Analyzing Vocabulary in MEXT-Approved High School Textbooks

TUPLS
Tokyo University of Pharmacy and Life Sciences

Shunsuke Ito

Laboratory of Language Sciences – Tokyo University of Pharmacy and Life Sciences

Background

▶**English Vocabulary in the *Course of Study (2021)***
- *Course of Study (2021)* defines the national curriculum guidelines in Japan.
➔High school English textbooks were revised in 2021 based on the *Course of Study (2021).*
  - The vocabulary was not controlled.
    - No clear guidelines on vocabulary.
    - No minimum vocabulary requirement.

▶**Previous Studies**
- Textbooks based on the previous guideline mainly covered vocabulary from the *New General Service List (NGSL)*, but their lexical range was limited, with only 12–38% of *NGSL* words appearing in the textbooks (Nakayama, 2022).
- One Japanese high school textbook series, for example, covered about 91.2% of the National Center Test vocabulary, indicating sufficient exam preparation (Tani, 2008).

▶**Research Questions**
1. To what extent is there variation among the new textbook series in the amount and range of vocabulary they include?
2. To what extent do the new textbook series cover the vocabulary used in the *Kyotsu Test* (the exam for University Admissions)?

Materials & Methods

▶**Corpora**
① Textbooks
- Textbooks for English Communication I, II, and III (ECI, ECII, ECIII)
- 70 MEXT-approved English textbooks published by 12 publishers in 24 series (Book 1–3, though two series have only 2 volumes).
- Sections: Reading and Optional reading
② Tests
- Entrance exam reading sections (main and supplementary) from the *Kyotsu Test* (2021–2025): 10 sets

	All	ECI	ECII	ECIII	Tests
Texts	952	313	332	307	64
Type	15,251	6,768	8,949	10,962	4,506
Token	601,924	145,998	197,037	258,889	57,636

Table 1 Textbook Corpus Composition

▶ **Word Lists for Coverage Calculation**
Five vocabulary lists were selected to represent four categories: basic, academic, test-related, and vocabulary for L2 learners.

Wordlist	Details
NGSL	- *New General Service List* 1.2 - 2,809 word families: Core general and daily English for L2 - Browne, C., Culligan, B., & Phillips, J. (2023)
NAWL	- *The New Academic Word List* - 957 word families: academic focus - Browne, C., Culligan, B., & Phillips, J. (2013)
UWL	- *University Word List* - 836 word families: Academic focus - with overlap with the NGSL - Xue, G. & Nation, P. (1984)
TSL	- *TOEIC Service List* 1.2 - 1,250 word families - *NGSL* + *TSL* = 98.5% TOEIC coverage - Browne, C., Culligan, B., & Phillips, J. (2016)
New Jacet 8000	- *New JACET List of 8000 Basic Words* - 8,000 words: for Japanese learners - JACET: Japan Association of College English Teachers (2016)

Table 3: Word Lists for Coverage Calculation

Procedure & Software

▶
- Step 1: A corpus of the bounds of 70 MEXT-approved English textbooks was created using *FileMaker Pro 2024.*
- Step 2: The corpus was lemmatized using *TagAnt* (Anthony, 2024).
- Step 3: Word lists were generated using *AntConc* (Anthony, 2024).
- Step 4: Vocabulary coverage was calculated using multiple wordlists (cf. Table 2).

$$Coverage\ [\%] = \frac{Textbook\ word\ types\ in\ Word\ List}{Word\ List\ size} \times 100$$

Results

▶**Comparison of word types and tokens by textbook series**

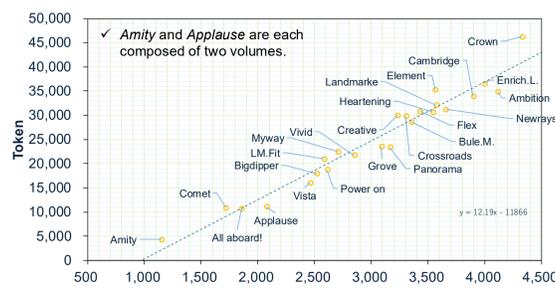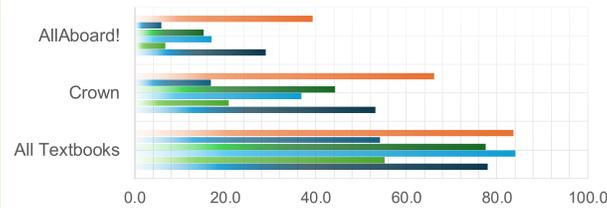✓ *Amity* and *Applause* are each composed of two volumes.

