

# A software application for assessing readability in the Japanese EFL context

Toshiaki Ozasa<sup>1</sup>, George R. S. Weir<sup>2</sup>, Masayasu Fukui<sup>1</sup>  
tozasa@heisei-u.ac.jp; george.weir@cis.strath.ac.uk; fukui@heisei-u.ac.jp  
<sup>1</sup>Fukuyama Heisei University, Japan

<sup>2</sup>Department of Computer and Information Sciences, University of Strathclyde, UK

## Abstract

We have been engaged in developing a readability index and its application software attuned for Japanese EFL learners. The index program, Ozasa-Fukui Year Level Program, Ver. 1.0, was used in developing the readability metric Ozasa-Fukui Year Level Index but tended to assume a high level of computer knowledge in its users. As a result, the application proved to be difficult to operate for 'lay' users. Since the majority of its expected users are not computer specialists, it is of crucial importance to develop a user-friendly program that can be used easily without any professional knowledge or skills. The goal of the present study was to provide a readability application that was informed by our index development work and sufficiently user-friendly for use by EFL professionals, i.e., Ozasa-Fukui Year Level Program, Ver. 2.0. The present paper describes the background context in which this work was undertaken and details our approach both to readability measurement and to the design of this software application for assessing readability in the Japanese EFL context.

## Introduction

The study of English as a foreign language (EFL) has a long history in Japan - with major periods of growth and pedagogical developments interspersed with the occasional downplay of English teaching (Imura, 2003; Fujimoto-Adamson, 2006; Sasaki, 2008). During the twentieth century, there were notable contributions from foreign academics. Among these was Harold Palmer (1877-1949), who spent fourteen years in Japan and served as 'Linguistic Adviser' to the Department of Education. In 1915, Palmer established the Institute for Research in English Teaching in Tokyo where the first Japanese conference on English Language Teaching was held in 1924 (Yamamoto, 1978; Smith, 1998).

Palmer's contribution was especially noteworthy for its emphasis upon practical methods of learning, but he also stressed the need to manage the 'gradation' of learning materials: 'In the ideal course [...] the course itself will be divided into appropriate stages, each of which will be marked by an increased capacity on the part of the student for assimilating and using language-material' (Palmer, 1921:67). Palmer appreciated that 'the rate of progress on the part of the student will depend very largely on the manner in which the vocabulary is graded' (Palmer, 1921:68) but went beyond vocabulary to recognize that 'gradation can and must also be observed in the study of the semantic aspect of a language' (Palmer, 1921:69). Such emphasis upon gradation in the vocabulary and semantic aspects of Japanese EFL learning materials, underpins the work reported in the present paper. A key requirement in support of appropriate gradation of language learning materials is the facility to gauge the comparative readability of texts.

## Readability

In previous work, we completed a quantitative diachronic analysis of fourteen sets of Japanese EFL textbooks and submitted a report entitled "A Quantitative Analysis of the English Textbooks of Meiji, Taisho and Early Showa Era," for research funded by the Grants-in-Aid for Scientific Research ((c)1, No. 19520535) of the Japan Society for the Promotion of Science. This was a quantitative analysis of Japanese EFL textbook corpora in which indices of their vocabulary and readability were computed, compared and interpreted from an historical perspective, using corpus analysis techniques. The fourteen sets of English textbooks consisted of five books each, giving a total of 70 books (see Table 1 for details).

Our analysis of this corpus, focused on seven features: (1) Overall tokens, (2) Cumulative overall tokens, (3) Types, (4) Cumulative types, (5) New types, (6) New Types / Tokens Ratio and (7) Readability. In the readability analysis, in particular, Flesch Reading Ease and Flesch-Kincaid Grade Level were employed as analytical tools.

