



Adaptation of Classic Readability Metrics to Czech

Klára Bendová^(✉)  and Silvie Cinková 

Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics,
Charles University, Malostranské nám. 25, 118 00 Praha 1, Czechia

{bendova,cinkova}@ufal.mff.cuni.cz

<http://ufal.mff.cuni.cz>

Abstract. We have fitted four classic readability metrics to Czech, using InterCorp (a parallel corpus with manual sentence alignment), CzEng 2.0 (a large parallel corpus of crawled web texts), and the `optimize.curve` fit algorithm from the SciPy library. The adapted metrics are: Flesch Reading Ease, Flesch-Kincaid Grade Level, Coleman-Liau Index, and Automated Readability Index. We describe the details of the procedure and present satisfactory results. Besides, we discuss the sensitivity of these metrics to text paraphrases and correlation of readability scores with empirically observed reading comprehension, as well as the adaptation of Flesch Reading Ease to Czech from Russian.

Keywords: Readability · Machine learning · Czech · Flesch Reading Ease · Flesch-Kincaid Grade Level · Coleman-Liau Index · Automated Readability Index · Parallel corpus

1 Introduction

This study describes a machine-learning based adaptation of classic readability formulas to Czech, using the parallel corpora InterCorp [26] and CzEng 2.0 [29] (see Sect. 3). Readability is “the ease of reading created by the choice of content, style, design, and organization that fit the prior knowledge, reading skill, interest, and motivation of the audience” [1] (p. 6). Especially in the English-speaking community, readability has been extensively researched [1], and many metrics have been established to assess readability automatically. The most classic examples are **Flesch Reading Ease** [7], **Flesch-Kincaid Grade Level** [8], **Coleman-Liau index** [9], and **Automated Readability Index** [10].

The Flesch Reading Ease was reported to have a good correlation with the reading comprehension: (“0.7 with the 1925 McCall-Crabbs reading tests and 0.64

Supported by the Czech Science Foundation grant 19-19191S: Linguistic Factors of Readability in Czech Administrative and Educational Texts. The work described herein has also been using data/tools/services provided by the LINDAT/CLARIAH-CZ Research Infrastructure (<https://lindat.cz>), supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2018101).

© Springer Nature Switzerland AG 2021

K. Ekštejn et al. (Eds.): TSD 2021, LNAI 12848, pp. 159–171, 2021.

https://doi.org/10.1007/978-3-030-83527-9_14

with the 1950 version of the same tests” [1], p. 58). At the time of its origin, it was known among publishers to increase readership by 40 to 60 per cent [1], p. 58.

The classic metrics certainly do not seem to contain any language-specific features, since they consider mainly word length (in characters or syllables) and sentence length (in tokens). However, the distributions of these lengths are language-specific, and so are syllable definitions. To keep the score scales comparable across languages, the function parameters must be tailored to each language individually.

Although neural-network based readability formulas are emerging [2, 4], these traditional metrics are still widely used in professional writing as well as in language teaching and assessment [3, 5]. They are even integrated in the reviewing functionalities of MS Word and Office Libre. Therefore we find it appropriate to provide their Czech adaptations as long as the traditional formulas have not been generally replaced by other readability assessment methods.

The paper is structured as follows: first we give a brief overview of the selection of the metrics we have adapted (Sect. 2), leaving aside more linguistically informed metrics such as Coh-Matrix [6] as well as the neural-network based approaches. Then we describe the data sets we used for the actual adaptation and a correlation measurement of these adapted metrics with reading comprehension (individual subsections of Sect. 3). Then we explain the adaptation method (Sect. 4), and eventually we report and discuss the results of the adaptation as well as the correlation of the adapted readability formulas with reading comprehension (Sects. 5 and 6).

