

Comparisons of Word Lists on New Word Level Checker

Atsushi Mizumoto^a, Geoffrey G. Pinchbeck^b, and Stuart McLean^c
^a*Kansai University*; ^b*Carleton University*; ^c*Momoyama Gakuin University*

Abstract

This paper introduces a novel online vocabulary profiling application called the New Word Level Checker (<https://nwlc.pythonanywhere.com/>) and word list resources used by the application. First, the rationale for developing another web vocabulary profiler and the word lists included in the application are described. Next, the lexical units (i.e., how words are counted) and rules (e.g., case sensitivity, contractions, abbreviations with periods, hyphenated words, and compounds) employed in the application are explained. Then, the word lists adopted for the application are compared to show which lists are best used for different purposes. Pedagogical implications of the use of the application and word lists are discussed, especially focusing on matching learners with vocabulary-level appropriate tests.

Keywords: word list, word counting unit, vocabulary profiling, tool development, web application

1 Introduction

New Word Level Checker (NWLC) (<https://nwlc.pythonanywhere.com/>) is a web application for vocabulary profiling. Using the core concept of its predecessor, Word Level Checker (<http://someya-net.com/wlc/>), developed by Someya (2006), NWLC analyzes English texts submitted by users and produces a coverage profile based on the built-in, user-selected word lists.

NWLC is an online application specifically created to provide vocabulary profiling that has been optimized for Japanese English as a Foreign Language (EFL) learning contexts. It was necessary to develop NWLC because word counting methods in the word lists adopted in other vocabulary profiling applications such as VocabProfiler (<https://www.lex tutor.ca/vp/comp/>) are not appropriate for Japanese EFL learners (e.g., McLean, 2018). Furthermore, vocabulary lists tailored for Japanese EFL learners are not included in other profilers. Another feature of NWLC is that it includes a knowledge-based word list, Scale of English Word Knowledge—Japanese (SEWK-J) (see below), in addition to conventional frequency-based word lists. Thus, it is possible, with NWLC, to match learners' actual vocabulary knowledge with texts to be analyzed.

2 Word Lists Included in NWLC

As of September 2021, NWLC features five research-based, trustworthy word lists: SEWK-J, New JACET8000, SVL12000, the New General Service List

(NGSL), and Common European Framework of Reference for Languages (CEFR-J). It is advantageous to have these five word lists because each is built on different principles, and the number of words in the list differs from list to list. Users can select and compare the word lists depending on their purposes.

3 SEWK-J

The SEWK-J was developed to estimate the difficulty that the vocabulary in a text presents to Japanese learners of English. Thus, the SEWK-J list estimates the likelihood that a word is known to Japanese university students. The probability of knowledge of a word is based on a multiple regression performed by Pinchbeck (manuscript in preparation) using vocabulary test data of the 149-item New Vocabulary Levels Test (McLean & Kramer, 2016) administered to Japanese University EFL students as the criterion (dependent) variable. The regression formula includes the following predictive variables to provide estimates for about 75,000 lemma headwords:

1. English L2 vocabulary yes/no test data—Accuracy
2. English L2 vocabulary yes/no test data—Reaction time
3. English-Japanese loan words—Identity
4. English-Japanese loan words—Frequency in Japanese
5. Age of acquisition
6. Age of exposure
7. Word frequency in a large general corpus of English

4 New JACET8000

The New JACET List of Basic Words (New JACET8000) (JACET, 2016) is the updated version of JACET8000 (JACET, 2003), compiled by the Japan Association of College English Teachers (JACET). Based on the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), New JACET8000 serves as an educational word list for Japanese learners of English, especially university students. The list has 8,000 words, and for each 1,000 words, the level (i.e., from 1 to 8) is provided by the NWLC profiler. The New JACET8000 is available from Dr. Shin Ishikawa's website (<http://language.sakura.ne.jp/s/voc.html>).

