Assigning CEFR-J Levels to English Texts Based on Textual Features

Satoru Uchida, Masashi Negishi
Kyushu University, Tokyo University of Foreign Studies
744 Motooka Nishi-ku Fukuoka Japan, 3-11-1 Asahi-cho Fuchu-shi Tokyo Japan

Abstract

The present study attempts to assign CEFR-J levels (Pre-A1 to C2) to English texts based on textural features. Based on a coursebook corpus which consists of EFL/ESL English textbooks that claim to be based on CEFR, four textual indexes are calculated. The indexes are ARI (a readability measure), VperSent (an average number of verbs included in each sentence), AvrDiff (the average of word difficulties) and BperA (the ratio of B level content words to A level content words). Regression models are created for each index to predict the level of the input text which is then implemented as an online application called CVLA (CEFR-based Vocabulary Level Analyzer). To show how CVLA works, experiments are conducted using major English language ability tests.

Keywords: CEFR-J, text level, readability

1. Introduction

CEFR (Common European Framework of Reference for Languages) has been widely used as a guideline for assessing the language ability of learners, originally across Europe in reaction to plurilingualism, but currently also employed in the context of EFL and ESL countries such as Japan for educational purposes. Against this backdrop, lists of English grammatical items and words for each CEFR level (A1-C2) have been created (e.g., English Grammar Profile [1] and CEFR-J Wordlist [2]). However, a limited number of attempts have been made to assign CEFR(-J) levels to English passages, and hence teachers and learners are uncertain about the difficulty of the reading and listening materials they encounter.

This study attempts to build a system that assigns CEFR-J levels (Pre-A1 to C2) to English texts based on textual features calculated from the input text. Using four regression models built on the textual features related to sentence structure and vocabulary, our system estimates the level of English passages. One of the characteristics of our approach is that the system is purely data-driven based on the corpusbook corpus we created for our project. Also, it provides 12 levels in CEFR-J scale for the input text, which are more precise than other online systems currently available.

The rest of this paper is organized as follows: Section 2 reviews related works including Flesch-Kincaid Grade Level, Coh-Matrix, Lexile® Measures, Text Inspector, and Reading Level Text Tool (for Dutch). Section 3 provides the description of the framework of CVLA and Section 4 shows the results of experiments using National Center Test for University Admissions, the trial version of the National Standardized Test for University Admissions, and official practice tests of TOEFL-iBT. Section 5 covers our conclusion and future tasks.

2. Related Works

The readability of a text has attracted academic attention and several readability scores have been proposed. One of the most widely-used scores is the Flesch Reading Ease, which is calculated using a regression-like formula. Since readability is the one of the important factors in education, the Flesch-Kincaid Grade Level was invented based on Flesch Reading Ease, assigning a score of a U.S. grade level. The following is the formula for the Flesch-Kincaid Grade Level (cf. Kincaid et al., 1975):

$$0.39\left(\frac{total\ words}{total\ sentences}\right) + 11.8\left(\frac{total\ syllables}{total\ words}\right) - 15.59$$

This index basically indicates the complexity of the text based on sentence length and word complexity but does not consider word levels (e.g. “cat” (A1 in the CEFR-J Wordlist) and “paw” (B2) are both one-syllable words but the difficulty is not the same). Also, it does not provide links to CEFR levels.

Coh-Metrix (McNamara et al., 2010, Graesser, McNamara, and Kulikowich, 2011) is a computational tool that provides linguistic and discourse indices of a text. It produces as many as 108 indexes including lexical diversity, referential cohesion, syntactic complexity and readability. Although some attempts have been made to use the indices for educational purposes (e.g. Crossley, Salsbury, and McNamara, 2012), it is still an open question how CEFR levels are related to those indices calculated by Coh-Metrix.

A noteworthy attempt to assess textual levels is Lexile® Measures, which consist of the Lexile reader measure and the Lexile text measure. The former measure is for the learners of English and assesses his/her reading ability in as a Lexile Scale. The latter scale is for evaluating text levels using textual information such as sentence length and word frequency, which can be measured by using the Lexile Analyzer®[3]. Lexile reader and text measures are correlated; if a learner has 1000 Lexile Scale, he/she is likely to understand 75% of the text in 1000 Lexile text measure. These measures are surely useful for learners and teachers of English, but they are not linked to CEFR levels and how each measure is calculated is left unexplained.

