

Building Personalized Readability Equation and English Vocabulary List for Continued Reading Study in Comparison among Differently Proposed Algorithms

Yoshinori Miyazaki ^{*}, Shumma Suzuki ^{*}

Abstract

The authors have developed REX, a web application for facilitating English reading study, for non-native speakers of English. REX has a function that personalizes their individual readability equations based on their study logs and that predicts the ease or difficulty of the texts provided. For a parameter of such equations, the rate of difficult words, which does not appear in the specified vocabulary list, is often used, as in some of the well-known readability formulas. On the other hand, we believe that the use of a static vocabulary list to produce its rate does not contribute to personalization, because each learner has different vocabularies. This triggered the authors to incorporate a new function to let learners click their unknown words while they are reading texts, play “English Word Game”, and further predict the status of known/unknown of the words in the designated vocabulary list. In this study, we will focus on new labeling algorithms and an experiment to evaluate whether the quality of labeling has been improved. Among the devised several algorithms (A to E), making comparisons is also attempted.

Keywords: E-Learning, Gamification, Personalized Wordlist, Readability, Similarity

1 Introduction

The authors have developed REX application to facilitate reading study on the Web. In REX system, we continue to learn about reading ability and add some new features in English texts suitable for the English ability of each user and in English game.

In estimating the English ability of users, the readability of English texts, which is defined from how difficult the users evaluate, is calculated in advance. The readability of the texts in REX is determined by using parameters such as the number of words in a text, the length of the sentence, and the degree of difficulty (proportion of unknown words). Functions which define readability are called readability equations.

A variety of readability formulas have been developed in English-speaking countries and elsewhere. Well-known formulas include Flesch Reading Ease (FRE) [1], Flesch–Kincaid Grade

^{*} Shizuoka University, Shizuoka, Japan

Level (FKGL), the original Dale–Chall (1948) [2], the revised Dale–Chall (1995), Zakuluk–Samules, Bormuth, Coleman–Liau, Gunning Fog, etc. (see [3] for more details).

On the other hand, because of fluctuating results for Japanese learners of English (JLEs), some Japanese researchers have pursued better solutions for them, based on the premise that more suitable parameters can be explored. For example, the formula created by Kiyokawa [4] (let this be called the Kiyokawa formula) was shown to be more accurate than the two Flesch counterparts (FRE and FKGL) as a measure of text difficulty for JLEs. The Kiyokawa formula is expressed as follows:

$$R_{\text{Kiyokawa}} = 0.288 N + 0.17 X + 1.739,$$

where N is the ratio of words not included in the word list defined by Kiyokawa (X is the number of words in a text).

In REX, a readability equation is personalized based on the study logs of each user. Since the readability equations influence the reading features of users, it is possible to provide English texts suitable for them.

REX has prepared a wordlist summarizing English words, which are useful for English learning. All the English words in the wordlist are recorded whether users understand the meaning. The function of recording is defined as marking a “label.” In research on English learning, we have used marking labels in many ways. [5] marked labels to check unknown English words in the reading texts and in multiple choice gamification. The final purpose to use labels in REX is reflecting users’ knowledge of English vocabulary. It will be possible to calculate the degree of difficulty in the text from label information, which is used as a parameter of readability equations.

However, in the “labeling” method, it takes a huge amount of time to judge whether all English words are known or unknown, while many of them cannot be marked a label. In order to solve this problem, the authors considered new labeling algorithms to improve the current method, in which all English words are marked labels.

This study aims to develop REX and give an explanation on the improvement points. In particular, we will focus on new labeling algorithms and an experiment to evaluate whether the quality of labeling has been improved.

2 Literature Review

Ehara et al., [6] have developed SocialDict which predicts unknown words from the history of English words clicked by the users. When predicting unknown English words in the texts, it will display the meaning to support users’ reading in advance. In SocialDict, it is judged whether users can read the texts by combining the function of known (unknown) word prediction and readability.

Fukui et al., [7] have developed CheckRead to create dictionaries of words and idioms with their difficult levels. CheckRead can form the readability equations and evaluate readability as well. In addition to the number of words and syllables, the difficulty of words and idioms is also considered as parameters of readability equations.

Reference [8] showed that the web contains a wealth of texts appropriate for reading study. When readability formulas were applied to the texts on the web, it was found that many of them were in the range of the readability values for ESL (English as a Second Language) texts. This shows the usefulness of web pages for learners’ reading study.

