

The application of readability formulas for the Tibetan teaching textbooks

Ruosong Zhou^{1, a}, Zhijuan Wang^{1, 2, b}

¹The College of Information Engineering, Minzu University of China, Beijing, 100081, China

²Minority Languages Branch, National Language Resource Monitoring & Research Center, Beijing, 100081, China

^azhouruosg@163.com, ^bwangzj.muc@gmail.com

Keywords: readability, formula, Tibetan, textbooks

Abstract. The aim of this article is to utilize different readability formulas proposed by some authoritative sources to determine the readability levels of the Tibetan teaching textbooks from 2th to 12th grade. The research about readability of textbooks can not only indicate the specific readability level of each book quantitatively but also provide advisable direction of improvement. For this purpose, we adopted four formulas including Flesch Reading Ease(FRE)、Gunning Fog Index(GFI)、Automated Readability Index(ARI)、Flesch-Kincaid Formula(FK) to conduct an experiment. As a result, the readability formulas showed good application for the Tibetan teaching textbooks on a whole.

1 Introduction

Readability refers to the degree that how easy or hard to understand a text briefly. Longman Dictionary of Language Teaching & Applied Linguistics gives the definition of readability in detail: Readability is about how easily written materials can be read and understood. Readability depends on many factors, including (a) the average length of sentence in a passage; (b) the number of new words a passage contains; (c) the grammatical complexity of the language used. Procedures used for measuring readability are known as “readability formula” [1].

Readability study plays an important part in reading theory, reading teaching and learning, reading assessment, text book compilation and evaluation, etc. The study of readability is of great practical as well as theoretical significance in selecting material of reading texts, and even the selection of teaching method on reading [2]. The appropriate reading materials will help to promote students’ studying. On the contrary, the mismatch texts will discourage students to make progress, even lose confidence in certain subject. Therefore, the research on readability of teaching materials becomes more and more significant.

Because readability study first rose in west countries in last century, it had made great achievements on the research of readability based on English texts. For example, Flesch-Kincaid Formula had been appointed as the standard readability formula for United States Department of Defense [3]. In practice, it had also been applied to the Microsoft Office Word. Leah conducted an experiment to analyze the readability of Programming Textbooks [4]. Umit also compared different readability formulas on science-technology and social science textbooks [5]. As for readability based on Chinese texts, although it was being paid attention lately, great progress has been made by Chinese linguistic scholars. Lin Zheng and Li Shaoshan created new readability formulas suiting for Chinese texts [6]. Lee constructed a novel Chinese readability classification model to measure the score of readability [7].

As one of China’s ethnic minorities, Tibetan education is an indispensable branch of national education [8]. Meanwhile, The Tibetan teaching textbooks are a useful tool for students to study Tibetan well. So the importance of measuring readability for the Tibetan teaching textbooks is not less than the Chinese or English one. The research about Tibetan textbooks is wide involving with vocabulary statistics and shallow parsing analysis [9,10,11,12]. But the readability research based

on Tibetan texts is still blank, so we intended to measure the readability index of primary Tibetan teaching textbooks by means of some famous formulas.

2 Methodology

2.1 Readability formulas

a. Flesch Reading Ease

Flesch Reading Ease (FRE) is regarded as one of the most accurate formulas among those ones [13]. It is proposed in 1948, which is one of the oldest as well as widely used formulas. The computational formula is showed in Eq.1.

$$FRE=206.835-(1.015*ASL)-(84.6*ASW) \quad (1)$$

In above equation, ASL= Average Sentence Length; ASW=Average Number of Syllables per Word.

The result of FRE is a number ranging from 0 to 100. According to Flesch's designing, the higher the number is, the easier the text is [14].

b. Gunning Fog Index

Gunning Fog Index (GFI) is named after its creator. He thought that the problem of hard to read or understand was caused by the improper complexity in the writing [15]. The computational formula is showed in Eq.2.

$$GFI=0.4*(ASL+PHW) \quad (2)$$

In above equation, ASL= Average Sentence Length; PHW=Percentage of Hard Words, the author defined that the syllables of Hard Words must be equal or more than three.

c. Automated Readability Index

The Automated Readability Index (ARI) produces a grade level needed to understand the text [16]. The ARI also relies on the average characters per word rather than the average syllables per word, which is more convenient to count the words. The computational formula is showed in Eq.3.

$$ARI=4.71*AWL+0.5*ASL-21.43 \quad (3)$$

In above equation, AWL=Average number of character per word; ASL= Average Sentence Length.

d. Flesch-Kincaid Formula

The Flesch-Kincaid (FK) Grade Level is mostly used in the field of education. The score got from this formula also indicates an U.S. grade level [17]. It is easier and more convenient for scholars, educators, teachers to measure the reading ease of books and texts. As described before, it had been appointed as the standard readability formula for United States Department of Defense and applied to the Microsoft Office Word [18]. The computational formula is showed in Eq.4.

$$FK=(0.39*ASL)+(11.8*ASW)-15.59 \quad (4)$$

In above equation, ASL= Average Sentence Length; ASW=Average Number of Syllables per Word.