[1] http://www.englishprofile.org/wordlists

[2] The CEFR-J Wordlist Version 1.3. Compiled by Yukio Tono, Tokyo University of Foreign Studies. Retrieved from http://www.cefr-j.org/download.html

[3] https://la-tools.lexile.com/free-analyze/

Text Inspector (cf. Savile, 2012) is another attempt that is relevant in this context. It is an online text level analyzer based on English Vocabulary Profile (EVP). It offers key statistics of a text such as lexical diversity, lexical distributions based on EVP and wordlists from major corpora, and metadiscourse information. In addition, the SCORECARD section (although only available for subscribers) provides the estimated CEFR levels of the input text selectively using the indices mentioned above. However, the calculation process is not open, and it does not provide detailed CEFR levels (i.e. CEFR-J levels) although it includes original D (academic) levels.

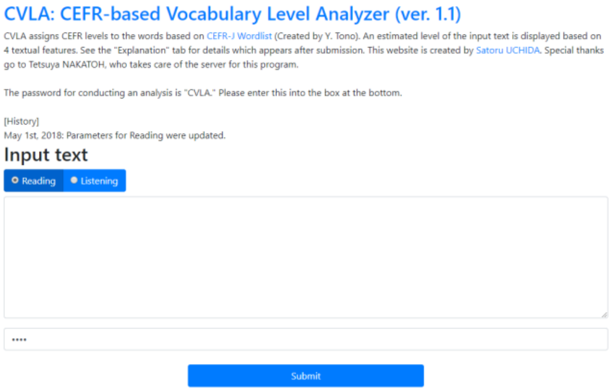
Finally, a Dutch text level analyzer based on CEFR-levels created by Velleman and van der Gees (2014) should be noted here (although it is not for English). Their system, which is called Reading Level Text Tool, employs the average number of simple words, average number of words per sentence, average number of pronouns in a sentence, average number of syllables per word, average number of prepositions per sentence, and the number of names and terms for calculating 10 scale scores, which are the basis for the text level judgement. This system is used for making easy-to-understand documents issued by governmental agencies and is proved its usefulness.

In summary, although there have been several ambitious attempts to investigate text levels, some lack explanation for the assessment and others lack links to CEFR levels. Without explanation for the estimation of the text level, it is difficult to adjust the level of the input text. For example, if a teacher wants to make a B1-level text into A2 level, he/she would not know what changes to make to lower the level of the text. Also, to make the system fit for Japanese educational situation, CEFR-J levels should be employed for level judgment since more than 80% of Japanese learners of English belong to A level (Negishi, Takada, and Tono, 2013) and hence more detailed description is needed for low levels.

3. Framework of CVLA

CVLA (CEFR-based Vocabulary Level Analyzer) is a free online system[4] that assigns CEFR-J levels (see 3.2 for details) to a text. With a very simple interface (Figure 1), users can choose from "Reading" and "Listening" modes (see 3.5 for differences) for their own analysis. The program is written in Perl and TreeTagger[5] is used for part-of-speech tagging.

This section elaborates on how CVLA is constructed. First, we will briefly overview the coursebook corpus which is used for building regression models, followed by a brief explanation of CEFR-J and CEFR-J Wordlist. Then, four text features (ARI, VperSent, AvrDiff, and BperA) used in CVLA are introduced. Average scores of each feature is then calculated with respect to the subcorpora of coursebook corpus, which are in turn used to build regression models. Finally, the result of a sample analysis is demonstrated.



CVLA: CEFR-based Vocabulary Level Analyzer (ver. 1.1)

CVLA assigns CEFR levels to the words based on CEFR-J Wordlist (Created by Y. Tono). An estimated level of the input text is displayed based on 4 textual features. See the "Explanation" tab for details which appears after submission. This website is created by Satoru UCHIDA. Special thanks go to Tetsuya NAKATOH, who takes care of the server for this program.

The password for conducting an analysis is "CVLA." Please enter this into the box at the bottom.

[History]
May 1st, 2018: Parameters for Reading were updated.

