

EDYTA CHARZYŃSKA^{1,A} & ŁUKASZ DĘBOWSKI^{2,B}

¹University of Silesia, Katowice, Poland

²Polish Academy of Science, Warsaw, Poland

^Aedyta.charzynska@us.edu.pl ; ^Bldebowski@ipipan.waw.pl

EMPIRICAL VERIFICATION OF THE POLISH FORMULA OF TEXT DIFFICULTY

Abstract

The aim of the study was to verify the accuracy of the formula for assessing the difficulty of texts written in Polish. The study involved 1,309 persons aged between 15 and 84. 15 texts were used, each approximately 300 words long, representing different subjects and varied difficulty level. Text comprehension was checked with multiple choice tests, the cloze procedure and open-ended questions. Significant correlations between the difficulty of a text and its comprehension were found ($r_{mc}(15) = -0.529, p = 0.043$; $r_{open} = -0.519, p = 0.047$; $r_{cloze} = -0.656, p = 0.008$). The results confirmed relative accuracy and usefulness of Pisarek's readability formula. The discussion includes proposed ways of improving the current form of the formula.

Keywords: psycholinguistics; readability; text difficulty; readability formula; text comprehension; cloze procedure

1. Introduction

It is a truism that the text's author wants to be properly understood by the recipients. Thus, the ability to write clearly and understandably is one of the basic writing skills. So as to increase the chances that the recipients will understand the text in conformity with the author's intention, the author needs to pay attention to text readability.

In literature of the subject various definitions and ways of understanding of readability can be found (Klare, 1963; Samson, 1993). First, some definitions point out the legibility of the printed material due to its layout or typographic features. Thus, readability understood this way is affected by the layout of the text, font, the presence of graphics etc. Second, some other definitions connect readability with the easiness of reading thanks to interesting content or the aesthetic value of the writing style. Third, a definition of readability may be simply focused on easiness of

text comprehension due to the style of writing (Klare, 1963). Readability described this way is connected with the comprehensibility of the text: the higher readability of the text, the more comprehensible it is (Samson, 1993, p. 58). This third way of understanding readability has been adopted in this text.

One objective of studies on readability is to develop tools to measure it (Klare, 1963). Apart from classic methods (such as dichotomous questions, open-ended questions, single/multiple choice tests etc.), the cloze procedure developed by Taylor (1953) is very popular. This method is based on the assumption drawn from Gestalt psychology of the human tendency to perceive the entirety even if only a part is visible. So a human strives for a mental closure of the figure. According to Taylor (1956), the cognitive process involved in doing a cloze test is similar: people try to complete the text which has a 'defect' using predictions based on the available contextual hints. The 'defect' means that every *n*th (usually every 5th) word is removed from the text and the reader is asked to complete each gap with one word that he or she thinks was used by the author. In order to determine text readability with the cloze procedure, readability index is used, being the sum or percentage of correctly filled gaps. The validity of this method, estimated on the basis of correlation with other readability measures, has been confirmed (Rankin, 1959; Bormuth, 1966). Its great advantage is that it not only takes into account the factors affecting text readability that have been already known but also the ones that have not been discovered or measured yet; furthermore, it does not disregard non-linguistic factors such as the reader's knowledge, linguistic skills or even motivation and interest in the text (Pisarek, 2007, p. 249).

Despite many advantages of the cloze method, it also has some limitations. It is time consuming and laborious, as it involves finding the participants, their engagement in the study, and then calculating the results. One of the ways of estimating text readability which do not have the limitations connected with the cloze test are analytical methods readability formulas (Flesch, 1948; Gunning, 1952; Chall & Dale, 1995). Most of the formulas take into account two main factors: lexical or semantic features (such as the length of words, their similarity and frequency of occurrence (popularity)) and sentence or syntactic features (average sentence length; Chall & Dale, 1995). Despite justified criticism of readability formulas (Bailin & Grafstein, 2001; Klare, 1974), thanks to their practical values they are still appreciated as a method of readability measurement.

The vast majority of readability formulas were developed for the English language, which is a considerable limitation. So it is not obvious whether they can be effectively applied to other languages, and if so, whether they should be used in the form identical to that for the English language or should be changed substantially.