Miyazaki et al., [9] added modules that measure the number of grammatical and speech objects

by using the result of syntax analysis. Those modules can also be used for the parameters of English readability equations.

Miyagishi et al., [5] added gamification to insert games into REX study. After finishing readability, you can get an item called a battle ticket to challenge a multiple choice English word game. Therefore, it can mark labels in the game as well as in reading.

3 Web Application REX

One of the primary objectives of this study is to develop a computer program that allows program users to obtain one or more tailor-made readability equations after a sequence of simple operations, as explained below. The overall study flow is illustrated in Figure 1.

The main feature of REX (Reading EXercise) is to provide learners with learning materials appropriate for their reading comprehension levels, by producing their own readability equations. While learners are using REX, they are required to rate each text read during their study. REX builds readability equations for each learner, which are linear equations created by a multi-regression analysis between the rated values (independent variable) and the parameters of the texts (dependent variables). REX automatically chooses the best combination of parameters for the equation in terms of the coefficient of determination.

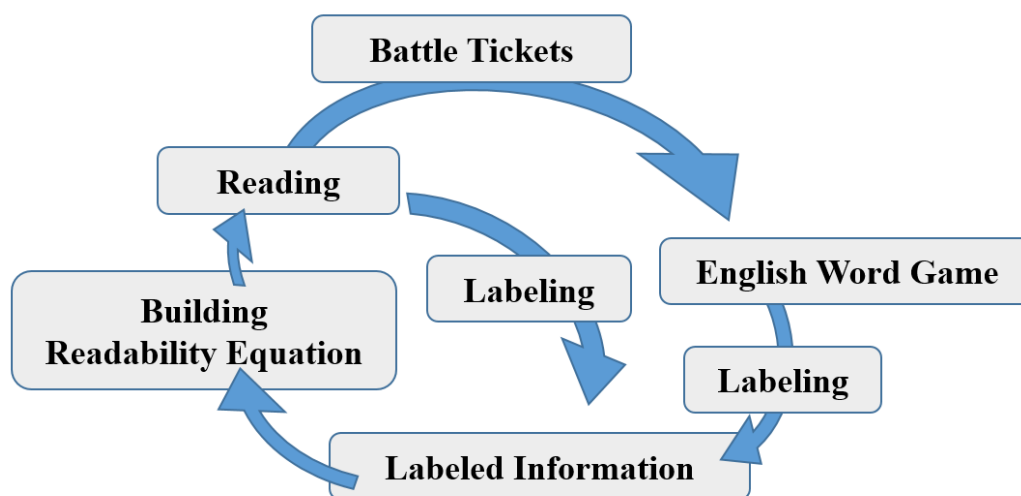


Figure 1: Study Flow of REX

3.1 Update the Vocabulary List by Reading

REX is a web application intended for English reading study. Based on the vocabulary list of SVL12000 [10], it records whether each user understands the meaning of the English words.