2 Related Work

2.1 Readability Metrics

Flesch Reading Ease. The Flesch Reading Ease scales between 0 (most difficult) and 100 (easiest). The easiest level approximately corresponds to four school years of education, whereas texts below 30 require reading skills at the level of a college-graduate. The formula considers the mean of syllables per token and the mean of tokens per sentence. The scale is interpreted as follows from Table 1.

$$ReadingEase = 206.935 - 1.015(Tokens/Sentences) - 84.6(Syllables/Tokens)$$

Table 1. Scale of the Flesch Reading Ease [7]

Reading ease score	Style description	Estimated reading grade
0–30	Very difficult	College graduate
30–50	Difficult	13th to 16th grade
50–60	Fairly difficult	10th to 12 grade
60–70	Standard	8th to 9th grade
70–80	Fairly easy	7th grade
80–90	Easy	6th grade
90–100	Very easy	5th grade

Flesch-Kincaid Grade Level. The Flesch-Kincaid Grade Level is derived from the Flesch Reading Ease. It is simplified and converted to grade level (according to the U. S. education system) – roughly as years of education (0–15), considering the same variables as the Flesch Reading Ease:

$$GradeLevel = 0.39(Tokens/Sentences) + 11.8(Syllables/Tokens) - 15.59$$

Automated Readability Index. The Automated Readability Index renders readability as the U. S. grade level (years of education), considering the number of tokens per sentence and the number of characters per token. The advantage of this formula over those considering syllables is that tokens are more easily retrieved (OCR suffices to gain the entire input to this formula).

$$GradeLevel = 0.5(Tokens/Sentences) + 4.71(Characters/Tokens) - 21.43$$

Coleman-Liau Index. The Coleman-Liau Index also approximates the U. S. grade level (years of education) by considering the mean number of characters per 100 tokens and the mean number of tokens per 100 sentences.

$$GradeLevel = 0.0588(Characters/Tokens \times 100) - 0.296(Sentences/Tokens \times 100) - 15.8$$

2.2 Language-Specific Adaptations of Readability Metrics

Šlerka and Smolík [11] tentatively applied several readability metrics (among them Flesch Reading Ease, Flesch-Kincaid Grade Level, and Automated Readability Index) to selected Czech texts with assumed readability differences (textbooks and reference books for different Czech grade levels and a selection of prose by Karel Čapek, spanning childrens’ books, press columns, short stories, and novels. Šlerka and Smolík demonstrated that the selected metrics were yielding sensible information even without any adaptation to Czech: their ranking of the texts corresponded to the researchers’ assumptions, although, as expected, the scores were clearly on different scales. For instance, the Flesch Reading Ease considers even simple Czech texts extremely difficult. Even mainstream press prose often sinks under zero (the English scale spanning 0–100).

So far, the formula most adapted to other languages has been the Flesch Reading Ease [12]: Italian, French (cf. also [13–16]), Spanish, German (cf. [17]), Russian ([18, 19, 22]), Danish, Bangla and Hindi [23], and Japanese.

3 Data

To adapt the originally English readability metrics, we used two types of parallel English-Czech corpora (see Table 2):

1. InterCorp;
2. CzEng 2.0.

The former is a high-quality, but smaller, linguistic resource entirely consisting of manually translated and manually sentence-aligned digitized texts originally published in print; the latter is a huge text bulk acquired by web-crawling, with an unspecific portion of texts translated automatically, and a completely automatic alignment.

We were also interested in the correlation between the Czech formula and measured reading comprehension. For this experiment, we used the LiFR data set of Czech paraphrased administrative texts.

InterCorp. InterCorp is an entirely manually translated parallel corpus [26, 27], manually sentence-aligned, with Czech as the pivot language (foreign languages are never directly aligned with each other, but over Czech). The Czech texts occur as original texts as well as translations. Among foreign texts, originals or translations from Czech were preferred during the acquisition, but translations from other languages are present as well. The corpus primarily comprises fiction, but also non-fiction and legal texts from the multilingual official production of the EU bodies. The Czech-English pair contains 348 texts totalling to 2,364,684 sentences or 33,190,659 tokens in the English counterpart. To augment the data, we split the texts into 100-sentence chunks, totalling to 19,722 samples. Before the sampling, we filtered out 1:n and n:1 aligned sentences, keeping only the 1:1 aligned sentences.