Input text

Figure 1: The front page of CVLA

3.1 Coursebook Corpus

To examine textual and grammatical features of each CEFR level and to find criterial features that distinguish CEFR levels, a coursebook corpus was created (a project supported by JSPS KAKENHI JP24242017). This corpus consists of five subcorpora (A1, A2, B1, B2 and C) using EFL/ESL textbooks which were created under CEFR framework[6]. Also, each file is marked with "skill" labels such as Reading, Listening, Writing, and Speaking.

Table 1 is a summary of the coursebook corpus with the information of Reading and Listening sections (C level is excluded since it is not used in this study).

CEFR	# of textbooks	# of words	Reading	Listening
A1	17	164,585	51,455	9,370
A2	21	278,750	103,417	21,503
B1	26	486,787	234,982	37,747
B2	23	582,763	248,173	42,516
Total	87	1,512,885	638,027	111,136

Table 1 : An overview of the coursebook corpus

3.2 CEFR-J and CEFR-J Wordlist

CEFR-J refers to the adapted version of CEFR, which is specially tailored to English education in Japan. The CAN DO lists in five skills (listening, reading, spoken interaction, spoken production, and writing) are the main components accompanied by a wordlist called the CEFR-J Wordlist. Based on careful investigations, the original CEFR levels are further divided into 12 categories: Pre-A1, A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2, C1, and C2. This is mainly to accommodate the needs of Japanese users, most of whom belong to A level. Negishi, Takada, and Tono (2013) show that more than 80% of Japanese learners of English are at A level. This means that the six levels in CEFR (A1, A2, B1, B2, C1 and C2) are too broad to assess the ability of learners in Japan.

[4] http://dd.kyushu-u.ac.jp/~uchida/cvla.html

[5] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

[6] Since there are few C level textbooks, C1 and C2 are combined as one category (C). See Uchida (2015) for more information about the coursebook corpus.

The CEFR-J Wordlist is also created for the use in the context of Japanese English education. Using corpora of Japanese English textbooks, entrance examinations, EVP etc., words are selected and assigned four levels (A1, A2, B1, and B2) with part of speech (see Tono (ed.) 2013 for more information). The list contains 7815 words (A1: 1165, A2: 1416, B1: 2451, B2: 2783) and this information is utilized in CVLA (e.g. ability.n (A2), abroad.adv (B1), abolish.v (B2)).

3.3 Textual Features

In order to assign CEFR-J levels to a text, four textual features (ARI, VperSent, AvrDiff, and BperA) are employed in CVLA. These features can be further divided into sentence features (ARI and VperSent) and vocabulary features (AvrDiff and BperA).

ARI (Automated Readability Index) is a readability index which is calculated using the following formula (Senter and Smith, 1967):

$$4.71\left(\frac{characters}{words}\right)+0.5\left(\frac{words}{sentences}\right)-21.43$$

This index does not utilize the number of syllables unlike the Flesch-Kincaid grade level, which makes it easier for a computer program to calculate. As the formula shows, ARI is sensitive to sentence and word lengths. According to Kincaid et al. (1975), the correlation between ARI and Flesch Reading Ease is 0.87. If this index is higher than expected, users can lower the text level by separating sentences or using shorter words.

VperSent, which stands for "Verbs per Sentence", is an average rate of verbs included in each sentence. If this index is high, users can lower the text level by using simpler constructions (e.g. by avoiding passive, gerund, and past particles). For example, the sentence "The article was written by a scientist," which includes two verb elements (*was* and *written*) can be changed into "A scientist wrote the article." By such a treatment, the score of VperSent can be lowered.

AvrDiff shows the average of word difficulties when A1 is 1, A2 is 2, B1 is 3, and B2 is 4. Word levels are determined based on CEFR-J Wordlist (hence the system does not consider C1 and C2 words). Functional words are excluded from the calculation. If this index is high, users can lower the text level by replacing higher level words with easier ones (e.g. "inform" (B1) -> "tell" (A1)).

BperA signifies the ratio of B level content words against A level content words (nouns, verbs, adjectives, and adverbs). If this index is high, users can lower the text level by avoiding B level words or using fewer B level words.

