Proceedings Paper:

Kazakov, Dimitar Lubomirov orcid.org/0000-0002-0637-8106, Minkov, Stefan, Margova, Ruslana et al. (2 more authors) (Accepted: 2025) Towards Creating a Bulgarian Readability Index. In: Proceedings of the Workshop on on Advancing NLP for Low-Resource Languages (LowResNLP) at RANLP 2025. Workshop on Advancing NLP for Low-Resource Languages at RANLP 2025, 11 Sep 2025 Association for Computational Linguistics (ACL), BGR. (In Press)

Towards Creating a Bulgarian Readability Index

Dimitar Kazakov
University of York
dlk2@york.ac.uk

Stefan Minkov
GATE Institute
stefanminkov2003@gmail.com

Ruslana Margova
GATE Institute
ruslana.margova@gate-ai.eu

Irina Temnikova
GATE Institute
irina.temnikova@gate-ai.eu

Ivo Emanuilov
GATE Institute
ivo.emanuilov@gate-ai.eu

Abstract

Readability assessment plays a crucial role in education and text accessibility. While numerous indices exist for English and have been extended to Romance and Slavic languages, Bulgarian remains underserved in this regard. This paper reviews established readability metrics across these language families, examining their underlying features and modelling methods. We then report the first attempt to develop a readability index for Bulgarian, using end-of-school-year assessment questions and literary works targeted at children of various ages. Key linguistic attributes, namely, word length, sentence length, syllable count, and information content (based on word frequency), were extracted, and their first two statistical moments, mean and variance, were modelled against grade levels using linear and polynomial regression. Results suggest that polynomial models outperform linear ones by capturing non-linear relationships between textual features and perceived difficulty, but may be harder to interpret. This work provides an initial framework for building a reliable readability measure for Bulgarian, with applications in educational text design, adaptive learning, and corpus annotation.

1 Introduction

The importance of text comprehensibility is undeniable and even crucial in cases, such as the medical domain and emergency situations (Temnikova et al., 2015; Friedman et al., 2008). The most straightforward way to estimate how comprehensible a text is is to understand how easy it is to read by a target group of readers. The linguistic characteristics that make a text easier or harder to read are referred to as *readability*. Its significance is shown by the fact a substantial number of languages have seen work on quantifying readability and improving it through measures, such as manual or automatic text simplification (Alfear et al., 2024; Al-Thanyyan and Azmi, 2021; Siddharthan, 2014; Saggion and Hirst, 2017).

A text readability index is a language-specific tool that requires appropriate resources for its creation. Bulgarian NLP has a half century-long tradition yet in this aspect it ranks among the lower-resourced languages, as it lacks such an index or the specific age-appropriate text corpora that would support its creation. This motivates our work, which represents the first steps towards building a readability index for Bulgarian.

2 Background

Readability is defined as the effect of all elements that make a text more or less comprehensible to a group of readers. Some scholars consider that readability is also linked to how interesting the text is (Dale and Chall, 1949; DuBay, 2004; Klare, 1963; McLaughlin, 1969; Hargis et al., 1998). Text complexity can be affected on multiple levels, from morphology to pragmatics, some of which are hard to evaluate automatically. Most frequently, readability is estimated through a combination of surface linguistic features, such as the average length of words in characters or syllables, the average sentence length in words, the words' difficulty estimated by their frequency in large corpora or their mere presence in a corpus of a certain size. These features are then typically used as attributes of some type of regression, where the predictor aims to approximate the quantified reading level of the training texts, and the resulting formula is evaluated in terms of goodness of fit, through the coefficient of determination r^2 on unseen data. The regression equation, which contains simple arithmetic operations and language-specific numerical parameters, is then used as "readability index". These indices were originally created to find reading material that matched the reading abilities of students of a certain grade or age.

These traditional readability indices have been criticised for their somewhat simplistic approach, which does not take into account factors at the syntactic, semantic, and pragmatic levels, the logical order of ideas, etc. Some of these shortcomings have been addressed in recent readability tools and resources.

