

Identifying the CEFR-J Levels of the Reading Texts Introduced in a Course for *Current English 1 (Reading)*

Aika Miura

Abstract

The study investigates the difficulties of the reading materials covered in an optional course for reading current news articles, and examines the validity of introducing authentic materials without making pedagogical amendments. The Common European Framework of Reference for Languages (CEFR)-based Vocabulary Level Analyzer (CVLA), which analyses the text level according to the four indexes based on the CEFR-J Wordlist, was used to assign the CEFR-J levels to the texts from the two types of course materials: (i) news articles in the assigned textbook and (ii) current online articles regarding the Sustainable Developmental Goals and the COVID-19 pandemic selected by the enrolled students. While the length of the articles was controlled in the edited textbook accompanied by various scaffolding activities to ensure a deep understanding of the materials, no pedagogical adjustments were made to the authentic articles. The quantitative analyses indicated that there were no major differences in the difficulties between these two text groups; most of the texts were assigned as C1 and C2, the most advanced levels in the CEFR-J.

Keywords: Common European Framework of Reference for Languages (CEFR), CEFR-J, CEFR-based Vocabulary Level Analyzer (CVLA), text level

Introduction

This paper investigates the difficulties of the reading texts introduced in a course entitled *Current English 1 (Reading)* in the spring semester of the 2020 academic year. The course targeted 24 sophomore, junior, and senior students at Rikkyo University, and was conducted completely online via Zoom. In this course, along with nine units from the assigned textbook titled, *Meet the World: English Through Newspapers 2020* published by Seibido (Wakaari, 2020), each student was asked to choose a current newspaper article on the Internet, present an oral summary of the article using PowerPoint slides, and submit a written summary at the end of the course. Although the reading materials from the assigned textbook were edited to control the number of words of the text in each unit, and various scaffolds (e.g., Japanese translations for some vocabulary, a short summary with listening activities, and true-or-false quizzes) were provided to help students understand the content, the texts derived from the online news articles selected by the students were authentic and not pedagogically controlled by the teacher. To assess the validity of introducing authentic materials to the course, this study identifies the difficulty level of every text the students encountered in this course using a web-based tool called the Common European Framework of Reference for Languages (CEFR)-based Vocabulary Level Analyzer (ver. 1.1) (CVLA) (Uchida, n.d.; Uchida & Negishi, 2018), and examines the differences between the controlled texts in the published textbook and authentic texts from online news sites.

Course Description

The target course is *Current English (1) Reading*, which was taught by the author in the spring semester of the 2020 academic year. In 2020, Rikkyo University took the special measure to conduct all English courses online due to the COVID-19 pandemic. As such, this course was taught remotely,

using the online video communication tool Zoom. This course was one of the minor (i.e., optional) subjects for sophomore, junior, and senior students, and 24 students were enrolled. Table 1 describes the course objective and contents, which follow the unified syllabus of this course provided by Rikkyo University.

Table 1

Objective and Contents of the Course

| Objective | Contents |
|---|---|
| The aims of this course are for students to read, understand, and then discuss both domestic and international English-language news articles while learning about a variety of topical issues. | This is a low-intermediate English reading course. Students will learn to read and understand English-language news articles, either online or via print media, building on the reading strategies learned in R&W1 ¹ . Students will also build their vocabulary and further enhance the discussion skills learned in their first year while learning about a variety of topical issues, both domestic and global. |

Following the standard course objective and contents provided by the unified syllabus described above, two types of reading materials were given, as follows.

1. The textbook titled, *Meet the World: English Through Newspapers* (Wakaari, 2020) was assigned, as it was a suggested coursebook in the unified syllabus provided by Rikkyo University. Nine out of 20 units were covered. Each unit contains an article published by Jiji Press, and so on, in January 2019. As written in Table A1 in the Appendix, the token, which is the total number of words contained in each text, ranges from 239 to 355.
2. Twenty-four news articles selected from the Internet by the students were shared in class. Each student was asked to select a news article on current issues, especially regarding the Sustainable Developmental Goals (SDGs) and/or COVID-19 pandemic. Twenty-three articles were published between March and July 2020, except for the one released in January 2018, as shown in Table A2 in the Appendix. The main sources of the articles include The Japan Times, BBC, CNN, National Geographic, and NHK. The token of each article differs greatly, ranging from 421 words to more than 2,000 words. The author made a corpus containing these texts (i.e., the *focus* corpus explained below) in the Sketch Engine (Lexical Computing CZ s.r.o., 2020), which is an online text analysis tool. Using this tool, keyword analyses (i.e., identifying individual words appearing more frequently in the *focus* corpus than in the *reference* corpus) were conducted. The English Web Corpus 2015 (enTenTen15), which is a web text corpus containing 13 billion words created in 2015, was used as a reference corpus. The top four single keywords in the *focus* corpus were *coronavirus* (appearing in 17 texts), *lockdown* (in nine texts), *pandemic* (in 15 texts), and *preprint* (in one text), and the top four multi-word keywords were *social distancing* (in 10 texts), *labor shortage* (in two texts), *coronavirus pandemic* (in three texts), and *coronavirus crisis* (in three texts).

The articles from the first group were given for the detailed reading activity with various activities accompanied as mentioned in the Introduction section, while those from the second group were used to let the students have opportunities to read for gist (i.e., skimming). The Results and Discussion

1 “R&W1” is an abbreviation of the course titled “Reading & Writing 1,” which was one of the compulsory courses for first year students at Rikkyo University.

section describes detailed textual features of both groups. Tables A1 and A2 in the Appendix provide the source, headline, token, and date of publication of the articles.

