
Practical Research Reports

Building Essential Vocabulary in First-Year University English

Paul WADDEN^{1)*}, Sakae ONODA²⁾, Marcel VAN AMELSVOORT³⁾, Toru TOYAMA⁴⁾, Hiroaki UMEHARA⁵⁾

【Abstract】

This report describes a first-year English research-and-curriculum-development project of a liberal arts faculty. To “fill the gaps” in first-year students’ general English vocabulary, the researchers attempted to identify the words which students’ most need to learn to improve their academic reading and listening comprehension ability. They began by focusing on the 2,000 to 2,800 most frequently used words identified by the New General Service List (NGSL), eliminated words appearing in junior high school textbooks, English loan words, and vocabulary students could be expected to know. Researchers then asked lower-intermediate students to identify the words they “Don’t know,” “Know a little,” and “Know.” They hoped the results would confirm their hypothesis that focus on a targeted group of 400 words was optimal for students, and thus they could focus on a specific study plan and materials for intermediate- and lower-level students. What they found instead was that the vocabulary gaps were much larger and the variance of unknown words among students much greater than anticipated. One broader implication of the research is that many first-year university students in Japan could significantly benefit from systematically learning the 2,800 most commonly used words in English as they begin their university study.

Key words: Academic vocabulary, New General Service List (NGSL), First-Year English, Japan higher education

調査・実践報告

大学一年英語科目における必須英単語の構築

ポール ワーデン^{1)*}, 小野田 榮²⁾, マルセル ヴァン アメルズフォート³⁾, 外山 徹⁴⁾, 梅原 洋陽⁵⁾

【要 旨】

この研究報告は、某大学国際教養学部初年次英語プログラムにおける研究およびカリキュラム開発プロジェクトの一部を記述したものである。本研究の目的は、学術的文献の読み聞き能力と理解力の向上のために学生が優先的に学ぶべき語彙の特定である。英文頻出単語を集めた NGSL (New General Service List) の 2,000 ~ 2,800 位の単語から、中学英語教材の頻出単語や高校生でも知っているような単語を除外し、目標語彙リストとした。そして、初級と中級の英語クラスの参加者に、これらの単語を「知らない」「なんとなく知っている」「知っている」の三段階に分類してもらった。著者らは、作成した 400 語の目標単語リストは学習者にとって最適であり、各レベルに応じた単語学習計画を作成することが望ましいと考えた。しかし、調査から、同じレベルのクラスに在籍しているにもかかわらず、習得語彙に大きな差があり、未習得語に対する意識にもかなりの差があることが判明した。このことから、日本の大学 1 年生の多くは、大学での英語学習開始時に、NGSL に含まれる 2,800 語を体系的かつ包括的に学習することが有効であると言えるだろう。

キーワード : アカデミックな単語、New General Service List (NGSL)、大学 1 年の英語、日本の高等教育

¹⁾ Faculty of International Liberal Arts, Juntendo University (E-mail: wadden@juntendo.ac.jp)

²⁾ Faculty of International Liberal Arts, Juntendo University (E-mail: s-onoda@juntendo.ac.jp)

³⁾ Faculty of International Liberal Arts, Juntendo University (E-mail: amelsvoort@juntendo.ac.jp)

⁴⁾ Faculty of International Liberal Arts, Juntendo University (E-mail: t.toyama.er@juntendo.ac.jp)

⁵⁾ Faculty of International Liberal Arts, Juntendo University (E-mail: h.umehara.yz@juntendo.ac.jp)

* Corresponding author: Paul WADDEN

[Received on September 14, 2022] [Accepted on January 19, 2023]

