



## Development of a Readability Index Attuned to the New English Course of Study of Japan (3) Development of Ozasa-Fukui Year Level, Ver. 3.4.2nhnc1-5 Title of Paper

Toshiaki Ozasa<sup>1</sup>, Masayasu Fukui<sup>2</sup>, Kiyomi Watanabe<sup>3</sup>

Fukuyama Heisei University<sup>1</sup>, Fukuyama Heisei University<sup>2</sup>, Saniku Gakuin College<sup>3</sup> (Japan)

### Abstract

Since 2007, we have been engaged in developing readability indices attuned to the English textbooks based on the English Course of Study in Japan, and have developed readability indices, Ozasa-Fukui Year Level, Ver. 1, 2 and 2.1. However, the Course of Study was revised and put implemented from 2012 on, which made it necessary to update these English readability indices. This paper reports on a project that aims to develop a new readability index that is attuned to the new English teaching system in Japan, Ver. 3.4.2nhnc1-5. In the first phase, three sets of representative English textbooks in Japan were digitized in order to develop readability criterion, textbook-based word dictionary and textbook-based idiom dictionary. In the second phase, first, a multiple regression analysis, a linear analysis, was computed, using sentence length, word length, textbook-based word difficulty and textbook-based idiom difficulty as independent variables and the year of passages as a dependent variable, which yielded a linear function (Diff). Then, non-linear analyses were computed, using experts' readability judgment of selected English sentences as a dependent variable and the obtained linear function (Diff) as an independent variable. The computation yielded a non-linear function expression for Ver. 3.4.2nhnc1-5 with an unexpectedly high explanation rate. Ver. 3.4.2nhnc1-5 =  $4.2579 \cdot \exp(-60.1150 \cdot 0.2613^{\text{Diff}}) + 1$  (Diff =  $0.0915 \cdot \text{Words} + 0.5621 \cdot \text{Syllables} + 1.6230 \cdot \text{WordDiff} + 0.0822 \cdot \text{IdiomDiff} - 0.2836$ ) ( $r^2 = 0.8802$ ).

### 1. Introduction

Since 2007, we have been engaged in developing readability indices for measuring readability, based on the textual data of the English textbooks for junior and senior high schools in Japan. In a series of these attempts, we have so far developed three readability indices, Ozasa-Fukui Year Level, Ver. 1, 2, and 2.1, which are attuned to the English Course of Study in Japan (effective 2002-2011 for junior high school, 2003-2012 for senior high school). The main features of the three readability indices developed under this scheme are outlined in Table 1, along with other newly developed ones. The Course of Study was revised and put implemented from 2012 on, which made it necessary to update these English readability indices. This paper reports on a project that aims to develop a new readability index that is attuned to the new English teaching system in Japan, Ver. 3.4.2nhnc1-5. [1] [2] [3] [4] [5] [6] [7].

Table 1 Versions of O-F Year Level Developed

Version	r <sup>2</sup>	Dependent variable			Independent variable	Analysis
		Evaluation	Level Number	Datum size		
1	0.413	Objective	5	Big <sup>*1</sup>	4 <sup>*2</sup>	linear
2	0.824	Empirical (3)	50	126	4	nonlinear
2.1	0.822	Empirical (3)	50	916	4	nonlinear
3.1nh	0.7902	Empirical (3)	40	546	4	nonlinear
3.1nc	0.8373	Empirical (3)	40	151	4	nonlinear
3.1nhnc	0.7502	Empirical (3)	40	697	4	nonlinear
3.2nh	0.9144	Empirical (3)	40	157	4	nonlinear
3.2nhnc	0.8236	Empirical (3)	40	308	4	nonlinear
3.2s	0.7702	Empirical (3)	40	126	4	nonlinear
3.3nhsnc	0.7829	Empirical (3)	40	434	4	nonlinear
3.4.1nhnc1-5wus	0.6400	Empirical (3)	50	399	4	nonlinear

\*<sup>1</sup>All sentences contained in 15 vols. Of the textbooks.

\*<sup>2</sup>Sentence length + word length + textbook word difficulty + textbook idiom difficulty

## 2. Aim

The present paper aims to develop a new readability index, i.e., Ver. 3.4.2nhnc1-5, of higher quality in prediction or explanation that is attuned to the newly revised English Course of Study effected in 2012 and 2013, based on the digital data of the two sets of Japanese representative English textbooks for junior and senior high schools.

## 3. Method

First, a multiple regression analysis, a linear analysis, was computed, using sentence length, word length, textbook-based word difficulty and textbook-based idiom difficulty as independent variables and the year of all the passages of the textbooks used in the analysis as a dependent variable, which yielded a linear expression (Diff).

Then, non-linear analyses were computed, using experts' empirical judgment of English sentences as a dependent variable and the linear function (Diff) as an independent variable. The measure used for the dependent variable was a new criterion specifically developed for this analysis, which is based on the intuitive judgment of sentences selected from the two sets of the Japanese EFL texts by three experienced Japanese EFL teachers including the first author, who are well familiar with the status quo of the English textbooks and teaching of middle-grade schools in Japan. The textbooks used were the following English textbooks for junior and senior high schools.