Pisarek (1969) attempted to develop a readability formula for the Polish language. In the creation of his formula, he adopted one factor — average sentence length — directly from English. In the case of the other factor, he chose another way, taking into consideration the specificity of the Polish language. On the basis of linguistic analyses, he decided that the best indicator of lexical features was the percentage of difficult words, i.e. words having 4 or more syllables in the dictionary form. The text difficulty formula proposed by Pisarek (1969) is:

$$T = \frac{\sqrt{T_s^2 + T_w^2}}{2} \quad (1)$$

where T — text difficulty, T_s — syntax difficulty index, meaning the average sentence length (in words), T_w — vocabulary difficulty index, meaning the percentage of “difficult” (four-syllable and longer) words. In addition, Pisarek (2007, p. 258) presented ranges describing text difficulty: 4 – 7 — very easy texts, 7.1 – 10 — easy texts, 10.1 – 13 — average texts, 13.1 – 16 — difficult texts, 16.1 – 20 — very difficult texts.

In spite of its potential practical value, Pisarek’s formula has one important flaw: it has never been empirically verified. It means that so far nobody has attempted to answer the question whether it can really accurately determine the difficulty of a text, thus being a predictor of comprehensibility for the potential recipient group.

Considering the above mentioned facts, the aim of this study was to check empirically the predictive validity of Pisarek’s formula of text difficulty.

2. Method

2.1. Participants

The study involved 1,321 individuals whose native language was Polish. 12 persons returned tests without completing them, so finally data from 1,309 individuals were used (including 844 women and 464 men; 1 person did not provide the information on sex). The sociodemographic characteristics of the studied sample is presented in Table 1. The participants’ mean age was 34.87 years, $SD = 16.33$, min. = 15, max. = 83. Approximately 2/3 of the sample were people with secondary or higher education. Most of them (75.7%) lived in towns with up to 500 thousand residents. More than half of the participants were professionally active, whereas nearly 1/3 were school or university students.

2.2. Measures

15 non-literary tests in Polish, each about 300 words long, were used in the study. The texts were selected so as to ensure a variety of types, topics and the level of difficulty initially determined with the Gunning fog index (Gunning, 1952). The texts included: 2 texts from school handbooks, 2 academic texts, 2 acts, 2 official letters, 2 manuals, 2 legal brochures and 3 newspaper/magazine articles (see Table 2).

All participants received a study set including a demographics section and 3 texts. For each text, 3 kinds of reading comprehension tests were prepared: multiple-choice, cloze and open-ended questions. The multiple choice test was made up of 4 single choice questions, each having 4 alternative answers. The cloze test, consisting of 50 gaps, was prepared by removing every 5th word from the text, beginning with the second sentence. The open-ended questions test included 5 questions.

The sets were created randomly: one text was drawn from each group of tests and they were joined together. If the same text was drawn, the drawing was repeated. In order to control the impact the order of text types might have had on the dependent variable, the order of tests in a set was regularly changed.

Table 1: Sociodemographic characteristics of the sample.

	N = 1309	%
Sex		
Women	844	64.5
Men	464	35.4
N/A	1	0.1
Age (in years)		
15-18	264	20.2
19-29	351	26.8
30-40	225	17.2
41-51	224	17.1
52-62	143	10.9
63-73	68	5.2
74-84	21	1.6
N/A	13	1.0
Education		
Elementary	85	6.5
Lower secondary	235	17.9
Vocational	90	6.9
Secondary	413	31.6
Higher	486	37.1
Place of residence		
Village	288	22.0
Town up to 100 thousand residents	470	35.9
Town between 100 and 500 thousand residents	437	33.4
Town over 500 thousand residents	84	6.4
N/A	30	2.3
Professional status		
Pupil	280	21.4
Student	89	6.8
Employed	714	54.5
Unemployed	72	5.5
Pensioner	115	8.8
N/A	39	3.0

Table 2: Characteristics of texts: linguistic features and comprehension level.