**Table 1. Details of the textbook Corpus**

<b>Title</b>	<b>Author(s)</b>	<b>Short name</b>
Sanders' Union Readers	Sanders (1861-63)	Union
New National Readers	Barnes (1883-84)	National
English Readers: The High School Series	Dening (1887)	Dening
Seisoku Mombusho Eigo Tokuhon (Education Ministry English Readers)	Education Ministry (1889)	Seisoku
Standard Choice Readers	Shobido (1902)	Choice
Kanda's New Series of English Readers (Revised Edition)	Kanda (1903)	Kanda NSER
The Globe Readers	Okakura (1907)	Globe
New English Drill Books	Kumamoto (1907)	Drill
Girls' Taisho New Readers	Tsuda & Kumamoto (1916)	Taisho
The Standard English Readers	Palmer (1926-27)	Standard (P)
The Standard English Readers	Takehara (1932)	Standard (T)
Girls' Pacific Readers	Sawamura (1939)	Pacific
New Jack and Betty: English Step by Step (1,2,3), New High School English (1,2)	Hagiwara et al. (1952)	Jack & Betty
Sunshine English Course (1,2,3), Sunshine English Course (1,2)	Shimaoka et al. (1996), Tsuchiya et al. (1997)	Sunshine

Figures 1 and 2 show what we obtained from the Flesch Reading Ease and Flesch-Kincaid Grade Level analyses. A detailed critical evaluation of the results of these readability analyses revealed that they were not accurate enough to portray various features of the English text for the textbooks in focus. These indices were not considered adequate for differentiating the sentential difficulty of each textbook; the results did not seem to match professional insight on the readability of the analyzed textbook corpora. In addition, the results of these 'simple' readability analyses were somewhat different from those obtained through analyses using other indices, such as overall tokens, etc. While part of the reason for this result may be the nature of readability itself, as essentially orthogonal to lexical indices, the unsatisfactory outcome of the analysis may also have been due to the inherent deficiency in the index tools themselves. In large measure, this consideration motivated our present study into readability.