- New Crown English Series 1, 2 & 3. [Junior high school] 2012, Tokyo: Sanseido.
- Crown English Communication 1 & II. [Senior high school] 2013. Tokyo: Sanseido.
- New Horizon English Course 1, 2 & 3. [Junior high school] 2012, Tokyo: Tokyo Shoseki.
- Prominence English Communication, 1 & II. [Senior high school] 2013. Tokyo: Tokyo Shoseki.

The computation was carried out using Fukui's College Analysis, a statistics computer program developed for social studies by Masayasu Fukui at Fukuyama Heisei University, Japan.



## 4. Results and Discussion

Since the results of the least squares analysis for the development of the Ver. 3.4.1nhnc1-5wus proved to be unsatisfactory to our expectation, another new attempt was made to develop a readability index of higher validity, Ver. 3.4.2nhnc1-5, by reexamining and improving the data used for developing the Ver.3.4.1nhnc1-5wus. The basic process of the analysis was the same as that used for the development of the former three versions.

First, a multiple regression analysis, a linear analysis, was computed, using sentence length, word length, textbook-based word difficulty and textbook-based idiom difficulty as independent variables and the school years of all the passages of the English textbooks as a dependent variable. The textbooks used for the analysis were two sets totaling 10 volumes, i.e., New horizon English course 1, 2 & 3, Prominence English communication 1 & II and New crown English series 1, 2 & 3 and Crown English Communication 1 & II. The following is the results of the linear regression analysis.

$$\text{Diff} = 0.0915 * \text{Words} + 0.5621 * \text{Syllables} + 1.6230 * \text{WordDiff} + 0.0822 * \text{IdiomDiff} - 0.2836 \text{ (Diff)}$$

Using the obtained regression expression, the coefficient correlations were computed between the readability values of the experts' empirical criterion (Year) and the estimation of the obtained regression expression (Diff), between Year. Figure 1 shows the relationship between the experts' readability evaluation (Year) and regression expression (Diff) with Year on the y-axis and Diff on the x-axis. In the figure the linear line represents its regression line.

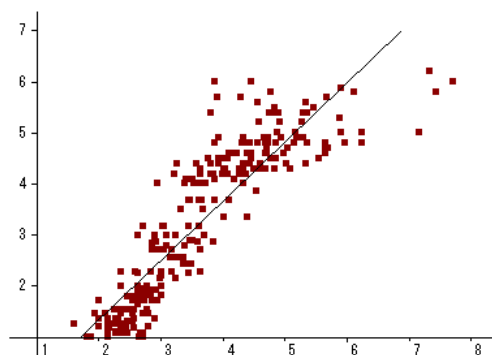


Figure 1 Coefficient correlations between experts' evaluation and Diff's estimation

At the second phase of the analysis, nonlinear, least squares analyses were computed, using the experts' judgment of 434 English sentences as a dependent variable and the above linear function (Diff) as an independent variable. In the analysis, two kinds of least squares analyses were computed, yielding two nonlinear function expressions, the Logistic solution and the Gompertz solution. The results of the solution are detailed in the following analysis results and in Figure 2.

Logistic solution  
 Object Variable Year  
 Expression  $\text{NewDiff} = a / (1 + b * \exp(-c * \text{Diff})) + 0.91$   
 Optimal Solution  
 A 4.1481  
 B 932.8258  
 C 2.0805  
 Num. of Cases 265  
 Num. of Solutions 1  
 Least Sq. Value 73.0800  
 Obs/Pred R 0.9372  
 Obs/Pred R<sup>2</sup> 0.8784



Gompertz solution

Object Variable	Year
Expression	NewDiff = $a \cdot \exp(-b \cdot c^{\text{Diff}}) + 1$
Optimal Solution	
A	4.2579
B	60.1150
C	0.2613
Num. of Cases	265
Num. of Solutions	1
Least Sq. Value	72.0142
Obs/Pred R	0.9382
Obs/Pred R <sup>2</sup>	0.8802

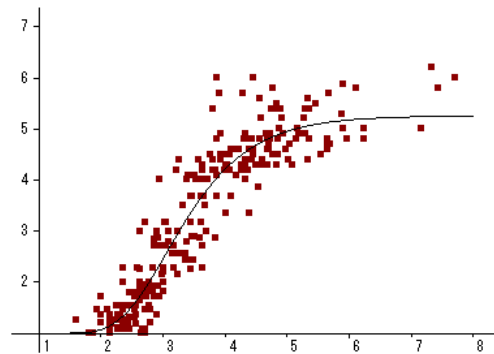


Figure 3 Prediction of Gompertz solution

As is shown in the analysis data above, the computation yielded the following nonlinear prediction expression (NewDiff).