No. of text	Kind of text	Average sentence length	% of difficult words	Text difficulty	n MC	% of correct MC answers	n cloze	% of correct cloze answers	n open	% of correct open-ended answers
1.	Text from a secondary school biology handbook	13.75	10.18	8.55	105	80.87	77	27.17	73	42.50
2.	Simplified paraphrase of an official letter	12.62	7.32	7.29	91	76.37	102	58.30	73	84.72
3.	Fragment of a secondary school history handbook	15.67	6.03	8.39	85	88.92	89	46.45	71	71.59
4.	Drug administration instruction	11.64	8.59	7.23	79	84.49	69	61.67	92	43.96
5.	Liberal arts academic text	20.21	12.72	11.94	79	70.57	117	30.31	99	27.14
6.	Scientific academic text	22.62	11.56	12.70	75	69.56	97	29.00	93	30.67
7.	Daily newspaper article	16.28	6.14	8.70	110	74.24	75	47.33	103	68.63
8.	Official decision	19.69	11.75	11.46	89	51.78	81	36.30	76	53.87
9.	Device manual	9.50	12.98	8.04	86	66.86	81	43.38	123	67.50
10.	Official decision	13.59	10.03	8.45	73	72.95	101	49.90	77	37.63
11.	Text from a legal brochure	16.72	7.64	9.19	100	84.00	68	57.49	87	83.72
12.	Text from a legal brochure	15.26	16.90	11.38	89	51.40	104	47.03	88	43.29
13.	Economic magazine article	17.44	7.53	9.50	70	57.80	87	30.74	88	38.82
14.	Accounting Act	28.64	9.52	15.09	76	62.83	94	28.30	85	39.76
15.	Polish Language Act	15.52	9.82	9.18	102	60.78	67	56.39	81	42.78

n — number of participants who completed a particular kind of test

The answers to multiple choice questions and the words inserted in the cloze test were checked by the persons who collected data against the answer key. The answers to open-ended questions were checked by a team of 2 linguists, other than the team creating the tests.

2.3. Procedure

The paper-and-pencil questionnaires were distributed with the snowball method: they were handed to 50 persons coming from different environments and having different education levels, who were asked to reach possibly varied groups of people.

Before completing the tests, the supervisor informed the respondents of the objective of the study, as well as of its anonymous and voluntary nature. The time of test completion was on average 30–40 minutes.

3. Results

Table 2 presents the linguistic features of the texts calculated with the Jasnopis application (Broda, Nitoń, Gruszczyński, & Ogrodniczuk, 2014) and the results of comprehension tests. The scale of difficulty of the texts measured with Pisarek's formula ranged from $T = 7.23$ for text 4 (Drug administration instruction) to $T = 15.09$ for text 14 (Accounting Act).

The correlations between the variables were calculated with the use of Pearson's correlation coefficient. No significant relationship was found between the two components of the formula: average sentence length and percentage of difficult words ($r(15) = 0.043$; $p = 0.880$). The analysis of correlation between linguistic variables and text comprehension gave the following results: the average sentence length correlated positively with comprehension assessed with the cloze test ($r(15) = -0.607$; $p = 0.016$). The negative relationship was also noted for the average sentence length and the result of multiple choice tests and open-ended questions, although the correlation coefficients for these variables did not prove to be statistically significant ($r_{mc}(15) = -0.322$; $p = 0.243$; $r_{open}(15) = -0.401$; $p = 0.139$). The percentage of difficult words correlated negatively with comprehension in the multiple choice test ($r(15) = -0.607$; $p = 0.016$) and the open-ended questions test at the statistical tendency level ($r(15) = -0.477$; $p = 0.072$). In the case of the cloze test, no significant relationships were noted $r(15) = -0.276$; $p = 0.320$).

The most important analyses concerned the correlations between the difficulty of texts calculated on the basis of Pisarek's formula and their comprehension. The relationships proved to be significant for all the types of test. Correlation coefficients were, respectively: -0.519 for the open-ended questions test ($p = 0.047$), -0.529 for the multiple choice test ($p = 0.043$) and -0.656 for the cloze test ($p = 0.008$).

Discussion of results

To sum up the results of the study, Pisarek's formula in the present form turned out to be a relatively good predictor of difficulty of Polish non-literary texts. It can be used e.g. in editors' work to estimate the difficulty of a text, and thus to preliminarily evaluate the match between the style in which the text was written and its potential readers.

