



IJLLALW

International Journal of Language Learning and Applied Linguistics World  
(IJLLALW)

Volume 31 (1), September 2022; 1-15  
EISSN: 2289-2737 & ISSN: 2289-3245

Yujiko Ohashi, et al  
www.ijllalw.org

## VOCABULARY ANALYZER BASED ON CEFR-J WORDLIST FOR SELF-REFLECTION (VACSR): FROM CLASSROOM CORPUS COMPILATION TO SELF-REFLECTION

**Yukiko Ohashi**

*(Yamazaki University of Animal Health Technology y\_watanabe@yamazaki.ac.jp)*

**Noriaki Katagiri**

*(Hokkaido University of Education katagiri.noriaki@a.hokkyodai.ac.jp)*

**Takao Oshikiri**

*(Toita Women's College oshikiri@toita.ac.jp)*

### ABSTRACT

We present a tool to compare the occurrence of vocabulary items included in classroom corpora for each level of the CEFR-J wordlist. Vocabulary Analyzer for Self-Reflection (VACSR) identifies each vocabulary item's frequency in the transcribed texts, indicating the English tokens' distribution on the CEFR-J's four scales as well as unused vocabulary items and those not belonging to either scale. The resulting values will provide a vocabulary benchmark for classrooms, facilitating reflection in teacher training. The study shows the VACSR development procedure, programming design, and how to use it, followed by a pilot study to examine its feasibility. The findings showed three benefits of VACSR: (1) word frequencies in corpora displayed by CEFR-J wordlist level, (2) letting users compare different classes taught by different teachers, and (3) providing resources to adjust classroom speech to match the word levels of teachers' utterances and students' comprehension. Teachers and teacher trainees can reflect on their classroom speech by examining word levels in the spoken corpora found to match students' language levels. VACSR result output tables help teachers match learners' classroom speech levels. This paper addresses multiple appearances of the same headword tokens despite divergent parts-of-speech and has additional implications for researchers.

### KEYWORDS

Vocabulary analyzer, classroom corpus, CEFR-J, teacher training, reflective practice

1

The international journal of language teaching and applied linguistics world  
www.ijllalw.org  
Copyright 2022



IJLLALW



## INTRODUCTION

A wide variety of corpora have been presented so far and have played a large part in developing frequency-based vocabulary lists such as the General Service List (GSL) (West, 1958) and the Academic Word List (AWL) (Coxhead, 1988, 2000) based on abundant corpus evidence. Both are basic vocabulary items to be taught in primary schools, the GSL comprising 2,000-word families and the AWL containing 570. These vocabulary lists, comprised of word families occurring frequently across a wide range of academic fields, reveal basic lexical items to learn in English for academic purposes.

The availability of essential vocabulary lists as educational benchmarks provides awareness of lexical items to teach in the classroom. While language teachers can access linguistic information they are expected to teach in class, related research involving classroom corpus compilation is still scarce. To address the complexity of conducting vocabulary analysis, the present study aims to develop a vocabulary analyzer applicable to classroom corpora, aiming to facilitate the vocabulary analysis procedure for language teachers to reflect on their vocabulary usage in the classroom.

## LITERATURE REVIEW

### *CEFR-J Wordlist*

Since the onset of the Common European Framework of Reference for Languages (CEFR), language educators have adjusted language syllabuses and curriculum guidelines to the CEFR standard, launching new programs and materials. While the CEFR is designed as a comprehensive reference tool to promote educational transparency, applicable in any language learning situation, the application of CEFR has revealed challenges for the countries that teach English as a foreign language (EFL) (Tono, 2013). Japan is one such country, where students learn English as a foreign language. To address challenges that Japanese learners face, CEFR-J (the Common European Framework of Reference–Japan), by Negishi, Takada, and Tono (2013), was created as a Japanese version of the CEFR, for Japanese students learning English (Negishi, Takada & Tono., 2013; Tono, 2012; 2013; Negishi & Tono, 2016).