Tables 2 and 3 describe how each unit or authentic article was taught and covered.

Table 2

Standard Lesson Plan Using the Assigned Textbook

| Activity | Details |
|--------------------|---|
| Pre-study at home | <ul style="list-style-type: none"> • The sections from “Before reading 1” (i.e., introduction to the target topic in Japanese) and “Before reading 2” (i.e., keywords introduction in Japanese and English) • Reading a given article • The sections² from “While reading 1” (i.e., taking notes instructed in Japanese), “While reading 2” (i.e., matching the topic of each paragraph with phrases in Japanese), “While reading 3” (i.e., gap filling exercise of the summary of the article), and “While reading 5” (i.e., true-or-false quiz to check the understanding of the content) |
| Class | <ul style="list-style-type: none"> • The teacher gives feedback on the pre-study at home. |
| Post-study at home | <ul style="list-style-type: none"> • Online quiz via Blackboard (i.e., the university’s Learning Management System): The sections from “After reading 1” (i.e., completing sentences by changing the orders of words with Japanese translations) and “After reading 2” (i.e., matching given words and the definitions) |

Note. Wakaari (2020)

Table 3

Standard Lesson Plan Using the Students’ Selected Articles on Current Issues

| Participant | Activities |
|---|---|
| Presenter (once per student) | <ol style="list-style-type: none"> 1. Find an online news article on current issues regarding the SDGs and/or COVID-19, and post the URL on the forum (<i>kejiban</i> in Japanese, or discussion board) on Blackboard, where everyone can share comments with other classmates. 2. Present an oral summary of the article using PowerPoint slides. 3. Submit a written summary (plus their own opinions, if necessary) of the article in more than 450 words by the end of the course. |
| Audience (every class except for when they are the presenters) | <ol style="list-style-type: none"> 1. Scan/browse the selected articles before class. 2. After class, post comments/thoughts/opinions about the presentation on the forum on Blackboard in more than 50 words of English. |
| Presenter & Audience (every class) | The teacher gives a supplementary explanation on the content and vocabulary after the presenter has completed their presentation. |

² The “While reading 4” section contains a listening activity to check the answers for “While reading 5,” which was covered in class.

Preceding the lessons described above, the teacher provided the following lessons as a series of introductory lectures:

1. Lesson 1: Introduce various websites of the world news (e.g., BBC and CNN), news in Japan (e.g., The Japan Times and The Japan News by The Yomiuri Shimbun), and world science news (e.g., National Geographic and Science News for Students), totaling 17 sites, and explain the key concepts and vocabulary of the SDGs, such as *17 goals, five Ps (people, prosperity, planet, peace, and partnership)*, and keywords (e.g., *sustainable, inclusive, and resilient*) (United Nations, n.d.).
2. Lesson 2: Review the various reading skills learned in R&W courses (e.g., previewing, scanning, skimming, and annotating) and introduce a sample article regarding the SDGs and COVID-19 (Solberg & Akufo-Addo, 2020).
3. Lesson 3: Instruct how to give a presentation online using Zoom and review various reading skills (e.g., identifying the main ideas, summarizing, etc.)
4. Lesson 4: Instruct how to write a summary based on the presentation (e.g., writing an essay and formatting)

Background to the Study

The CEFR describes what language learners can do at different stages of their learning, and essentially divides language proficiency into six levels, A1 and A2 (i.e., Basic User), B1 and B2 (i.e., Independent User), and C1 and C2 (i.e., Proficient User), and has been widely used worldwide as a framework for language learning, teaching, and assessment (English Profile, n.d.; Council of Europe, 2020). For anyone involved in English language education, such as material writers, test developers, teachers, and teacher trainers, the English Profile (n.d.) offers online tools, such as the English Vocabulary Profile (EVP) and the English Grammar Profile (EGP), providing information about the CEFR level of words, phrases, idioms, collocations, and grammatical forms. In Japan, the CEFR-J was developed by adapting the CEFR for English language teaching in Japan (Tono, 2013; Tono, 2020; Tono & Negishi, 2020). The A and B levels were subdivided, and the Pre-A1 level was added to the original CEFR as follows: Pre-A1, A1 (A1.1, A1.2, and A1.3), A2 (A2.1 and A2.2), B1 (B1.1 and B1.2), B2 (B2.1 and B2.2), C1, and C2 (Tono, 2013; Tono, 2020). Several resources based on the CEFR-J are available on the website, including the whole CEFR-J package, the CEFR-J Wordlist, and the CEFR-J Grammar Profile. In the present study, the CEFR-J Text Profile, which is an online application tool called CVLA (Uchida, n.d.; Uchida & Negishi, 2018), was used to assign the CEFR-J levels to the reading texts introduced in the course.

Method and Procedure

First, the articles of the assigned textbooks and the ones selected by the students from online news sites were all transformed into TXT files. The pages of the target units from the textbook were scanned using Optical Character Recognition, and only the main articles were manually extracted. Regarding the online articles, the headline, date of publication, name of the author, captions of photos and pictures, and links and headlines of related articles were deliberately excluded by the author, but the subtitles and words inserted in tables and figures were included in the TXT files.

According to Uchida and Negishi (2018), CVLA assigns one of the 12 CEFR-J levels (Pre-A1 to C2) based on four textual indexes calculated from the input text using regression models, concerning

the characteristics of the sentence structure and vocabulary. This system is based on the Corpusbook Corpus compiled by the CEFR-J project, which is composed of data from EFL/ESL textbooks created under the CEFR framework (Uchida & Negishi, 2018; Tono, 2013; Tono, 2020; Tono & Negishi, 2020). While the aforementioned EVP devised by the English Profile only allows the user to identify the CEFR level of the vocabulary, CVLA provides the estimated difficulty of English passages of listening and reading materials.