1. Introduction (Background and Purpose)

To understand a foreign language—at any level or in any form—learners must know the meaning of the words. To comprehend written English in particular, second-language scholars and vocabulary researchers have established that a learner needs to know about 95% of the vocabulary in a text. As Nation and Waring (1997) observe, “With a vocabulary size of 2,000 words, a learner knows 80% of the words in a text which means that one word in every five (approximately two words in every line) are unknown.” Nation and Waring further point out that studies by Na and Nation (1985) have “shown that this ratio of unknown to known words is not sufficient to allow reasonably successful guessing of the meaning of the unknown words. At least 95% coverage is needed for that.” According to Laufer and Ravenhorst-Kalovsk (2010), achieving this 95% coverage requires that readers know 4,000–5,000 of the most frequent word families in English. This allows them to grasp the text’s overall meaning and make reasonable inferences about the meaning of words they don’t know. In university study, reaching the 95% vocabulary comprehension level for academic texts and lectures (not only general English texts) is among the key learning tasks for students, and it is especially crucial for those engaged in content-based classes, EMI (English Medium Instruction) courses, preparation for overseas study, and for those demonstrating their English academic proficiency on the TOEFL test.

One important means to help students achieve this level of vocabulary knowledge is to integrate academic word lists into university curricula to help students build a solid foundation that will enable them to understand readings and lectures, and eventually to begin to use high-frequency academic vocabulary in speaking and writing as well. Learning only the vocabulary present in an EFL (English as a Foreign Language) textbook, or words

appearing in specific lessons in an English course, is insufficient. In Japan, teachers at Sophia University, International Christian University, and Waseda University, among others, draw upon a combination of four well-established word lists to build students’ vocabulary for liberal arts study, as do select programs of Trinity College at the University of Melbourne and New York University (Wadden, Ferreira, & Rush, 2016). In 2022, teachers of the first-year English courses in a liberal arts faculty at a university in Japan (University A) implemented the same series of 23 lessons from the Global Academic Vocabulary Lexicon (GAV for short) to build students’ word knowledge in order to boost their general English ability, support their liberal arts study, prepare them for future study in English, and improve their performance on the TOEFL exam. The GAV combines the headwords from the four major corpus-based academic word lists compiled to date: The Academic Word List (AWL) (570 words, Coxhead, 2000), The New Academic Word List (NAWL) (963 words, Browne et al., 2013), The University Word List (UWL) (836 words, Xue & Nation, 1984), and the English for Academic Purposes (EAP) Word List (874 words, Masuko et al., 1997). The words in these lists were identified by scanning millions of pages of academic texts from university courses across traditional fields, from biology to sociology and economics to health. Each of these wordlists (from 570 to 963 words) was developed independently and many of the words identified in them overlap. By itself, each list is insufficient for students to attain the 95% comprehension level, much less the even more desirable 98% level cited above. However, when the four lists are combined in the GAV, the approximately 1,750 frequently appearing academic words provide the most comprehensive learning materials in academic vocabulary yet created (see Wadden, Ferreira, & Rush, 2016 for a fuller discussion of the

GAV, its coverage, its use, and its compilation).

In 2022, the teachers in the first-year English program of University A began to implement a Content and Language Integrated Learning (CLIL) curriculum. Smit and Dafouz describe a CLIL approach as one in which “Learners are engaged in a joint learning practice of subject matter and foreign language” (2012, p. 1). At University A, this means as students learn English, they also acquire basic knowledge in liberal arts fields such as Sociology, Economics, Earth Science, Health and Medicine, Zoology, Natural History, and other areas. Understanding the basic concepts of these areas of knowledge increases students’ readiness to take university courses in English and helps increase their score on the TOEFL. In addition, systematically studying the approximately 1,750 most frequently appearing academic words—collected and featured on the GAV—supports students’ listening and reading comprehension of English and improves their overall academic skills and knowledge. However, teachers of the lower-level classes at University A discovered what is becoming a widespread challenge throughout Japan: As the number of high school students declines due to demographics, universities are competing for a smaller group of college-bound high school graduates. For example, in 1988 nearly 12 million students were enrolled in high school; in recent years this number has dropped to about 7 million. At the present time, however, a much higher proportion of high school graduates enter university: around 55%. (WENR, 2021). Many of these college applicants have a lower level of English achievement than in past years. As a result, universities and English teachers must now further adapt and simplify their materials and their courses to optimally teach these lower English proficiency university students (Kitano & Chiba, 2020).

