



A Corpus-Based Analysis of the Year 5 Malaysian Primary School Textbook

Fatin Qaisara Rozaimi¹*, Azwin Hasyana Azhar², Nur Fadhila Ahmad³,
Muhammad Fazli Abdullah⁴ and Norwati Roslim⁵

^{1,3} Faculty of Education, Social Science and Humanities, University Poly-Tech Malaysia,
Cheras, Malaysia

² School of Education, MAHSA University, Jenjarom, Malaysia

⁴ School of Liberal Studies, Universiti Kebangsaan Malaysia, Bangi, Malaysia

⁵ Academy of Language Studies, Universiti Teknologi MARA, Rembau, Malaysia
qaisara@uptm.edu.my

Abstract. Textbooks are important teaching resources that are commonly utilized by English language teachers in Malaysian primary schools. It is imperative that lexical items presented in the textbooks are aligned well with English as a second language (ESL) students' proficiency level to support language learning. Therefore, the evaluation of these lexical items has become increasingly important in the development of English textbooks. This study examines the lexical features of the Year 5 Malaysian primary school English textbook through a corpus-based approach. The objectives of the study are to identify the textbook's CEFR-aligned vocabulary items and to explore the textbook's thematic and lexical patterns, by conducting keyword and wordlist analyses. Data analysis includes the compilation of a textbook corpus and the analyses of word frequency, lexical composition, and content themes; all of which are used to determine the textbook's linguistic appropriateness for young ESL learners. The findings revealed that the textbook contains predominantly high-frequency function words and beginner-level vocabulary items. Additionally, the thematic analysis indicated the need for the integration of global competencies into English language instruction. However, notable advanced or content-specific vocabulary in the textbook points to challenges for CEFR Level A1 and A2 learners. Without appropriate linguistic scaffolding and textbook adjustments, their vocabulary acquisition might be significantly impaired. These findings suggest the need for strategic vocabulary input to ensure effective language acquisition. The study highlights the importance of balancing content richness with lexical alignment in the development of English textbooks that can support meaningful language learning experiences.

Keywords: English Textbook, Vocabulary Acquisition, Corpus Analysis, CLIL.

1 Introduction

Textbooks are a key resource in the language learning process, especially for vocabulary acquisition. They provide structured exposure to new words in meaningful contexts, supporting both language recognition and usage. Through consistent

© The Author(s) 2026

N. A. Ishak et al. (eds.), *Proceedings of the International Conference on Cross-Disciplinary Academic Research 2025 - Track 3 Advances in Humanities, Education, Teaching, Learning and Arts (ICAR-T3 2025)*, Advances in Social Science, Education and Humanities Research 1006,

https://doi.org/10.2991/978-94-6239-662-3_10

interaction with textbook content, students gradually build a foundational vocabulary that is crucial for language development. Corpus-based analyses such as keyword extraction and wordlist comparisons enable researchers and educators to systematically assess the textbook's lexical coverage, frequency distribution, thematic relevance, and alignment to the Common European Framework of Reference (CEFR) proficiency levels (Loewen & Sato, 2018; Martínez, 2023). These approaches not only reveal whether the vocabulary load is developmentally appropriate but also inform the degree to which the lexical input supports learners' progression from receptive to productive competence.

The integration of corpus-based methods in textbook evaluation is vital to English language curriculum planning, especially in contexts where international standards such as the CEFR are adopted (Zou et al., 2021). As Malaysia continues its CEFR-aligned English language education reform, the lexical content of ESL textbooks warrants critical scrutiny to ensure that it matches the cognitive and linguistic readiness of learners, especially at the primary level (Kementerian Pendidikan Malaysia, 2020).

In this study, the researchers aim to investigate the lexical demands of a Year 5 Malaysian ESL textbook by conducting corpus-based analyses of wordlist frequency and keyword distribution. By benchmarking the lexical items against CEFR level A1 and A2, the study evaluates whether the textbook provides sufficient scaffolding for learners who are expected to operate within the range of beginner-level English language proficiency. Specifically, the following research questions guide the study:

1. What vocabulary in the textbook corpus aligns with CEFR level A1 and A2 based on wordlist analysis?
2. What are the thematic and lexical patterns revealed through keyword analysis?