Figure 1 shows the interface of CVLA. The user simply pastes a text in a space, entering the specified password. Note that the text should not exceed 2,000 words for analysis.

Figure 1

The Interface of CVLA

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" (case-sensitive). Please enter this into the box at the bottom.

You can download [our paper on CVLA](#) presented at APCLC2018@Takamatsu. The following is the bibliographic information:

Uchida, S. and M. Negishi (2018) Assigning CEFR-J levels to English texts based on textual features. In Y. Tono and H. Isahara (eds.) Proceedings of the 4th Asia Pacific Corpus Linguistics Conference (APCLC 2018), pp. 463-467.

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

Input text

Reading Listening

Please enter the password here.

Submit

The CVLA outputs the result in four types of information: (i) text with the colored CEFR-J level assigned to each word³ (see Figure 2); (ii) a table with the result of the estimated text level and scores of the four indexes (see Figure 3); (iii) a bar chart showing the proportion of CEFR-J levels of the content words, such as nouns, verbs, adjectives, and adverbs (see Figure 4); and (iv) a table showing the distribution of the raw frequencies of the content words according to the CEFR-J levels (see Figure 5).

3 EVP is used for C level words (Uchida & Negishi, 2018) as the CEFR-J Wordlist does not contain any C1 or C2 words (Tono, 2020).

Figure 2

A Sample Result of Article No. 4: A Text With the CEFR-J Level Assigned to Each Word

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]
 Every day₉₂ we learn₃₅₇ something₃₃₂₀₅ new₉₁ about Covid-19 -- like how the virus₂₇₂₉ can damage₃₃₆₃ organs₃₆₃₂ **outside**₃₄₂₂, or how some kids₂₅₆ can become₁₄₃ severely₅₂₂₈ **ill**₃₂₄₄. One day₉₂, there 's excitement₃₇₆₁ over a **drug**₄₆₅ like hydroxychloroquine. **Seemingly**₃₇₉₁ the **next**, **several studies**₂₁₄ find₁₀₃ it has no benefit₈₂₃. Adding₃₂₆ to the deluge₁₇₄₁₀ of information₃₀₃ are numerous₂₃₇₈ **studies**₂₁₄, many of which have not₂₂₇₂₇ gone₃₅ through the **standard**₁₈₃₇, **scientific**₁₆₆₇, **vetting**₁₃₈₇₇ process₃₇₅ from **experts**₉₁₄ in the same field₄₅₆ who **evaluate**₂₂₅₇ the **quality**₇₆₅ of a **study**₂₁₄ before it can **publish**₁₂₁₆. This process₃₇₅, **called**₁₃₁ **peer**₂₃₉₆ review₁₃₁₈, helps₁₇₀ **weed**₄₆₆₈ out results₃₄₈ that are **misleading**₉₈₉₉ -- or **wrong**₆₇₇. " I have **colleagues**₁₅₀₆ that have **patients**₅₅₂ come₇₉ into their office₃₅₂ **waving**₂₅₁₀ articles₇₁₇ that they **printed**₃₅₃₁ from a preprint server₆₂₄₆ **promising**₃₇₅₉ ... **miracle**₃₄₉₀ **cures**₄₉₃₇ for Covid, which have never₁₄₀ been studied₉₀₅, could be **dangerous**₁₅₅₉, and that 's a big₁₆₄ **problem**₁₇₄, " said₂₃ Dr. Rita Redberg, a **professor**₈₆₉ of medicine₁₅₃₀ at the **University**₂₇₇ of California, San Francisco and the **editor-in-chief**₁₆₂₂₀ of the journal₁₂₃₉ JAMA Internal Medicine₁₅₃₀. " You have preprints **suggesting**₄₂₈ that things₉₇ worked₁₁₃

Figure 3

A Sample Result of Article No. 4: The Estimated Text Level and Measure of the Four Indexes

Mode: R

| 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 | 10.68 | 3.59 | 1.90 | 0.43 |
| Estimated level | B1.2 | C2 | C1 | C2 |

Estimated Text Level:C1

Figure 4

A Sample Result of Article No. 4 Selected by Student D: A Bar Chart Showing the Proportion of CEFR-J Levels of the Content Words