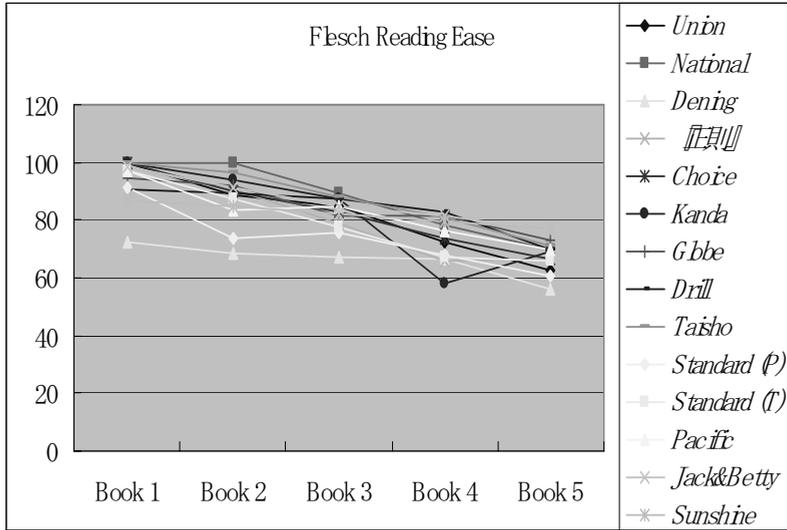


Figure 1. Flesch Reading Ease

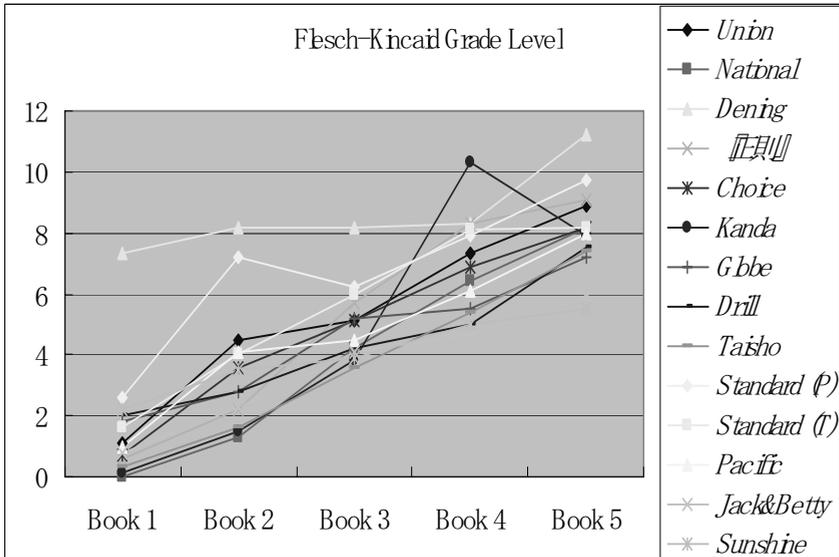


Figure 2. Flesch-Kincaid Reading Ease

The common presumption in readability measurement is that texts with short sentences and short words are more easily read and comprehended by the average reader. With this in mind, readability measures usually focus on features such as average sentence length (ASL) and average word length (AWL). While for many texts this assumption is appropriate, there are also many instances of texts with short but complex words and short but obscure sentences. For such examples, most readability measures would award 'good' readability scores, although most human readers may consider these texts 'difficult'. According to Connaster (1999), 'readability formulas fail to predict text difficulty' and he suggests that "text difficulty" is 'a perception of the reader and therefore cannot be objectively calculated by counting syllables, word length, sentence length, and other text characteristics'.

The perceived readability of any text can be understood as a complex function of the textual content and the reader's knowledge and language experience. The apparent need to accommodate the human reader's perspective in such measures, leads Oakland and Lane (2004) to recommend the use of 'readability methods that consider both quantitative and qualitative variables and are performed by seasoned professionals'.

Such comments challenge the notion that readability may be gauged independently of any individual reader, yet Palmer's comments on the need for 'graded materials' remind us of the educational requirement to gauge EFL reading materials. Indeed, this may plausibly be achieved, without recourse to individual readers, by reference to recognized educational standards. In the following, we propose such a principled approach as a method of rating texts relative to 'difficulty criteria' that are tied to the Japanese EFL teaching context.

In developing a readability index attuned for Japanese EFL learners, we followed an approach recommended in Weir & Ritchie (2006), and Anagnostou & Weir (2006; 2007). In the Japanese context, this work has led to the development of a 'year level index' measurement (presently, represented as the Ozasa-Fukui Year Level Index, Ver. 2.1). Unlike 'traditional' readability measures (such as Flesch Reading Ease and Flesch-Kincaid Grade Level), we employ a non-linear function for this readability index, using the 'Gompertz curve approximation method'. This was computed, using a newly developed subjective readability evaluation of one set of Japanese EFL textbooks as a dependent variable and a previously developed linear function as an independent variable. As a dependent variable, the following expression [1] was used:

$$[1] \text{Diff} = 0.0995 \times \text{Words}/S + 0.4302 \times \text{Syllables}/W + 0.9800 \times \text{WordDiff}/W + 0.0633 \times \text{IdiomDiff}/S + 0.2815$$

In previous work, we used the quantitative or objective textbook-based criterion measure as a dependent variable. However, a detailed qualitative examination of the results through multiple regression analyses revealed that this objective measure was not optimal as a criterion to guarantee coverage for the computation function. In order to improve the effectiveness of this expression, we developed a new readability criterion which is based on the intuitive judgment of 126 sentences from one set of Japanese EFL textbooks by four experienced Japanese EFL teachers. Each of these ‘informants’ is familiar with EFL textbooks and is practiced in teaching at middle-grade schools in Japan.

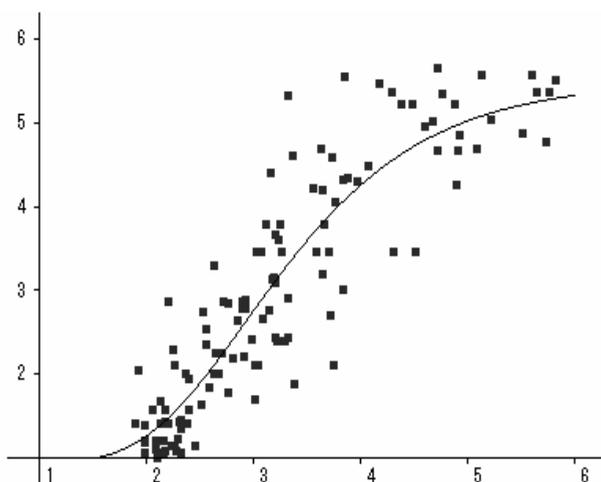
The following solution of the ‘Gompertz curve’ approximation method computed in this study, proved to be highest in predictive power ( $r^2=.824$ ) of the four non-linear methods simultaneously computed (see Table 2 and Figure 3). Using regression analysis, the resultant non-linear function *NewDiff* [2] was derived, using the previous function *Diff* as an exponent, i.e.,  $\text{NewDiff} = a \exp(-bc^{\text{Diff}}) + d$ .