$$\text{NewDiff} = 4.2579 \cdot \exp(-60.1150 \cdot 0.2613^{\text{Diff}}) + 1$$

$$(\text{Diff} = 0.0915 \cdot \text{Words} + 0.5621 \cdot \text{Syllables} + 1.6230 \cdot \text{WordDiff} + 0.0822 \cdot \text{IdiomDiff} - 0.2836)$$

Since the function expression of the Gompertz solution was higher than that of the Logistic solution in explanation ( $r^2$ ), the Gompertz solution was formally adopted as the index of the Ver. 3.4.2nhnc1-5 at the present stage of this project. [8] [9]

## 5. Conclusion

In the present study, the following function expression was adopted as an index of the Ozasa-Fukui Year Level, Ver. 3.4.2nhnc1-5. [10] [11]

$$\text{Ver. 3.4.2nhnc1-5} = 4.2579 \cdot \exp(-60.1150 \cdot 0.2613^{\text{Diff}}) + 1$$

$$(\text{Diff} = 0.0915 \cdot \text{Words} + 0.5621 \cdot \text{Syllables} + 1.6230 \cdot \text{WordDiff} + 0.0822 \cdot \text{IdiomDiff} - 0.2836) \quad (r^2 = 0.8802)$$

Currently, this could be the most reliable, authentic readability index that can be used in the formal education setting of Japan since it enjoys a satisfactorily high prediction value and is based on the data of the English textbooks that are widely used by the majority of students studying English in the junior and senior high schools of Japan. A word of caution would be needed, however, in using this index; since this index is based on the textbook data of five years from junior high first year to senior high first year, it could not be safely used in measuring readability beyond that year level.

## References

- [1] Anagnostou, Nikolaos K. & Weir, George R. S., 2006. From corpus-based collocation frequencies to readability measure. *ICT in the Analysis, Teaching and Learning of Languages* (ICTATLL Workshop Preprints), 21-22 August, 2006, 33-46.
- [2] Fukui, Masayasu, Ozasa, Toshiaki. 2008. Methodology for developing an English readability index with word and textbook-based idiom difficulty variables [Tango to jukugo no nanido o kouryosita eibun ri-dabiriri shihyou no sakuseihou]. *Educational Information Studies [Kyouiku Jouhou Kenkyuu]*, 24 (3), 15-22.

- [3] Fukui, Masayasu & Hosokawa, Mitsuhiro, 2004. Development of integrated program for social system analysis 8: Canonical coefficient correlation analysis, factor analysis and utility [Shakai sisutem bunseki notameno tougouka puroguramu 8: seijunsoukanbunseki, insibunseki and yutiriti]. *Business Information Studies*, 9, 23-35.
- [4] Leech, Geoffrey, Rayson, Paul & Wilson, Andrew. 2001. *Word frequencies in written and spoken English based on the British National Corpus*. Harlow: Pearson Education.
- [5] Readability formulas.[http://csep.psyc.memphis.edu/cohmetrix / readabilityresearch.htm](http://csep.psyc.memphis.edu/cohmetrix/readabilityresearch.htm)
- [6] Weir, George & Ozasa, Toshiaki. 2010. Learning from analysis of Japanese EFL texts. *Educational Perspectives*. 43, 1& 2, 56-66.
- [7] Weir, George R. S. & Ritchie, Calum, 2006. Estimating readability with the Strathclyde readability measure. *ICT in the Analysis, Teaching and Learning of Languages* (ICTATLL Workshop Preprints), 21-22 August, 2006, 25-32.
- [8] Fukui, Masayasu, Ozasa, Toshiaki & Hosokawa, Mitsuhiro. 2013. Development of new indices for measuring English readability based on the 2013 Course of Study of Japan: Developing Ozasa-Fukui Year Level, Ver. 3.1nhnc. [2013nendo gakushusidouyouryou ni junkyosita sin eibun ri-dabiritei sokutei shihyou no kaihatu: Ozasa-Fukui Year Level, Ver. 3.1nhnc no kaihatu]. *Fukuyama Heisei University Business Administration Studies*, 10, 49-62.
- [9] Ozasa, Toshiaki, Weir, George R. S. & Fukui, Masayasu. 2008. Toward a readability index for Japanese learners of EFL. *Conference Program of the 13th Conference of Pan-Pacific Association of Applied Linguistics*, 59-62.
- [10] Ozasa, Toshiaki & Watanabe, Kiyomi. 2014. Development of a readability index attuned to the new English course of study of Japan; Development of Ozasa-Fukui Year Level, Ver. 3.3nhsnc. *Proceedings of the 19th Conference of Pan-Pacific Association of Applied Linguistics*, 17<sup>th</sup> 18<sup>th</sup> August, 2014, Tokyo, Japan.
- [11] Ozasa, Toshiaki & Weir, R. S. George & Fukui, M.asayasu. 2014. Development of a readability index attuned to the new English course of study in Japan. *Proceedings of ICER2014 Conference*, 17<sup>th</sup>-19<sup>th</sup> November, 2014, Seville, Spain, 2448-2453.