One such example is the English-language Medical Research Council (MRC) Psycholinguistic Database (Coltheart, 1981) and Coh-Metrix (Graesser et al., 2004). MRC is a lexical database containing values for several psychological features for more than 150,000 English words. The features include familiarity, age of acquisition, concreteness, word length, etc. Coh-Metrix is an automatic text analysis tool that detects deeper text complexity and comprehensibility features, such as cohesion, word frequency, concreteness, familiarity, and sentence structure complexity.

Coh-Metrix has also been adapted for Spanish (Quispesaravia et al., 2016) and Brazilian Portuguese (Scarton and Aluísio, 2010). A similarly complex tool, called ErreXail, was created for Basque (Gonzalez-Dios et al., 2014). Machine Learning (ML) models have also begun being used to estimate text readability, making use of more complex representations, such as embeddings. Such models have been created for several languages including English, Spanish, Basque, French, Catalan, Italian, French, and Slovene, variously based on regression, classifiers, random forests, neural networks, or transformers (Vajjala and Meurers, 2012; Vajjala and Lučić, 2018; Madrazo Azpiazu and Pera, 2021; Martinc et al., 2019).

2.1 English Readability Indices

While readability indices have their limitations, they constitute the first step towards estimating text comprehensibility. Unsurprisingly, the language best supported with readability resources is English, and its readability indices are often adapted to other languages by obtaining a new set of language-specific numerical parameters. Here are some of the best known readability indices for English:

- **Flesch Reading Ease** (Flesch, 1948)

  $\text{FRE} = 206.835 - 1.015\frac{\#\text{words}}{\#\text{sentences}} - 84.6\frac{\#\text{syllables}}{\#\text{words}}$

- **Flesch-Kincaid Grade Level** (Kincaid et al., 1975), which outputs a U.S. school grade level.

  $\text{FKGL} = 0.39\frac{\#\text{words}}{\#\text{sentences}} + 11.8\frac{\#\text{syllables}}{\#\text{words}} - 15.59$

- **Gunning Fog Index** (Gunning, 1952)

  $\text{Fog Index} = 0.4\frac{\#\text{words}}{\#\text{sentences}} + 100\frac{\#\text{complex words}}{\#\text{words}}$

- SMOG Index (McLaughlin, 1969), designed for health literacy texts.

  $\text{SMOG Grade} = 1.0430\sqrt{\#\text{polysyllabic words}\frac{30}{\#\text{sentences}}} + 3.1291$

- **Coleman-Liau Index** (Coleman and Liau, 1975)

  $\text{CLI} = 0.0588L - 0.296S - 15.8$

  where L = average letters per 100 words and S = average sentences per 100 words

- **Automated Readability Index (ARI)** (Senter and Smith, 1967), which also outputs a U.S. grade level.

  $\text{ARI} = 4.71\frac{\#\text{characters}}{\#\text{words}} + 0.5\frac{\#\text{words}}{\#\text{sentences}} - 21.43$

2.2 Readability Beyond English

All features used above are easy to compute, which has led to efforts to adopt these indices for many other languages. Some of them are shown in Table 1. There are also readability formulae for several Slavic languages, which are mostly adapted from English (see Table 2).

To the best of our knowledge, there is no Bulgarian readability index. Both research and practical solutions on that topic remain limited. In fact, the Bulgarian official educational regulations do not mention readability.[1] There is not even a universally accepted Bulgarian term for this concept. Within the limited body of existing literature in Bulgarian on

[1] https://www.mon.bg/nfs/2018/01/naredba_6_11.08.2016_bg_ezik.pdf

the subject, readability is variously referred to as “четимост” (Sharkova and Garov, 2015) and “четивност” (Borisova, 2017), despite certain authors arguing in favour of the latter term (Angelova, 2018).

References to readability indices (RI) in Bulgarian publications are rare and typically pertain to educational curricula up to the fourth grade. In such cases, the methodologies mentioned are often based on adaptations of indices originally developed for the Russian language (Yocheva, 2017).[2]

There has beeen research to create primary education texts in Bulgarian annotated with reading difficulty. This resource was created by translating Italian children's texts into Bulgarian, calculating several of their readability characteristics, and correlating them with finger-tracking results from 73 Bulgarian children (Pirelli and Koeva, 2024; Lento et al., 2024; Koeva et al., 2023). However, the aim of this research was never to create a Bulgarian readability index, and the use of translated texts is a limitation of the corpus. This leaves our article as the first to present efforts towards creating a readability formula for Bulgarian.