$$[2] \text{NewDiff} = 4.535 \times \exp(-22.223 \times 0.345^{0.0995 \times \text{Words}/S + 0.4302 \times \text{Syllables}/W + 0.9800 \times \text{WordDiff}/W + 0.0633 \times \text{IdiomDiff}/S + 0.2815}) + 0.953$$

In Ozasa & Weir (2009), the coefficients of correlations were computed between Ozasa’s readability judgment (criterion) and the practical measurements of *Ozasa-Fukui Year Level*, *Flesch-Kincaid Grade Level* and *Flesh Reading Ease*, and the results were compared across the three coefficients. This revealed that *Ozasa-Fukui Year Level* was the most powerful in differentiating the readability of the selected 84 EFL sentences ( $r = 0.9041$ ). It also emerged that the coefficient between Ozasa’s criterion and *Flesch-Kincaid Grade Level* was remarkably high ( $r = 0.7137$ ), much higher than that between the criterion and *Flesch Reading Ease* ( $r = -0.4396$ ).

**Table 2. ‘Gompertz curve’ approximation**

Description	Optimal solution
a	4.535
b	22.223
c	0.345
d	0.953
No. of data used	126
Least square value	45.568
r (evaluation/prediction)	0.908
$r^2$	0.824



**Figure 3. ‘Gompertz curve’ approximation**

A comparative analysis of the standardized measurement data conducted in Ozasa & Weir (2009) also revealed that *Ozasa-Fukui Year Level* was more powerful in differentiating Book 1 texts, and that *Flesch-Kincaid Grade Level* was generally more powerful in differentiating the texts of higher years. In general, however, there were not clear quantitative differences observed between the measurements of the two indices.

The analyses carried out in Ozasa & Weir (2009) were two-fold, a criterion based analysis and criterion-less comparison of the two types of measurements. The criterion-based analysis succeeded in producing a clear-cut conclusion as to the relative efficiency of the three indices in question. In contrast, the criterion-less comparison failed to produce a persuasive or consistent conclusion as to the relative efficiency of the two analyzed indices.

The approach to readability measurement that we have developed combines ‘traditional’ factors such as average words per sentence and average syllables per word, with Japanese specific factors such as a measure of word difficulty, based upon the JACET list of basic English words (JACET, 2003), with the addition of two further Japanese-specific factors, viz., a measure of idiom difficulty and a sentence difficulty factor based upon the judgment of our panel of Japanese EFL experts. The resultant program allows the analysis of sentences of more extensive texts and produces an index value that indicates the approximate year level difficulty, relative to Japanese

High School standard. This is intended to be of direct application in the Japanese EFL context, with a specific view to enabling direct assessment of target texts for use in EFL teaching.

Having evaluated the new readability index from an empirical perspective, our attention turned to delivering a version of the associated software application that would be useful in applying the readability index. This entailed a series of design changes, especially with regard to the system's user interface, since the original purpose for the software system had been development and testing of our technique for readability measurement rather than the general application of such an index in the field of EFL.

The prototype software application, i.e., *Ozasa-Fukui Year Level Program, Ver. 1.0*, which was used in developing *Ozasa-Fukui Year Level Index*, was not originally intended for use by EFL teachers. In consequence, while it could be used effectively by its developers, it was difficult to operate for 'lay' users. With a view to future application of readability assessment in the Japanese EFL context, we sought to provide a readability application that was informed by our index development work and sufficiently user-friendly to be employed by EFL professionals.

## Software application

With a view to providing a usable tool for professional EFL readability assessment in the Japanese EFL context, we have developed the *Ozasa-Fukui Year Level Program v.2*. In the following, we outline the features of this program and detail some of our recent developments that aim to enhance the accessibility of the system. The main program display is illustrated in Figure 4.

This display shows the four windows that make up the analysis program. Top right is the set of main controls whereby the user may specify the required operations on a specified sample text. The text being analysed is displayed in the top left window, while the results of analysis are shown in tabular form (bottom left) and in graph form (bottom right). Following this initial design (Version 1), we sought to enhance the system in three respects: (1) simplification of the main menu, (2) multi-lingualization of the main menu, and (3) addition of an on-line help function.