The obtained results support the opinion of authors of readability formulas for the English language that text readability is determined by its two main components: lexical/semantic features and syntax/syntactic features (Chall & Dale, 1995). Hence, these factors may be good predictors of text readability regardless of the language. Differences between languages may only occur at the level of details, concerning the selection of indicators for these features. However, the hypothetical universal character of readability components must be further investigated.

Despite confirming relationships between the difficulty of texts measured with Pisarek's formula and their comprehension, it should be noted that these relationships, although significant, were not strong. In the case of two components of the difficulty formula, i.e. the average sentence length and the percentage of difficult words, significant correlations were only found for some comprehension tests. The obtained results may have been influenced by the small number of texts used in the study. Because the size of the sample affects the significance of correlation coefficient (Rubin, 2013, p. 214), considering a greater number of texts might influence the number of significant correlations between the variables.

Another important issue is whether and how it would be possible to improve the accuracy of Pisarek's formula. At least three potential ways of its improvement can be proposed:

1. modification by adding coefficients before the components,
2. changing the definition of "a difficult word" — only considering those 4-syllable or longer words which are not frequently used (e.g. on the basis of a frequency dictionary),
3. adding new variables to the equation, which may significantly affect the readability level of the text, such as abstract nouns or official terms. The above-mentioned proposals need to be empirically verified.

Moreover, it would be recommended to check the predictive validity of the modified formula not only with the use of classic comprehension tests or cloze procedure but also with other complementary methods: the assessment of reading fluency and accuracy or objective assessment of text difficulty by competent judges or subjective evaluations by ordinary users (Klare, 1981; Fuchs, Fuchs, Hosp, & Jenkins, 2001).

References

- Bailin, A. & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language and Communication*, 21(3), 285–301. [http://doi.org/10.1016/S0271-5309\(01\)00005-2](http://doi.org/10.1016/S0271-5309(01)00005-2)
- Bormuth, J. (1966). Readability: A new approach. *Reading Research Quarterly*, 1(3), 79–132. <http://doi.org/10.2307/747021>
- Broda, B., Nitoń, B., Gruszczyński, W., & Ogrodniczuk, M. (2014). *Measuring readability of Polish texts: Baseline experiments*. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland.
- Chall, J. & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge: Brookline Books.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <http://doi.org/10.1037/h0057532>
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239–256. http://doi.org/10.1207/S1532799XSSR0503_3
- Gunning, R. (1952). *The technique of clear writing*. New York: McGraw-Hill.

- Klare, G. R. (1963). *The measurement of readability*. Ames, Iowa: Iowa State University Press.
- Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, 10(1), 62–102. <http://doi.org/10.2307/747086>
- Klare, G. R. (1981). *Practical aspects of readability*. Milton Keynes, England: Institute of Educational Technology, The Open University.
- Pisarek, W. (1969). Jak mierzyć zrozumiałość tekstu? *Zeszyty Prasoznawcze*, 4, 35–48.
- Pisarek, W. (2007). *O mediach i języku*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych Universitas.
- Rankin, E. F. (1959). The cloze procedure — Its validity and utility. In O. Causey & W. Eller (Eds.), *Eighth Yearbook of the National Reading Conference* (pp. 131–142). Milwaukee: National Reading Conference.
- Rubin, A. (2013). *Statistics for evidence-based practice and evaluation*. Austin: University of Texas.
- Samson, D. C. Jr. (1993). *Editing technical writing*. New York: Oxford University Press.
- Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, 30, 415–433.
- Taylor, W. L. (1956). Recent developments in the use of ‘cloze procedure’. *Journalism & Mass Communication Quarterly*, 33(1), 42–99. <http://doi.org/10.1177/107769905603300106>

Acknowledgment

This work was supported by a grant from the National Science Centre of the Republic of Poland, titled: Mierzenie stopnia zrozumiałości polskich tekstów użytkowych (poza-literackich)/ Measuring the degree of readability of nonliterary Polish texts (no. DEC-2011/03/B/HS2/05799).

The authors declare that they have no competing interests.

The authors' contribution was as follows: concept of the study: first author; data analyses: second author; the writing: first and second author.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (<http://creativecommons.org/licenses/by/3.0/pl/>), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

© The Authors 2015

Publisher: Institute of Slavic Studies, PAS, University of Silesia & The Slavic Foundation