The establishment of the CEFR-J scale levels, from the primary levels of Pre-A1, A1.1-1.3, A2.1-2.2, to the intermediate levels of B1.1-1.2, and B2.1-2.2, contributed to the development of the CEFR-J wordlist, which has been repeatedly revised since its first version, in 2016, 2019, and 2020. We used the latest version, ver. 1.6, which is downloadable from the website. The CEFR-J wordlist comprises four levels: A1 (1,166-word types), A2 (1,411), B1 (2,445), and B2 (2,779), for a total number of 7,801-word types. According to its part-of-speech (POS), several vocabulary items repeatedly pertain to different levels of POS. The availability of CEFR-J has served as a benchmark to examine and compare the validity of vocabulary usage in a classroom. The application of CEFR-J vocabulary standards to classroom corpora reveals vocabulary usage variation in each level





(Ohashi & Katagiri, 2020), implying that the level and the total number of vocabulary items students are exposed to in a classroom vary depending on class circumstances, because of different teachers or materials.

Applying the CEFR-J vocabulary standards lets us expand vocabulary-related research and find accompanying issues to address. For example, Treffers-Daller and Milton (2013) summarize one of the problems observed in monolingual cultures, stating that monolingual speakers' vocabulary exposure in classrooms is much smaller than expected. Classroom corpora reveal the coverage of both vocabulary items taught in class and neglected ones that teachers are unlikely to use in a classroom. Combined with vocabulary analysis based on certain benchmarks such as the CEFR-J wordlist, creating classroom corpora will provide teachers with the opportunity to reflect on their vocabulary usage, helping them improve their own classes. As the complexity of vocabulary analysis involving corpus compilation may hinder research, this study attempts to develop a vocabulary analyzer complying with the CEFR-J wordlist, with the aim of facilitating corpus-based vocabulary analysis.

#### ***Vocabulary Analysis Tools Developed up to the Present***

To help language teachers become aware of their students' current status as regards vocabulary education, vocabulary analyzers have been developed, for example, AntConc developed by Anthony, (2022). Also, CEFR-J wordlist-oriented analyzers have been presented. The CEFR-based Vocabulary Level Analyzer 2 (CVLA 2) (Uchida & Negishi, 2019) enables us to examine the textual and grammatical features based on the CEFR-J standard. Ohashi, Honda and Katagiri (2021) developed a CEFR-J based vocabulary analyzer combined with a CEFR-J wordlist (CCVA). Along with CCVA, Ohashi et al. (2021) developed a web browser-based tool to help attach tags for classroom corpus compilation: Classroom Corpus Tagger (CCT), which realizes semi-automatic tagging. Table 1 shows the main functions and drawbacks of this tool.





Table I: CEFR-J Based Vocabulary Analysis Tools Developed up to the Present

Developed tools	Main functions	Drawbacks
Vocabulary analyzer combined with a CEFR-J wordlist (CCVA) (Ohashi et al.,	Analyze the levels of lexical items in the file based on the CFRJ-wordlist.	Cannot read the file unless the text is annotated. Cannot analyze plural corpora at the same time.
Classroom Corpus Tagger (CCT) (Ohashi et al., 2022)	Automatically annotate speaker and language tags.	Other than speaker and languages tags, detailed tags should be attached manually.

CCVA can read an annotated classroom corpus compiled by CCT, enabling it to reveal the vocabulary items covered or not covered in classrooms at each level. As the application of CCVA is limited to only one classroom corpus, which is a limitation of CCVA, this study attempted to develop a new tool to realize a comparison among plural classroom corpora, naming it Vocabulary Analyzer Based on CEFR-J Wordlist for Self-Reflection (VACSR).

### METHODOLOGY

This section describes the design of the VACSR Version 1.0 (hereafter "VACSR") this study developed. The VACSR enables comparing multiple classroom corpora. While CCVA (shown in Table I) can read only one corpus for analysis, VACSR can read several corpora at the same time and compare word occurrences and vocabulary levels between different files according to the CEFR-J wordlist. The analysis results are displayed in one column per file, making it easy to compare the differences among multiple files (Table II). Taking "after" in Table 2 as an example, it is easy to find that the word appeared in two files and the total occurrences were thirteen. Furthermore, it appeared nine times in a file compared with four times in another and zero in the rest. The tool can also count words not appearing in the wordlist but are present in the files.