In “Teaching and Learning Vocabulary in the Japanese University,” Wadden, Browne, & Nation

(2019) observe, “The typical Japanese student enters university with an English vocabulary of 2,000 to 4,000 words” and “beyond the initial 1,500 or so words the students have learned in elementary and middle school, individual students have not acquired the same core vocabulary” (p. 84). This is precisely the challenge that the first-year curriculum designers at University A, as well as designers and classroom teachers at universities throughout Japan now face. Many high school graduates have acquired too little English vocabulary, the number of words they have acquired varies widely, and because MEXT-approved middle- and high school English textbooks do not focus on a particular list of high-frequency words, students do not systematically learn the most frequently used words in English, such as by focusing on the 2,801 words in the NGSL (New General Service List), or the first 3,000 words of the JACET (The Japan Association of College English Teachers) 8000, which is another useful corpus-based frequency word list. As a result, different students know different words. Nakayama’s insightful 2022 article “A Close Examination of Vocabulary in Japanese EFL textbooks” provides an in-depth explanation of this disparity. These widespread differences in the words students know and don’t know is a significant obstacle for university English teachers in general and for those teaching CLIL courses in particular. The purpose of this study was to identify the crucial vocabulary which students have or have not yet learned of the most frequently occurring words in general English, focusing initially on the 2,000 to 2,801 most common words. The importance of knowing these words for university study in English cannot be over-emphasized or exaggerated: Lack of knowledge of these words impedes reading and listening comprehension of all but the most basic English texts and spoken communication and impedes the ability of students to learn in a CLIL-based liberal

arts curriculum.

2. Methods

Browne, one of the compilers of the NGSL and NAWL (New Academic Word List), suggests that for optimal learning the general university English class should begin by checking students' word knowledge of the NGSL (the most common 2,801 words in written English) to identify the "bandwidth" of words the students don't yet know in order to individually address the "significant gaps" in students' vocabulary (Wadden, Browne, & Nation, 2019, p. 86). To do this, Browne suggests administering the New General Service List Test (NGSLT) which can identify the set of words the students should begin studying (for example, the 100 words from the 2,400 to 2,500 most common words in written English). He then recommends teachers use smartphone applications such as WordEngine and Word Learner to create virtual classrooms and to help students develop their individualized vocabulary learning program by studying flashcards, taking practice tests, and playing educational word games while using spaced repetition to optimize retention. However, this method requires teachers and students to purchase a software app—an institutional hurdle at many schools—and requires the testing of the vocabulary levels of all students with the NGSLT to tailor their individualized vocabulary programs. Short on instructional time and facing challenges with requiring all students to pay for smartphone apps, the authors decided upon a different though related approach.

To identify the vocabulary most needed by the students to advance their English skills and academic study, the authors did the following:

1. To begin, they focused on the 2001-2801 bandwidth of the NGSL, believing that even the least proficient college-bound high school graduates

were familiar with the 2,000 most frequently used words in English. For example, "hunt," "storm," "thick," "achievement," and "negotiate" are the 1996-2,000 most common words. The authors strategically assumed that nearly all high school graduates who entered university already knew these 2,000 most widely used words.

2. Next, they eliminated from that last group of 800 words (the NGSL vocabulary from words 2001 to 2801) the vocabulary that appears in middle school textbooks as identified in "A Survey of the Vocabulary Used in the Junior High School English Textbooks" (Chu-o Institute for Educational Research, 2021). They reasoned that vocabulary studied in middle school would surely have been acquired by high school graduates.

3. They then reduced the list further by having three Japanese faculty (co-authors of this report) with extensive experience teaching English in high school and university eliminate words that in their judgment nearly all high school graduates already knew. This trimmed the list of foundational vocabulary to 400 high-value targeted words.