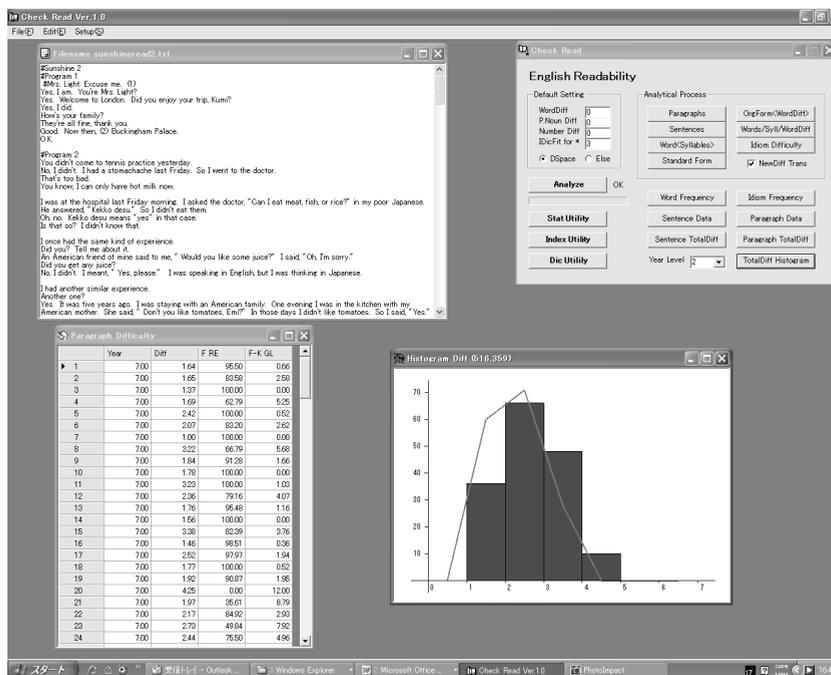


Figure 4. Main display of Ver. 1

### Menu simplification

Simplification of the main menu was considered a crucial factor in determining the usability of the *Ozasa-Fukui Year Level Program*. An over-complex main menu, with an excessive range of functional options, might confuse general users. In part, this complexity had arisen as a result of developing the system to meet two distinct, but associated needs. In the first place, the software was developed to permit experimentation with readability measurement. This required a variety of functions that permitted the user to modify the underlying measurement criteria. Subsequently, the system was directed more toward the application of the derived readability measurement technique. In this latter purpose lies its greatest potential user population, but such users are unlikely to require the underlying functionality that allows modification of the underlying measurement. Therefore, we sought in the system re-design to achieve a greater separation between the functionality of readability index development and the functionality of readability measurement.

This principle of separate functionality, allowed us to focus on simplifying the main menu, in order to limit the number of buttons to those essentially required for readability measurement, and to isolate the other functions – those more closely related to readability index development – from the main menu. We also introduced coloured button labels to distinguish less commonly accessed options from the more commonly used control buttons. Our efforts in this direction resulted in a reduction from eighteen buttons, four text fields, three checkboxes and one pull-down menu, in the initial version, to thirteen buttons, one checkbox and two pull-down menus in Version 2 (Figure 5).