Table II: A Sample of Report Files of CEFR-J A1

	A	B	C	D	E	F	G
	headword	CEFR-J	RANGE	FREQ	Corpus/Text file 1 Freq(2021- 0914_id3.txt)	Corpus/Text file 2 Freq(2021- 0919_id1.txt)	Corpus/Text file 3 Freq(2021- 0928_id2.txt)
1							
2	a	A1	2	29	14	15	0
3	a.m./A.M./am/AM	A1	2	2	1	1	0
4	about	A1	3	12	8	2	2
5	about	A1	3	12	8	2	2
6	action	A1	1	6	6	0	0
7	activity	A1	1	5	5	0	0
8	actor	A1	1	1	1	0	0
9	after	A1	2	13	9	4	0
10	again	A1	1	1	1	0	0

*Note.* Column A is the same as the “headword” column of the “ALL” sheet in the “CEFR-J Wordlist.xlsx” file. Column B is equal to the “CEFR-J” column in the “ALL” sheet in the “CEFR-J Wordlist.xlsx” file. Column C, “RANGE” displays the number of files containing the word. Column D, “FREQ,” displays the total number of words that appeared in all analyzed files. Columns E and the following display the number of words that appear in each analyzed file.

### Report File Output

The order in which the rows appear is the alphabetical order of words that appeared in each CEFR-J word category. Words that did not appear are displayed. For example, Table 2 displays *action* with “A1” in the adjacent horizontal cell under the CEFR-J column head. Since it appeared in the analysis files. Words that do not appear in Level A1 are marked as “(A1),” so it is clear which words did not appear. These words that do not appear are indicated as “0” in the corresponding cell. The same applies to words in “A2”, “(A2)”, “B1”, “(B1)”, “B2”, and “(B2)” levels. Words unregistered in the CEFR-J wordlist are listed as “(other)” under “(B2)” (Table 3).





Table III. A Sample of Report Files of CEFR-J B2 and others

	A	B	C	D	E	F	G
7793	yawn	(B2)	0	0	0	0	0
7794	yawn	(B2)	0	0	0	0	0
7795	yearn	(B2)	0	0	0	0	0
7796	yell	(B2)	0	0	0	0	0
7797	yield	(B2)	0	0	0	0	0
7798	youthful	(B2)	0	0	0	0	0
7799	yummy	(B2)	0	0	0	0	0
7800	zebra	(B2)	0	0	0	0	0
7801	zip	(B2)	0	0	0	0	0
7802	zoom	(B2)	0	0	0	0	0
7803	(audio)	(other)	1	2	0	0	2
7804	(overlapping)	(other)	1	1	0	0	1
7805	(pair	(other)	1	1	0	0	1
7806	(phrase	(other)	1	1	0	0	1
7807	1	(other)	2	18	16	2	0
7808	10;45	(other)	1	2	2	0	0

The CEFR-J contains words that change levels when their part-of-speech changes. For example, the adjective *next* is A1, and A2 for an adverb. Such words are counted in both levels; if *next* occurs twice, VACSR will regard it as a multiple entity, and yield an output “2” in both A1 and A2.

Regarding capitalization of the same word, words in the headword column are counted as case-insensitive. For example, “Okay” and “okay” count as the same word regardless of capitalization. Hyphenation due to line-spacing such as *knowl-edge* will be regarded as “(other)” although knowledge is ranked as Level A2 in the CEFR-J wordlist. As for “well-known,” a hyphen is correctly inserted between “well” and “known.” Therefore, “wellknown” without a hyphen is classified as “(other).”

### How to Run VACSR

As a precondition, it is necessary to name folders and files with single-byte characters (one-byte alphanumeric characters) since they may not work if they contain multibyte characters such as Japanese and Chinese. You can run VACSR (downloadable from URL: [https://drive.google.com/file/d/1CNvykjivw3y3ufrRaxINLWUc\\_bkFhkW8/view?usp=sharing](https://drive.google.com/file/d/1CNvykjivw3y3ufrRaxINLWUc_bkFhkW8/view?usp=sharing))





by following a procedure described below. The procedure differs slightly between Windows and macOS.

- 1) To use, unzip “vacrs.zip” and place the txt or xml file you wish to examine in the folder.

For Windows

- 2) Click “vacrs-win.exe” to create report-year-month-day-hour-minute-second.csv.