5 SVL12000

SVL12000 (Standard Vocabulary List 12000) was developed and published in 2001 by ALC Press, Inc. It is based on the BNC. Like New JACET8000, the word list is intended to be used for educational purposes, and many ALC-published materials use this list. As the name shows, the list has 12,000 words and can be divided into 12 levels for each 1,000-word band.

6 New General Service List

The NGSL was conceived as a modern update of the General Service List (West, 1953). It was created from the Cambridge English Corpus (CEC), and it

covers about 92% of general texts of English with a list of approximately 2,800 high frequency words (Browne et al., 2013). The NGSL is optimal for Japanese EFL learners because they learn approximately 3,000 words through junior and senior high school textbooks. Browne and his colleagues subsequently produced a secondary series of word lists for learners who have mastered the first 2,801 words with NGSL. These include the New Academic Word List (NAWL: 963 words), the TOEIC Service List (TSL: 1,259 words), and the Business Service List (BSL: 1,754 words).

The secondary word lists (i.e., NAWL, TSL, and BSL) exclude all words on the NGSL, but some words appear in more than one secondary list. For example, the word “impact” is in all three lists (NAWL, TSL, and BSL). Also, while the word “quit” is in both TSL and BSL, the word “syndicate” only appears in BSL. For this reason, NWLC considers the overlapping of words in the three lists and produces the output using the criteria shown below. In total, NWLC has 5,621 words for word profiling.

- Level 1: NGSL = 2,801 words
- Level 2: NAWL and TOEIC and BSL = 183 words
- Level 3: NAWL and TOEIC, NAWL and BSL, or TOEIC and BSL = 790 words
- Level 4: Only in NAWL, TOEIC, or BSL = 1,847 words

7 CEFR-J Wordlist

CEFR-J Wordlist was based on an English textbook corpus consisting of textbooks for primary and secondary schools in China, Korea, and Taiwan (Tono, 2019). The word levels were classified according to the CEFR levels, and they were aligned with the English Vocabulary Profile (<https://www.englishprofile.org/wordlists>). All headwords, as a result, have part-of-speech information and their corresponding CEFR(-J) levels as shown in Table 1. NWLC uses the CEFR-J Wordlist Version 1.5.

8 Word Counting Units and Rules

8.1 Proper Nouns and Numerals

In the NWLC, proper nouns and numerals (numbers) are first identified using an open-source Part of Speech tagger, *spaCy* (<https://spacy.io/>) and are then

Table 1. CEFR-J Level, Number of Headwords and School Levels in Japan

CEFR-J Level	Number of Headwords	School Level in Japan
A1	1,164	Elementary
A2	1,411	Junior High
B1	2,446	Senior High
B2	2,778	University
Total	7,799	

treated as possibly *known* words because they can be assumed to be understood by learners. The possessive's (e.g., Todd's dog) is also put into the *known* category in word profiler output. For the remaining words in the text, the following lemmatization (tokenization) rules are applied.

8.2 Lemmatization

A variety of word-counting methods have been used in previous studies (for a review, see McLean, 2018). In some word counting methods, “happy,” “happily,” “happiness,” and “unhappy” can be counted as one word family, with the headword “happy” (e.g., Nation, 2012). Four of the five word lists in NWLC adapt flemma counting (i.e., a base form as a headword and its inflected forms as a one word). For example, for the headword “study,” the following word forms are included and counted as one word: study, studies, studied, and studying. The flemma—a portmanteau of “family” and “lemma”—was first introduced by Pinchbeck (2014, 2017) to distinguish between word lists that include part of speech (POS) information and those that do not and is a recommended word counting method in the field of applied linguistics and in EFL teaching contexts (see McLean, 2018). Flemma counting combines inflections of lemma groups but does not distinguish the POS. That is, with flemma counting, the verb “study” and the noun “study” are both counted under the same headword “study.”