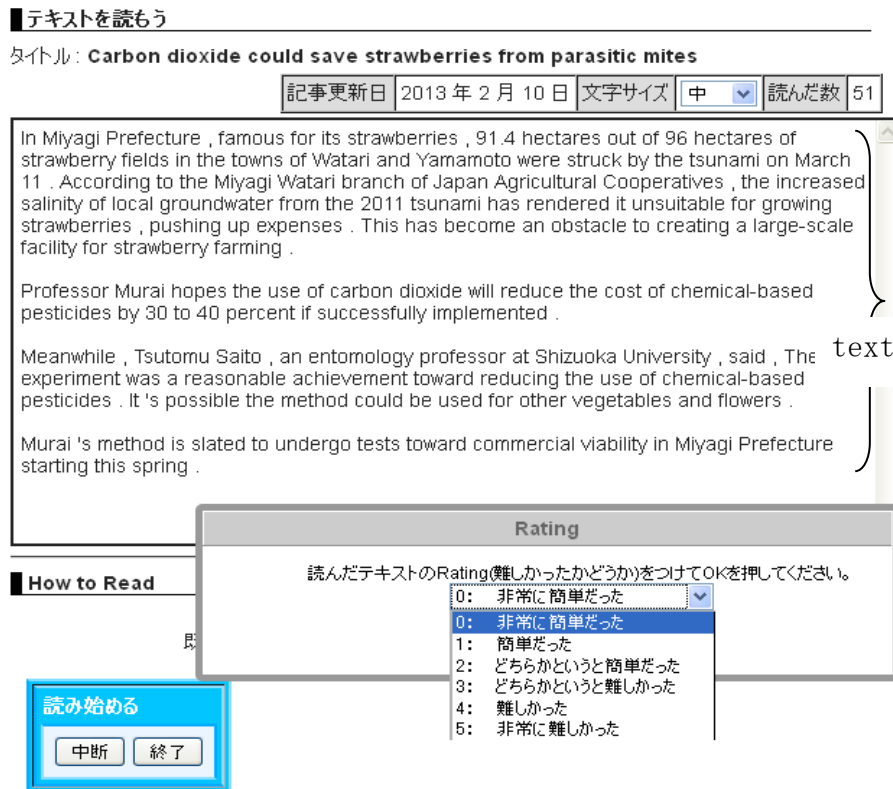


Figure 2: Screenshot of REX Reading Study

Each user reads and rates the English articles for them. “Rating” refers to the self-evaluation of the difficulty of reading texts, in which users choose one of the six following grades (Table 1).

Table 1: List of Rating Value

0 : Very easy	1 : Easy
2 : Moderate (more or less easy)	3 : Moderate (more or less difficult)
4 : Difficult	5 : Very difficult

Readability equation is created from each text parameter and rated value every multiple 10 times of reading. The method to select the parameters and texts using readability equations is based on deriving multiple regression equations. Refer to [11] for more details.

While reading texts, the background color of the English word is changed to yellow if it was judged as unknown words by marking label information. In addition, if users click on an English word while reading, its background color will be changed to orange. In the meantime, this English word is marked a label as an unknown word (Figure 2). English words that were not clicked are marked labels as known words after rating.

All the English words having the same original form as the clicked ones are changed the background color at the same time. We use TreeTagger [12] to annotate the original form of the words. If yellow or orange English words are clicked again, they will become colorless and are marked labels as known.

In this study, users can not only re-check English words after rating (Figure 3) but can also re-mark labels to click unknown English words or uncheck known English words. Those English words are derived from evaluating the English ability from users' label history, and only displayed at an appropriate level. After reading, users will get a battle ticket.

偶然にふと				
accidentally	brehtaking	brutal	bustle	caricature
catalyst	compelling	convey	corruption	crooked
dedicate	disbelief	effectively	espionage	exacting
guarded	inferior	insurgent	lawsuit	lone
motivate	motivation	pawn	protagonist	sensitivity
sequence	stunt	suspension	taunt	thriller
ultimate	uncover	undo	unpredictable	villain
yearning				

Figure 3: Re-checking Words After Reading

3.2 Update Vocabulary List by English Word Game

After finishing reading, users can play an English Word Game with a battle ticket (Figure 4). They will choose a correct answer from 4 options. The selected English words are “noun,” “verb,” “adjective” or “adverb,” which are the most popular in the vocabulary list based on SVL12000. In this function, we use English-Japanese Weblio dictionary [13] to get the meaning and Japanese translation.

英単語ゲーム
ゲームスタート!

36 / 36

あなた: レベル21

7

36 / 36 36 / 36 36 / 36

ワサビくん レベル 22

ワサビくん レベル 22

ワサビくん レベル 22

items: 12 strawberries, 9 oranges, 4 purple items, 2 carrots

word provided: pry

問題: pry の意味を以下の4つから選んでください。

- なだめる, なだめすかす, 慰撫する
- のぞく, せんさくする
- 激しく攻撃する, 猛攻する
- 逆にする, 反対にする, 転倒させる

confidence: 4 自信あり, 3 やや自信あり, 2 やや自信なし, 1 自信なし

Figure 4: Screenshot of English Word Game

A new element added in this study is confidence. Users will rate how self-confident they are when answering a question. The degree of confidence is divided into four levels, and labels are made according to it and correction as well (Table 2).

Table 2: List of the Level of Confidence and Correction

	<i>Confident</i>	<i>More or Less Confident</i>	<i>More or Less Unconfident</i>	<i>Un-confident</i>
Correct answer	4	3	2	1
Incorrect answer	unchanged	unchanged	0	0

If the incorrect answer was considered as misunderstanding with the high degree of confidence, the vocabulary will not change. In the table, numerical values indicated 0 are unknown, and higher numerical values than 0 are known. If users choose the correct meaning of the English word, the enemy character is beaten. In case of incorrect answer, he is attacked by the enemy character. The life point will decrease when being attacked, then English Word Game is over if it becomes 0. Users get the life point according to the level of confidence, which is either beating a great damage for a correct answer with high confidence or receiving damage heavily for an incorrect answer. English Word Game will continue until the life point becomes 0 with one battle ticket.