CzEng 2.0. CzEng 2.0 [29] is a large Czech-English corpus of texts harvested on the web, primarily used for shared translation tasks. It contains several sections of news texts: a Czech monolingual corpus with a machine-translated English counterpart and an English monolingual corpus with a machine-translated Czech counterpart. Besides, there is a corpus of web-crawled parallel texts, for which there is no guarantee that they are human-translated, but most of them are probably at least post-edited by a human. The translation direction is never indicated. All CzEng 2.0 corpora are automatically sentence-aligned.

For our experiments we used random samples of CzEng 2.0 documents, sometimes in combination with the InterCorp data (see Table 2).

Table 2. List of used datasets

Name	Size in texts	Description
InterCorp	19,722	Manual translation, both directions
csnewsCS small	1,997	Czech origin, automatically translated
csnewsEN small	1,997	English origin, automatically translated
csnewsBOTH small	3,996	csnewsCS small + csnewsEN small
ALL small	23,718	csnewsBOTH small + InterCorp
csnewsCS big	20,905	Czech origin, automatically translated
csnewsEN big	20,905	English origin, automatically translated
csnewsBOTH big	41,810	csnewsCS big + csnewsEN big
ALL big	61,532	csnewsBOTH big + InterCorp

LiFR. LiFR is a corpus of paraphrased administrative and legal texts with reading comprehension measured on readers across age groups and education levels [32]. LiFR comprises 300–500 token documents on six topics: a contract, house rules, two court decisions and two ombudsman’s reports. Each topic is represented by three different text versions: an original (“legalese”) and two paraphrases. The paraphrases were written by two domain experts instructed to make the original texts maximally comprehensible but preserve all information.

To compare the writing styles of the experts, a reading-comprehension test was designed and administered for each topic (the original and the two paraphrases); i.e. each triple of texts. Each test consisted of multiple-choice as well as open questions. Each text was read by 30–60 readers, with no reader seeing different versions of the same topic. Their success was recorded as the proportion of correct choices. The resulting score for each text was computed as the mean success of all readers in all questions. Therefore, the comprehension scale spans 0–1 (the y-axis in the plot in Figs. 1 and 2).

3.1 Pre-processing

The data of both corpora (InterCorp [26] and CzEng 2.0 [29]) came already split to sentences. InterCorp was also tokenized, while CzEng 2.0 was not. We tokenized it with UDPipe [28].

Besides token and sentence counts, the readability formulas require syllable and character counts. Hence, before fitting the functions, we also had to extract syllable and character counts for each token in the texts in a separate step.

Character Counts. The Coleman-Liau and ARI consider the token length in characters, originally conceived as typewriter strokes. Their numbers were retrieved by the `len` function in Python, with no respect to the mapping of characters to phonemes. Hence, e.g., the Czech phoneme “ch” counted as two characters.

Syllable Counts for Czech. The phonotactic rules as well as phoneme distributions are language specific. The syllable-counting scripts for Czech were based on a syllable-counting script by David Lukeš from the Institute of the Czech National Corpus, which considers the pitch (a vowel, diphthong, or a syllabic consonant), rather than syllable boundaries. Compared to using the PyHyphen library [25], the rule-based script was giving better results in manual sample checks.

Syllable Counts for English. The English script also focuses on the syllable pitch represented by a vowel or a diphthong approximated by rules for the written language, especially with respect to vowel sequences (so, e.g., the word *employee* and its derivations is perceived as having three syllables, while *eyeing* as having two syllables.)

4 Method

4.1 Determining Language-Specific Function Parameters

Each of the selected readability metrics is a function. Given an English-Czech parallel corpus, we assume that the translations (in either direction) preserve roughly the same readability as the originals.