The distinction in the terms *flemma* and *lemma* is often not always made. For example, a type of resource used by the AntConc software that is termed a “lemma list” has no requirement for POS tag information. In fact, the resource labeled as, “lemma list,” as used in the NWLC (see below) is all based on *flemma*-grouped word lists.

In contrast to the family and flemma grouping methods, lemma counting can detect the POS differences. The CEFR-J Wordlist adopts lemma counting since the original CEFR word lists are also lemmatized. Thus, in the CEFR word lists, the *verb* “study” is classified as A1 and the *noun* “study” is A2.

When a user selects the New JACET8000 or the SVL12000, NWLC uses the flemma list, *AntBNC Lemma List* (<https://www.laurenceanthony.net/software/antconc/>), which is based on all words in the BNC corpus with a frequency greater than two, for flemmatization. Modifications were manually made to match the headwords of New JACET8000 and SVL12000. For example, the words “interesting” and “interested” are listed as two headwords in both New JACET8000 and SVL12000, so they were excluded from the flemma entry “interest” in these lists (i.e., interest = interest, interests). In addition, words with British spellings in New JACET8000 and SVL12000 are included in the revised flemma list (e.g., advertise = advertise, advertises, advertised, advertising, advertize, advertizes, advertized, advertizing).

For the NGSL, flemmatization is simple because all NGSL lists are provided as flemma groupings. NWLC uses the flemma lists available at the NGSL website. For the CEFR-J Wordlist, NWLC utilizes spaCy to assign a POS and its lemma form. For SEWK-J, NWLC refers to a flemma list developed by Pinchbeck (manuscript in preparation).

Figure 1 shows a flowchart representing the algorithm of (f)lemmatization in NWLC. As can be seen, the application does its best to provide, by referring to the selected word list, the vocabulary profile of the text uploaded by the users.

8.3 Capitalized Letters

As the headwords in New JACET8000, SVL12000, and CEFR-J Wordlists include capital letters (Table 2), they are treated as they are (e.g., “I” not “i”). However, when the NGSL or the SEWK-J is used, NWLC treats all words as lowercase because all headwords in these lists are lowercase. In other words, the New JACET8000, SVL12000, and CEFR-J profilers are case sensitive, whereas those using the NGSL and SEWK-J are case insensitive.