4 New "Labeling" Method

In previous studies, there were two methods to mark labels – Reading and English Word Game. However, the number of English words is enormous while the vocabulary list is based on SVL12000 that extends to 12,000 words. The problem is that it requires much time to label all the English words in the vocabulary list just by Reading and English Word Game.

For this reason, our research group has developed a new labeling algorithm that can mark labels to all the English words. Although this algorithm is a trial method, it can label all the English words (defined and undefined English words, which are marked labels in Reading and English Word Game, are called “appeared English words”) in Reading and English Word Game. However, in order not to influence users’ intention, we developed this method as an algorithm that can reflect users’ English vocabulary ability as much as possible.

4.1 Algorithm A (The use of the known ratio in appeared English words)

According to the algorithm, it predicts that if 90% of appeared English words at a certain level are known, 90% of unappeared English words of the remaining level will be also known. Then, we calculate the ratio of English words labeled “known” in the total of appeared English words at a certain level. It means we use the latest label information to calculate the known label out of the total number of appeared English words either in Reading or English Word Game. When label information is used in Reading, it is summed with “known” as 1 and “unknown” as 0. In English Word Game, if a label is used as the value shown in Table 2, in case the level of non-value is from 1-4, you will divide the label by 4 to normalize the value. In addition, we can define the label of the unknown English words for each level by calculating the label of known English words. However, it is necessary to decide which word should be extracted to mark labels from the unknown wordlist. To solve this problem, two following algorithms are examined.

4.1.1 Algorithm A-1

The label of the number of unknown English words are randomly defined from the English word-list.

4.1.2 Algorithm A-2

According to this algorithm, since the degree of difficulty of English words is different even at the same level, it is preferable to label the simple English words as "known" first. To define which one is simple, label information is used. After reviewing it, how many words are marked label and the number of known label are calculated. Then, English words are rearranged in levels by giving first priority to "descending order of the number of known labels," and second priority to "descending order of the number of English words that have been marked its label already." This process is defined as "ranking."

4.2 Algorithm B (Input the known ratio of appeared English words)

Similar to Algorithm A, this algorithm uses the known ratio in appeared English words. In Algorithm A, the label of unappeared English words is defined by the known ratio of appeared English words. Meanwhile, the known ratio of appeared words is used as the input value of unappeared word while marking a label. For example, if the known ratio of English words at some level is 0.9, 0.9 is used as the input value to mark labels to all unappeared English words.

4.3 Algorithm C (Use label information to define the degree of similarity of English ability)

4.3.1 Algorithm C-1

This algorithm does not mark labels to unappeared English words for each level that are defined by SVL12000 like Algorithm A and Algorithm B, but marks labels to 12,000 English words at a time. If we want to label an English word, it refers to the label information of other users. With this method, it is easier to predict a known or unknown word (defined known words as known value, unknown words as unknown value).

However, the accuracy of one's vocabulary list may be lowered when using the label information of other users whose English ability greatly differs. To improve the accuracy, we develop the algorithm with the user who has similar English ability. The coincidence in English ability (defined as English ability similarity) is required. Moreover, English ability similarity can be used to label appeared English words that are common between users. By extracting the label information from the commonly appeared English words from two users, we can create an n-dimensional vector of both label information. Therefore, the similarity English ability of two users can be defined by using the two label information vectors.

We consider the known value and unknown value to calculate the English ability similarity. If the known value is greater than the unknown one, 1 is used as input for label; otherwise, 0 is used. The detailed stipulation of Algorithm C-1 is omitted.

4.3.2 Algorithm C-2

This algorithm is basically the same as Algorithm C-1. The difference is that the wider range of input value can be taken. The value considered as the input value is defined by the ratio of the total of known and unknown values, not just 1 or 0.

4.4 Algorithm D (Combined method of Algorithms B and C2)

Algorithm C2 should work finely in the long run, but one drawback is that, without accumulation of study logs of other users, it will not, because this algorithm refers to other users depending on their similarities. To deal with this, the authors came up with the idea of combining C2 with Algorithm B, by the equation which weights B and C2 by the number of other users to whom the word in question appeared. The more students to whom the specific word appears, the higher weight applied on Algorithm C2 (less weight on Algorithm B).

