



Text Complexity: Periods of Study in Russian Linguistics

Igor O. Guryanov¹, Iskander E. Yarmakeev², Aleksandr S. Kiselnikov³, Elena V. Harkova⁴

1 Kazan Federal University, Institute of Philology and Intercultural Communication, igor.goor@gmail.com

2 Kazan Federal University, Institute of Philology and Intercultural Communication

3 Kazan Federal University, Institute of Philology and Intercultural Communication

Kazan State University of Architecture and Engineering

4 Kazan Federal University, Institute of Philology and Intercultural Communication

ABSTRACT

Linguistic studies dealing with the text analysis are up-to-date. The present article is a descriptive one aimed to give information on the periods of the text complexity studying in Russian linguistics. As the pre-survey showed unlike the foreign experience there is a lack of information on this topic in Russian studies. Text complexity is a subject of many foreign researches as well as there are many researches on what parameters should be considered as text complexity parameters. The fulfilled researcher allows to point out the three main periods. The first period that is the second half of the XX century is concentrated primarily on quantitative parameters of the text (text length, sentence length, word length etc) that is readability. The second period is the turn of the XX and XXI centuries tends to solve the problem of comprehensibility and the quantitative parameters are enlarged with some qualitative (words acquaintance, word abstractness etc.). The third period is the modern stage of text complexity studying which supposes the development of different computer programs to evaluate the text complexity automatically.

Keywords: text complexity, quantitative parameters of text complexity, qualitative parameters of text complexity, readability, text comprehensibility.



1. INTRODUCTION

Any text can be described in the terms of the scientific approach it is being studied. The idea of the academic text complexity evaluation that should be of certain level is getting topical nowadays. There is no certain list of text complexity parameters as well as there is no any range of parameters: which have influence on the text complexity. First, the periods of text complexity studying and their key parameters should be found out. As the survey showed foreign practice is much wider than the Russian one.

Readability, comprehensibility, complexity and difficulty are texts characteristics evaluated in applied linguistics with mathematical formulas and computer programs. But still now they can be confused as the four mentioned text characteristics have blurred borders in Russian linguistics and are not always exactly distinguished while researches. E. Puskina (Razumovskiy,1999) uses the term “complexity” to give information of the text parameters as well as to say how difficult the text for readers’ comprehension is. Text “complexity” (Biryukov , 1967; Lesskis , 1964 ; Zil'bergleyt et al, 2012) can be defined as the characteristic of the text parameters only while the text “difficulty” (33, 34) evaluation relates to the readers’ experience based on vocabulary and syntaxes acquaintance, semantic links understanding, the text genre and its discourse model etc (Crossley et al, 2008; Just,1987; Kosova,2006). Speaking of the text difficulty readers’ background knowledge based on their social, psychological, historical, cultural age and other individual features should be taken into consideration (Alderson,2000; Gal'perin,1981).

O. Razumovskiy (Shpakovskiy,2007) says that Russian linguistics still does not have a range of text complexity parameters as well as there is still no instrument for its evaluation.

2. METHODS AND MATERIAL

The present article is a descriptive one. The target aim is to analyze Russian linguists’ works, papers and studies and to point out the periods of text complexity study in Russian linguistics and to identify the key text complexity parameters at all stages. That is why the material for study is published works of Russian linguists of the XX-XXI centuries.

3. RESULTS



Text readability is measured based on text quantitative parameters (text, sentence and word length, number of long or difficult words etc.). Readability index demonstrates the target group of the text and correlates the reader's age, not biological but academic.

Nowadays there are over 200 formulas for different languages English, Spanish, French, German, Dutch etc (Al-Khalifa,2010).

First readability formulas were pragmatic, they served for military (Flesch Reading Ease (Flesch,1949) and Flesch-Kincaid Readability Test (Koda,2005)) and civil (McLaughlin formula (Mikk,1970)) purposes. For example, Flesch index is measured based on three constants and two non-constants which are average length of the sentence (*als*) and average number of syllables in the word (*asw*): $Flesch\ Index = 206.835 - (1.015 \times als) - (84.6 \times asw)$. The less syllables the word has the less informative it is; the shorter the sentence is the less cohesions it has. If the Flesch Index is between 90 and 100 then it is for the 4th grade students, but if the index is between 0 and 30 – the text is for the college graduates. Flesch Formula was adapted for Russian language texts by I.