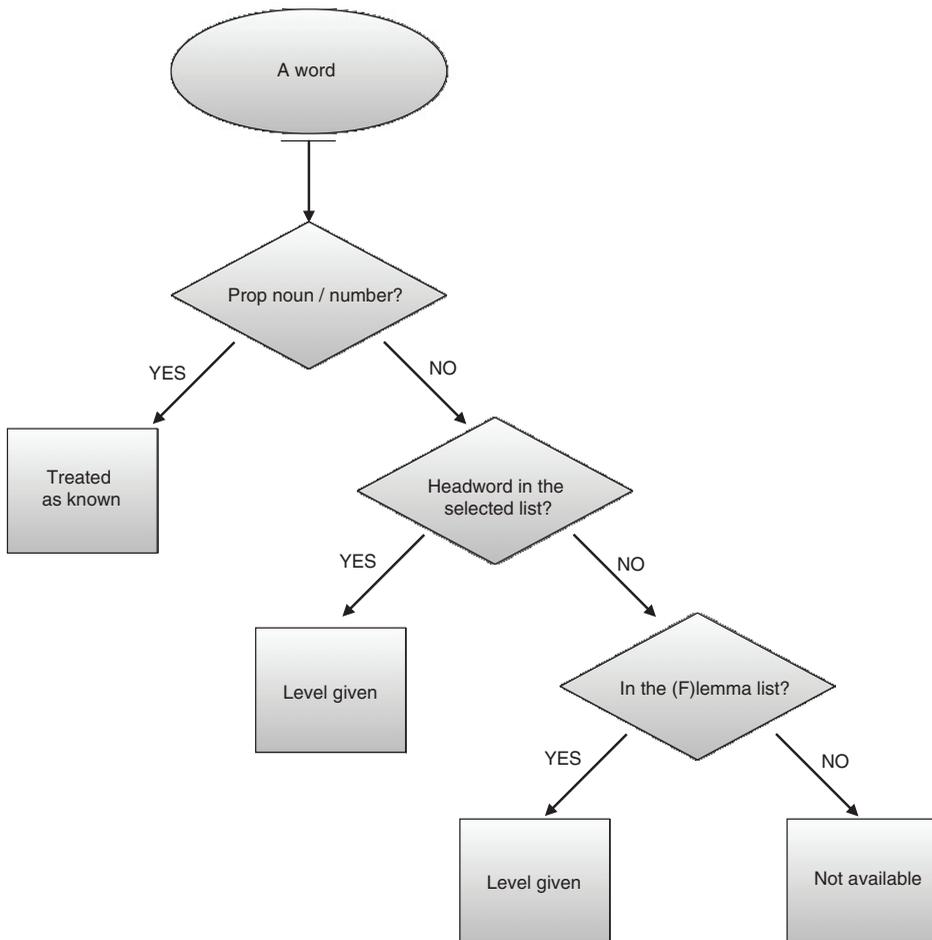


Figure 1. Flowchart representing the algorithm of lemmatization in NWLC.

Note. For CEFR-J, spaCy is used to lemmatize the word.

Table 2. List of Words with Capitalized Letters

New JACET8000 (20 words)
 April, August, December, English, February, Friday, I, January, Japanese, July, June, Monday, November, October, Saturday, September, Sunday, Thursday, Tuesday, Wednesday

SVL12000 (44 words)
 April, August, Bible, Buddhism, Buddhist, Catholic, Christ, Christian, Christianity, Christmas, December, Easter, Fahrenheit, February, Friday, God, I, Islam, January, July, June, March, May, Messrs., Miss, Monday, Mr., Mrs., Ms., Muslim, November, OK, October, Protestant, Renaissance, Satan, Saturday, September, Shinto, Sunday, Thanksgiving, Thursday, Tuesday, Wednesday

CEFR-J (72 words)
 A.M., AM, Apr, April, Atlantic, Aug, August, CD player, CD, CD-ROM, CV, Christian, DJ, DNA, DVD, Dec, December, Dr, Dr., E-mail, Englishman, Feb, February, Friday, HIV, I, ID card, ID, IT, Internet, Jan, January, July, June, MP3 player, Mar, March, May, Mediterranean, Miss, Monday, Mr, Mr., Mrs, Mrs., Ms, Ms., Nov, November, OK, OK, Oct, October, Olympia, Olympiad, Olympic, Olympics, PC, P.M., PM, Saturday, Sept, September, Shakespearean, Soviet, Sunday, T-shirt, TV, Thursday, Tuesday, Wednesday, X-ray

Table 3. List of Words with Apostrophes

SVL12000 (1 word)
 o'clock

NGWL (2 words)
 o'clock, ma'am

CEFR-J (6 words)
 driver's license, 'm, o'clock, 're, 's, 'd

8.4 Contracted Forms

NWLC detects the contracted forms by using spaCy and reverts them to the uncontracted forms (e.g., I'm => I am). Note that, as contracted forms such as “s” as in “he’s” (he has, he is) and “d” as in “we’d” (we had or we would) are not distinguished by the spaCy tagger, NWLC regards “s” and “d” as they are. We consider this lack of precision acceptable because the headwords, “be,” “is,” “have,” “has,” “would,” etc. are all among the most frequent words. For users who need to accurately distinguish between these contracted forms, the results may need to be checked carefully.

The SVL12000, NGSL, and CEFR-J contain words with apostrophes (Table 3). If those words are in the input text, NWLC treats them as they are.

8.5 Abbreviations with Periods

The SVL12000 and CEFR-J include abbreviations that use periods. If those word lists are selected and the input text has those words in Table 4, NWLC treats them as they are.