To test this assumption, we computed the Flesch Reading Ease with the original English formula on the English as well as Czech documents and measured Pearson’s product moment correlation between the corresponding language counterparts. The Czech scores strongly correlated with the English scores. As expected, the correlation was highest on the manual translations in InterCorp (0.9, p-value $< 2.2e^{-16}$, 95% conf. interval 0.897–0.902). On the unspecified mix of manually and machine-translated texts in CzEng, the correlation was 0.84 (p-value $< 2.2e^{-16}$, 95% conf. int. 0.823–0.849) between Czech originals and English translations and 0.79 (p-value $< 2.2e^{-16}$, 95% conf. int. 0.769–0.802) between English originals and Czech translations. That proves our assumption that the readability of translated texts is comparable to their originals, and therefore we can fit the function parameters for the parallel Czech texts to obtain the Czech scores as similar to the corresponding English scores as possible. The error permitting, this will make the adapted Czech readability scores interpretable on the same scale as the original English scores.

To determine the Czech-specific parameters to replace the original English-specific parameters in the English FRE function, we have used the non-linear `optimize.curve` fit algorithm from the SciPy library [24].

5 Results

We evaluated the fits by RMSE (Root Means Square Error). Table 3 shows the values of RMSE (Root Means Square Error) of the individual metrics trained on the individual datasets, as they were evaluated on 15% of each dataset. The first table row indicates the scale on which the function values can lie. The best results were obtained by fitting the metrics functions on InterCorp.

The grade levels would typically span 6–18 years of human age, corresponding to years spent in the education system, but the scale is not rigid (we observe values between -5 (*sic!*) and 20). At the first glance, the most realistic grade-level range is presented by the Flesch-Kincaid Grade Level, whose minimum values lie, for our Czech as well as English texts, around the kindergarten age, and the maximum at nineteen years of age (corresponding to college studies). The Automated Readability Index (ARI) reaches even below the infant age, and so does, even more, the Coleman-Liau index. The Coleman-Liau index appears to be less sensitive, using a shorter range than ARI and Flesch-Kincaid. All RMSEs are quite small, given the range of the scales (see also Table 3): below one year in all metrics on the Grade Level scale and 4.6 on the 0–100 scale.

Table 3. Root means square errors for datasets

Dataset	Flesch Reading Ease	Flesch-Kincaid Grade Level	Coleman-Liau Index	ARI
Scale	0–100	1–15	1–15	1–15
InterCorp	4.639	0.755	0.697	0.734
csnewsCS small	8.705	1.527	1.449	1.626
csnewsEN small	9.115	1.527	1.436	1.801
csnewsBOTH small	8.775	1.449	1.249	1.821
ALL small	5.673	0.982	0.840	0.976
csnewsCS big	8.825	1.727	1.370	1.820
csnewsEN big	9.988	1.829	1.369	2.032
csnewsBOTH big	9.791	1.753	1.437	1.874
ALL big	8.489	1.652	1.208	1.798

These are the resulting adaptation of the four classic readability metrics to Czech:

Flesch Reading Ease

$$= 206.935 - 1.672 \times \text{Tokens/Sentences} - 62.18 \times \text{Syllables/Tokens}$$

Flesch-Kincaid Grade Level

$$= 0.52 \times \text{Tokens/Sentences} + 9.133 \times \text{Syllables/Tokens} - 16.393$$

Coleman-Liau Index

$$= 0.047 \times \text{Characters/Tokens} \times 100 - 0.286 \times \text{Sentences/Tokens} \times 100 - 12.9$$

Automated Readability Index

$$= 3.666 \times \text{Tokens/Sentences} + 0.631 \times \text{Characters/Tokens} - 19.491.$$

To examine the association between the readability formulas and reading comprehension, we computed the scores (FRE, Flesch-Kincaid, Coleman-Liau, and ARI) for each text from the LiFR corpus. Figures 1 and 2 illustrate the results. Figure 1 shows the Flesch Reading Ease scores on the x-axis and the reading comprehension scores on the y-axis. The plot is divided into three facets representing the three different text versions. Figure 2 renders the scores of the other three formulas, which are supposed to span approximately the same scale (the U.S. grade levels).