For macOS

- 2) Even if a warning about the downloaded file appears, choose “Open” without regard to the warning.
- 3) A black window will appear. “...completed” appears on the screen. Close it.
- 4) In the folder report-year-month-day-hour-minute-second.csv will be created.  
You can run it a second time by simply clicking on “vacrs-mac.”

#### **Notes on VACSR Use**

Once VACSR resolves the file of the classroom corpus, it will output “report.csv.” The analyzable corpus file types are .txt and .xml extensions. VACSR can analyze a maximum of 10 files at a time. The maximum size of one file should be less than 1 MB, which is 1,000,000 single-byte UTF-8 letters. With an average of 5 characters per single byte word, VACSR can support up to 200,000 words.

VACSR reads text file corpus. VACSR can analyze a file with or without tags. If there are <eng> and <s> tags in the file, the English words between them (between <eng>...</eng> and <s>...</s>) are retrieved. In other words, any string between <eng>...</eng> and <s>...</s>, whether in English, in Japanese, or in any other language, is eligible as an English word, even if it is enclosed between “<” and “>”. <eng> and <s> are not counted in the analysis results. In other words, <eng> is skipped and not counted as “(other)”. If there are no tags <eng> and <s> in the file, it is analyzed as plain text. Text enclosed with “<” and “>” other than <eng> and <s> is displayed as “(other)”. The hierarchy of tags can be as deep as 10 layers.

French and German can also be used without being garbled if saved in UTF-8 encoding. If using Windows “Notepad”, it is possible to select the character encoding when saving the file, so if UTF-8 is selected there, the characters will not be garbled. If the file is garbled when opened in Excel and can be opened in UTF-8 without garbled characters.



**Pilot Study**

This section shows the conduct of a sample study by test-running VACSR, and displays sample results and analyses. It describes the objective, materials, i.e., three small classroom spoken corpora, results that VACSR yielded followed by analyses (Section 4.3), and a discussion of the pilot stud.

*Objectives*

The objective of this pilot study was to demonstrate (1) running VACSR with classroom spoken corpus, obtaining results, i.e., reports filed output, (2) analysis samples of the results, and (3) possible implications to self-reflection of instructors of English language.

*Materials and methods*

The pilot study utilized elementary school English classroom corpora used in Katagiri and Ohashi (2018) compiled for their study. They compiled an elementary school English classroom corpus from 16 classes at three different elementary schools. The classroom spoken corpus contained 15,100 tokens deriving from 732 types (headwords). Table 4 shows the corpus profile broken down into three sub-corpora.

Table IV: Summary of the Elementary School English Corpora (Adapted from Katagiri & Ohashi, 2018, p. 69)

Corpora	Word count		
	Tokens	Headword	Token mean
A	1,745	259	436.3
B	8,188	307	2,729.3
C	5,167	496	574.1
$\Sigma$	15,100	732	943.8

Note. Corpora A, B, and C were compiled from three different English teaching types in Japanese elementary schools. Corpus A is made up of grades 5 and 6, where English was taught as a foreign language activity; Corpus B is from grades 2, 3, and 4, where English was taught as a school subject; and Corpus C is from grades 5 and 6, where English taught as a school subject (Adapted from Katagiri & Ohashi, 2018, p. 68).

**RESULTS AND DISCUSSION**

Table V shows the top 20 headwords resulting from running VACSR with an elementary school English corpus (Katagiri & Ohashi, 2018) in the order of combined frequency. The third column (Table V) indicates headwords' CERF-J wordlist levels such as A1 and A2 as Table 2 showed, and



those not in the CEFRJ wordlist shown as “(other)” as in Table III displayed. The RANGE column indicates the coverage of the headwords in the corpora that VACSR received as input in a single running. The frequency column shows the accumulated frequency of each headword appearing in the corpora that VACSR analyzed. The right three columns display each frequency of the headwords in the corpora that VACSR counted.