Table 4. List of Words with Periods

 SVL12000 (4 words)

Mr., Mrs., Ms., Messrs.

CEFR-J (8 words)

a.m., A.M., Dr., Mr., Mrs., Ms., p.m., P.M.

Table 5. List of Hyphenated Words

 NGWL (3 words)

e-book (in TSL), e-book (in BSL), by-law

CEFR-J (62 words)

bad-tempered, brand-new, brother-in-law, CD-ROM, check-in, check-in counter, check-in desk, daughter-in-law, duty-free, easy-going, e-mail, E-mail, face-to-face, father-in-law, film-maker, first-floor, full-time, good-looking, grown-up, half-price, hand-held, hard-working, heart-warming, high-tech, hip-hop, hi-tech, last-minute, left-hand, long-distance, long-term, make-up, middle-aged, mother-in-law, next-door, non-smoking, old-fashioned, out-of-date, part-time, right-hand, second-hand, self-confidence, self-confident, self-service, semi-final, short-term, sister-in-law, so-called, son-in-law, take-off, tee-shirt, T-shirt, up-to-date, washing-up, well-balanced, well-built, well-dressed, well-known, well-organised, well-organized, well-paid, worn-out, X-ray

8.6 Hyphenated Words

Hyphenated words are first divided into two words (e.g., “Osaka-based” is treated as two words, “Osaka” and “based”) in all word lists except for the cases where the selected list has hyphenated words as headwords (Table 5).

8.7 Compounds/Multi-Word Units

If a headword in the selected word list consists of more than one word (e.g., “bank account” and “mobile phone” in CEFR-J), it is counted as one word (unit), and NWLC returns the word profiling accordingly. NGSL and its three secondary lists (NAWL, TSL, and BSL) have only one compound, “ice cream,” while CEFR-J has 145 compounds (Table 6).

8.8 Comparisons of Word Lists: Which One to Use?

Table 7 summarizes the word counting units and rules, as described in the previous section, for the word lists adopted in NWLC.

Because all word lists in NWLC vary in terms of their rules and word definitions, direct comparisons are not straightforward. Therefore, a comparison of the coverage rate of each word list was made to provide a rule-of-thumb guideline for users to select which word list for what purposes.

Two English texts were analyzed with NWLC using all five word lists. The first English text was composed of (a) the Center Test in 2020 (6,309 words) and (b) the Kyotsu Test in 2021 (7,905 words), and both the reading and listening sections including the script for the listening section were used. The Kyotsu Test (Kyotsu means “common” in English), which was called the Center Test until 2020 with

Table 6. List of Compounds

NGWL (1 word)
ice cream

CEFR-J (145 words)
according to, air conditioning, air force, alarm clock, all right, bank account, because of, board game, bus station, bus stop, capital letter, car park, carbon dioxide, carbon footprint, carbon monoxide, CD player, central heating, chat show, check-in counter, check-in desk, chest of drawers, chewing gum, classical music, climate change, common sense, credit card, de facto, debit card, definite article, department store, digital camera, dining room, disc jockey, disk jockey, driver's license, driving licence, each other, environmentally friendly, exchange rate, exclamation mark, extreme sports, face to face, fast food, fed up, fire brigade, fire station, first floor, first lady, first language, first name, first person, frying pan, full stop, global warming, good afternoon, good morning, good night, grocery store, ground floor, hard drive, have to, health care, heart attack, high school, hip hop, human rights, ice cream, ice hockey, ice skating, ID card, identity card, indefinite article, instead of, inverted commas, junk food, last minute, last name, living room, main course, martial art, message board, mineral water, mixing bowl, mobile phone, modal verb, MP3 player, native speaker, navy blue, net surfer, next door, next to, no one, olive oil, on to, ought to, out of, owing to, pen friend, per cent, petrol station, phrasal verb, pocket money, point of view, polar bear, police officer, police station, post office, primary school, prime minister, public transport, question mark, real estate, remote control, rush hour, science fiction, second person, secondary school, shop assistant, sitting room, soap opera, social networking, soft drink, sports center, sports centre, steering wheel, stock market, swimming costume, swimming pool, table tennis, text message, third person, tour guide, traffic jam, traffic light, travel agent, upside down, used to, vice president, video clip, video game, virtual reality, washing machine, weather forecast, web page, worn out