3 Data

The initial dataset consisted of a collection of 68 texts of national external assessment exams for grades 4, 7 and 10,[3] as well as the end of grade 12 Bulgarian matriculation exam.[4] The texts are published on the website of the Ministry of Science and Education, and we have only used those parts that test language comprehension for our purposes. The texts used in the matriculation exams are a balanced, 50:50 sample from Bulgarian modern classics and journalistic publications. The final dataset also incorporates 49 excerpts of fiction books in Bulgarian listed as recommended reading for grades 1–12. For each grade, several excerpts of approximately 1000 words have been selected: 6 or 7 for grades 1–4, and 3 for grades 5–12.

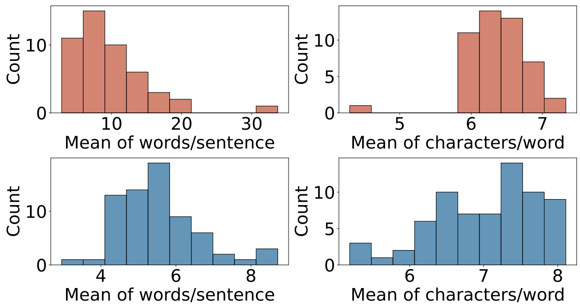

Figure 1: Descriptive statistics of data: Assessment texts (top), Literary prose (bottom)

The texts included both Bulgarian originals and translations. The year of publication was chosen to be around the middle of the 20th century as the best trade-off between representing Bulgarian as currently spoken and the lack of copyright. Note there is one mediaeval text adapted to modern Bulgarian. The texts are only in prose – due to the chosen focus of the task poetry was not included. Figure 1 contains descriptive statistics for each of the two parts of the corpus.

4 Methodology

For the purposes of this study, a number of choices need to be made and the outcomes compared. To begin with, one needs to find a suitable and representative dataset of texts corresponding to various levels of readability. Ideally, each text will have its readability level assigned on an ordinal or absolute (numerical) scale. With such explicit annotation, one can train an ordinal, linear, etc. regression model expressing the readability level as a function of a salient set of features defined over the text, typically relating to such statistics as word and sentence length or word frequency.

Optional preprocessing steps, such as stemming, mapping words to lexemes, removing words from the closed lexicon (also known as stop-words) may be carried out. After that, the chosen statistics $x_1, \ldots, x_n$ are calculated for each text and a regression model is fitted in order to express the readability score as a function of these statistics.

A linear model $\hat{y} = f(\mathbf{x})$ is the obvious first stop in the search for the best model due to its simplicity and interpretability, e.g. *the longer the average length of the words and the sentences, the less readable the text.*

[2]The links to these adapted formulas are currently inaccessible for analysis due to restrictions on access to Russian websites.

[3]https://www.mon.bg/obshto-obrazovanie/natsionalno-vanshno-otsenyavane-nvo/

[4]https://www.mon.bg/obshto-obrazovanie/darzhavni-zrelostni-izpiti-dzi/izpitni-materiali-za-dzi-po-godini/

Language	Index	Predicted	Features
German	Amstad (Amstad, 1978)	Reading-ease score (higher = easier)	Average sentence length (ASL); Average word length in syllables (AWL)
German	Wiener Sachtextformel (Bamberger and Vanecek, 1984)	age or education level	Proportion of words with three or more syllables; ASL; Proportion of words with six or more characters; Proportion of one-syllable words
Swedish, Danish, Norwegian	LIX (Lesbarhetsindex) (Björnsson, 1968)		ASL; Percentage of long words (>6 letters)
Dutch	Brouwer Leesindex (Brouwer, 1963)	ideal Dutch proficiency	ASL; AWL
French	Flesch Douma (Douma, 1960)		ASL; AWL
Romanian	Dascălu's adaptation (Dascălu et al., 2015)		Sentence/syllable/word length; Lexical complexity, parts of speech
Japanese	JARI (Fujita et al., 2012)		kana/kanji counts, word/sentence length, character/word complexity
Chinese	CRIE (Chinese Readability Index Explorer) (Sung et al., 2015)	School-year level classification (Year1–12)	Lexical (word length, stroke count); Syntactic, semantic features
Arabic	(AbuShaira, 2011)		Morphological, lexical, syntactic text features
Persian	(Behzadi and Mohammadi, 2017)		Sentence length, word length, lexical density
Hindi	(Kumar et al., 2020)		sentence length, word length, syllable/character counts
Indonesian	Dwiyanto's Score (Pranowo, 2011)		Average number of paragraphs, # sentences per paragraph; sentence length; share (in %) of: extended sentences, compound sentences, sentences with polysemy, passive sentences, unfamiliar words, abstract words, specialised terms, conjunctions, loan words and phrases)