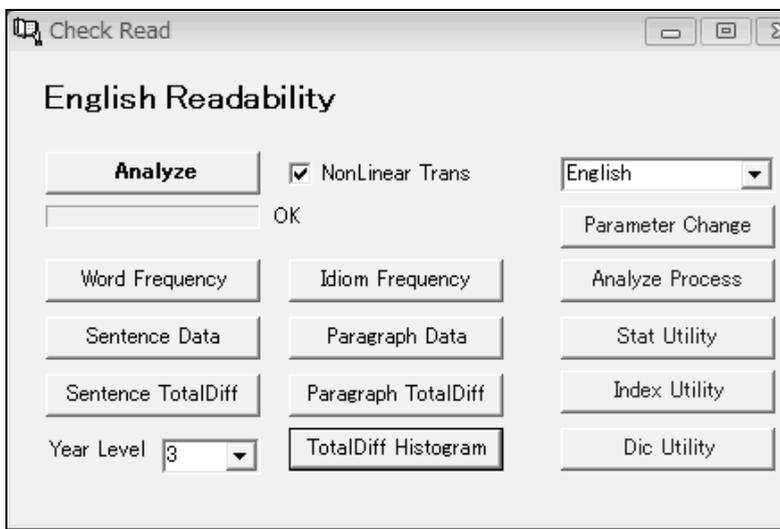


Figure 5. Revised Main menu

### ***Multilingual menu***

The interface for the initial version of the *Ozasa-Fukui Year Level Program* was implemented in English. However, since the readability analysis method applied in our program is potentially applicable to texts in any language with users whose first-language is not English, a user-selectable multi-lingual main menu was considered a desirable addition to Version 2 of the system. This was achieved by means of a separate configuration file that specifies language specific menu data. Thereby, the interface language can be switched without affecting other aspects of the program. The first extension in this vein was provision of a Japanese language option to display all main menu labels in Japanese rather than English (Figure 6).

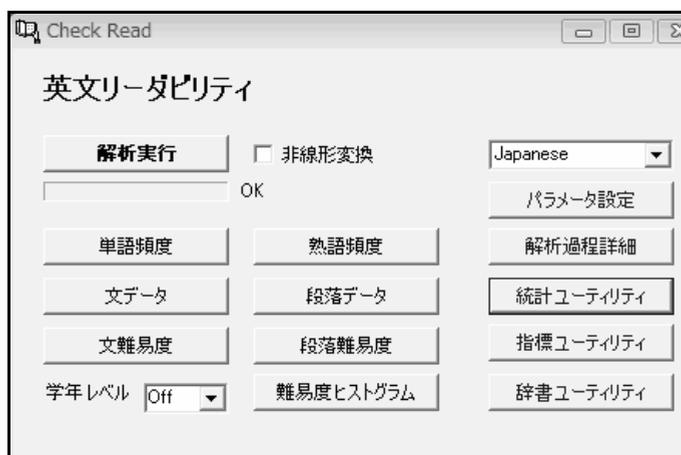


Figure 6. Japanese Main Menu

### On-line help facility

Finally, to enhance program usability, we integrated an on-line 'help' function. While this is presently available in Japanese, in due course, this will also reflect the user's choice of local language. Figure 7 shows the image picture of the current Japanese help display.

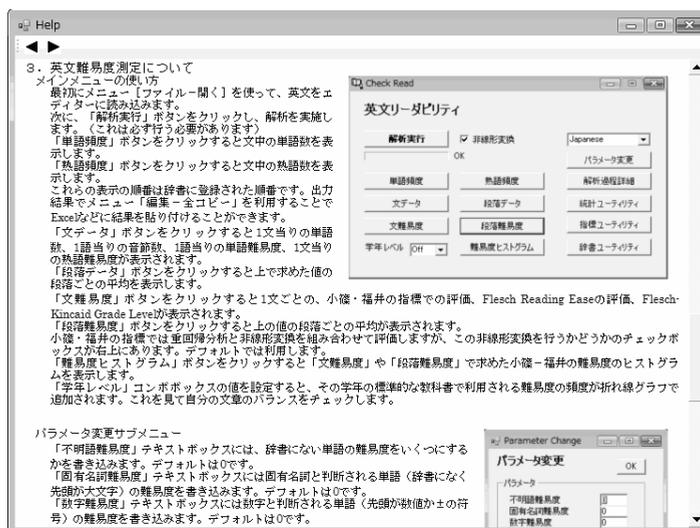


Figure 7. Japanese Help System

## ***Usability testing***

As a further step toward enhancing the usability of our developing system, we engaged the assistance of several academics with a research background in human-computer interaction. These experts have conducted a usability study on the prototype of version 2 and provided helpful feedback on how best to accommodate users engaging in readability analysis. Input from these analytical studies of the interface used in the *Ozasa-Fukui Year Level Program* will be reflected in the forthcoming revision of this program.

## **Conclusions**

In the context of our work on readability index development, we have produced the *Ozasa-Fukui Year Level Program*. This was the basis for our index development and is now being revised in order to afford a general readability measurement facility for use in the Japanese EFL context.

To this end, we have taken the initial version of the program and simplified its main menu. Support for end-user languages has been extended through multilingualization of the main menu, and the addition of a help function. The number of buttons on the user interface has been reduced to the essentials and other functions have been isolated from the main menu. We have also aimed to make the main menu more colourful and efficient for the general user.