One of the characteristics of these features is that the score is stable across various types of texts since each score uses general textual features such as number of words and levels of content words. Another advantage is that they are easy to understand and provide clear direction on how to make the text easier or harder. This would be helpful in adjusting the difficulty of a text to the target level.

3.4 Average Scores of Each CEFR Levels

Using the subcorpora of each CEFR level (A1, A2, B1, and B2) of the coursebook corpus, the four textual features were calculated using Reading and Listening sections respectively. The results are shown in Table 2 and Table 3. It is clear from these tables that the scores of the Listening section is generally lower than that of the Reading section. This implies that the texts used for listening are simpler and easier in terms of sentence construction and vocabulary level.

Reading	ARI	VperSent	AvrDiff	BperA
A1	5.73	1.49	1.31	0.08
A2	7.03	1.82	1.41	0.12
B1	10.00	2.37	1.57	0.18
B2	12.33	2.88	1.71	0.26

Table 2: Average scores of reading section

Listening	ARI	VperSent	AvrDiff	BperA
A1	4.52	1.44	1.27	0.06
A2	6.56	2.05	1.41	0.11
B1	7.99	2.35	1.50	0.13
B2	11.72	3.16	1.67	0.19

Table 3: Average scores of listening section

3.5 Regression models

Based on the scores above, regression models were built for each feature respectively for Reading and Listening when A1=1, A2=2, B1=3, and B2=4. The lm function of R (ver. 3.5.1) was used for this purpose. The following are the formulae:

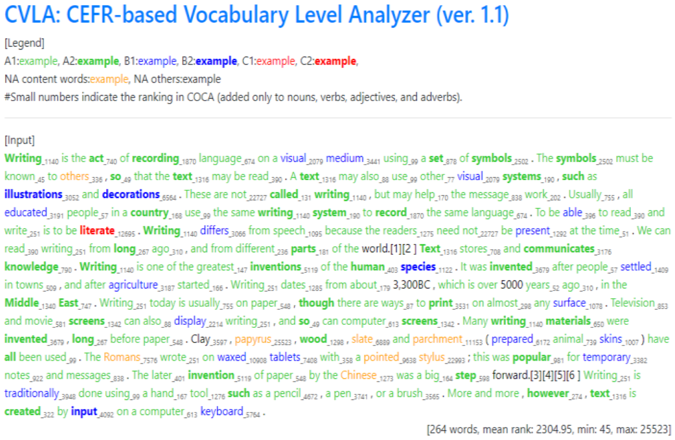
[Reading]
$lm_ARI=0.4298*$ARI-1.27085
$lm_VperSent=2.1075*$VperSent-2.01
$lm_AvrDiff=7.2961*$AvrDiff-8.4442
$lm_BperA=16.3043*$BperA-0.1087

[Listening]
$lm_ARI=0.41621*$ARI-0.70358
$lm_VperSent=1.7751*$VperSent-1.4942
$lm_AvrDiff=7.6131*$AvrDiff-8.6299
$lm_BperA=23.9936*$BperA-0.4847

In order to obtain CEFR-J levels, each score is converted into 12 levels using the following criteria: Pre-A1 (<0.5), A1.1 (<0.84), A1.2 (<1.17), A1.3 (<1.5), A2.1 (<2), A2.2 (<2.5), B1.1 (<3), B1.2 (<3.5), B2.1 (<4), B2.2 (<4.5), C1 (<5), and C2 (<6). The final text level is judged using the average CEFR level of each textual features. For example, when $lm_ARI=1.1 (A1.1), $lm_VperSent=1.5 (A1.2), $lm_AvrDiff=2.1 (A2.2), and $lm_BperA=2.6 (B1.1), the final estimated level is 1.825 (A2.1).

3.6 An example of Analysis

For a sample analysis, the entry of “writing” in Simple English Wikipedia[7] was employed and the passage (264 words) was analyzed using the “Reading” mode of CVLA. The results are shown in Figure 2 and Table 4.



CVLA: CEFR-based Vocabulary Level Analyzer (ver. 1.1)

[Legend]
A1:example, A2:**example**, B1:example, B2:**example**, C1:example, C2:**example**,
NA content words:example, NA others:example
#Small numbers indicate the ranking in COCA (added only to nouns, verbs, adjectives, and adverbs).