Table 1: Non-English readability indices

Language	Formula	Features
Polish	Jastrzębski's Index (Jastrzębski, 1981)	ASL, AWL
Czech	Flesch adaptation (Čech, 2013)	ASL, AWL (syllables)
Slovak	Flesch adaptation (Ivanová, 2010)	ASL, AWL (syllables)
Serbian & Croatian	Flesch / LIX (Mihaljević and Skelin, 2017)	ASL proportion of long words (>6 chars), syllable counts
Russian	(Solovyev et al., 2018)	sentence length, word length, syntactic complexity, vocabulary metrics
Russian	(Solovyev et al., 2023)	ASL, AWL, frequency list of the Russian elementary school textbooks

Table 2: Readability indices for Slavic languages

Additional features derived from the original ones (and known as *basis functions*) can be added to the data in order to search for non-linear relationships, e.g. if it appears that the growth of the readability index is faster than linear with respect to the sentence length (sl), one could add the feature (sl^2) in order to better approximate that relationship. Similarly, interaction terms, that is, the product of two original features could be added as a new feature to capture the fact that doubling both the average word length and the average sentence length results in more than double the growth of the readability score. All of the above can easily be achieved through the use of polynomial regression, which combines all original features into all possible terms of up to a certain order n, e.g. for $n = 2$, all terms of type $x_i \times x_j \quad \forall i, \forall j$ will be added.

The result of the regression is evaluated on unseen data using the so called coefficient of determination, r^2. A value of 1 indicates all unseen data fits perfectly the model, $r^2 = 0$ corresponds to a model that is no better than simply predicting the average score of the texts in the training data set in all cases, without considering any of the attributes. Negative values of r^2 are possible despite the somewhat oddly chosen name of this evaluation metric, and would suggest a fit that is even worse, e.g. the model predicts trends that are opposite to the ones observed in the data.

The most common features used in the majority of related indices are the average word length expressed as a number of characters and the average number of words in the sentence. We are also adopting these here. In addition, we consider the average number of syllables per word, which is calculated as the number of vowels or graphemes containing a vowel. Bulgarian orthography is mostly phonetic with a few infrequent digraphs (дж, дз) and complex graphemes, such as ч = ch, щ = sht, ю = iu/yu and я = ia/ya, yet such a feature may prove useful if readability is related to, say, the length of prosody patterns within a word.

Adding the standard deviation σ of at least some of the features is another attempt better to represent the underlying distributions: while two texts may have the same average number of words per sentence, a greater σ would mean the text is more likely to have sentences of extreme length, which may prove more challenging to the reader.

It may be helpful to mention that the default expectation for word count is to find an overdispersed distribution, with variance σ^2 greater than the mean, e.g.:

$$Variance = Mean + \frac{Mean^2}{k}$$

while the number of letters per word produces tighter distributions with variance closer to the mean, which is modelled well by a Poisson ($\mu = \sigma^2$) or negative Binomial distribution.

We have also experimented with features reflecting how common or rare a given word is in the data. This was either quantified, using Shannon entropy, $log_2 P(w_i)$ or represented as a Boolean feature expressing whether the word appeared at all in the training data.