4.5 Algorithm E (Modified algorithm of Algorithm D)

Algorithm E is basically the same as Algorithm D except some modification. In Algorithm D, similarities of different learners are computed, and the word is marked only when both of the learners know the words (this is when similarity is augmented). In Algorithm E, the word is marked also when none of the learners know the word, too, because the learners coincide in that the word is unknown to both of them.

5 Experiment and Discussion

Experiment was conducted to study the vocabulary list generated by the proposed algorithms which is related to practical English words of users. In this experiment, 10 users (including undergraduate and graduate students) were asked to finish REX 100 times, an English Word Game by the tickets they received after reading. The algorithms were applied every 10 times of reading, and the vocabulary list was updated. We used a list fixed to 12 levels by SVL12000, and the known English words at level 4 or less were assumed initially. Next, users were asked to finish 100 texts and 240 questions of English words test. This omitted updating words and vocabulary list as described. 240 questions were selected and divided into 20 levels from “noun,” “verb,” “adjective” and “adverb.”

We observed the accuracy of 240 question-English-word-test and whether an English word is known or unknown. For example, in a vocabulary list, known English words should be selected correctly by selecting the right meaning in the English word test. Otherwise, it is possible to predict that you had given a wrong answer to the unknown word in the vocabulary list. According to the result, we say “Matched” if it is as predicted, and “Unmatched” if it is not. The greater the number of matches, the more vocabulary lists will be updated in this experiment, which can define the users’ ability of English words.

In this experiment, we compared the number of matches of vocabulary list when each of the algorithm is used and the vocabulary list without them. Three vocabulary lists for reading, English Word Game are generated, and defined as known or unknown vocabulary list. See Table 3 and Table 4 for the results (Table 3 for all words, and Table 4 for unappeared words only, respectively).

Table 3: Result of the Experiment by Each Algorithm

		Subject ID										Sum
		255	256	257	258	259	261	262	263	264	265	
Algorithms	None	95.0	111.0	140.0	138.0	131.8	127.0	135.0	153.0	134.0	134.0	1298.8
	A-1	139.9	142.1	141.4	147.5	198.3	152.9	140.7	152.5	132.6	134.2	1482.1
	A-2	89.0	108.0	135.0	132.0	120.0	106.0	132.0	152.0	130.0	148.0	1252.0
	B	137.4	140.0	141.6	149.2	196.7	153.7	139.6	151.9	131.0	134.5	1475.6
	C-1	111.0	126.0	143.0	141.0	138.8	135.0	140.0	160.0	131.0	131.0	1356.8
	C-2	137.2	140.9	145.4	147.5	181.2	150.9	143.5	155.4	129.0	134.2	1465.1
	D	136.0	139.4	142.9	148.7	192.9	152.9	140.5	152.6	131.6	135.1	1472.7
	E	129.2	136.8	149.1	139.9	144.0	138.7	140.4	153.9	142.7	137.1	1411.8

Table 4: Result of the Experiment by Each Algorithm (for Unappeared Words)

		Subject ID										Sum
		255	256	257	258	259	261	262	263	264	265	
Algorithms	None	32.0	56.0	89.0	90.0	28.0	61.0	94.0	108.0	90.0	85.0	733.0
	A-1	76.9	87.1	90.4	99.5	94.5	86.9	99.7	107.5	88.6	85.2	916.3
	A-2	26.0	53.0	84.0	84.0	16.3	40.0	91.0	107.0	86.0	99.0	686.3
	B	74.4	85.0	90.6	101.2	93.0	87.7	98.6	106.9	87.0	85.5	909.8
	C-1	48.0	71.0	92.0	93.0	35.0	69.0	99.0	115.0	87.0	82.0	791.0
	C-2	74.2	85.9	94.4	99.5	77.5	84.9	102.5	110.4	85.0	85.2	899.4
	D	73.0	84.4	91.9	100.7	89.2	86.9	99.5	107.6	87.6	86.1	906.9
	E	66.2	81.8	98.1	91.9	40.3	72.7	99.4	108.9	98.7	88.1	846.0
# of un-appeared		164	169	165	179	124	157	181	178	174	169	1660

5.1 Discussion

It can be seen from the first table that the known value of users with id = 255, 256, 259 and 261 has increased greater than others when compared with “None” case when no treatments were made. In fact, user with id = 256 and id = 259 have a high English ability and TOEIC score greater than 750 points, so they are considered to have lots of known English words. Also, the number of coincidences was increased when the algorithm was generated, especially for Algorithms A-1, B, C-2, and D. In the algorithm of this experiment, for users having small numbers of reading, if the English words at level less than or equal to 4 are unknown, the unappeared English words at level less than or equal to 4 are also unknown. In this way, the English ability of users is initially estimated.