From the formulas we can see that FRE and FK were calculated with two same factors. The difference between them is the scope of their outcome and the meaning of outcome.

2.2 Measuring the readability

Whether the above-mentioned formulas can measure the readability of Tibetan teaching materials precisely is connected with the tested data. To make sure of the quality of data, here we select the standard primary Tibetan teaching textbooks published by Tibetan language writing committee of five provinces as the tested data. According to Cao's paper, there are 12 books totaling 274 articles with 9224 different words in the primary Tibetan teaching textbooks [19]. Here we choose 2th to 12th volume as the measured object(the 1th volume has no articles).

In order to measure the readability, we need to count the number of words, sentences, syllables and characters, which is used to compute ASL, ASW, PHW and AWL. The procedure of measuring the readability is showed in Fig.1.

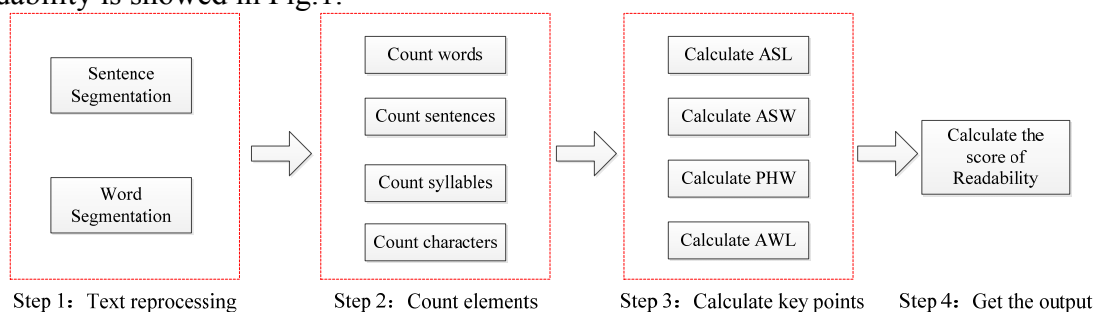


Fig.1 The procedure of measuring the readability

From Fig.1 we can see, the first of all, we need to reprocess the articles which include two parts, sentence segmentation and word segmentation. Here we finished Word Segmentation with the system of Tibetan Word Segmentation Based on Word-Position Tagging made by Congjun Long [20]. Secondly, we count those important elements that will be useful in Step 3. Then we calculate the key points with the outcome of those elements. The last thing we need to do is calculating the score of readability according to formulas. All of these Steps are consistent, only with right order can we get the accurate output.

3 Results and analysis

3.1 Statistic results

In this part, we mainly show the results of the counted elements as well as calculated key points described above from the macroscopic and microcosmic angle.

Table.2 shows the detail average counts of each article in macroscopic angle.

Table.2 The average value of each article

statistics No	Words per article	Sentences per article	Syllables per article	Characters per article
2th volume	59.3	14.90	77.90	330.40
3th volume	100.74	19.52	130.91	555.26
4th volume	134.42	24.25	175.96	735.25
5th volume	206.54	33	268.88	1126.38
6th volume	198.96	31.38	266.21	1107.67
7th volume	254	35.29	345.13	1421.96
8th volume	260.39	42.70	358.17	1481.22
9th volume	389.48	55.30	521.74	2149.78
10th volume	439.54	65.25	605.13	2526.04
11th volume	662.55	102.18	877.09	3624.64
12th volume	569.33	96.33	757.10	3166.05

From the above Table.2, we can find the regularity that the numerical value is higher and higher with the increase of grade on the whole, which reflect that students in high grade have to master more complex articles compared with those in low grade.

Table.3 shows the detail average frequency of each article in microcosmic angle. This results will be applied to those formulas directly.

Table.3 The average frequency of each article

statistics No	ASL	ASW	PHW	AWL
2th volume	3.98	1.31	0.027	5.57
3th volume	5.16	1.30	0.023	5.51
4th volume	5.54	1.31	0.022	5.47

5th volume	6.26	1.30	0.032	5.45
6th volume	6.34	1.34	0.029	5.57
7th volume	7.20	1.36	0.045	5.60
8th volume	6.10	1.38	0.042	5.69
9th volume	7.04	1.34	0.034	5.52
10th volume	6.74	1.38	0.041	5.75
11th volume	6.48	1.32	0.030	5.47
12th volume	5.91	1.33	0.030	5.56

The same as Table.2, the numerical value is nearly higher and higher. No matter in which angle, the articles in Tibetan teaching textbooks are designed with more words as students' grade increase.

