Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 104 (2017) 477 – 484

Procedia Computer Science

ICTE 2016, December 2016, Riga, Latvia

Eye Movement Analyses for Obtaining Readability Formula for Latvian Texts for Primary School

Aigars Atvars[a,b,*]

[a]*OptimEyz, Ltd., Liepajas street 5-23, Priekule, Priekules nov., LV-3434, Latvia*
[b]*University of Latvia, Raina blv 19, Riga, LV-1586, Latvia*

Abstract

To determine the difficulty of text, readability formulas can be used. The research was made to find readability formula for Latvian. Readability formulas for English were used as guidelines. The novelty was the use of eye movement tracking during reading to obtain quantitative data that lead to readability formula. Eye fixation durations were gathered during readability measurements. Average values of fixation durations were calculated to characterize texts and readers. 15 texts with various difficulty levels were composed for exposing them to readers. More than 300 children of grades 1 - 4 were participating in measurements. Average values of eye fixation durations of "readers from a certain grade while reading a certain text" were used to find their correlation with values of several English readability formulas. It was found that good correlation occurs with Spache readability formula (R=0.74), which is used for evaluation of primary-grade reading materials in English. It was found that percentage of hard words is the central parameter that determines the difficulty of texts for primary school.

Peer-review under responsibility of organizing committee of the scientific committee of the international conference; ICTE 2016

Keywords: Eye movements; Readability; Readability formula

1. Introduction

Reading is a significant way how people obtain new information. Some texts are written in a manner that is easy to read, but some are hard to read. Since the end of 19th century[1] quantitative analyses of texts were performed to

* Corresponding author. Tel.: +371-25986724.
E-mail address: Aigars.Atvars@lu.lv

Peer-review under responsibility of organizing committee of the scientific committee of the international conference; ICTE 2016
doi:10.1016/j.procs.2017.01.162

find those parameters that determine the difficulty of texts. Various parameters were formulated that may influence the difficulty of texts (in 1934 E. Dale and R.W. Tyler[2] analyzed 29 parameters that can characterize the text), but then it was found that many of them correlate between themselves. Therefore most obvious parameters of texts were selected to characterize texts - average number of words used per sentence, average number of syllables per word, percentage of difficult words in a text and other. These parameters do not correlate between themselves and can be easily calculated. They were used to obtain simple formulas that characterize the difficulty of texts and their appropriateness for certain grade readers. These are readability formulas. Some readability formulas for English are given in Table 1.

The peak for the development of readability formulas was the middle of 20th century. Although there is criticism on the ability of readability formulas to characterize the difficulty of texts[3,4], these formulas are in use. For example, Flesch-Kincaid reading ease and Flesch-Kincaid grade level formula is incorporated in MS Office and appears after spell-check. Readability formulas and their applications are still in the field of interest of researchers[5,6,7].

Table 1. The list of various readability formulas for English.

Readability formula	Formula description	The range of values
The Flesch Reading Ease Readability formula[8]	206.835 – (1.015 x [average sentence length]) – (84.6 x [average number of syllables per word])	0 – difficult text; 100 – simple text;
The Flesch - Kincaid Grade Level Readability formula[9]	(0.39 x [average sentence length]) + (11.8 x [average number of syllables per word]) – 15.59	A value corresponds to a grade for whom this text is appropriate
The Gunning's Fog Index (FOG Readability formula)[10]	0.4 ([average sentence length] + [percentage of difficult words]); difficult words in this case are those who have 3 or more syllables.	A value corresponds to a grade for whom this text is appropriate
SMOG formula[11]	$3 + \sqrt{[\quad 30\ )}$ difficult words in this case are those who have 3 or more syllables.	A value corresponds to a grade for whom this text is appropriate
Coleman – Liau Index[12]	0.0588 *[average number of letters per 100 words] – 0.296 [average number of sentences per 100 words] -15.8	A value corresponds to a grade for whom this text is appropriate
Automated readability index[13]	4.71 * [average number of letters per word] + 0.5 * [average sentence length] - 21.43	A value corresponds to a grade to whom this text is appropriate
New Dale - Chall readability formula[14]	0.1579 *[percentage of rare words in sentence (who are not in the list of 3000 most popular words))] + 0.0496 * [average sentence length] + 3.6365	Their own scale which can be recalculated to grade level by the help of a certain table
Spache formula[15]	0.141 * [average sentence length] + 0.086 *[percentage of difficult words] + 0.839	A value corresponds to a grade for whom this text is appropriate

average number of syllables per word= (number of syllables in a text)/(number of words in a text)

average sentence length = (number of words in a text)/(number of sentences in a text)

Readability formulas determine for which grade a certain text is appropriate. For average person from the lower grades this text would be too hard to read and understand. The task for the educator is to match the appropriate text to a certain grade. Readers in each grade have different reading abilities - some are reading better than others. Therefore for each person his own reading grade level has to be determined. For this various reading tests are used. A person would like to read texts that correspond to his reading grade level. Therefore the task is to match texts to a reading grade level of a reader. Problems arise when texts exposed to readers are too hard for them to understand (texts are for higher grade than the actual reading grade of readers). In this case texts have to be simplified to match