures for First and Second Language Texts. In: *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*. Rochester, USA: Association for Computational Linguistics, pp. 460–467.
- Hernández-Murillo, R. and Shell, H. G.** (2014). The Rising Complexity of the FOMC Statement. *Economic Synopses, Federal Reserve Bank of St. Louis*, N 23.
- Imperial, J. M. and Ong, E.** (2021). *A Simple Post-Processing Technique for Improving Readability Assessment of Texts using Word Mover's Distance*. arXiv Preprint, arXiv:2103.07277.
- International Monetary Fund** (2018). *World Economic Outlook: Challenges to Steady Growth*. Washington, DC, October.
- Isakov, A., Grishin, P. and Gorlinsky, O.** (2018). Fear of Forward Guidance. *Russian Journal of Money and Finance*, 77(4), pp. 84–106. <https://doi.org/10.31477/rjmf.201804.84>
- Ivanov, K. V.** (2013). Avtomatizatsiya otsenki blagozvuchiya tekstov [Computerized Euphony Assessment]. *Novie informatsionnye tehnologii v avtomatizirovannykh sistemakh*, 16, pp. 253–256. [In Russian].
- Jansen, D.-J.** (2011a). Has the Clarity of Central Bank Communication Affected Financial Markets? Evidence from Humphrey-Hawkins Testimonies. *Contemporary Economic Policy*, 29(4), pp. 494–509.
- Jansen, D.-J.** (2011b). Mumbling with Great Incoherence: Was It Really So Difficult to Understand Alan Greenspan? *Economics Letters*, 113(1), pp. 70–72.
- Karpov, N., Baranova, J. and Vitugin, F.** (2014). Single-Sentence Readability Prediction in Russian. In: *Analysis of Images, Social Networks and Texts. AIST 2014*. Springer, pp. 91–100. https://doi.org/10.1007/978-3-319-12580-0_9

- Kauchak, D., Mouradi, O., Pentoney, C. and Leroy, G.** (2014). Text Simplification Tools: Using Machine Learning to Discover Features that Identify Difficult Text. In: *Proceedings of the 47th Annual Hawaii International Conference on System Sciences, HICSS 2014*. IEEE Computer Society, pp. 2616–2625. <https://doi.org/10.1109/HICSS.2014.330>
- Kim, Y.** (2014). Convolutional Neural Networks for Sentence Classification. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, pp. 1746–1751. <https://doi.org/10.3115/v1/D14-1181>
- Kincaid, J. P., Fishburne, R. P. Jr., Rogers, R. L. and Chissom, B. S.** (1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) For Navy Enlisted Personnel*. Naval Technical Training Command: Research Branch Report, N 8.
- Klobukova, L. P., Bitekhtina, G. A., Noreyko, L. N., Andryushina, N. P. and Odintsova, I. V.** (2014). *Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Basoviy Uroven'. Obshchee vladenie* [Russian as Foreign Language Basic Dictionary. First Level. General Knowledge]. Saint-Petersburg: Zlatoust. [In Russian].
- Korobov, M.** (2015). Morphological Analyzer and Generator for Russian and Ukrainian Languages. In: *Analysis of Images, Social Networks and Texts. AIST 2015*. Springer, pp. 320–332. https://doi.org/10.1007/978-3-319-26123-2_31
- Krioni, N. K., Nikin, A.D. and Filippova, A.V.** (2008). Avtomatizirovannaya sistema analiza slozhnosti uchebnykh textov [The Automated System Analysis of the Difficulties of Training Texts]. *Vestnik UGATU*, 11(1), pp. 101–107. [In Russian].
- Kuznetsova, O. S. and Merzlyakov, S. A.** (2015). Kommunikatsionnye kanaly Banka Rossii v kontekste mirovogo opyta [Communication Channels of the Bank of Russia in the Context of the World Experience]. *Den'gi i Kredit*, 12, pp. 34–39. [In Russian].