4. Finally, the researchers had about 70 of the lower intermediate students (mainly students with a TOEFL ITP of 399 or below) in their 240-student cohort identify, one by one, the words they already knew. To do this, the researchers administered four Google Forms surveys of 100 words which asked students to respond to each of the 400 words by labeling it as "don't know," "know a little," "know."

Only the response "know" was counted as "knowing" the target word; responses "don't know" and "know a little" were both categorized as "not knowing." The 70 or so respondents were nearly all Japanese high school graduates of about 18-19 years of age and divided fairly evenly between female and male. Seventy-three students responded to the first 100-word survey, 72 to the second 100-word survey, 68 to the third 100-word survey, and 61 to the fourth

NGSL Survey of 3rd 100 Words (201-300)

Please answer whether you "don't know," "know a little," or "know" a word.

indication *

don't know

know a little

know

inflection *

don't know

know a little

know

Fig. 1. Excerpt from Survey 3.

100-word survey.

The researchers hoped that the results would confirm their hypothesis that focus on these targeted 400 words was optimal for students, and thus allow the researchers to reduce the word list even further.

Overall, this four-step process was intended to provide the first-year English program and its teachers and students with a more precisely focused word list that could be covered in lower-level and mid-level classes at the beginning of the academic year so that students' fundamental vocabulary and reading and listening skills could receive an immediate boost upon beginning university study. This in turn would enable their CLIL-based study of liberal arts contents during their first year. The authors defined the 400-word vocabulary list bilingually in English and in Japanese, drawing primarily upon Browne, Culligan, & Phillips' initial work on NGSL Project website, but then further refining and revising the definitions. They labeled the words for part of speech (noun, verb, etc.), and uploaded them to the free Quizlet learning website. This site is among the most widely used online vocabulary learning sites of its kind in the world—according to its website it has more than 60 million active users (Quizlet)—and it offers a rich variety of learning activities, review practice, and games for

individual study as well as for group use in the classroom. The plan was for intermediate and lower intermediate students to initially fill their “significant gaps,” to use Browne's earlier phrase. Further plans were even made to give matriculating students the vocabulary list in February so that during March and early April they could prepare in advance and begin their university English classes knowing at least the most common 2,801 vocabulary used in written English. This would empower their first-year English study.

3. Results

The first three steps of the process outlined above proceeded without incident. The 800-target vocabulary were downloaded from the NGSL website, words currently used in middle school textbooks were eliminated, and three Japanese faculty in the English program reviewed the list for words they believed were widely covered in high school and *juku* (cram school) study, or were loanwords in Japanese and known to students through their katakana equivalents; this allowed the researchers to further reduce the list. The remaining 400 words were then uploaded to Quizlet in bilingual form and readied for both class and individualized use (the website allows teachers to create classes and, if they have a teacher's account, to track their students' practice and progress). This approach promised to provide students with a free (no cost) highly focused group of vocabulary to study. However, the results of Step 4—the four surveys asking the 70 or so lower intermediate students to identify the words they already knew or didn't know—were wholly unexpected.

Survey 1 (abortion to dedicate) 73 responses

Survey 2 (deficit to index) 72 responses

Survey 3 (indication to province) 68 responses

Survey 4 (provision to withdraw) 61 responses

To the researchers' surprise, there was not a single

word in the group of 400 words covered by the surveys that all or nearly all of the students knew. Only 3 words out of 400—“academic,” “award,” and “ban”—were indicated as “known” by more than 80% of the students. In addition:

- Only about 15 words out of 400 were known by 70% to 80% of the respondents. These were “achieve,” “adapt,” “decline,” “define,” “demonstration,” “domestic,” “due,” “effective,” “element,” “expand,” “foundation,” “manufacture,” “numerous,” “objective,” and “option.” This means that 20%-30% of the students lacked confidence in their understanding of these extremely basic vocabulary or did not know them at all.
- About 50% of the respondents indicated knowing words such as “consistent,” “dedicate,” “distribute,” “flexible,” “interpret,” “insight,” “notion,” “participant,” “principle,” “reasonably,” “recognition,” “specialize,” “substance,” “temporary,” and “withdraw.” This conversely means that 50% lacked confidence in their understanding of these fundamental vocabulary or didn’t know them at all.
- The majority of the 400 words were known by fewer than 50% of the students (47% overall), though there was a great deal of variance.
- About 75% or more of the students did not know fairly common general English words in the NGSL and often used in academic writing such as “govern,” “component,” “complexity,” “conservative,” “firmly,” “indication,” “intellectual,” “liberal,” “occupy,” “obligation,” “qualification,” and “theoretical.”
- Another surprise in the results was that fewer than half of the students (less than 50%) knew the meaning of extremely common function words (adverbs and conjunctions) such as “altogether,” “consequently,” “furthermore,” “meanwhile,” “nevertheless,” and “subsequently.”

The results provide dramatic support for the assertion that “beyond the initial 1,500 or so words the students have learned in elementary and middle school, individual students have not acquired the same core vocabulary” (Wadden, Browne, & Nation, 2019, p. 84).

4. Discussion and Conclusion

As a matter of perspective, it was never the intention of the researchers to find fault with students and their language ability or to criticize their English education in high school. Language learning is difficult and whatever the level of the students is, teachers and curricula should adapt to teach them effectively from the optimal starting point. After all, language learning in life is a marathon, not a sprint, and learners who study consistently over a long period (not fast learners or crammers) are the ones who over time are likely to achieve the greatest proficiency.

In addition, it should be remembered that the best action-based research uncovers important facts and then adopts changes based upon them. The results of the present study suggest the authors must return to the drawing board and take action by expanding the vocabulary that mid- and lower-proficiency students study when they arrive in university. Likely, the entire 1,000-word bandwidth from NGSL 1,800 words through 2,801 will need to be covered (this band starts with “vast,” “exhibition,” and “producer” and it ends with “ashamed,” “found,” and “thirst”). Possibly even the NGSL words starting from 1,001 (“shoot,” “announce,” “unless”) should be reviewed. Nakayama’s recent (2022) analysis of the NGSL words covered in Japanese high school textbooks made the astonishing discovery that the senior high Eigo Komyunikeishon (English Communication) textbooks from the principal publishers cover only a relatively small number of words from the NGSL—from 15% to 30% (see

Figure 2 below). And the number of NGSL words NOT included at all in the texts (remember, these are the 2801 most frequently appearing words in general English texts) range from 1,732 words to 2,457 words. In other words, 67% to 88% of the words in this foundational vocabulary are not covered—the vast majority. Only outside of their high school English communication texts, for example in juku (cram school) or personal study, would students encounter the words.

To create materials to introduce students to such a large number of basic vocabulary words will require an enormous amount of labor in writing and revising English and Japanese definitions, identifying parts of speech, organizing words into groups, and uploading them to an effective language-learning website such as Quizlet. Another alternative is to turn to pay-to-use educational websites and software such as Word Learner and WordEngine mentioned above, despite the cost and the additional test administration required.

There are some shortcomings to the current study that should be noted. First, its measure of word knowledge is based upon students’ self-report. Inaccuracies could be built in. That noted, the four word-knowledge surveys were administered to four different summer intensive classes of 1st-year

students with remarkably similar results. This corroborates to some degree their accuracy and consequently the researchers are reasonably confident in the data collected. In addition, it should be noted that the surveys were conducted in July after the students had already had nearly four months of university study of English—four 90-minute periods or “koma” a week for 14 weeks—so their word-recognition of vocabulary they “knew” (overall 47%) was presumably higher than when they initially began university. This suggests that many entering students have a significant need for a vocabulary “flood,” to use Nation’s term, of the NGSL 1,000-word vocabulary bandwidth from 1,800 to 2801. Review of the 1,000 to 1,800 bandwidth may be worthwhile, too, as Nakayama’s study below suggests.