$y = 12.19x - 11866$

Figure 3: Token–Type Relationship

- Tokens and types are correlated. (r=0.98)
- *Crown* had the most types and tokens
- *All Aboard!* had the fewest tokens.
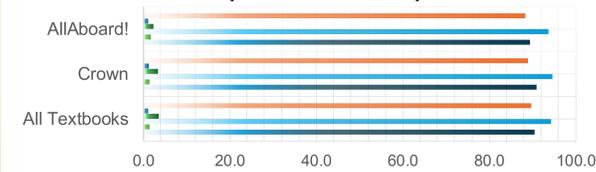 ➔ΔTokens: 35,474

▶ **Coverage Five Word list & Tests**
◆ **Word Types (lemma-based)**

	All Textbooks	Crown	All Aboard!
NGSL	83.7	66.2	39.2
NAWL	54.1	16.8	5.9
UWL	77.5	44.3	15.2
JACET	84.1	36.8	17.0
TSL	55.2	20.7	6.8
Tests	77.9	53.1	29.0

Figure 4: Type Coverage of Textbooks by Word List

◆ **Word Tokens (lemma-based)**

	All Textbooks	Crown	All Aboard!
NGSL	89.7	89.0	88.2
NAWL	1.0	1.2	1.0
UWL	3.6	3.4	2.3
JACET	94.4	94.6	93.8
TSL	1.4	1.4	1.6
Tests	90.5	90.6	89.4

Figure 5: Token Coverage of Textbooks by Word List

▶ **Word List Words Not Found**

Wordlist	Examples
NGSL	Academic Vocabulary: expansion, cite Social/Political/Legal Vocabulary: agenda, poll Emotive/Ethical Vocabulary: abortion, emphatic Formal/Archaic Expressions: whilst, wed Mathematical/Scientific Vocabulary: formula, criteria
NAWL	Natural Sciences: chloride, amino, vector Linguistics / Cognitive Science: syntax, discourse, lexical Psychology / Medicine: syndrome, psychiatric, fetus Social / Political Sciences: nationalism, communism
UWL	Abstract & Logical Concepts: ambiguity, intrinsic Academic Argumentation: formulate, refute, stipulate Disciplinary Terms: cylinder, inflation, syntax Derivational / Formal Vocabulary: emancipate, suffice Evaluative Adjectives: empirical, maternal, pertinent
TSL	Business & Office: payroll, invoice, contractor Formal Everyday Items: cookbook, concierge, wristwatch Procedures & Systems: registration, validation, warranty Practical Evaluation: applicable, quarterly, overdue Compounds & Derivatives: paperback, photocopy

Figure 6: Word List Words Not Found

Vocabulary Less Likely to Appear in Textbooks
- Abstract, academic, or technical terms
- Words related to society, politics, medicine, or business
- Archaic or formal expressions
- Less frequent vocabulary items

Discussion

1. Large differences in vocabulary size and coverage among textbooks
   → Teachers can select materials suited to students' needs, but such differences may affect students' vocabulary knowledge.

2. All textbooks show high *NGSL* coverage and low *UWL/NAWL* coverage
   → NGSL is a useful measure for evaluating textbooks.
   → Differences in the number of NGSL types and the proportion of NGSL tokens may lead to differences in vocabulary acquisition among L2 learners.

3. In *All Aboard*!, overall coverage exceeds 80% in tokens for the *NGSL*, but Type coverage is around 40%, indicating a limited range of basic vocabulary exposure.
   → Supplemental materials should be provided according to students' future paths.

4. Words not appearing in the textbooks
   → Textbook development may prioritize high frequency and practical vocabulary.

Conclusion

- The analysis of multiple vocabulary lists suggests that the NGSL can serve as an indicator for evaluating textbook vocabulary.
  → The NGSL is also an effective list for L2 learners.
- Existing textbooks do not adequately support acquisition of academic or technical vocabulary.
  → Supplementary materials and instruction are essential.

Reference

- Anthony, L. (2024). *AntConc* (v4.3.1) [Computer software]. Waseda University.
- Anthony, L. (2024). *TagAnt* (v2.1.1) [Computer software]. Waseda University.
- Browne, C., & Culligan, B. (2016). *TOEIC service list (TSL) 1.1* [Word list]. Licensed under Creative Commons Attribution-ShareAlike 4.0 International License. https://www.newgeneralservicelist.com/toeic-service-list
- Browne, C., Culligan, B., & Phillips, J. (2013). *New academic word list (NAWL)* [Word list]. Licensed under Creative Commons Attribution-ShareAlike 4.0 International License. https://www.newgeneralservicelist.com
- Browne, C., Culligan, B., & Phillips, J. (2023). *The new general service list (NGSL)* [Online database]. https://www.newgeneralservicelist.com
- Ministry of Education, Culture, Sports, Science and Technology (Ed.). (2023). Explanation of the Course of Study for Senior High Schools: Foreign Languages – English.
- JACET Basic Word Revision Committee (Ed.). (2016). *The new JACET list of 8000 basic words*. Kirihara Shoten.
- Nakayama, S. (2022). *A close examination of vocabulary in Japanese EFL textbooks*. JALT2021 Proceedings: Reflections & New Perspectives, 24–33. https://jalt-publications.org/sites/default/files/pdf-article/jalt2021-pcp-024.pdf
- Tani, K. (2008). A comparative analysis of vocabulary in the National Center Test for University Admissions and high school English textbooks: From the viewpoint of coverage. Practical English Studies, 14, 47–55.
- Tono, Y. (2014). *CEFR-J wordlist (Version 1.6)* [Database]. Tokyo University of Foreign Studies, Laboratory for Linguistic Studies. https://www.cefr-j.org/
- Xue, G., & Nation, P. (1984). *The university word list (UWL)* [Word list]. https://webhome.auburn.edu/~nunnath/engl6240/wlistuni.html

Contact

- Language Science Laboratory, Tokyo University of Pharmacy and Life Sciences
- 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan
- E-mail: s218006@toyaku.ac.jp
- Shunsuke Ito (M1)