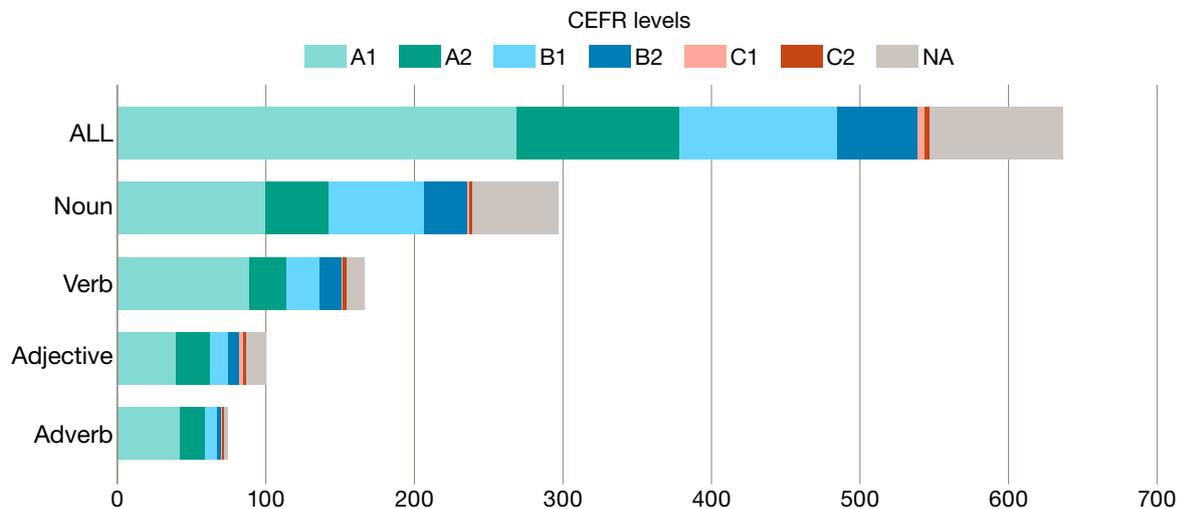


Figure 5

A Sample Result of Article No. 4: The Distribution of Raw Frequencies of Content Words According to the CEFR-J Levels

| POS/CEFR | A1 | A2 | B1 | B2 | C1 | C2 | NA |
|-----------|----|----|----|----|----|----|----|
| Noun | 99 | 43 | 65 | 29 | 1 | 0 | 60 |
| Verb | 89 | 25 | 22 | 15 | 1 | 1 | 13 |
| Adjective | 39 | 24 | 12 | 7 | 2 | 2 | 14 |
| Adverb | 42 | 17 | 8 | 3 | 1 | 0 | 3 |

The estimated text level, shown below the table in Figure 3 (i.e., C1), was determined according to the four indexes of textual features: Automated Readability Index (ARI), VperSent (i.e., verbs per sentence), AvrDiff (i.e., the average of word difficulties), and BperA (i.e., the ratio of B-level content words to A-level content words). The scores of the four indexes are further described in the Results and Discussion section.

The ARI produces an appropriate representation of the US Grade Level (from scores 1 to 14) (Wikipedia, 2018). For example, Article No. 4 selected by Student D in Figure 3 indicates a score of 10.68, which corresponds to a US Grade Level of between the 10th (aged 15 to 16) and 11th (aged 16 to 17) grade. CVLA output the estimated level as B1.2 based on the ARI measure.

The score of VperSent is the average rate of verbs contained in each sentence. A high score for this index means that the target sentences are composed of complex constructions, such as the use of passive tense, gerund, and past participle, and that the level of the text can be lowered using simple constructions (Uchida & Negishi, 2018). For example, Article No. 4 shows a score of 3.59, which was assigned as C2 level.

The AvrDiff index shows the average word difficulties when content words assigned as A1 level are given a score of 1, A2 words are given a score of 2, B1 words score 3 points, and B2 words score 4, based on the CEFR-J Wordlist, which was created in the CEFR-J project and contains 7,815 words in total (Uchida & Negishi, 2018).⁴ The score of Article No. 4 was 1.9, and the estimated level was C1.

The BperA indicates the ratio of B-level content words to A-level content words, and the text level can be lowered using fewer B level words (Uchida & Negishi, 2018). In Figure 3, the score of Article No. 4 was 0.43, which was assigned as C2 level.

Results and Discussion

Overall Results of the Estimated CEFR-J Levels

Of the articles in the assigned textbook, five were identified as C1 level, and four articles as C2 level. Among the collection of online articles selected by the students, one article was assigned as B2.2, nine as C1, and 12 as C2. Since CVLA does not accept texts exceeding 2,000 words, the articles chosen by Students I and L in Table A2 were excluded from the analyses. Figures 6 and 7 show the indexes of ARI, VperSent, AvrDiff, and BperA, as well as the estimated CEFR-J level of the

⁴ Words assigned as C1 or C2 level are regressively estimated.

textbook and online articles. The number of the x -axis indicates the serial number of each article. The scores of the CEFR-J levels were determined according to the calculations where B2.2 received a score of 4.5, C1 scored 5, and C2 scored 6. The ARI scores in both plots fluctuate compared to the other indexes, which is likely because the index is sensitive to sentence and word lengths (Uchida and Negishi, 2018). The ARI measures tend to correspond with the VperSent scores, especially in Figure 6, indicating that readability could be influenced by the structures of sentences. Regarding the AvrDiff and BperA indexes, there were no big differences among all the texts in either group.