Oborneva (Piotrovskiy,1977) in 2006: $Flesch\ Index = 206.836 - (1.52 \times als) - (65.14 \times asw)$. I. Oborneva analyzed 6 mln. words of 100 English literary texts and their Russian equivalents and compared the average word length in Russian and English languages using “Slovar russkogo yazyka pod redaktsyey Ozhegova” – 39174 words and “Muller English-Russian dictionary” – 41977 words. As the survey showed average English word consists of 2.97 syllables, while the Russian one – of 3.29 syllables.

First Russian readability formulas appeared in 60s-70s of the XX century as a result of K. Bektaeva, G. Lesskis, M. Matskovskiy, Ya. Mikk, R. Piotrovskiy and A. Piotrovskaya researches.

G. Lesskis (Matskovskiy,1976) was focused on the syntactic complexity on the number of simple and complex sentences, their structures, number of words etc.

Ya. Mikk was concentrated on studying the text comprehensibility which he described as “text ability to assist its understanding¹” (Nevdakh,2008), while the text difficulty he defined as “text ability to prevent its understanding²” (Nevdakh,2008). Parameters of text comprehensibility are: 1) number of words in the sentence, 2) number of acquainted words in the text and 3) ratio of concrete and abstract nouns in the text. The parameter

¹ Translated by Aleksander Kisel'nikov

² Translated by Aleksander Kisel'nikov



of the word acquaintance was obtained empirically: a group of students ranged the words into six groups from the well known to the unknown ones.

Two approaches were used for the abstractness identification. First is to subdivide nouns into three groups: a) animate and inanimate objects comprehensible by the sensory organs, b) phenomena comprehensible by the sensory organs and c) non-comprehensible by the sensory organs. The second approach is a morphological one and is based on the nouns with abstract affixes identification. Involving not only quantitative parameters but the qualitative ones in 1970 Ya. Mikk made the text comprehensibility formula (MCF) for the Estonian language: $MCF = 0.131 \times asl + 9.84 \times \text{average abstractness of the text vocabulary} - 4.59$.

In 1976 M. Matskovskiy (20) presented his formula: $X = 0.62 \times asl + 0.123 \times \% \text{ of 3 and more syllable words in text} + 0.051$. To make the constants for his formula M. Matskovskiy involved a group of sixty 7th grade students who empirically subdivided texts into seven groups from the easiest to the most difficult.

New objects of studies and new parameters started the new stage of statistical method appliance for the text complexity evaluation. K. Shannon information theory was further used by Russian linguists suggested to analyze not only quantitative parameters but qualitative: letters, syllables, grammatical morphemes, syntactical composition etc (Pushkina,2004).

As the first steps of text studying Yu. Tuldava was interested in quantitative text parameters and suggested a formula: $K = asw \times lg als$ (Tuldava,1979). Later he turned to the polysemy studying. He discovered that a word in average has 3.7 meanings: a verb – 4.6, a noun – 3.1 (Ushakov,1980). The more polysemantic words text has the more complex it is.

At modern stage linguists concentrate on academic texts (Krioni,2008; Nikin et al,2007; Osborneva,2006; Razumovskiy,1999 38 etc) to produce the proper academic texts for different purposes. Another feature of the stage is 1) the more parameters involvement and 2) correlation between quantitative and qualitative parameters discovering (Lerner,1974; Nikin et al,2007; Solnyshkina ,2014 etc).

Automatic homonym meaning identification was in focus of A. Ermakov and V. Pleshko (Ermakov et al,2002) who declared that context analysis is an integral part and made an automatic syntactical analyzer for the Russian language – Russian Context



Optimizer for Oracle Data Base. This instrument points out the nominal group to recognise the homonym meaning.

P. Tolpegin (Tomina,1985) tried to create a computer program for co-reference links between antecedent and anaphor that is the change of a noun with a third person pronoun.