We measured the correlation (Pearson product moment) of the reading comprehension with the individual readability scores for the entire text collection. The effects were heavily statistically insignificant (most p-values above 0.3), and the estimated effects were anyway extremely weak (mostly below 0.2). Therefore we can report no correlation of readability scores and reading comprehension on this data.

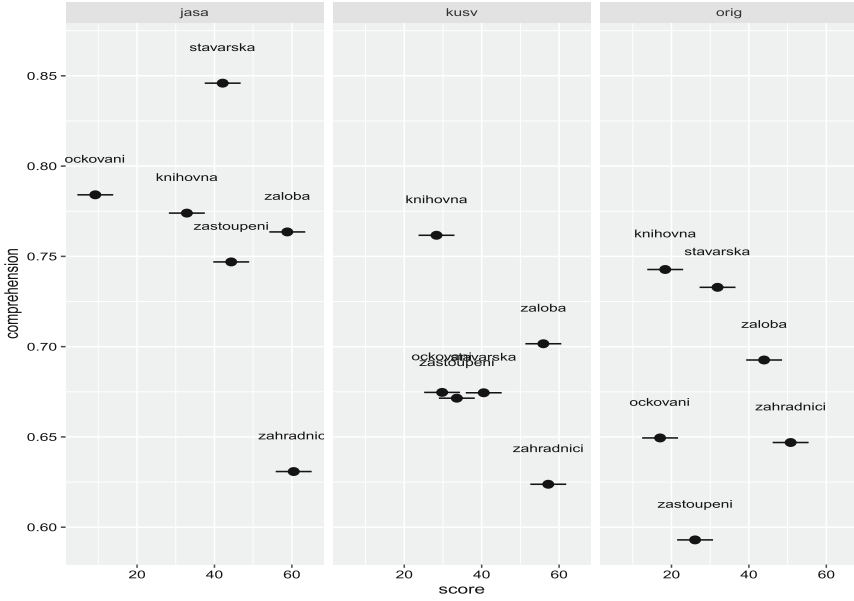


Fig. 1. Average reading comprehension by Flesch Reading Ease in different document versions by different authors.

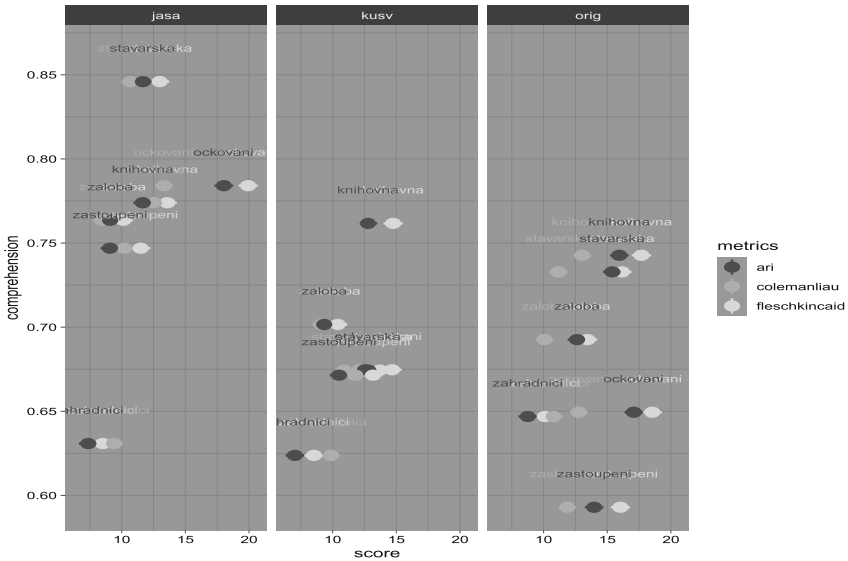


Fig. 2. Average reading comprehension by Flesch-Kincaid Grade Level, Coleman-Liau Index, and Automated Readability Index in different document versions by different authors. A thin white path connects the different scores for each text.