It should also be noted that roughly 80% of the first-year curriculum reform in University A’s first-year English program revolves around reading and listening to liberal arts content, and yet the 10% of the curriculum based upon learning high-frequency English vocabulary provides the foundation for all of it. In short, without knowing the words, students will be unable to grasp the content. Therefore, it is precisely Content and Language Integrated Learning that is most needed. The purpose of this study was

Textbook	Number of NGSL words covered in the textbook	Number of NGSL words not covered in the textbook	Coverage (%)
VISTA I	431	2,370	15
All Aboard! I	404	2,397	14
COMET I	344	2,457	12
VISTA II	425	2,376	15
ELEMENT II	913	1,888	33
COMET II	547	2,254	20
LANDMARK Fit III	859	1,942	31
ELEMENT III	1,069	1,732	38
All Aboard! III	650	2,151	23

Fig.2. Coverage of NGSL vocabulary in high school textbooks (from Nakayama, 2022, p.213)

to better identify the crucial high frequency vocabulary students need to first enable and then to make progress in their university English study.

Finally, the results of this current research and its implications are noteworthy because they have applicability far beyond the liberal arts faculty of University A where the research was conducted. Universities throughout Japan enroll students who have similar backgrounds and comparable language proficiency from the same public and private high schools. We are, in other words, in the same large higher education boat together. There is a pressing need—nationwide—for entering university students to focus on and intensively learn a much larger number of English vocabulary than previously assumed.

5. Acknowledgements

The authors gratefully acknowledge the support of the Ministry of Education's multi-year Grant-in-Aid for Scientific Research (21K00714) for the curriculum reform within which this action-research project was undertaken.

6. References

- Browne, C., Culligan, B., & Phillips, J. (2013). New General Service List. Retrieved from <http://www.newgeneralservicelist.org/>, accessed September 10, 2022.
- Chu-o Institute for Educational Research. (2021). *A survey of the vocabulary used in the junior high school English textbooks*. Tokyo: Chu-o Institute for Educational Research.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Kitano, M., & Chiba, K. (2020). Analyzing gaps of knowledge in high frequency words by university freshman in Japan. *Bulletin of the Faculty of International Studies*, Bunkyo University 31(1).
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15-30.
- Masuko, M. S., Mizoguchi, M. Sano, K. Shiina, R. Thrasher, & Yoshioka, M. (1997). *English Vocabulary for Academic Purposes: Analysis of Data from the Physical and Social Sciences*. Tokyo: Liber Press.
- Na, L., & Nation, I. S. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16(1), 33-42.
- Nakayama, S. (2022). A close examination of vocabulary in Japanese EFL textbooks. In P. Ferguson, & R. Derrah (Eds.), *Reflections and New Perspectives*. JALT. <https://doi.org/10.37546/JALTPCP2021-24>.
- Nation, I. S., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. Retrieved from <http://www.fltr.ucl.ac.be/fltr/germ/etan/bibs/vocab/cup.html/>, accessed September 10, 2022.
- Quizlet. (n.d.) Quizlet. Retrieved from <https://quizlet.com/> and <https://quizlet.com/mission>, accessed September 10, 2022.
- Smit, U., & Dafouz, E. (2012). Integrating content and language in higher education: An introduction to English-medium policies, conceptual issues and research practices across Europe. *AILA Review*, 25, 1-12.
- Wadden, P., Browne, C., & Nation, P. (2019). Teaching and learning vocabulary in the Japanese university. In Wadden P. & Hale C. (Eds.) *Teaching English at Japanese Universities: A New Handbook*. Abingdon, UK: Routledge. 84-96.
- Wadden, P., Ferreira, D., & Rush, E. (2016). The Global Academic Vocabulary lexicon: A new

- ELT resource. *Accents Asia* 8(2), 135-142.
- WENR. (2021). Education in Japan. Retrieved from <https://wenr.wes.org/2021/02/education-in-japan>, accessed Sep 10, 2022.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215-229.