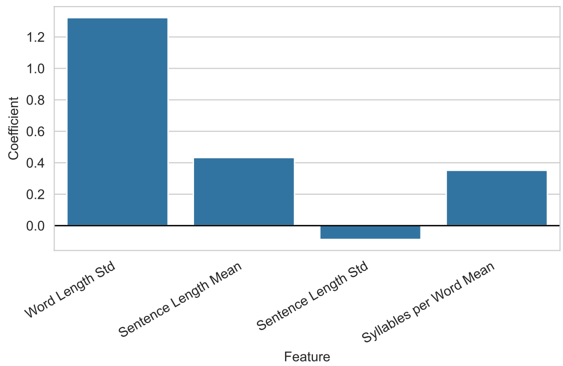

Figure 2: Coefficients for best-performing Linear Regression

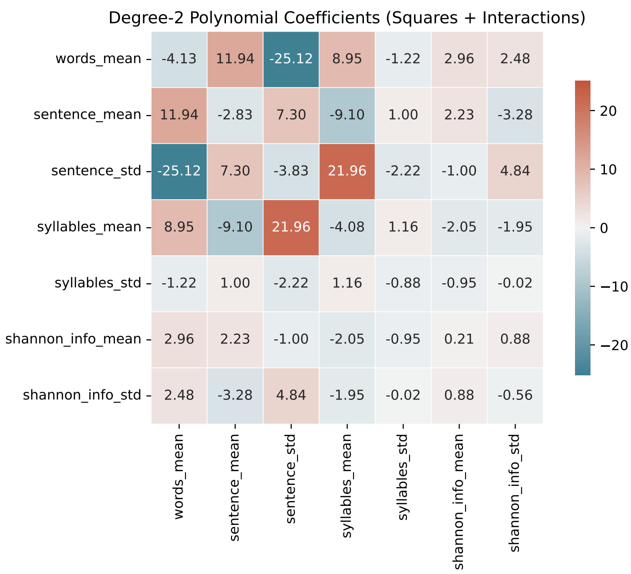

Figure 3: Coefficients for best-performing Polynomial-2 Regression

5 Results

We found that linear regression offered the easiest to interpret model, but performed worse than the polynomial regression with degree two. We also discovered that removing the words of the closed vocabulary always improved the results, so all reported findings in the rest of this section make use of this preprocessing step. The following results are based on what we referred to as the main dataset earlier in the text.

The coefficients for the best-performing linear and quadratic regression are displayed in Figures 2 and 3.

Considering the linear regression approach, we found that the input features that result in the best fit are the standard deviation of word length, the mean and the standard deviation of sentence length, and the mean number of syllables per word (x_1 to x_4):

$$\hat{y} = 1.32x_1 + 0.43x_2 - 0.09x_3 + 0.35x_4$$

The $r^2 = 0.41$ goodness of fit is an encouraging, if not exciting, result. The fact that word length standard deviation is present in the equation, but the mean is not, appears less puzzling when we are reminded of the Poisson-like distribution of this property, where $\mu \approx \sigma$.

The polynomial regression with degree 2 outperformed the linear regression by almost 44% with $r^2 = 0.59$. Seven features were used in the best model, namely, word length mean, sentence length mean and standard deviation, syllables per word count mean and standard deviation, and Shannon entropy mean and standard deviation, x_1 to x_7, respectively.

$$\begin{aligned}
\hat{y} = \\
&- 2.6372\,x_1 - 0.7867\,x_2 + 1.7984\,x_3 \\
&- 4.1278\,x_1^2 - 2.8260\,x_2^2 - 3.8334\,x_3^2 \\
&+ 2.0329\,x_4 + 1.8077\,x_5 - 0.5070\,x_6 + 0.2704\,x_7 \\
&- 4.0808\,x_4^2 - 0.8801\,x_5^2 + 0.2061\,x_6^2 - 0.5575\,x_7^2 \\
&+ 11.9439\,x_1x_2 - 25.1250\,x_1x_3 + 8.9497\,x_1x_4 \\
&- 1.2169\,x_1x_5 + 2.9605\,x_1x_6 + 2.4798\,x_1x_7 \\
&+ 7.2979\,x_2x_3 - 9.0990\,x_2x_4 + 0.9970\,x_2x_5 \\
&+ 2.2256\,x_2x_6 - 3.2815\,x_2x_7 + 21.9574\,x_3x_4 \\
&- 2.2181\,x_3x_5 - 1.0012\,x_3x_6 + 4.8432\,x_3x_7 \\
&+ 1.1557\,x_4x_5 - 2.0520\,x_4x_6 - 1.9476\,x_4x_7 \\
&- 0.9511\,x_5x_6 - 0.0247\,x_5x_7 + 0.8845\,x_6x_7
\end{aligned}$$