Table V. Top Twenty Headwords of VACSR Analyzing Three Sub-corpora of Elementary School English Class

Rank <sup>a</sup>	headword	CEFR-J <sup>b</sup>	RANGE <sup>c</sup>	FREQ <sup>d</sup>	Corpus/Text file 1 Freq (Corpus- A.txt)	Corpus/Text file 2 Freq (Corpus- B.txt)	Corpus/Text file 3 Freq (Corpus- C.txt)
1	I	(other)	3	848	12	693	143
2	to	A1	3	673	13	529	131
3	to	A1	3	673	13	529	131
4	you	A1	3	573	52	239	282
5	a	A1	3	515	11	426	78
6	have	A1	3	430	5	408	17
7	want	A1	3	328	2	233	93
8	want	B2	3	328	2	233	93
9	like	A1	3	252	3	194	55
10	like	A1	3	252	3	194	55
11	like	B1	3	252	3	194	55
12	go	A1	3	249	8	191	50
13	go	B1	3	249	8	191	50
14	's	A1	3	232	30	107	95
15	the	A1	3	232	16	154	62
16	in	A1	3	220	6	158	56
17	in	A2	3	220	6	158	56
18	OK	(other)	3	220	15	5	200
19	do	A1	3	190	10	90	90
20	Ah	(other)	3	177	1	67	109

Note. <sup>a</sup>. Though the VACSR result file does not contain the rank column, it is added for clarity.

<sup>b</sup>. CEFR-J = Common European Framework of Language Reference – adjusted for Japanese learners of English. A1 = a level equivalent to third grade up to seventh grade in Japanese school, A2 = a level equivalent to eighth grade to tenth grade in Japanese school, B1 = a level equivalent to eleventh and twelfth grades in Japanese school, B2 = a level equivalent to twelfth grade to a sophomore in a Japanese college, and (other) indicates the headword is not included in the CEFR-J wordlist. <sup>c</sup>. RANGE column numerals indicate the coverage of the headwords appearing in the target corpora. For example, “3” indicates the headword that appears in the three corpora that VACSR analyzed. <sup>d</sup>. FREQ = frequency. Cells in the FREQ column show the total frequency of the headword in all the corpus text files that VACSR analyzed.



be the steps for the trainees to include in their classroom speech relatively frequently when they exhibit the unused words in classroom discourse.

Table VI: Samples of 20- Wordlist-Headwords (in alphabetical order) in A1–B2, not Appearing in the Target Corpora

Category <sup>a</sup>	headword	CEFR-J <sup>b</sup>	RANGE <sup>c</sup>	FREQ <sup>d</sup>	Corpus/ Text file 1 Freq (Corpus- A.txt)	Corpus/ Text file 2 Freq (Corpus- B.txt)	Corpus/ Text file 3 Freq (Corpus- C.txt)
A1	above	(A1)	0	0	0	0	0
	above	(A1)	0	0	0	0	0
	address	(A1)	0	0	0	0	0
	action	(A1)	0	0	0	0	0
	actor	(A1)	0	0	0	0	0
A2	ability	(A2)	0	0	0	0	0
	abroad	(A2)	0	0	0	0	0
	accept	(A2)	0	0	0	0	0
	acceptable	(A2)	0	0	0	0	0
	accident	(A2)	0	0	0	0	0
B1	abandon	(B1)	0	0	0	0	0
	able	(B1)	0	0	0	0	0
	abnormal	(B1)	0	0	0	0	0
	aboard	(B1)	0	0	0	0	0
	aborigine	(B1)	0	0	0	0	0
B2	abandoned	(B2)	0	0	0	0	0
	abnormally	(B2)	0	0	0	0	0
	abolish	(B2)	0	0	0	0	0
	aboriginal	(B2)	0	0	0	0	0
	abruptly	(B2)	0	0	0	0	0

Note. <sup>a</sup>. The category column shows the CEFR-J wordlist levels from A1 (*primary*) to B2 (*advanced*). <sup>b</sup>. CEFR-J = Common European Framework of Language Reference – adjusted for Japanese learners of English. A1 = a level equivalent to third grade up to seventh grade in Japanese school, A2 = a level equivalent to eighth grade to tenth grade in Japanese school, B1 = a level equivalent to eleventh and twelfth grades in Japanese school, B2 = a level equivalent to twelfth grade to a sophomore in a Japanese college, and (*other*) indicates the headword is not included in the CEFR-J wordlist. <sup>c</sup>. RANGE column numerals indicate the coverage of the headwords appearing in the target corpora. Thus “0” indicates the headword did not appear in any of the three corpora analyzed by VACSR. <sup>d</sup>. FREQ = frequency. Since “RANGE” indicates “0,” the frequency is also “0.”