3.2 Readability analysis

After finishing the data preparatory work, we combine Equation 1,2,3,4 with the value showed in table.3 to get the score of readability. The outcome of readability for Tibetan teaching textbooks is showed in Table.4.

Table.4 The score of readability for Tibetan teaching textbooks

formula No	FRE	GFI	ARI	FK
2th volume	91.65985445	1.602738889	6.802496124	1.463328089
3th volume	91.65740231	2.073292302	7.111080202	1.756936904
4th volume	90.46314709	2.226109595	7.104856461	2.018559769
5th volume	90.35038475	2.516446388	7.385404667	2.212133171
6th volume	87.20289315	2.548415872	7.962774291	2.67163113
7th volume	84.57879336	2.896976824	8.536393274	3.250261065
8th volume	84.2755969	2.456475636	8.411888585	3.019680543
9th volume	86.35804233	2.830689549	8.088764928	2.963656403
10th volume	83.52732781	2.710813164	9.006459389	3.282420027
11th volume	88.25858431	2.605696392	7.579337833	2.559830189
12th volume	88.33586322	2.376091459	7.717202815	2.406465853

Besides FRE, the other three formulas have the same rule that the higher the numerical value is, the article is harder to be understood.

In order to reflect the detail readability levels tested by those formulas between different volumes intuitively, we illustrate them by line chart. Chart1-4 shows the readability levels by each formula.

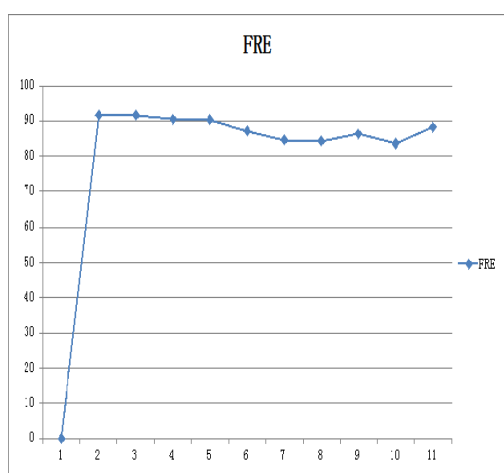


Chart.1 The readability level tested by FRE

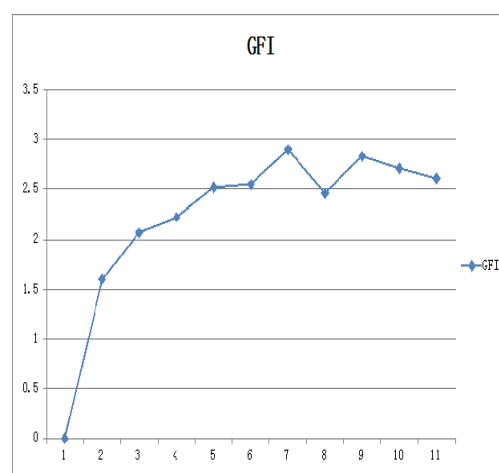


Chart.2 The readability level tested by GFI

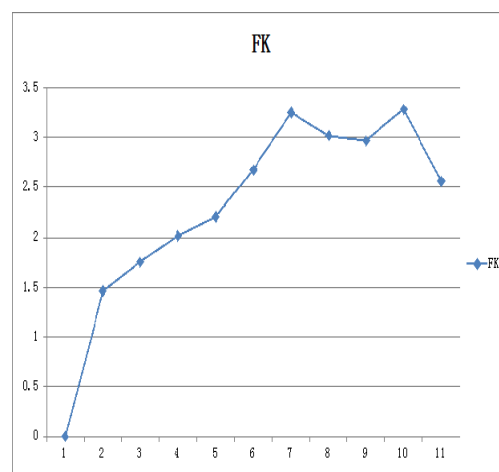
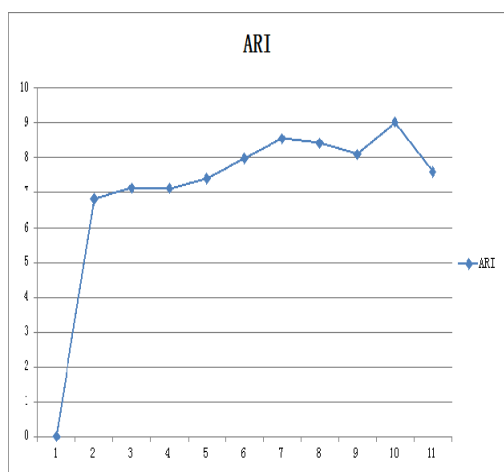


Chart.3 The readability level tested by ARI

Chart.4 The readability level tested by FK

From the above four Charts, we can get the information about the readability of the primary Tibetan teaching textbooks as follows.