Figure 6
Four Indexes and CEFR-J Level of the Articles in the Assigned Textbook

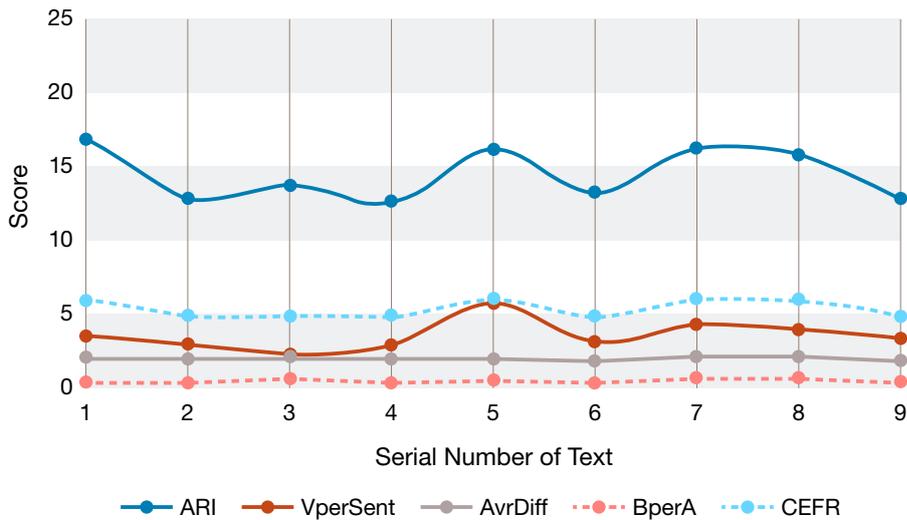
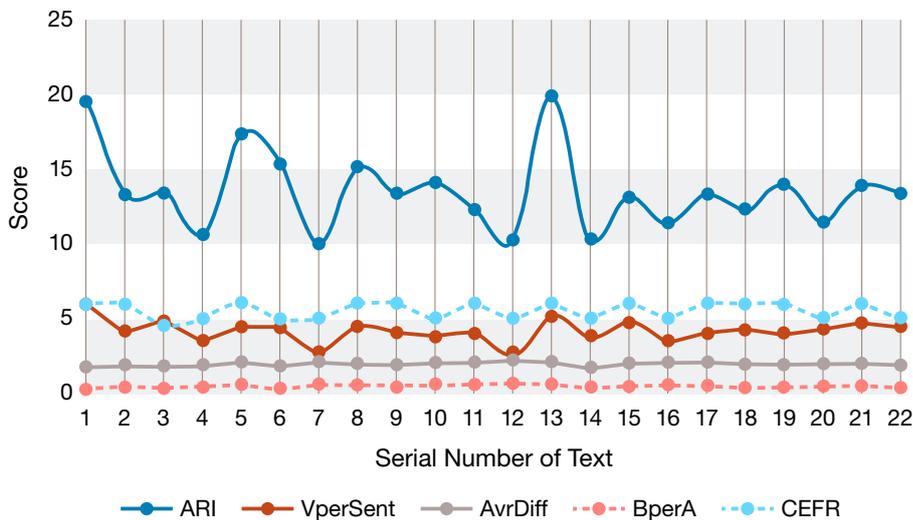


Figure 7
Four Indexes and CEFR-J Level of the Articles Selected by the Students



Comparing Articles No. 3 (with the headline “Virus Forcing Rethink of Japanese Way of Business at Toyota, CEO Says” from The Japan Times, totaling 469 words) and No. 4 (with the headline “Science Speeds Up During Coronavirus Pandemic—But at What Cost?” from CNN, totaling 1,236 words), the estimated CEFR-J levels were B2.2 for No. 3 and C1 for No. 4. The length of No. 3 was approximately

2.6 times shorter than that of No. 4. However, according to the scores of VperSent and ARI, Article No. 3 is likely to be more difficult than Article No. 4; the scores of VperSent and ARI for Article No. 3 were 4.75 and 13.4, respectively, while those of Article No. 4 were 3.59 and 10.68, respectively. In fact, the No. 4 text was written with simpler structures than No. 3, but the content intuitively seemed much more challenging and unfamiliar to the students than that of No. 3. The topic of No. 4 was *preprint servers*, such as *bioRxiv* and *medRxiv*, which was a hot topic in the news of COVID-19, while that of No. 3 was the Japanese corporate culture (e.g., *genchi genbutsu*, or go and see for yourself) influenced by COVID-19.

Distribution of Vocabulary According to the CEFR-J Levels

Figures 8 and 9 show the distribution of vocabulary (i.e., content words, such as nouns, verbs, adjectives, and adverbs) according to the CEFR-J levels of the articles in the assigned textbook and the ones selected by the students. The ratio of A1 vocabulary was the highest, followed by A2, B1, NA (not applicable), B2, C1, and C2 in the texts in both groups. This tendency is evident in the distribution of content words in each part-of-speech category in Figures 10 and 11. Therefore, it can be assumed that the vocabulary assigned as A1, A2, and B1 levels accounted for a major portion of the newspaper articles in general, regardless of the length of texts.

Figure 8

Distribution of Vocabulary According to the CEFR-J Levels of the Articles in the Assigned Textbook

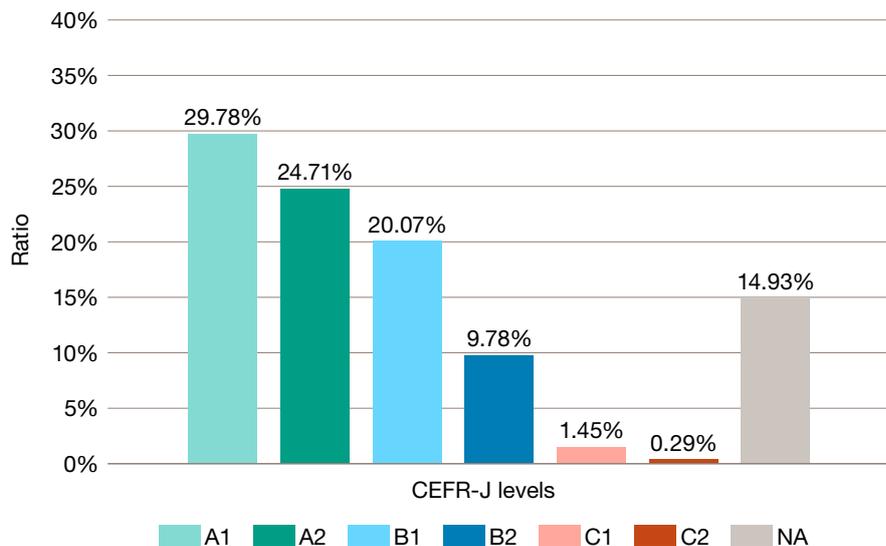


Figure 9

Distribution of Vocabulary According to the CEFR-J Levels of the Articles Selected by the Students