- Laxton, D. and Freedman, C.** (2009). *Inflation Targeting Pillars – Transparency and Accountability*. IMF Working Paper, N 262.
- Lesskis, G. A.** (1962). O razmerakh predlozheniy v russkoy nauchnoy i khudozhestvennoy proze 60-kh gg. XIX v. [On the Length of Sentences in Russian Academic Literature and Fictional Prose in 1860s]. *Voprosy Yazykoznaniya*, 2, pp. 78–95. [In Russian].
- Lesskis, G. A.** (1964). O zavisimosti mezhdu razmerom predlozheniya i ego strukturoy v raznykh vidakh teksta [The Relation Between the Size of the Sentence and Its Structure in Different Types of Text]. *Voprosy Yazykoznaniya*, 3, pp. 99–123. [In Russian].
- Lively, B. A. and Pressey, S. L.** (1923). A Method for Measuring the ‘Vocabulary Burden’ of Textbooks. *Educational Administration and Supervision*, 9, pp. 389–398.
- Large, I.** (1939). Predicting Reading Difficulty of Selections for Children. *Elementary English Review*, 16(6), pp. 229–233.
- Luan, Y. and Lin, S.** (2019). Research on Text Classification Based on CNN and LSTM. In: *2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. IEEE, pp. 352–355. <https://doi.org/10.1109/ICAICA.2019.8873454>
- Lyashevskaya, O. N. and Sharov, S. A.** (2009). *Chastotny slovar' sovremennogo russkogo yazyka (na materialakh Natsionalnogo korpusa russkogo yazyka)* [Frequency Dictionary of Modern Russian Language]. Moscow: Azbukovnik. [In Russian].

- Martinc, M., Pollak, S. and Robnik-Šikonja, M.** (2021). Supervised and Unsupervised Neural Approaches to Text Readability. *Computational Linguistics*, 47(1), pp. 141–179. https://doi.org/10.1162/coli_a_00398
- Matskovskiy, M. S.** (1976). Problemy chitabelnosti pechatnogo materiala [Issues of Readability of Print Material]. In: T. M. Dridze, A. A. Leontiev, eds. *Smyslovoe vospriyatie rechevogo soobshcheniya v usloviyakh massovoy kommunikatsii*. Moscow: Nauka, pp. 126–142. [In Russian].
- Meng, C., Chen, M., Mao, J. and Neville, J.** (2020). ReadNet: A Hierarchical Transformer Framework for Web Article Readability Analysis. In: *Advances in Information Retrieval. ECIR 2020*. Springer, pp. 33–49. https://doi.org/10.1007/978-3-030-45439-5_3
- Mikk, Ya. A.** (1970). O faktorakh ponyatnosti uchebnogo teksta: avtoreferat dissertatsii [On Clarity Factors of Academic Texts: PhD. thesis abstract]. Tartu. [In Russian].
- Mikk, Ya. A.** (1981). Optimizatsiya slozhnosti uchebnogo teksta: v pomoshch avtoram i redaktoram [Optimisation of the Academic Text Difficulty: In Aid of Authors and Editors]. Moscow: Prosveshchenie. [In Russian].
- Mohammadi, H. and Khasteh, S.** (2019). *Text as Environment: A Deep Reinforcement Learning Text Readability Assessment Model*. arXiv Preprint, arXiv:1912.05957.
- Nevdakh, M. M.** (2008). Issledovanie informatsionnykh kharakteristik uchebnogo teksta metodami mnogomernogo statisticheskogo analiza [Multivariate Statistical Analysis in Research of Information Characteristics of Academic Texts]. *Prikladnaya informatika*, 4, pp. 117–130. [In Russian].
- Oborneva, I. V.** (2006). *Avtomatizirovannaya otsenka slozhnosti uchebnykh tekstov na osnove statisticheskikh parametrov* [Automatic Assessment of the Complexity of Educational Texts Based on Statistical Parameters: PhD Thesis]. Moscow. [In Russian].