We experimented with Polynomial-2 regression not containing any interaction terms to see if the better results stem in the non-linear relationship between individual features and the predicted output, but the results dropped substantially, to $r^2 = 0.43$.

$$\begin{aligned}
\hat{y} = \\
&- 9.4568\,x_1 + 0.0155\,x_2 + 0.8595\,x_3 + 5.9938\,x_4 \\
&+ 10.8226\,x_5 - 1.9557\,x_6 + 2.9524\,x_7 \\
&+ 7.8462\,x_1^2 - 0.0820\,x_2^2 - 0.6443\,x_3^2 - 4.6842\,x_4^2 \\
&- 9.3665\,x_5^2 + 1.0714\,x_6^2 - 3.1744\,x_7^2
\end{aligned}$$

6 Discussion

The results so far indicate that more data may be needed, as the strong contribution of interaction terms, which do not appear in any of the indices of Section 2.1, is suggestive of overfitting. Our unreported results on the assessment texts alone, which are essentially end of school year comprehension questions, showed that the authors of these questions did not make an effort to adjust their style to the age of the reader in any meaningful way. We also discovered that removing the words of the closed lexicon (the so-called stop list) improves the outcome.

7 Conclusions

In conclusion, this study reviewed established readability metrics across multiple language families and introduced the first attempt to develop a readability index for Bulgarian. By analyzing end-of-school-year assessment materials and children's literature, key linguistic features–word length, sentence length, syllable count, and information content–were extracted and statistically modelled against grade levels. The findings indicate that polynomial regression more effectively captures the non-linear relationship between these features and text difficulty compared to linear models, though with reduced interpretability. This research lays the groundwork for a Bulgarian readability index, with promising applications in educational content creation, adaptive learning systems, and the legal domain.

8 Work Limitations

We are aware that the surface linguistic characteristics of our choice do not reflect all aspects of text comprehensibility. At the same time, using only features that are expected to have a bearing on the level of readability was a helpful way to gauge the suitability of texts used. We expect to see additional features, such as embeddings, included in our future experiments as we gradually expand our corpus.

9 Ethical and Legal Considerations and Broader Impact

We are only using data in the public domain in this study. Publishing a readability index can only contribute to social goals, such as providing accessible, easy to understand information to the public.

Our work sheds light on the surface linguistic complexity and readability characteristics of Bulgarian exam materials and Bulgarian literature books recommended to specific school age groups. Our finding that the materials for different school classes cannot be distinguished on the basis of psycholinguistic characteristics known to affect text comprehension (DuBay, 2004) should probably lead to more in-depth experiments to test whether such materials are appropriate for the Bulgarian school grades they were designed for. In such a way, our findings may assist in improving school education.

Our specific interest is creating a formula to provide a measurable way to estimate the readability of Bulgarian laws. The Law on Normative Acts and Decree No 883 of 24.04.1974 on the implementation of the Law on Normative Acts represent the Bulgarian legal framework that ensures new laws are clear, complete, and easy to interpret. Its main principles are: Precision of Norms, Interpretation of Ambiguities, Prohibition of Extensive Interpretation, and Filling Legal Gaps. These principles are designed to ensure clarity, completeness, and legal predictability, protecting citizens' rights and maintaining consistency in the legal system. The ability to quantify these desirable properties of a text would provide support to the strive for high quality legislation that meets the requirements of the rule of law and ensures legal certainty.

10 Acknowledgments

This research was part of the GATE project B-CLEAR funded by the Horizon 2020 WIDESPREAD2018-2020 TEAMING Phase 2 Programme under grant agreement No. 857155, the Programme "Research, Innovation and Digitalisation for Smart Transformation" 2021–2027 (PRIDST) under grant agreement No. BG16RFPR002-1.014-0010-C01, and the project BROD (Bulgarian-Romanian Observatory of Digital Media) funded by the EU Digital Europe Programme under contract No. 101083730. We also thank the national infrastructure CLaDA-BG (https://clada-bg.eu/bg/) for making available the word stop list and word frequency lexicon used here, and Prof. Tatyana Angelova for her suggestions.