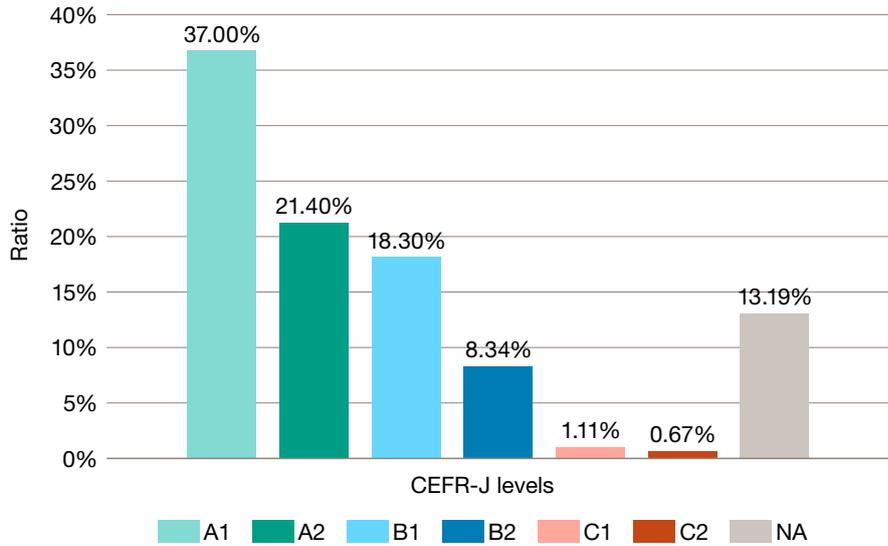


Figure 10

Distribution of Content Words (Nouns, Verbs, Adjectives, and Adverbs) According to the CEFR-J Levels of the Articles in the Assigned Textbook

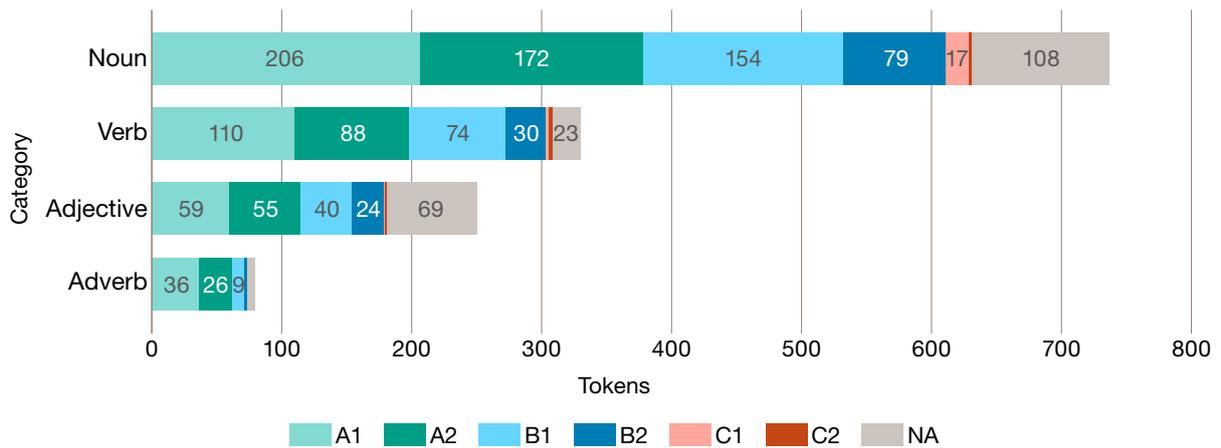
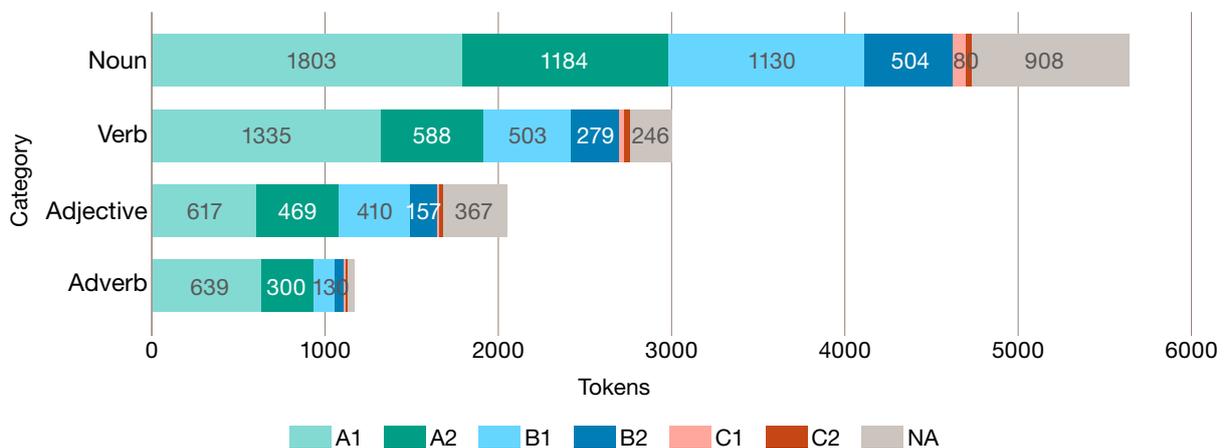


Figure 11

Distribution of Content Words (Nouns, Verbs, Adjectives, and Adverbs) According to the CEFR-J Levels of the Articles Selected by the Students