Table 7. Summary of Word Counting Units and Rules

	SEWK-J	NewJ8	SVL	NGSL	CEFR-J
Total Words	74,810	8,000	12,000	5,621	7,799
Lemma or Flemma	Flemma	Flemma	Flemma	Flemma	Lemma
(F)lemma List	Yes	No	No	Yes	No
Capitalized	0	20	44	0	72
Case	Insensitive	Sensitive	Sensitive	Insensitive	Sensitive
Apostrophe	0	0	1	2	6
Abbreviation with Period	0	0	4	0	8
Hyphen	0	0	0	3	62
Compound	0	0	0	1	145

a different test format, is a standardized exam for students who intend to enter a national, public, or private university in Japan. Over 470,000 test takers sat for the Kyotsu Test in 2021. As the Center and Kyotsu test items reflect the Course of Study (national curriculum guidelines) in Japan, the test items from 2020 and 2021 used in this comparison can serve as a yardstick for evaluating the coverage rate, with each word list, of the standard English text that high-school graduates in Japan are likely to encounter.

The second English text analyzed with the word lists on NWLC was Test of English for International Communication (TOEIC). The text was collected and randomly sampled from materials such as TOEIC Official Test Preparation

Guide. Both listening and reading sections are included. Because many Japanese university students and businesspeople take the TOEIC test, it can be regarded as a representative text for EFL learners at those levels.

After ignoring question numbers, symbols, and Japanese text, the Center and Kyotsu Tests and the TOEIC comprised 14,214 and 104,299 words, respectively.

Figure 2 shows the results of the coverage rate of five word lists on NWLC. The level or category marked by the red boxes indicates 95% coverage of the text by each word list. As it is clear from the results, all five lists reach the 95% coverage threshold, which indicates that the vocabulary included in these word lists is within the range of practical use for these important, high-stake tests in Japan.

The blue boxes in the figure indicate 98% coverage of the text by each word list. This result demonstrates that, to reach the 98% coverage threshold which is