A. Grechikhin (Grechikhin,2007) considered text information, sentence complexity, narrativity abstractness and the text structure clarity as the factors of text complexity. To identify the word acquaintance, he suggests 1) to distinguish real-life words and scientific terms, 2) to discover the text vocabulary variety and 3) to find long words. He also stressed the text coherence and cohesion studying importance.

Yu. Shpakovskiy is interested in academic chemistry texts. His researches aimed to study the text complexity by a range of parameters: length of words, word combinations and sentences, percent of simple and complex sentences in the text, number of links and their types. Moreover, he designed a formula: $Y = 20.24 + 0.48 \times \% \text{ of nine- and more letter long words in the text} + 0.58 \times \% \text{ of terms} + 0.41 \times \% \text{ of symbolic notations in chemical reactions}$. The auxiliary program for it is “Statistika” (Solnyshkina et al, 2014).

M. Nevdakh studying the high school texts on philosophy and economics theory made a program “Readability analysis” (Nikin et al, 2007), the key parameter of which is the amount long words with their prefixes and affixes.

Narration abstractness and linguistic constructions based on number of long words (three and more syllables), number of sentences containing long words, average sentence length, number of Participles 1 and Participles 2 and number of sentences containing them, number of complex sentences etc. are got being considered as text complexity parameters (Lerner,1974). For the text complexity evaluation they designed a computer program “Text parameters complexity evaluation”. The terms in the test increase its complexity (Razumovskiy,1999) as they are very abstract (Nevdakh,2008). Number and types of derivation structures in the word also influence the text complexity (Razumovskiy,1999).

The summarized results of Russian linguists’ researchers for Russian language text readability evaluation by I. Begtin (2014).were turned into “ru.readability.io” web-resource.



The beginning of the 21st century in Russian linguistics and text complexity evaluation is concerned with the study of coherence, synonyms, hyponyms and hyperonyms, ellipsis, references etc. To relate anaphors and antecedent while the analysis of syntactic cohesions in the sentence as well as between different sentences of the text a program “Automatic text information rubrication (in Russian, English, German and French languages)” was made (Abramov,2008).

N. Karpov (Karpov,2014) started fundamental researches on vocabulary. He studied lexical minimums and analyzed texts to understand the per cent of words which are in and out of the lexical minimum.

V. Solovyev and V. Ivanov (Tolpegin,2008) faced a problem while computer proceeding of texts, data bases and other information in Russian.

M. Solnyshkina, E. Harkova and A. Kiselnikov (Solovyev,2016) adopt foreign experience of text complexity evaluation by means of Coh-Metrix and T.E.R.A. computer programs for Russian language text analysis.

4. RESUME

To summarize, the text complexity studying in Russian linguistics may be sub-divided into three stages. The first stage is time of quantitative parameters studying. The second stage is cooperation of quantitative and qualitative parameters. The third stage is some new parameters discovering, deeper studying of discovered parameters and computer programs design.

The target of Russian linguists is to make an instrument for text complexity evaluation alike foreign T.E.R.A.

5. ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

6. REFERENCES

- Abramov V.E. 2008. Avtomaticheskoe rubritsirovanie i referirovanie tekstovoy informatsii: v tom chisle na inostrannykh yazykakh: dis... kand. tekhnich. nauk. : 05.25.05 / Abramov Valeriy Evgen'evich. – M., 163 p.
- Alderson J.C. 2000. Assessing Reading. – New York: Cambridge University Press, 2000. – 398 p.