To examine whether statistically significant differences existed between the articles in the assigned textbook and the ones selected by the students from the Internet in terms of the distribution of the CEFR-J level vocabulary in each category, chi-square tests were conducted. The results of these tests showed that significant differences were evident in the categories of verbs ($\chi^2 = 21.9$, $df = 4$, $p < .001$, Cramer's $V = .05781$) and adjectives ($\chi^2 = 17.18$, $df = 4$, $p < .01$, Cramer's $V = .06156$), except for the C1 and C2 categories, whose expected values were less than five. In terms of verbs, the ratio of A1 vocabulary was higher in the articles chosen by the students (accounting for 44.38%) than that of the assigned textbooks (33.43%), but the vocabulary assigned as A2 and B1 tended to be more frequent in the textbook (26.75% and 22.49%, respectively) than in the articles chosen by the students (19.55% and 16.72%, respectively). NA verbs appeared slightly more often in the students' articles (8.18%) than in the textbook (6.99%). The students' articles also tended to contain more A1 adjectives (30.02%) than the textbook (23.69%), but the ratio of NA adjectives in the students' was lower (17.86%) than that of the textbook (27.71%).

Finally, the most frequently appearing nouns in both groups, which tend to be topic-sensitive,⁵ are described in relation to the assigned CEFR-J levels. The outcome derived from the Sketch Engine indicates that the top six nouns in the assigned textbook were *year* (16 occurrences), *China* (14), *ice* (13), *percent* (13), *hydrogen* (13), and *visitor* (11). The words except for *China* (NA other), *percent* and *hydrogen* (NA content words) were A1 or A2 vocabulary according to CVLA. By contrast, the top seven frequent nouns in the students' articles were *people* (119), *country* (73), *year* (70), *pandemic* (56), *health* (52), *government* (52), and *coronavirus* (50). The words except for *pandemic* and *coronavirus* (NA content words) were all assigned as A1 or A2.

Conclusion

This paper examined the validity of introducing authentic news articles selected by students compared with texts provided in the published textbook by identifying the text levels that CVLA assigned to them: that is, the CEFR-J levels. In the assigned textbook, the number of words in each text was controlled and/or articles of the same length were deliberately chosen for publication. On the other hand, as the texts selected by the students were completely authentic without any amendments made by the teacher, the length of the texts differed greatly. It was assumed that the authentic texts, which normally targeted advanced English-speaking readers, could have been more challenging to the students than the controlled texts in the textbook. However, according to the results retrieved from CVLA, most of the texts were identified as either C1 or C2 level, and the distribution of the CEFR-J levels in the content words of both groups turned out to be very similar according to the results shown in Figures 8 and 9. Between the texts from the assigned textbook and the authentic articles selected by the students, statistically significant differences were only observed in the ratio of verbs and adjectives except for the C1 and C2 vocabulary. The proportion of A1 verbs and adjectives tended to be higher in the authentic articles than in the textbook, as Figures 10 and 11 show.

In conclusion, based upon the quantitative analyses of the textual difficulties identified by CVLA, no major differences between both text groups of newspaper articles were observed in terms of the difficulties according to the assigned CEFR-J levels, even though the majority of the authentic articles selected by the students were published after the start of the COVID-19 pandemic and initially

⁵ The top three frequent verbs in both groups were *be*, *have*, and *say*.

assumed to contain more NA vocabulary than the textbook articles.

From a pedagogical viewpoint, with the author having been a teacher of this course, some of the topics of the authentic articles chosen by the students were challenging. The topic of *preprint servers* was one example as discussed in the Results and Discussion section. As almost most of the articles dealt with topics related to COVID-19, the content tended to be technical and varied, covering topics including business, economy, politics, health care, medical treatment, education, and society, which required certain background knowledge to have a full understanding. Nevertheless, each student was only asked to introduce their chosen article in class, and the other students (i.e., the audience) only needed to browse or skim the article beforehand (which was not compulsory) and to write short comments or state opinions on the presentations they heard afterwards. Doing so should have given the students sufficient opportunities to become familiar with the current topics in relation to the SDGs and the COVID-19 pandemic, and to identify their classmates' individual interests from their selections.

As for future pedagogical implications, teachers could instruct the students to choose articles of a certain length that are appropriate to their proficiency levels, to ensure every student has equal preparation time. The enrolled students in this course were initially instructed to submit a 450-word written summary of the selected article, but a few students chose articles containing less than 450 words. Therefore, a solution was made by instructing them to add their opinions and/or refer to their classmates' comments on the forum as part of their summary.

For future additional analyses, the text level of the students' written summaries as well as their comments on the forum could also be analyzed to examine how they managed to *paraphrase* the information given in the introduced articles in the course, which may represent a *mediation* aspect, the recent addition to the CEFR (Council of Europe, 2018).

References

- Automated Readability Index. (2018, August 23). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Automated_readability_index&oldid=856199533
- Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment: Companion volume with new descriptors*. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Council of Europe. (2020). *The CEFR levels*. <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>
- English Profile. (n.d.). *What is the CEFR?* <https://www.englishprofile.org/the-cefr>
- Lexical Computing CZ s.r.o. (2020). *Sketch Engine*. <https://www.sketchengine.eu/>
- Solberg, E., & Akufo-Addo, N. A. D. (2020, April 23). *Why we cannot lose sight of the Sustainable Development Goals during coronavirus*. World Economic Forum. <https://www.weforum.org/agenda/2020/04/coronavirus-pandemic-effect-sdg-un-progress/>
- Tono, Y. (Ed.). (2013). *Can-do risuto sakusei katsuyo: Eigo toutatsudo shihyou CEFR-J gaido bukku* [CEFR-J guidebook: Making and applying can-do lists]. Taishukan.
- Tono, Y. (2020). *CEFR-J*. <http://cefr-j.org/>
- Tono, Y., & Negishi, M. (Eds.). (2020). *Kyozaï tesuto sakusei no tame no CEFR-J risosu bukku* [CEFR-J resource book to develop materials and tests]. Taishukan.
- Uchida, S. (n.d.). *CVLA: CEFR-based Vocabulary Level Analyzer (ver. 1.1)*. <http://dd.kyushu-u.ac.jp/~uchida/cvla.html>
- Uchida, S., & Negishi, M. (2018). Assigning CEFR-J levels to English texts based on textual features. In Y. Tono & H. Isahara (Eds.), *Proceedings of the 4th Asia Pacific Corpus Linguistics Conference* (pp. 463–468). http://dd.kyushu-u.ac.jp/~uchida/APCLC2018_UchidaNegishi_web.pdf
- United Nations. (n.d.), *Sustainable Development Goals*. <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>
- Wakaari, Y. (2020). *Meet the world: English through newspapers 2020*. Seibido.