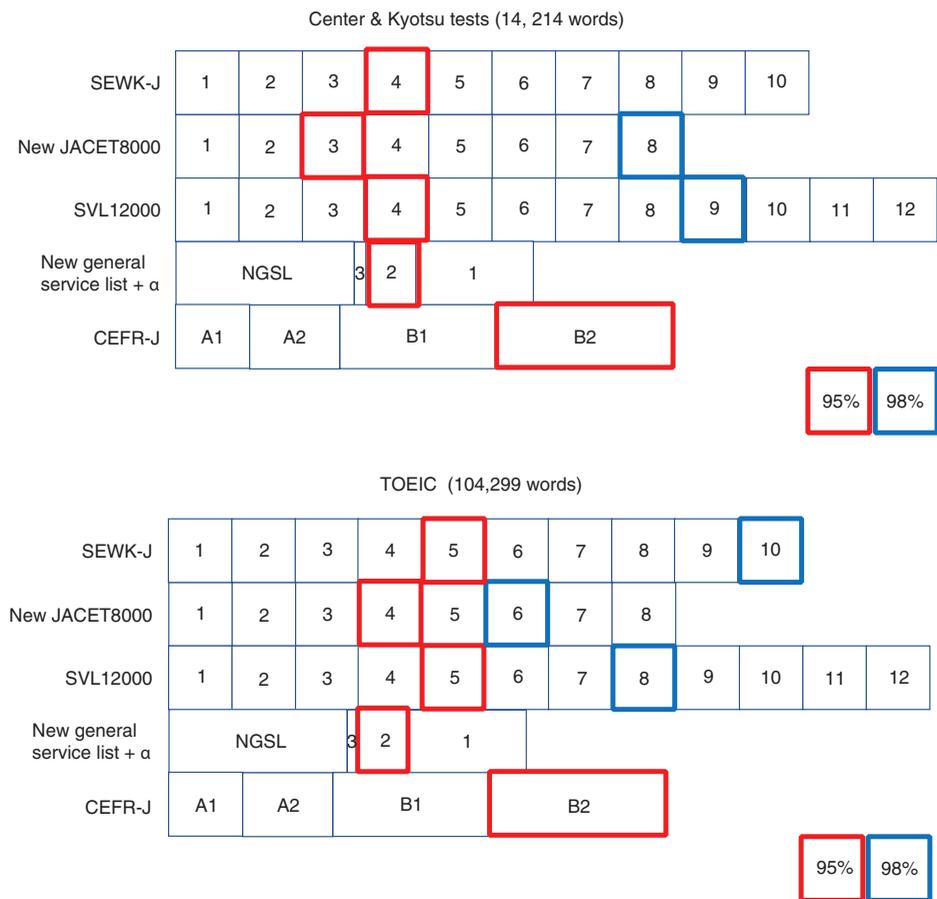


Figure 2. Coverage rates of five word lists on NWLC.

Note. For New General Service List + α (three other lists), 3 indicates NAWL and TOEIC and BSL (183 words), 2 indicates NAWL and TOEIC, NAWL and BSL, or TOEIC and BSL (790 words), and 1 indicates that the word appears only in NAWL, TOEIC, or BSL (1,847 words). In this analysis, SEWK-J has only 10 levels, each with 1,000 words, for comparison with other lists.

posited by vocabulary research literature (Schmitt et al., 2011), it requires several thousands more words in New JACET8000, SVL12000, and SEWK-J. In the case of the NGSL plus three secondary lists (NAWL, TOEIC, and BSL) and the CEFR-J, it was not possible to reach the 98% coverage threshold for either the Center and Kyotsu Tests or the TOEIC. This is not surprising given the fact that the number of words included in NGSL plus the three secondary lists is 5,621, and 7,815 for CEFR-J. It should be noted that when it comes to the coverage rate of text, CEFR-J has a disadvantage because it adopts the lemma counting unit, and the same word may have a different POS, which is not the case with the other four lists in NWLC; this means that many more lemmas are required to cover the same amount of text as compared to flemmas.

From the results of the comparisons of word lists used in NWLC, it can be claimed that all word lists could be used to examine the coverage rate of text because all the lists proved to have sufficient coverage of English texts, up to 95%. Although only the results of analyzing two English texts were reported in this article, the same tendency can be observed with other texts as well.

If the purpose of using the list is to learn the first 3,000 words in the flemma form, the NGSL may be a good choice. NGSL was reported to cover 92% of most general English texts (Browne et al., 2013). In this sense, the learning efficiency is much higher for learners. As CEFR-J distinguishes POS of the words, it could be used for assessing learners' productive vocabulary use in writing or speaking, which in turn implies that CEFR-J could be used to measure the depth of vocabulary knowledge.

Among New JACET8000, SVL12000, and SEWK-J, New JACET8000 had better coverage than the other two in the current comparisons. As New JACET8000 is an updated word list integrating other word lists derived from large corpora, the coverage performance seems to be better than other word lists. If the user needs to learn more than 8,000 words, which is what the New JACET8000 covers, the SVL12000 or SEWK-J could be utilized. Especially, the SEWK-J list contains 74,810 headwords, and thus it is possible to inspect how many words learners need to know to reach the 98% coverage threshold, which cannot be accomplished with the other word lists in NWLC.