The preceding sub-sections demonstrated the first two objectives of the pilot study by showing how to run VACSR, and reported sample analysis proposals using the results that VACSR yielded. Researchers with different interests might be interested in analyzing varying aspects of classroom corpora regarding the use of VACSR for analyzing multiple classroom corpora based on the CEFR-J wordlist. The final sub-section of the pilot study discusses the third objective of the VACSR pilot study, namely, the possible implication to self-reflection of instructors of English language, followed by a discussion on other analysis angles that the preceding sections did not cover: parts-of-speech (POS) and headwords.

Considering that the corpora (Katagiri & Ohashi, 2018) used in this pilot study were compiled from varying grades—Corpus A from grades 5 and 6, Corpus B from grades 2, 3, and 4, and Corpus C from grades 5 and 6 in Japanese elementary schools—the words appearing in levels A2 and higher constitute a substantial self-reflective practice with suggestions and implications for teachers. Theoretically, the words used in levels A2 and higher would not match the English word level of elementary school students. Thus, elementary school teachers using these corpora can consider lowering the word levels by reflecting on the VACSR results and are expected to reword their classroom English speech.

We must note that the CEFR-J wordlist contains approximately 8,000-word types classified in four levels. These word types are coined with part-of-speech (POS) information, resulting in multiple headwords with identical spelling, for example, Table 5 lists two *aboves* in the headword column. We must also note that headwords of the same word family are not lemmatized” *abandon* (CEFR-J wordlist B1) and *abandoned* (CEFR-J wordlist B2). We can find other combinations of the same headwords listed in different levels, such as modal auxiliary *can* (A1) and noun *can* (A2). Such multiple appearances of the seemingly identical headwords in the VACSR outcome table might decelerate detailed analyses for teacher reflective practice. The following section addresses this issue as future work for updating VACSR.

## CONCLUSION

We discussed the development of CEFR-J-based vocabulary analyzers, which promoted VACSR. This section summarizes the advantages of VACSR with suggestions on how we utilize VACSR for teachers’ reflective practices.

### *VACSR Benefit*

Referring to the findings and examples shown in the preceding section, this study revealed that the current version of VACSR (version1) can showcase at least three significant benefits highlighted below:





1. Reading the transcribed text, VACSR analyzes the level of each lexical item based on the CEFR-J wordlist and automatically counts the occurrences of each item included in the whole text (a part of the analysis results is shown in Table II). The results reveal overused vocabulary items and unused ones according to each level in the CEFR-J wordlist. As VACSR also reveals the unused vocabulary items (Table III), teachers can reflect on lexical items they overused in class as well as those they are likely to neglect.
2. VACSR can read multiple corpus files – text files after a single running, enabling the users to obtain results for teachers as well as teacher trainers and researchers to compare their vocabulary usage across different classes taught by different teachers. Comparing teachers' vocabulary usage across different classes underscores the variation in the level and occurrences of vocabulary items used in each class, leading the language teachers to conduct a reflection practice where they notice how to develop their own classes in terms of vocabulary usage.
3. Considering non-appearing headwords in the VACSR result output table would propose to teachers an additional word repertoire for improving classroom teacher talk and interaction with students. Teachers can adjust their classroom speech to match their students' English language levels, for example, trying to use words from the CEFR-J relatively higher levels for intermediate or advanced learners, and lowering the word levels for primary learners. Multiple results can suggest, for example, that novice teachers learn words uttered by experienced teachers from corpora, observing different lengths of vocabulary usage from theirs analyzed by VACSR.

The pilot study proposed various analysis possibilities using the VACSR result output table. Prospective VACSR users can utilize the above benefits by using them for various purposes. Teachers and teacher trainees can use VACSR for reflective practice, such as to better their classroom speech, and researchers can examine multiple classroom spoken corpora depending on their academic pursuits.