References

M. AbuShaira. 2011. Readability formulas: An overview and a proposed Arabic readability formula. *International Journal of Academic Research*, 3(1):79–84.

Suha S. Al-Thanyyan and Aqil M. Azmi. 2021. Automated text simplification: a survey. *ACM Computing Surveys (CSUR)*, 54(2):1–36.

Noof Abdullah Alfear, Dimitar Kazakov, and Hend Al-Khalifa. 2024. Meta-evaluation of sentence simplification metrics. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11229–11235, Torino, Italia. ELRA and ICCL.

Toni Amstad. 1978. *Wie verständlich sind unsere Zeitungen?* Ph.D. thesis, University of Zürich.

Tatyana Angelova. 2018. The child and his/her literacy based on international study PIRLS'2006 and national external assessment for primary education – 4th grade. *Детето и педагогиката*, pages 133–142.

Richard Bamberger and Eva Vanecek. 1984. *Lesen - Verstehen - Lernen - Schreiben*. Jugend Volk Wien.

Bahareh Behzadi and Mohammad Mohammadi. 2017. Measuring the readability of Persian texts: Development and evaluation of a new Persian readability formula. *Studies in Literature and Language*, 15(2):1–10.

Carl-Hugo Björnsson. 1968. *Läsbarhet*. Liber.

Tanya Borisova. 2017. *Fundamentals of Reading Literacy Grades 1–4*. Yambol: Diagal Print.

P. Brouwer. 1963. *Tekstanalyse: Een methode voor het bepalen van de moeilijkheidsgraad van teksten*. Wolters.

Meri Coleman and T. L. Liau. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2):283–284.

Max Coltheart. 1981. The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4):497–505.

Edgar Dale and Jeanne S. Chall. 1949. The concept of readability. *Elementary English*, 26(1):19–26.

Mihai Dascălu, Silviu Bîrleanu, and Scott A. Crossley. 2015. Analyzing Romanian texts: Proposing a readability formula for Romanian language. In *2015 International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 47–53.

J. Douma. 1960. Een Nederlandse leesbaarheidsindex. *Tijdschrift voor Taal en Letteren*.

William H. DuBay. 2004. The principles of readability. *Education Resources Information Center (ERIC)*. 76 pages, URL: https://eric.ed.gov/?id=ED490073.

Rudolf F. Flesch. 1948. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233.

Daniela B Friedman, Manju Tanwar, and Jane VE Richter. 2008. Evaluation of online disaster and emergency preparedness resources. *Prehospital and disaster medicine*, 23(5):438–446.

Atsushi Fujita, Satoshi Sato, and Tetsuro Ishikawa. 2012. Automatic readability assessment for Japanese text.

Itziar Gonzalez-Dios, María Jesús Aranzabe, Arantza Díaz de Ilarraza, and Haritz Salaberri. 2014. Simple or complex? assessing the readability of Basque texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 334–344, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Arthur C Graesser, Danielle S McNamara, Max M Louwerse, and Zhiqiang Cai. 2004. Coh-metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, 36(2):193–202.

Robert Gunning. 1952. *The Technique of Clear Writing*. McGraw-Hill.

Gretchen Hargis, Ann Hernandez, Polly Hughes, Jennifer Ramaker, Shannon Rouiller, and Elizabeth Wilde. 1998. *Developing Quality Technical Information: A Handbook for Writers and Editors*. Prentice Hall PTR.

Iveta Ivanová. 2010. Hodnotenie čitateľnosti slovenských textov. *Slovenská reč*, 75(1):1–11.

Jerzy Jastrzębski. 1981. *Czytelność tekstów polskich*. Wydawnictwo Uniwersytetu Gdańskiego.

J. Peter Kincaid, Roger P. Fishburne, Richard L. Rogers, and Brad S. Chissom. 1975. Derivation of new readability formulas for navy enlisted personnel. Technical Report 8-75, Naval Technical Training Command Millington TN Research Branch.

George R. Klare. 1963. *The Measurement of Readability*. Iowa State University Press, Ames, Iowa.

Svetla Koeva, Valentina Stefanova, Ivelina Stoyanova, and Maria Todorova. 2023. Assessing the reading literacy of early graders. *Bulgarian Language*, 70(4):92–102.