Appendix

Table A1

Articles from the Assigned Textbook

| Unit | Specified Source | Headline | Token | Date of Publication |
|------|------------------|---|-------|---------------------|
| 1 | Jiji | Japanese companies in rural areas facing difficulty in hiring graduates | 283 | January 13, 2019 |
| 2 | N/A | Foreign visitors go on record shopping spree | 278 | January 18, 2019 |
| 3 | N/A | Niigata rice exports to China start | 291 | January 9, 2019 |
| 4 | AFP-Jiji | India plans manned space mission by 2021 | 309 | January 13, 2019 |
| 5 | AFP-Jiji | Saudi teenager ‘under the care’ of U.N. agency | 316 | January 8, 2019 |
| 6 | AP | Shenzhen switches to electric cars | 325 | January 9, 2019 |
| 7 | N/A | Frog calls may help improve telecom technology | 239 | January 9, 2019 |
| 8 | N/A | Japan to power fishing boats with Toyota’s hydrogen fuel cells | 346 | January 4, 2019 |
| 9 | N/A | Study: Greenland ice melting four-fold faster than decade ago | 355 | January 26, 2019 |

Table A2

Articles Selected by the Students

| Student | Article No. | Source | Headline | Token | Date of Publication |
|---------|-------------|-----------------|---|-------|---------------------|
| A | 1 | The Japan Times | LDP panel considering five-year transition plan for September school year start | 421 | May 19, 2020 |
| B | 2 | CNN | Cats can infect other cats with coronavirus, researchers find | 477 | May 13, 2020 |
| C | 3 | The Japan Times | Virus forcing rethink of Japanese way of business at Toyota, CEO says | 469 | May 31, 2020 |
| D | 4 | CNN | Science speeds up during coronavirus pandemic—but at what cost? | 1,236 | May 15, 2020 |
| E | 5 | CNN | Hungarian leader’s outrageous power grab | 1,006 | April 3, 2020 |
| F | 6 | The Japan Times | COVID-19 crisis takes toll on children’s cafeterias for disadvantaged | 1,092 | May 8, 2020 |

IDENTIFYING THE CEFR-J LEVELS OF THE READING TEXTS INTRODUCED
IN A COURSE FOR CURRENT ENGLISH 1 (READING)

| Student | Article No. | Source | Headline | Token | Date of Publication |
|---------|-------------|---------------------|---|-----------------|---------------------|
| G | 7 | CNN | Climate change and coronavirus: Five charts about the biggest carbon crash | 1,664 | May 5, 2020 |
| H | 8 | The Japan Times | Europe's broken tourism industry struggles to save the summer | 1,017 | May 16, 2020 |
| I | N/A | Quarts | The coronavirus pandemic is reshaping education | Exceeding 2,000 | March 30, 2020 |
| J | 9 | Time | How South Korea's nightclub outbreak is shining an unwelcome spotlight on the LGBTQ community | 1,346 | May 14, 2020 |
| K | 10 | The ASEAN Post | Hate and discrimination in a pandemic world | 1,276 | May 12, 2020 |
| L | N/A | Daily Mail | More evidence emerges that smokers are protected from coronavirus | Exceeding 2,000 | May 11, 2020 |
| M | 11 | CNN | Coronavirus is causing a flurry of plastic waste. Campaigners fear it may be permanent | 1,184 | May 4, 2020 |
| N | 12 | The Japan Times | COVID-19 versus Japan's culture of collectivism | 856 | May 4, 2020 |
| O | 13 | UN News | UN leads call to protect most vulnerable from mental health crisis during and after COVID-19 | 1,299 | May 14, 2020 |
| P | 14 | National Geographic | Kids are having pandemic dreams too | 1,067 | May 11, 2020 |
| Q | 15 | CNBC | No lockdown here: Sweden defends its more relaxed coronavirus strategy | 1,113 | March 30, 2020 |
| R | 16 | NHK | Coronavirus hits Rohingya refugee camp | 445 | June 4, 2020 |
| S | 17 | The Japan Times | Abuses still abound in labor-strapped Japan's foreign 'trainee' worker system | 1,075 | January 2, 2018 |
| T | 18 | BBC | Coronavirus: How New Zealand relied on science and empathy | 1,381 | April 20, 2020 |
| U | 19 | National Geographic | Your daily commute won't ever be the same | 1,466 | May 11, 2020 |
| V | 20 | BBC | How coronavirus is driving a revolution in travel | 983 | May 16, 2020 |
| W | 21 | BBC | Coronavirus: Will we ever shake hands again? | 1,594 | May 6, 2020 |
| X | 22 | National Geographic | Education interrupted. Years lost. Students face 'cruelty' of new visa policy | 1,369 | July 19, 2020 |