[Input]

Writing$_{1140}$ is the **act**$_{740}$ of **recording**$_{1870}$ language$_{674}$ on a visual$_{2079}$ medium$_{3441}$ using$_{99}$ a **set**$_{878}$ of **symbols**$_{2502}$. The **symbols**$_{2502}$ must be known$_{45}$ to others$_{336}$, **so**$_{49}$ that the **text**$_{1316}$ may be read$_{390}$. A **text**$_{1316}$ may also$_{88}$ use$_{99}$ other$_{77}$ visual$_{2079}$ **systems**$_{190}$, **such** as **illustrations**$_{3052}$ and **decorations**$_{6564}$. These are not$_{22727}$ **called**$_{131}$ **writing**$_{1140}$, but may help$_{170}$ the message$_{838}$ work$_{202}$. Usually$_{755}$, all educated$_{3191}$ people$_{57}$ in a **country**$_{168}$ use$_{99}$ the same **writing**$_{1140}$ **system**$_{190}$ to **record**$_{1870}$ the same language$_{674}$. To be able$_{396}$ to read$_{390}$ and write$_{251}$ is to be **literate**$_{12695}$. **Writing**$_{1140}$ differs$_{3066}$ from speech$_{1095}$ because the readers$_{1275}$ need not$_{22727}$ be present$_{1292}$ at the time$_{51}$. We can read$_{390}$ writing$_{251}$ from **long**$_{267}$ ago$_{310}$, and from different$_{236}$ **parts**$_{181}$ of the world.[1][2] **Text**$_{1316}$ stores$_{708}$ and **communicates**$_{3176}$ **knowledge**$_{790}$. **Writing**$_{1140}$ is one of the greatest$_{147}$ **inventions**$_{5119}$ of the **human**$_{403}$ **species**$_{1122}$. It was **invented**$_{3679}$ after people$_{57}$ settled$_{1409}$ in towns$_{509}$, and after agriculture$_{3187}$ started$_{166}$. Writing$_{251}$ dates$_{1285}$ from about$_{179}$ 3,300BC , which is over 5000 years$_{52}$ ago$_{310}$, in the **Middle**$_{1340}$ **East**$_{747}$. Writing$_{251}$ today is usually$_{755}$ on paper$_{548}$, **though** there are ways$_{87}$ to **print**$_{3531}$ on almost$_{298}$ any surface$_{1078}$. Television$_{853}$ and movie$_{581}$ **screens**$_{1342}$ can also$_{88}$ display$_{2214}$ writing$_{251}$, and **so**$_{49}$ can computer$_{613}$ **screens**$_{1342}$. Many **writing**$_{1140}$ **materials**$_{650}$ were **invented**$_{3679}$, **long**$_{267}$ before paper$_{548}$. Clay$_{3597}$, papyrus$_{25523}$, **wood**$_{1298}$, slate$_{6889}$ and parchment$_{11153}$ (prepared$_{6172}$ animal$_{739}$ skins$_{1007}$) have **all** been used$_{99}$. The Romans$_{7576}$ wrote$_{251}$ on waxed$_{10908}$ tablets$_{7408}$ with$_{358}$ a pointed$_{9638}$ stylus$_{22993}$; this was **popular**$_{981}$ for temporary$_{3382}$ notes$_{922}$ and messages$_{838}$. The later$_{401}$ **invention**$_{5119}$ of paper$_{548}$ by the Chinese$_{1273}$ was a big$_{164}$ **step**$_{598}$ forward.[3][4][5][6] Writing$_{251}$ is traditionally$_{3948}$ done using$_{99}$ a hand$_{167}$ tool$_{1276}$ **such** as a pencil$_{4672}$, a pen$_{3741}$, or a brush$_{3565}$. More and more , **however**$_{274}$, **text**$_{1316}$ is **created**$_{322}$ by **input**$_{4092}$ on a computer$_{613}$ keyboard$_{5764}$.