- 1) From 2th to 7th volume, the score of each textbook's readability increases or decreases gradually, which means they become harder and harder to be understood for primary students.
- 2) From 8th to 12th volume, the score of each textbook's readability fluctuates unsteadily, even changing in opposite direction, which means the readability level tends to approach a maximum point relatively.
- 3) The score of last three textbook's readability increases or decreases gradually, which explained why the difficulty of textbook used by graduates was reduced slightly.

Conclusions

There are numerous mathematical formulas for the purpose of measuring the score of readability. Here we mainly adopt four widely used formulas to testify the effectiveness about their application for the Tibetan teaching textbooks. The results showed that the readability level tested by the four formulas was in accord with the real grade overall, and they got the similar trend. But whether these formulas can be applied to more complicated Tibetan textbooks is still unknown. Not all of the readability formulas can be adapt to each language rightly. We need to design the readability formulas for a special language (Tibetan) according to its literal features, which is our next research direction.

Acknowledgements

In this paper, the research was sponsored by Key Program of National Natural Science Foundation of China (No. 61331013), National Language Committee of China (No. WT125-46, No. WT125-11).

References

- [1]. Richards J C, Schmidt R W. Longman dictionary of language teaching and applied linguistics[J]. Longman Dictionary of Language Teaching & Applied Linguistics, 2010.
- [2]. Shao-Shan LI. A Survey of Studies on Readability[J]. Journal of PLA University of Foreign Languages, 2000.
- [3]. Kintsch W, Vipond D. Reading comprehension and readability in educational practice and psychological theory[J]. Perspectives on memory research: essays in honor of Uppsala University's 500th Anniversary, L. Nilsson, Editor. 1979, Lawrence Erlbaum Associates, 2014: 329-65.

- [4]. Schultz L. Readability Analysis of Programming Textbooks: Traditional Textbook or Trade Book[C]//Proceedings of the Information Systems Educators Conference ISSN. 2014, 2167: 1435.
- [5]. Izgi U, Seker B S. Comparing different readability formulas on the examples of science-technology and social science textbooks[J]. Procedia-Social and Behavioral Sciences, 2012, 46: 178-182.
- [6]. Xiao Liu. The review of the research about text readability[J]. Journal of Hubei University, 2015:141-146.
- [7]. Lee Y S, Tseng H C, Chen J L, et al. Constructing a novel Chinese readability classification model using principal component analysis and genetic programming[C]//Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on. IEEE, 2012: 164-166.
- [8]. Nima B. The choice of languages in Tibetan school education revisited[J]. Chinese Education & Society, 2008, 41(6): 50-60.
- [9]. GAO Lu, Yu Hongzhi. Statistical analysis of Tibetan language primary school textbooks character, word, sentence[C]//National minority youth Natural Language Processing Symposium.2008.
- [10]. Han Ying, Zhang JinShuang. The analysis of vocabulary statistics on primary Tibetan teaching textbooks[J]. Journal of Northwest University for Nationalities, 2010,31:79.
- [11]. Wang WenLing. Shallow Parsing on Tibetan textbook of Junior Middle school[D]. Northwest University for Nationalities,2012.
- [12]. Chai CL. The understanding and thinking about Tibetan teaching materials[J]. Qinghai education, 2004, 2(2): 37-38.
- [13]. Coleman M, Liao T L. A computer readability formula designed for machine scoring[J]. Journal of Applied Psychology, 1975, 60(2): 283.
- [14]. Woodmansey K. Readability of educational materials for endodontic patients[J]. Journal of endodontics, 2010, 36(10): 1703-1706.
- [15]. Crossley S A, Greenfield J, McNamara D S. Assessing text readability using cognitively based indices[J]. Tesol Quarterly, 2008, 42(3): 475-493.
- [16]. van Oosten P, Tanghe D, Hoste V. Towards an Improved Methodology for Automated Readability Prediction[C]//LREC. 2010.
- [17]. Burke V, Greenberg D. Determining Readability: How to Select and Apply Easy-to-Use Readability Formulas to Assess the Difficulty of Adult Literacy Materials[J]. Adult Basic Education and Literacy Journal, 2010, 4(1): 34-42.
- [18]. Stockmeyer N O. Using Microsoft Word's readability program[J]. Michigan Bar Journal, 2009, 88: 46.
- [19]. Hui Cao, HongZhi Yu. The statistical research about Tibetan lexicon[Z]. Beijing: People's Publishing House, 2013.11.
- [20]. Kang C, Jiang D, Long CJ. Tibetan word segmentation based on word-position tagging[C]//Asian Language Processing (IALP), 2013 International Conference on. IEEE, 2013: 239-242.