6 Conclusion and Research Plan

By using the proposed algorithms, it is possible to improve the number of coincidences to a certain extent, and the improvement of users' English ability can be estimated. However, in comparison to ranking algorithm, the number of coincidences were more numerous when generating randomly. It can be a better labeling algorithm, so the authors would like to further develop it in the future.

In addition, by using the relationship between English words, we are studying a method to judge whether an English word is known or unknown. For example, it is possible for users who know the meaning of "visible" to know the meaning of "invisible," which has prefix "in." If we can understand the relationship between English words and usage, it will be useful to evaluate the result.

One of the challenges is to clarify the definition of "known." Since words may have various meanings, to decide whether users only know the main or rare meaning is not defined by REX and that depends on users. Therefore, it is necessary to clarify the definition of "known," in other words, to separate the vocabulary list for each meaning of the same English word.

Acknowledgement

We owe special thanks to Huynh Nguyen Tra My (an ex-student from Graduate School of Integrated Science and Technology, Shizuoka University) for her help in the proofreading process of the manuscript with suggestions [14].

References

- [1] R. Flesch, "A new readability yardstick," *Journal of Applied Psychology*, vol. 32, 1948, pp. 221-233.
- [2] E. Dale and J. Chall, "A formula for predicting reliability: instructions," *Educational Research Bulletin*, vol. 27, 1948, pp. 37-54.
- [3] K. Norizuki, "In search of new dimensions for readability for Japanese learners of English," *Bulletin of Shizuoka Sangyo University*, vol. 6, 2004, pp. 167-180.
- [4] H. Kiyokawa, "Development of a readability formula for high school and university students," *Studies of Language and Literature, Wayo Women's University*, 1988, vol. 21, pp. 43-63.
- [5] Y. Miyagishi, Y. Miyazaki, Y. Hasegawa, and T. Oshiro, "Effect of Continuous Language Learning and Personalization with the Use of Gamification," *77th IPSJ Annual Conference*, 2015, pp. (4)-847-848.
- [6] Y. Ehara, T. Ninomiya, N. Shimizu, and H. Nakagawa, "A reading support tool with prediction capability and its extension to readability measurement," *IPSJ 2010-NL-196(18)*, 2010, pp. 1-7.

- [7] M. Fukui and T. Ozawa, "Methodology for Developing an English Readability Index with Word and Idiom Difficulty Variables," *Journal of Japan Society of Educational Information*, vol. 24, no. 3, 2009, pp. 15-22.
- [8] A.L. Uitdenbogerd, "Web readability and computer-assisted language learning," *Proc. of the 2006 Australasian Language Technology Workshop (ALTW2006)*, 2006, pp. 99-106.
- [9] Y. Miyazaki, T. Oshiro, and Y. Hasegawa, "Strategy toward Personalization with REX – Extracting Articles Suitable for Individual Learners," *52nd LET Annual Conference*, 2012, OP-46.
- [10] SVL12000: <http://www.alc.co.jp/vocgram/article/svl/> (date of access: May 20th, 2019)
- [11] T. Oshiro and Y. Miyazaki, "Proposing a Method to Personalize Readability Equations with REX - Development of a Parameter Estimation Method Suitable for Individual Learners and Its Effects –," *LET Chubu Bulletin*, vol.24, 2013, pp.25-34.
- [12] TreeTagger: <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTree-Tagger.html> (date of access: May 20th, 2019)
- [13] Weblio Dictionary: <http://ejje.weblio.jp/> (date of access: May 20th, 2019)
- [14] H. N. Tra My, S. Suzuki, Y. Miyazaki, *Building Personalized Readability Equation and Personalized English Vocabulary List for Continued Study*, *Proceedings of 6th IIAI International Congress on Advanced Applied Informatics, Learning Technologies and Learning Environments (LTLE2017)*, 2017, pp. 789-793.