- Al-Khalifa H.S., Al-Ajlan A.A. 2010. Automatic Readability Measurements of the Arabic Text: An Exploratory Study. *The Arabian journal for science and engineering*, 35 p.
- Begtin I.V. 2014. Chto takoe «Ponyatnyy russkiy yazyk» s tochki zreniya tekhnologii. Zaglyanem v metriki udobochitaemosti tekstov: Blog kompanii «Informatsionnaya kul'tura». URL: <http://habrahabr.ru/company/infoculture/blog/238875/>, свободный. Проверено 16.07.17.
- Biryukov B.V., Tyukhtin B.C. 1967. *O ponyatii slozhnosti*. V kn.: Logika i metodologiya nauki. Materialy IV Vsesoyuznogo simpoziuma. – M.: Prosveshchenie, Nauka, P. 219-231.
- Crossley S.A., Greenfield J., McNamara D.S. 2008. Assessing Text Readability Using Cognitively Based Indices. *Tesol Quarterly*, Vol. 42. No. 3. – P. 475-493.
- Ermakov A.E., Pleshko V.V. 2002. Sintaksicheskiy razbor v sistemakh statisticheskogo analiza teksta // *Informatsionnye tekhnologii*. : Izd. «Novye tekhnologii». no. 7. – P. 30-34.
- Filippova A.V. Upravlenie kachestvom uchebnykh materialov na osnove analiza trudnosti ponimaniya uchebnykh tekstov: avtoref. dis. ... kand. tekhn. nauk : 05.13.10 / Filippova Anastasiya Vladimirovna. – Ufa, 2010. – 20 p.
- Flesch R. 1949. *The Art of Readable Writing*. – Harper & Row, 1949. – 237 p.
- Gal'perin I.R. 1981. *Tekst kak ob"ekt lingvisticheskogo issledovaniya*. – M.: Nauka, 1981. – 140 p.
- Grechikhin A.A. 2007. *Sotsiologiya i psikhologiya chteniya: Uchebnoe posobie dlya vuzov*. – M.: MGUP, 383 p.
- Just M.A., 1987. Carpenter P.A. *The psychology of reading and language comprehension*. – MA, US: Allyn & Bacon, 1987. – 518 p.
- Karpov N.V. 2014. Identifikatsiya urovnya slozhnosti teksta i ego adaptatsiya. – URL: <http://www.slideshare.net/karpnv/ss-31225145#14356960593761&fbinitialized>, Kincaid J.P., Fishburne R.P., Rogers R.L., Chissom B.S. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel (Research Branch Report 8-75). – Memphis, TN: Naval Air Station, 1975. – 40 p.



- Koda K. 2005. *Insights into second language reading*. – Cambridge: Cambridge University Press, 2005. – 344 p.
- Kosova M.M., Zil'bergleyt M.A. 2006. *Opisatel'naya statistika uchebnykh tekstov po fizike // Trudy BGTU. Ser. VI. «Izdat. delo i poligrafiya», Vol. XIV. – P. 167-170.*
- Krioni N.K., Nikin A.D., Filippova A.V. 2008. *Avtomatizirovannaya sistema analiza parametrov slozhnosti uchebnogo teksta // Tekhnologiya i organizatsiya obucheniya : nauch. izdanie. – Ufa: UGATU, P. 155-161.*
- Lerner I.Ya. 1974. *Kriterii slozhnosti nekotorykh elementov uchebnika: Problemy shkol'nogo uchebnika. – M., Vol. I. P. 47-58.*
- Lesskis G.A. 1964. *O zavisimosti mezhdru razmerom predlozheniya i ego strukturoy v raznykh vidakh teksta. – VYa, no. 3. – P. 99-123.*
- Matskovskiy M.P. 1976. *Problemy chitabel'nosti pechatnogo materiala V kn.: Smyslovoe vospriyatie rechevogo soobshcheniya v usloviyakh massovoy kommunikatsii. – M.: Nauka, P. 126-142.*
- Mikk Ya.A. 1970. *O faktorakh ponyatnosti uchebnogo teksta: Avtoref. dis. kan. ped. nauk. / Mikk Yaan Arnol'dovich. – Tartu: Izd-vo Tart. in-ta, 1970. – 22 p.*
- Nevdakh M.M. 2008. *Issledovanie informatsionnykh kharakteristik uchebnogo teksta metodami mnogomernogo statisticheskogo analiza // Prikladnaya informatika.: Izd. "NOU «MFPU «Sinergiya»", – 2008. – no. 4. – P. 117-130.*
- Nikin A.D., Krioni N.K., Filippova A.V. 2007. *Informatsionnaya sistema analiza uchebnogo teksta // Telema-tika':tr. XIV Vserop. nauch.-metod. konf. T. 2. GosNII inform. tekhnol. i telekommunikatsiy «Informika», 2007. – P. 463-465.*
- Oborneva I.V. 2006. *Avtomatizirovannaya otsenka slozhnosti uchebnykh tekstov na osnove statisticheskikh parametrov: dis. ... kand. ped. nauk: 13.00.02 / Oborneva Irina Vladimirovna. – Moscow, 165 p.*
- Piotrovskiy R.G., Bektaev K.B., Piotrovskaya A.A. 1977. *Matematicheskaya lingvistika. Ucheb. posobie dlya ped. institutov. – M.: Vysshaya shkola, 383 p.*
- Pushkina E.P. 2004. *Teoretiko-eksperimental'noe issledovanie strukturno-semanticheskikh parametrov teksta: Avtoref. dis. kand. filol.nauk: 10.02.19 / Pushkina Elena Sergeevna. – Kemerovo, 155 p.*