Shiv Kumar, Sandeep Kumar, and Rajendra Singh. 2020. A readability index for Hindi language texts. *Procedia Computer Science*, 167:1658–1667.

Alessandro Lento, Andrea Nadalini, Marcello Ferro, Claudia Marzi, Vito Pirrelli, Tsvetana Dimitrova, Hristina Kukova, Valentina Stefanova, Maria Todorova, and Svetla Koeva. 2024. Assessing reading literacy of Bulgarian pupils with finger–tracking. In *Proceedings of the Sixth International Conference on Computational Linguistics in Bulgaria (CLIB 2024)*, pages 140–149.

Iñaki Madrazo Azpiazu and Sole Pera. 2021. Multilingual and cross-lingual readability: Novel resources and methods for the assessment of reading difficulty. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4661–4671.

Matej Martinc, Senja Pollak, and Marko Robnik-Šikonja. 2019. Supervised and unsupervised neural approaches to text readability. *Computational Linguistics*, 45(3):495–530.

G. Harry McLaughlin. 1969. Smog grading — a new readability formula. *Journal of Reading*, 12(8):639–646.

Bojan Mihaljević and Ivana Skelin. 2017. Readability of croatian texts. *Information and Communication Technology*, 5:59–64.

Vito Pirelli and Svetla Koeva. 2024. Developing materials for assessing reading literacy and comprehension of early graders in Bulgaria and Italy. *Foreign Language Teaching*, 51(1).

D. D. Pranowo. 2011. Instrument of Indonesian texts readability. http://staff.uny.ac.id/sites/default/files/Readability%20instrumentthesis.pdf. [Online].

Anibal Quispesaravia, Wilker Perez, Maria S. Cabezudo, and Fernando Alva-Manchego. 2016. Coh-metrixesp: A complexity analysis tool for documents written in Spanish. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4694–4698.

Horacio Saggion and Graeme Hirst. 2017. *Automatic text simplification*, volume 32. Springer.

Caroline Scarton and Sandra M. Aluísio. 2010. Coh-metrix-port: A readability assessment tool for texts in Brazilian Portuguese. In *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language, Extended Activities Proceedings (PROPOR)*, volume 10, pages 1–2.

Richard J. Senter and Edward A. Smith. 1967. Automated readability index. Technical Report AMRL-TR-66-220, Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio.

Denitza Sharkova and Kosta Garov. 2015. Приложения на облачни технологии в обучението. In *VIII Национална конференция „Образованието и изследванията в информационното общество"*, page 166.

Advaith Siddharthan. 2014. A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2):259–298.

Valery Dmitrievich Solovyev, VV Ivanov, and Marina Ivanovna Solnyshkina. 2023. Readability formulas for three levels of Russian school textbooks. *Записки научных семинаров ПОМИ*, 529(0):140–156.

Vladimir Solovyev, Marina Solnyshkina, Viktor Ivanov, and Andrey Kiselnikov. 2018. Assessment of reading difficulty levels in Russian texts. *Procedia Computer Science*, 136:234–239.

Yao-Ting Sung, Tsung-Hsien Chang, and Kuan-Eng Huang. 2015. Chinese readability index explorer (CRIE): A Chinese readability assessment system. *Behavior Research Methods*, 47(2):340–351.

Irina Temnikova, Sarah Vieweg, and Carlos Castillo. 2015. The case for readability of crisis communications in social media. In *Proceedings of the 24th international conference on world wide web*, pages 1245–1250.

Sowmya Vajjala and Ivana Lučić. 2018. One size does not fit all: Multi-task learning for genre-based readability assessment. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 173–179.

Sowmya Vajjala and Detmar Meurers. 2012. On improving the accuracy of readability classification using insights from second language acquisition. In *Proceedings of the Seventh Workshop on Innovative Use of NLP for Building Educational Applications*, pages 163–173.

Kalina Yocheva. 2017. Approaches to the formation of reading literacy in bilingual students · Подходи при формирането на четивна грамотност у ученици билингви. *Rhetorics and communication · Риторика и комуникации.*

Radek Čech. 2013. Automatické hodnocení čitelnosti češtiny. In *Proceedings of the 9th Workshop on Recent Advances in Slavonic Natural Language Processing*, pages 33–40.