[264 words, mean rank: 2304.95, min: 45, max: 25523]

Figure 2: An example of CVLA analysis

CEFR	ARI	VperSent	AvrDiff	BperA
A1	5.73	1.49	1.31	0.08
A2	7.03	1.82	1.41	0.12
B1	10.00	2.37	1.57	0.18
B2	12.33	2.88	1.71	0.26
Input	7.72	2.89	1.68	0.19
Estimated level	A2.2	B2.2	B2.1	B1.1

Table 4: Estimated levels of each feature

CVLA colors each word according to CEFR levels in CEFR-J Wordlist and EVP (only for C level words): A, B, and C level words are shown in green, blue, and red respectively. The bold face indicates that the words belong to the upper level, that is, they are A2 (green bold face), B2 (blue bold face), and C2 (red bold face). The table indicates raw scores of each textual feature and the estimated levels using the regression models. For the present passage, the final text level is judged as B1.2. One possible way to make this passage easier is to use simpler sentences because the estimated level for VperSent is the highest among these four measures. Also, it might be possible to decrease the text level by lowering AvrDiff score, that is, by using more A1 and A2 level words.

4. Experiments

To show how each textual feature behaves, we conducted experiments using some major English ability tests. The targets are National Center Test for University Admissions (Center), the trial version of the National Standardized Test for University Admissions (Trial), and official practice tests of TOEFL-iBT (iBT). Although the constructions of Center and Trial are different, it is expected that these two tests include English passages of similar level because both tests are intended for the same grade (mainly for the 3rd year high school students) who study under the same curriculum. iBT is a non-Japanese test used mainly for measuring the English language ability of non-native speaker who wish to study in universities where English is used for teaching. Also, iBT is one of the examinations that are available for Japanese high school students to submit to universities from 2020 (some universities already accept iBT scores for proving students’ English ability). Therefore, how iBT differs from Center and Trial is a hot topic for both high school teachers and students in Japan.

We used 9 passages from Center in 2015, 2016, and 2017 (D4A, D5, and D6), 10 passages from Trial (1A, 1B, 2A, 2B, 3A, 3B, 4, 5A, 5B, and 6), and 9 passages from practice tests in the TOEFL-iBT Official Guide (*The Official Guide to the TOEFL Test with DVD-ROM, Fifth Edition*, ISBN: 978-1260011210) for the current analysis.

Table 5 summarizes the averages of each textual feature and Table 6 shows a summary of CEFR-J levels of the English passages of each question in each test. In addition, to visualize the results, sentence features (ARI and VperSent) and vocabulary features (AvrDiff and BperA) are plotted in Figure 3 and Figure 4 respectively. Here, “C” and “Tr” stand for Center and Trial respectively, which are followed by question number in each set.

The results indicate that Trial has the lowest score of the three. This is because A level passages are included in Trial whereas they are not in Center (and iBT). This is one of the major changes from Center, in that all the questions in Trial are intended to check reading comprehension including A1 and A2 items (there are no pronunciation and grammar questions in Trial). This new policy seems to be reflected in the results of CVLA; in Table 6, Trial covers the widest range of CEFR-J levels. Overall, however, the distributions of Center and Trial are very close in both tables and figures showing that high school students can readily accept the Trial test in place of Center.

In addition, it should be noticed that iBT has a different tendency from Center and Trial. Table 5 shows that the average of every score of iBT is the highest of the three and Table 6 indicates that all the passages are judged as high in the CEFR-J scale. Furthermore, vocabulary scores (AvrDiff and BperA) are significantly high as can be seen in Figure 4. This may imply that students need extra preparation for iBT as far as the reading section is concerned (and perhaps for other sections as well).

Test	ARI	VperSent	AvrDiff	BperA	Final
Center	9.41	3.09	1.69	0.25	3.73
Trial	7.63	2.41	1.63	0.22	2.99
iBT	13.66	3.28	2.11	0.59	5.30

Table 5: Average of each score in Center, Trial, and iBT

[7] https://simple.wikipedia.org/wiki/Writing

Level	Center	Trial	iBT
A1.3		1	
A2.1	2	1	
A2.2		2	
B1.1	1	2	
B1.2		1	
B2.1	1	1	
B2.2	2		1
C1	3	2	5
C2			3