6 Discussion

The results were always better when trained on InterCorp than on different samples of CzEng 2.0. Surprisingly, more data (InterCorp combined with CzEng 2.0) were increasing the RMSE. We speculate that it is because the CzEng 2.0 data is on the one hand very noisy, but on the other hand it covers only one genre – news, which is not diverse enough to cover the entire scale. Besides, even high-quality machine-translated texts can differ from human-translated texts in ways that are not obvious to human readers but can affect readability scores.

In general, some noise is inevitable even when working with human-translated texts, as in the case of InterCorp. InterCorp primarily contains fiction, scholarly texts, and popular non-fiction. In all these genres, the translator primarily aims at the equivalence of content, cultural connotations, and possibly equivalence of the emotional response of the reader. Especially in artistic texts, structural equivalence is neither a necessary nor a sufficient condition for the translation to be perceived as optimal.

We were surprised by the Coleman-Liau Index and ARI reaching below zero also in their English version. However, these texts were indeed unnaturally simple. Most of them were dialog passages from dramas by V. Havel (Audience, Largo Desolato, Garden Party), which are known for their laconicism.

Knowing that the Flesch Reading Ease had many international adaptations, we experimented with the Russian Flesch Reading Ease formula by Osborneva [19]. Osborneva based her calculations on the difference in the number of syllables in Russian and English words, drawing on Slovar russkogo yazyka pod redaktsyey Ozhegova (39,174 words) [20] and Muller English-Russian dictionary (41,977 words) [21]. In addition, she analyzed six million words of parallel Russian-English literary texts. We used the Czech-Russian language pair in InterCorp, fitting the Russian formula to Czech counterparts of Russian texts.

Osborneva’s original formula had the following parameters:

$$FRE(Ru) = 206.835 - 1.3(Tokens/Sentences) - 60.1(Syllables/Tokens).$$

The adapted formula for Czech had the following parameters:

$$FRE(CsRu) = 206.835 - 1.388(Tokens/Sentences) - 65.09(Syllables/Tokens).$$

And the adapted formula for Czech from English had the following parameters:

$$FRE(CsEn) = 206.935 - 1.672(Tokens/Sentences) - 62.18(Syllables/Tokens)$$

The constant was always fixed, so the fitting algorithm was only working with the coefficients.

The RMSE of the Czech formula adapted from Russian outperformed the one fitted on English (4.639 vs. 3.748) [34]. However, when applied to the CzEng 2.0.

data, the RMSE was slightly higher than the one of the English-fitted formula. This suggests that the formula adapted from Russian be overfitted to InterCorp.

To examine the difference between the two FRE adaptations, we measured their correlation (Pearson's product moment) on the CzEng 2.0 Czech originals, CzEng 2.0 Czech translations, and the Czech InterCorp text samples, respectively, obtaining extremely high positive and highly significant correlations: 0.996, 0.994, 0.994 with 95% confidence intervals within 0.005.

We also performed the pairwise t-test. The means of the differences between scores given by the FRE adaptation from English and those given by the FRE adaptation from Russian were 1.89 (95% conf. int. 1.834–1.95), 0.88 (95% conf. int. 0.8–0.97), and 2.63 (95% conf. int. 2.62–2.65), respectively. All the differences were highly significant, which is not surprising, considering the high number of observations in each case. However, given that the RMSE of both adaptations are higher than the mean differences in the values they return, we conclude that this difference can be neglected.

Concerning the undetected correlation between the readability formulas and the reading comprehension in the LiFR corpus, the most likely reason is that the texts and their paraphrases were controlled for identical content. In legal texts this means a significant vocabulary overlap due to terminology and multi-word names of institutions, full personal names, etc. This constrains the variability of token length, which is a crucial distinction criterion for all discussed readability metrics.

On the other hand, we could clearly observe that one author (see Fig. 1 and Fig. 2, “jasa”), clearly wrote more comprehensible texts than the others. However, these texts were not significantly simpler in terms of readability scores.