### ***Limitations and Future Work***

The authors are now planning to update VACSR to address the multiple appearance issue that the pilot study has revealed. As VACSR presupposes using plain text files from classroom corpora, either without or with POS tags deleted or hidden from them, the VACSR result table displays multiple appearances of headwords in the same CEFR-J wordlist levels; for example, an adjective *fun* and a noun *fun* in level A1, or in different levels; a noun *lead* (A2) and a verb *lead* (B1). The multiple appearances are attributed to different POS attached to words with identical spelling. VACSR in the future needs to distinguish differing POSs to analyze the text files if the corpora are complete with POS attributes, that is, POS tags, which is usually the case with compiled corpora. The updated VACSR will be able to identify headwords with different POSs. If the classroom





corpora are tagged with POS information, the updated VACSR can output more detailed results in output tables. When the VACSR is equipped with a function of attaching POS attributes to the result output tables, teachers, teacher trainees, and researchers will be able to take more advantage of utilizing the VACSR.

### Note

- 1) The CEFR-J Wordlist Version 1.6. Compiled by Yukio Tono, Tokyo University of Foreign Studies. Retrieved from [http://www.cefr-j.org/download.html#cefrj\\_wordlist](http://www.cefr-j.org/download.html#cefrj_wordlist) on 03/03/2022.

### REFERENCES

- Anthony, L. (2022). AntConc (Version 4.0.11) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238.
- Coxhead, A. (1998). *The development and evaluation of an academic word list* (Master Thesis, Victoria University of Wellington, New Zealand).
- Katagiri, N., & Ohashi, Y. (2018). Instructor lexical analyses of English activities and English language as a subject in Japanese Elementary Schools. *ARELE* 29,65-80. DOI [https://doi.org/10.20581/arele.29.0\\_65](https://doi.org/10.20581/arele.29.0_65)
- McCarthy, M. (2001). Is There a Basic Spoken Vocabulary: Technology and Common Sense. *The Journal of TESOL France*, (8) (C4), 29-36. [https://www.tesol-france.org/uploaded\\_files/files/TESOL%20Vol%208%202001%20C4.pdf](https://www.tesol-france.org/uploaded_files/files/TESOL%20Vol%208%202001%20C4.pdf) (accessed 2022-05-16).
- Negishi, M., Takada, T., & Tono, Y. (2013). A progress report on the development of the CEFR-J. In E.D. Galaczi & C. J. Weir (Eds.), Exploring language frameworks, *Proceedings of the ALTE Krakow Conference, July 2001*, 135-163.
- Negishi, M., & Tono, Y. (2016). An update on the CEFR-J project and its impact on English language education in Japan. *Studies in Language Testing*, 44, 113-133.
- Ohashi, Y., & Katagiri, N. (2020). The Ratios of CEFR-J Vocabulary usage compared with GSL and AWL in elementary EFL classrooms and suggestions of vocabulary items to be taught. *Asia Pacific Journal of Corpus Research*, 1 (1), 61-94.
- Ohashi, Y., Honda, F., & Katagiri, N. (2021). Classroom vocabulary analyzer combined with CEFR-J wordlist (CCVA): tool development to examine vocabulary levels in classroom corpora based on the CEFR-J wordlist. *International Journal of Language Learning and Applied Linguistics World*, 27 (4), 1-12.
- Ohashi, Y., Katagiri, N., & Oshikiri, T. (2021). *Jyugyō corpus kouchiku notameono Jidou tag zuke tool* [Developing Classroom Corpus Tagger: A Spoken Language Tagger to Compile Classroom Corpora]. Proceedings of the JAECs 47th Conference, 43-48.





IJLLALW

International Journal of Language Learning and Applied Linguistics World  
(IJLLALW)

**Volume 31 (1), September 2022; 1-15**

**EISSN: 2289-2737 & ISSN: 2289-3245**

***Yujiko Ohashi, et al***

***www.ijllalw.org***

- Tono, Y. (2012). *Eigo Totatsudo Shihyo CEFR-J guidebook* [The CEFR-J handbook. A resource book for using CAN-DO Descriptors for English language teaching]. Tokyo: Taishukan.
- Treffers-Daller, J., & Milton, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4 (1), 151-172.  
<https://doi.org/10.1515/applirev20130007>
- Tono, Y. (Ed.). (2013). *The CEFR-J handbook*. Tokyo: Taishukan Shoten.
- Uchida, S., & Negishi, M. (2019). Assigning CEFR-J levels to English texts based on textual features. *Proceedings of Asia Pacific Corpus Linguistics Conference 4*, 463-467.  
<http://hdl.handle.net/2324/2244112>
- West, M. (1953). *A general service list of English words*. Longman, London.

WWW.IJLLALW.ORG