Although the degree of coverage of texts can be used as a measure of how efficiently a word list represents a target register, word lists can also be used for other purposes. Notably, the SEWK-J was not designed to increase the coverage performance; rather, it was intended to map the Japanese university students' vocabulary knowledge on the word list. In this sense, it is unique in that the rank order of the headwords in SEWK-J represents the likelihood that a word in the list is known to the learners in the target population (i.e., Japanese university students). Thus, while frequency-based lists are of great value in directing learners towards the most frequent—and so valued—words in a list, when matching learners with texts, we should consider basing tests and lexical profilers on what learners *do* know (i.e., by using word lists based on learners' actual vocabulary knowledge such as SEWK-J), rather than frequency-based lists, which indicate what learners *should* know (Paul Nation, personal communication, August 8, 2021). By matching learners with texts, vocabulary profiling will allow teachers

and materials developers to estimate how difficult candidate texts might actually be for learners at a given level of proficiency.

In terms of the assessment of vocabulary knowledge, word profilers and vocabulary tests should be based on the same word list scale to match the vocabulary level of a text with learners' lexical knowledge. To this end, the *Vocableveltest.org* website (<https://vocableveltest.org/>) can be utilized because it facilitates the creation and administration of meaning-recall vocabulary levels tests, which better measure the type of vocabulary knowledge that can be employed when reading than meaning-recognition (multiple-choice or matching) tests (McLean et al., 2020; Stewart et al., 2021), on several of the same word lists that are used in the NWLC (e.g., SEWK-J, NGSL, NAWL, TSL, and New JACET 8000).

Acknowledgements

This research was made possible by a grant-in-aid for Scientific Research (21H00553 and 20K00792) from the Japan Society for the Promotion of Science.

References

- Browne, C., Culligan, B., & Phillips, J. (2013). *New General Service List project*. <http://www.newgeneralservicelist.org/>
- Japan Association of College English Teachers (JACET). (2003). *JACET8000*. JACET.
- Japan Association of College English Teachers (JACET). (2016). *The new JACET list of 8000 basic words*. Kirihara Shoten.
- McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39(6), 823–845. <https://doi.org/10.1093/applin/amw050>
- McLean, S., & Kramer, B. (2016). The development of a Japanese bilingual version of the new vocabulary levels test. *Vocabulary Education and Research Bulletin (VERB)*, 5(1), 2–5. <https://doi.org/10.7763/IJLLL.2015.V1.26>
- McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing*, 37(3), 389–411. <https://doi.org/10.1177/0265532219898380>
- Nation, P. (2012). *The BNC/COCA word family lists*. https://www.wgtn.ac.nz/_data/assets/pdf_file/0004/1689349/Information-on-the-BNC_COCA-word-family-lists-20180705.pdf
- Pinchbeck, G. G. (2014, March). *Lexical frequency profiling of a large sample of Canadian high school diploma exam expository writing: L1 and L2 academic English* [Roundtable presentation]. American Association for Applied Linguistics Annual Conference 2014, Portland, OR. <https://app.box.com/s/2772oofxkg6jyzmd317j>
- Pinchbeck, G. G. (2017). *Vocabulary use in academic-track high-school English literature diploma exam essay writing and its relationship to academic*

- achievement*. Doctoral dissertation. University of Calgary. <https://prism.ucalgary.ca/handle/11023/3676>
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Someya, Y. (2006). *Word level checker* [Web app]. <http://someya-net.com/wlc/>
- Stewart, J., Stoeckel, T., McLean, S., Nation, P., & Pinchbeck, G. G. (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. *Studies in Second Language Acquisition*, 43(2), 462–471. <https://doi.org/10.1017/S0272263121000437>
- Tono, Y. (2019). *CEFR-J wordlist version 1.5*. https://www.cefr-j.org/download.html#cefrj_wordlist
- West, M. (1953). *A general service list of English words*. Longman.