Table 6: CEFR-J levels in Center, Trial, and iBT

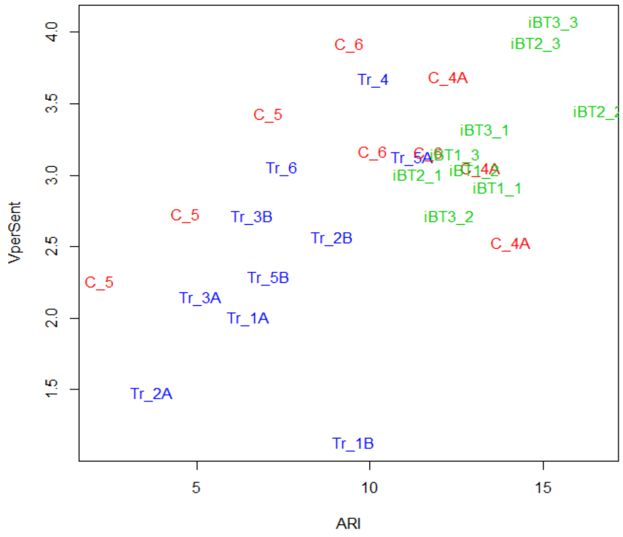


Figure 3: ARI and VperSent

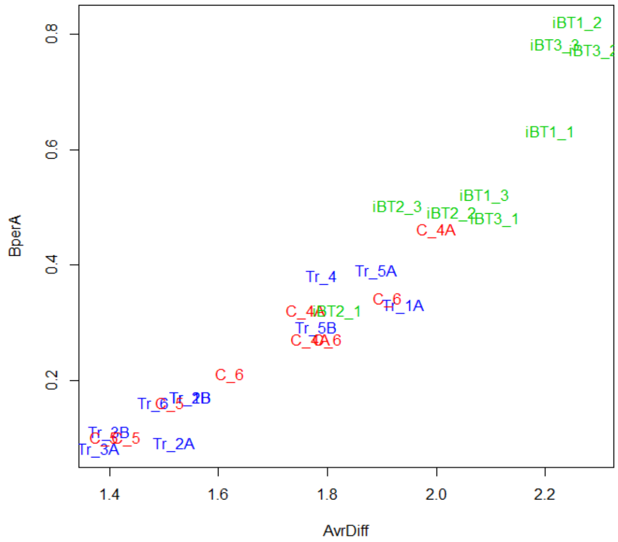


Figure 4: AvrDiff and BperA

5. Conclusion

Using the coursebook corpus, we have created an online system called CVLA that assigns CEFR-J levels to a text. The four textual features used for the judgment are intuitive and can be used for adjusting text levels for teaching purposes. The experiments using reading passages of English ability tests demonstrated the similarities and differences between Center and Trial. At the same time, it revealed the idiosyncrasy of iBT when compared with Center and Trial.

Future tasks in this line of research include using grammatical features of the input text to provide a more accurate estimation of CEFR-J levels. Also, listening sections of each test await further investigation.

6. Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP16H01935.

7. Bibliographical References

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, *29*(2), 243-263.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational researcher*, *40*(5), 223-234.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Research Branch Report 8–75. Chief of Naval Technical Training: Naval Air Station Memphis.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*(4), 292-330.

Negishi, M., Takada, T., & Tono, Y. (2013). A progress report on the development of the CEFR-J. In *Exploring language frameworks: Proceedings of the ALTE Kraków Conference* (pp. 135-163).

Savile, N. (2012). The English Profile: Using learner data to develop the CEFR for English. In Y., Tono, Y. Kawaguchi, M. Minegishi (eds.), *Developmental and crosslinguistic perspectives in learner corpus research* (pp. 17-26). John Benjamins Publishing.

Senter, R. J., & Smith, E. A. (1967). *Automated readability index*. AMRL-TR-6620. Aerospace Medical Division, Wright Patterson AFB, Ohio.

Tono, Y. (ed.) (2013). *The CEFR-J Handbook: A resource book for using CAN-DO descriptors for English language teaching* (original in Japanese). Tokyo: Taishukanshoten.

Uchida, S. (2015). A CEFR-based Textbook Corpus: An attempt to reveal linguistic features of each level (original in Japanese). *English Corpus Studies*, 22, 87-99.

Velleman, E., & van der Geest, T. (2014). Online test tool to determine the CEFR reading comprehension level of text. *Procedia computer science*, *27*, 350-358.