Also, most texts were lying between 30 and 50 points on the FRE scale, or 10 and 15 on the Grade Level scales, which is quite a narrow concentration. Although the RMSEs were quite low with their positions below 5 and 1, respectively, it can still have been too much with such a homogeneous data, and possible interesting differences may have been blurred by the RMSEs.

Last but not least, not even the differences in comprehension were particularly big between two of the three authors. The distributions of the comprehension values and the readability scores suggest that “jasa” must have had a writing strategy independent of length of sentences and words.

This observation is in accordance with DuBay, p. 116: “‘Don’t write to the formula’, because it is too easy to neglect the other aspects of good writing. Readers need the active voice, action verbs, clear organization and navigation cues, illustrations and captions [...]. More than anything else, they need texts that create and sustain interest.” [1].

On the other hand, we could at least see that the metrics were largely consistent with each other (the Flesch Reading Ease scale is reverted with respect to the others: the higher the score, the *easier* the text whereas the other approximately translate to “this many years at school this text takes to comprehend”).

7 Conclusion

We have adapted the following four classic readability formulas to Czech: Flesch Reading Ease, Flesch-Kincaid Grade Level, Coleman-Liau Index, and Automatic Readability Index, based on three available English-Czech parallel data sets, using a generic curve-fitting algorithm. The adaptations reached good RMSEs below one grade level on the interpretation scales (cf. Table 3). Despite historical records on a strong correlation between FRE and reading comprehension, we were not able to detect it on the Czech data with reading comprehension that we had at our disposal.

We will offer these and several more Czech-adapted metrics for incorporation into existing publicly available readability evaluation platforms where Czech is present, such as CTAP and EVALD [5, 33]. In the future, we intend to provide Czech adaptations of other classic readability formulas (e.g. SMOG [35]), especially those considering vocabulary (e.g. the Dale-Chall formula [30]) as a substantial readability feature [36, 37], using the language profiles for Czech as a foreign language.

References

1. DuBay, W.: Smart Language. Readers, Readability, and the Grading of Text. Impact Information, Costa Mesa, California (2007)
2. Deutsch, T., et al.: Linguistic features for readability assessment. In: Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications, pp. 1–17. Association for Computational Linguistics, Seattle (2020)
3. Batinić, D., et al.: Creating an extensible, levelled study corpus of Russian. In: Proceedings of the 13th Conference on Natural Language Processing (KONVENS), Bochum, Germany September, pp. 38–43. Ruhr-Universität, Bochum (2017)
4. Azpiazu, I.M., Pera, M.S.: Multiattentive recurrent neural network architecture for multilingual readability assessment. *Trans. Assoc. Comput. Ling.* **7**, 421–436 (2019)
5. Rysová, K., et al.: Introducing EVALD - software applications for automatic evaluation of discourse in Czech. In: RANLP Proceedings, Bulgaria, pp. 634–641 (2017)
6. McNamara, D.S., et al.: Automated Evaluation of Text and Discourse with Coh-Matrix. Cambridge University Press, Cambridge (2014)
7. Flesch, R.: The Art of Readable Writing, 2nd edn. Harper, New York (1974)
8. Kincaid, J.P., et al.: Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Defense Technical Information Center (1975)
9. Coleman, M., Liau, T.L.: A computer readability formula designed for machine scoring. *J. Appl. Psychol.* **60**, 283–284 (1975)
10. Senter, R.J., Smith, E.A.: Automated Readability Index. AMRL-TR. Aerospace Medical Research Laboratories (US), Wright-Patterson Air Force Base (1967)
11. Šlerka, J., Smolík, F.: Automatická měřítka čitelnosti pro česky psané texty. *Studie z aplikované lingvistiky* **1**(1), 33–44 (2010)
12. Garais, E.-G.: Web applications readability. *Romanian Econ. Bus. Rev.* **5**, 117–121 (2011)