Our short-term goal is to have a public release of the readability index (by means of the *Ozasa-Fukui Year Level Program*). Through wider use and the feedback thereby generated, we may better improve the usability of our program. This release will also include a user manual for general users, both in Japanese and English and should prove beneficial in helping EFL teachers apply our readability index for their professional use. In so doing, we believe that they can better determine the grade level for sample English texts and so achieve Palmer's requirement to 'observe the gradation in the study of the semantic aspect of the language'.

## **Acknowledgement**

The present study is funded by the Grants-in-Aid for Scientific Research ((c)1, No. 19520535) of the Japan Society for the Promotion of Science. The authors express their appreciation for this financial support to the Japan Society for the Promotion of Science.

## References

- Anagnostou, N. K., & Weir, G. R. S. (2006). From corpus-based collocation frequencies to readability measure. *ICT in the Analysis, Teaching and Learning of Languages (ICTATLL Workshop)* (pp. 33-46). Glasgow, U.K.
- Anagnostou, N. K., & Weir, G. R. S. (2007). Average collocation frequency as an indicator of semantic complexity. *ICT in the Analysis, Teaching and Learning of Languages (ICTATLL Workshop)* (pp. 43-48). Hiroshima, Japan.
- Barnes C. J. (1884). *New national readers*. New York: A.S. Barnes.
- Connaster, B.R. (1999). Last rites for readability formulas in technical communication, *Journal of Technical Writing and Communication*, 29(3) (pp. 271–287).
- Dening W. (1887). *English Readers: The High School Series*. Tokyo: Mombusho Publishing Bureau.
- Fujimoto-Adamson, N. (2006). Globalization and history of English education in Japan, *Asian EFL Journal*, 8(3), Article 13.
- Imura, M. (2003). *English education in Japan during the past 200 years*. Taishukan Shoten, Tokyo (In Japanese).
- JACET (2003). *List of 8000 basic words. basic word list revision committee*. Tokyo: Japan Association of College English Teachers (JACET).
- Kanda N. (1903). *Kanda's New Series of English Readers* (Revised Edition). Tokyo: Sanseido.
- Kumamoto, K. (1907). *New English Drill Books*. Tokyo: Kaiseikan.
- Oakland, T., & Lane, H.B. (2004). Language, reading, and readability formulas: Implications for developing and adapting tests. *International Journal of Testing*, 4(3), 239-252.
- Okakura, Y. (1907). *The globe readers*. Tokyo: Dainippon Tosho.
- Ozasa, T., & Weir, G. R. S (2009). Evaluating a newly developed index for readability measurement of Japanese EFL textbooks. *Proceedings of 14<sup>th</sup> Conference of Pan-Pacific Association of Applied Linguistics* (pp. 59-62). Tokyo, Japan.
- Palmer, H.E. (1921). *The principles of language study*. Harrup, London.
- Palmer, H. E. (1926). *The Standard English readers*. Tokyo: IRET.
- Saeki, Y. (1902). *Standard choice readers*. Tokyo: Shobido.
- Sanders, C. W. (1861). *Sanders' union readers*. New York: Iveson, Phinney, Blakeman & Co.
- Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing*, 25, 63-83.
- Sawamura, T. (1939). *Girls' pacific readers*. Tokyo: Kairyudo.
- Seisoku Mombusho Eigo Tokuhon (1889). *Education Ministry English readers*. Tokyo: Education Ministry.
- Shimaoka, T. (1996). *Sunshine English course (1, 2, 3)*. Tokyo: Kairyudo.
- Smith R. C. (1998). The Palmer-Hornby contribution to English teaching in Japan. *International Journal of Lexicography*, 11, 269-291.

- Takehara, T. (1932). *The Standard English readers*. Tokyo: Taishukan.
- Tsuchiya, S. (1997). *Sunshine English course (1, 2)*. Tokyo: Kairyudo.
- Tsuda, U., & Kumamoto, K. (1916). *Girls' Taisho New readers*, Tokyo: Kaiseikan.
- Weir, G. R. S., & Ritchie, C. (2006). Estimating readability with the Strathclyde readability measure. *ICT in the Analysis, Teaching and Learning of Languages (ICTATLL Workshop)* (pp. 25-32) Glasgow, U.K.
- Yamamoto, N. Y. (1978). The oral method: Harold E. Palmer and the reformation of the teaching of the English language in Japan. *ELT Journal*, XXXII, 151-158.