- Omotosho, B. S.** (2019). *Central Bank Communication in Ghana: Insights from a Text Mining Analysis*. MPRA Paper, N 98297.
- ООО 'INFOM'** (2021). *Izmerenie inflyatsionnykh ozhidaniy i potrebitel'skikh nastroyeniy na osnove oprosov naseleniya. Volna 95* [Estimation of Inflation Expectations and Consumer Sentiment Based on the Surveys. Round 95]. January. [In Russian]. http://cbr.ru/Collection/Collection/File/31949/inFOM_21-01.pdf [accessed on 26 February 2021].
- Piotrovskiy, R. G., Bektaev, K. B. and Piotrovskaya, A. A.** (1977). *Matematicheskaya lingvistika* [Mathematical Linguistics]. Moscow: Vyshaya shkola. [In Russian].
- Pitler, M. and Nenkova, A.** (2008). Revisiting Readability: A Unified Framework for Predicting Text Quality. In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 186–195.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J. and Manning, C. D.** (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, pp. 101–108.
- Reynolds, R.** Insights from Russian Second Language Readability Classification: Complexity-Dependent Training Requirements, and Feature Evaluation of Multiple Categories. In: *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*. Association for Computational Linguistics, pp. 289–300.

- Rubin, V. L., Liddy, E. D. and Kando, N.** (2006). Certainty Identification in Texts: Categorization Model and Manual Tagging Results. In: J. G. Shanahan, Y. Qu, J. Wiebe, eds. *Computing Attitude and Affect in Text: Theory and applications*. Springer, pp. 61–76. https://doi.org/10.1007/1-4020-4102-0_7
- Rybka, R. B., Sboev, A. G. and Ivanov, I. I.** (2015). Chislennoe modelirovanie protsedury sintaksicheskogo razbora s ispol'zovaniem neyronnykh setey [Computational Modelling of Sentence Analysis Procedure Using Neural Networks]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkul'turnaya kommunikatsiya*, 3, pp. 28–33. [In Russian].
- Santucci, V., Santarelli, F., Forti, L. and Spina, S.** (2020). Automatic Classification of Text Complexity. *Applied Sciences*, 10(20), 7285. <https://doi.org/10.3390/app10207285>
- Schwarm, S. and Ostendorf, M.** (2005). Reading Level Assessment Using Support Vector Machines and Statistical Language Models. In: *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pp. 523–530.
- Si, L. and Callan, J.** (2001). A Statistical Model for Scientific Readability. In: *Proceedings of the Tenth International Conference on Information and Knowledge Management*. New York: Association for Computing Machinery, pp. 574–576.
- Solnyshkina, S. I. and Kiselnikov, A. S.** (2015). Slozhnost' teksta: etapy izucheniya v otechestvennom prikladnom yazykoznanii [Text Complexity: Study Phases in Russian Linguistics]. *Tomsk State University Journal of Philology*, 6(38), pp. 86–99. [In Russian]. <https://doi.org/10.17223/19986645/38/7>
- Svensson, L. E. O.** (1997). Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets. *European Economic Review*, 41(6), pp. 1111–1146. [https://doi.org/10.1016/S0014-2921\(96\)00055-4](https://doi.org/10.1016/S0014-2921(96)00055-4)
- Tikhonov, A. N.** (2002) *Morfemno-orfograficheskiy slovar'* [Morphemically Spelling Dictionary]. Moscow: AST, Astrel. [In Russian].
- Tuldava, Yu. A.** (1975). Ob izmerenii trudnosti tekstov [On Measuring the Complexity of the Text]. *Uchenye zapiski Tartusskogo universiteta: Trudy po metodike prepodavaniya inostrannykh yazykov*, 345, pp. 102–120. [In Russian].
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I.** (2017). *Attention Is All You Need*. arXiv Preprint, arXiv:1706.03762.
- Vor der Brück, T., Helbig, H. and Leveling, J.** (2008). *The Readability Checker DeLite*. FernUniversität in Hagen Technical Report.