13. Kandel, L., Moles, A.: Application de l'indice de flesch à la langue française. *Cahiers Etudes de Radio-Télévision* **19**, 253–274 (1958)
14. De Landsheere, G.: Pour une application des tests de lisibilité de Flesch à la langue française. *Le Travail Humain*, pp. 141–154 (1963)
15. Henry, G.: Comment mesurer la lisibilité. Labor, Brussels (1975)
16. François, T., Fairon, C.: An AI readability formula for French as a foreign language, p. 477 (2012)
17. Amstad, T.: Wie verständlich sind unsere Zeitungen? *Studenten-Schreib-Service* (1978)
18. Guryanov, I., et al.: Text complexity: periods of study in Russian linguistics. *Revista Publicando* **4**, 616–25 (2017)
19. Osborneva, I.V.: Mathematical model for evaluation of didactic texts. *Proc. Moscow State Pedagogical Univ.* **4**, 141–147 (2006)
20. Ozhegov, S.I.: *Slovar russkogo yazyka. Russkij jazik*, Moskva (2010)
21. Muller, V.K.: *Anglo-russkij slovar. Russkij Jazyk*, Moscow (1985)
22. Solnyshkina, M., Ivanov, V., Solovyev, V.: Readability formula for Russian texts: a modified version. In: Batyrshin, I., Martínez-Villaseñor, M.L., Ponce Espinosa, H.E. (eds.) *MICAI 2018, Part II. LNCS (LNAI)*, vol. 11289, pp. 132–145. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04497-8_11
23. Sinha, M., et al.: New readability measures for Bangla and Hindi texts. In: *Proceedings of COLING 2012: Posters*, pp. 1141–1150. The COLING 2012 Organizing Committee, Mumbai (2012)
24. SciPy 1.0 Contributors, Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., et al.: SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Meth.* **17**, pp. 261–272 (2020)
25. Leo, Dr., Behmo, R.: Pyhyphen. The hyphenation library of LibreOffice and Firefox wrapped for Python. Version 4.0.0, 15 February 2021. <https://pypi.org/project/PyHyphen/>. Accessed 2 Apr 2021
26. Rosen, A.: InterCorp - a look behind the façade of a parallel corpus. In: Gruszczyńska, E., Leńko-Szymańska, A. (eds.) *Polskojęzyczne korpusy równoległe. Polish-language Parallel Corpora*, pp. 21–40. Instytut Lingwistyki Stosowanej, Warszawa (2016)
27. Čermák, F., Rosen, A.: The case of InterCorp, a multilingual parallel corpus. *Int. J. Corpus Ling.* **13**, 411–427 (2012)
28. Straková, J., et al.: Open-source tools for morphology, lemmatization, POS tagging and named entity recognition. In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 13–18. ACL, Stroudsburg (2014)
29. Kocmi, T., et al.: Announcing CzEng 2.0 Parallel Corpus with over 2 Gigawords (2020)
30. DuBay, W.H.: *The Classic Readability Studies* (2007)
31. Wickham, H.: *ggplot2: Elegant Graphics for Data Analysis*. Verlag, New York (2016). <https://doi.org/10.1007/978-0-387-98141-3>
32. Chromý, J., Cinková, S., Šamánková, J.: Srozumitelnost českého odborného a úředního textu - proč se jí zabývat a jak ji měřit. *Stud. Appl. Ling.* **12**(1) (2021)
33. Chen, X., Meurers, D.: CTAP: a web-based tool supporting automatic complexity analysis. In: *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, pp. 113–119. The COLING 2016 Organizing Committee, Osaka (2016)
34. Bendová, K.: Using a parallel Corpus to adapt the Flesch Reading Ease Formula to Czech. To appear in *Jazykovedný časopis*

35. McLaughlin, G.: SMOG grading - a new readability formula. *J. Read.* **12**, 639–646 (1969)
36. Falkenjack, J., Mühlenbock, K., Jönsson, A.: Features indicating readability in Swedish text. Presented at the (2013)
37. Gonzalez-Dios, I., Aranzabe, M.J., Díaz de Ilarraza, A., Salaberri, H.: Simple or Complex? Assessing the readability of Basque Texts. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 334–344. Dublin City University and Association for Computational Linguistics, Dublin (2014)