- Razumovskiy O.P. 1999. *Optimologiya, chast 1. Obshchenauchnye i filosofsko-metodicheskie osnovy.* – Novosibirsk: Izdatel'stvo IDMSh, 285 p.
- Shpakovskiy Yu.F. 2007. *Otsenka trudnosti vospriyatiya i optimizatsiya slozhnosti uchebnogo teksta: (Na materiale tekstov po khimii): avtoref. dis ... kand. filol. nauk: 10.02.19 / Shpakovskiy Yuriy Frantsevich.* – Minsk, 2007. – 21 p.
- Solnyshkina M.I., Harkova E.V., Kisel'nikov A.S. 2014. *Comparative Coh-Metrix Analysis of Reading Comprehension Texts: Unified (Russian) State Exam in English vs Cambridge First Certificate in English / English Language Teaching.* – Canada. Canadian Center of Science and Education, Vol.7, No. 12. – P. 65-76.
- Solovyev V, Ivanov V. 2016. *Knowledge-Driven Event Extraction in Russian: Corpus-Based Linguistic Resources.* – Computational Intelligence and Neuroscience, – vol. 2016, – Article ID 4183760, 11 pages. doi:10.1155/2016/4183760.
- Tolpegin P.V. 2008. *Avtomaticeskoe razreshenie koreferentsii mestoimeniy tret'ego litsa russkoyazychnykh tekstov: dis. ... kand. tekhnich. nauk: 05.13.17 / Tolpegin Pavel Vladimirovich.* – M, 238 p.
- Tomina Yu.A. 1985. *Ob'ektivnaya otsenka yazykovoy trudnosti tekstov (opisanie, povestvovanie, rassuzhdenie, dokazatel'stvo): dis. ... kand. ped. nauk: 13.00.02 / Tomina Yuliya Alekseevna.* – M, 226 p.



- Tuldava Yu.A.1979. O nekotorykh kvantitativno-sistemnykh kharakteristikakh polisemii // Uchenye zapiski Tartuskogo universiteta. – Tartu: Izd-vo Tart. un-ta., – Vol. 502. – P. 107-124.
- Ushakov K.M.1980. O kriteriyakh slozhnosti uchebnogo materiala shkol'nykh predmetov // Novye issledovaniya v pedagogicheskikh naukakh. no. 2 (36) / Sost. I.K. Zhuravlev, B.C. Shubinskiy. – M.: Pedagogika, P. 33-35.
- Zil'bergleyt M.A., Shpakovskiy Yu.F., Nevdakh M.M. 2012.Povyshenie kachestva uchebnoy literatury dlya vuzov // Izdatel'skoe delo i poligrafiya: tezisy 76-y nauch.-tekhnich. konferentsii professorsko-prepodavatel'skogo sostava, nauchnykh sotrudnikov i aspirantov, Minsk, 13-20 fevralya 2012 g. (Elektronnyy resurs) / otv. za izdanie I. M. Zharskiy; UO «BGTU». – Minsk: BGTU, P. 89-92.