These inquiries are crucial as previous studies have shown that a misalignment between textbook vocabulary and learners' language proficiency can impede language acquisition, especially when learners are not repeatedly exposed to high-frequency words or when thematic content lacks coherence (Dang & Webb, 2020; Nation, 2022). In Malaysia, where English is taught as a second language, ensuring lexical accessibility is not merely a pedagogical concern but a policy priority. By leveraging tools from corpus linguistics and referencing established lexical frameworks such as the English Vocabulary Profile (Council of Europe, 2020), this study contributes empirical evidence toward the validation and enhancement of CEFR-aligned teaching materials.

2 Literature Review

2.1 Corpus-based Textbook Evaluation

Over the past two decades, corpus-based methods have become increasingly valuable for evaluating the lexical and thematic content of ESL textbooks. Textbook corpora, when analyzed using tools like frequency lists, keyword analyses, and lexical profiling, can reveal whether the materials support incremental vocabulary acquisition

and learner progression (Nation, 2022; Dang & Webb, 2020). For example, Martínez (2023) emphasizes that corpora not only help measure the lexical load of materials but also allows for comparison with benchmarked levels, such as those in the CEFR.

In recent years, several studies have adopted this methodology to assess the linguistic appropriateness of textbooks for ESL learners, particularly in countries aligning their national curricula with the CEFR (Zou et al., 2021). In the Malaysian context, this becomes especially relevant as the Ministry of Education emphasizes CEFR-aligned instruction from primary to tertiary education (Kementerian Pelajaran Malaysia, 2020). Corpus-based evaluations thus offer empirical validation of whether the vocabulary demands of ESL textbooks are consistent with learners' actual proficiency levels.

Moreover, multiple studies examining lexical frequency highlighted that learners need multiple exposures to high-frequency words to achieve mastery. Nation (2022) argues that a vocabulary size of at least 2,000–3,000-word families is required for effective reading comprehension at the A2–B1 level. Therefore, the textbook designed for Year 5 students which is typically situated at CEFR level A1 and A2 must carefully select and recycle core vocabulary to build learner readiness. Additionally, studies by Alali and Schmitt (2012) and Beglar (2005) which utilized the WordSmith tool to analyze learners' vocabulary profiles through their essay writings concluded that ensuring high frequency of vocabulary that is cognitively appropriate for a particular CEFR level is fundamental in improving students' overall language proficiency.

2.2 CEFR-alignment and Lexical Benchmarks

The CEFR provides internationally recognized guidelines for defining language proficiency across six levels (A1 to C2). Recent expansions of lexical resources such as the English Vocabulary Profile (Council of Europe, 2020) have enabled more precise measurement of lexical items associated with each level. This has facilitated corpus-based evaluations of whether textbooks match CEFR expectations in terms of wordlist complexity, thematic range, and communicative intent (Martínez, 2023).

Several Malaysian studies (e.g., Hamid et al., 2021) have noted gaps between textbook claims of CEFR alignment and the actual linguistic content. These findings reinforce the need for systematic corpus analysis to verify alignment, particularly for textbooks used at lower primary levels where vocabulary breadth and cognitive load must be tightly calibrated. For instance, Zou et al. (2021) found that many CEFR-aligned textbooks still contain low-frequency lexical items that exceed the threshold expected of A1–A2 learners, risking learner demotivation and comprehension failure.

2.3 Vocabulary Acquisition and Keyword Profiling

Vocabulary remains one of the strongest predictors of second language proficiency. As Schmitt et al. (2022) highlighted, learners require not just recognition, but deep knowledge of vocabulary meaning, usage, collocation, and register. Keyword profiling through corpus-based methods allows researchers to identify high-frequency thematic words, assess lexical coverage, and examine whether vocabulary repetition is

sufficient for retention (Webb & Nation, 2017). In a textbook evaluation, keyword and wordlist analysis can reveal not only which words are most prominent, but also which thematic areas dominate the content. This contributes to an understanding of semantic density and content distribution, both of which impact learner engagement with the textbook. According to Loewen and Sato (2018), when keyword analysis is integrated with CEFR-linked profiling, educators can gain insights into whether lexical demands promote cognitive development at appropriate levels. The vocabularies found in the keyword and wordlist analysis can be categorized into function words and content words. Function words refer to words that do not carry their own meanings but act as grammatical words that help a sentence make sense, while the latter refers to words that carry semantic meanings and can stand alone (Fromkin, Rodman, and Hyams, 2003). Interestingly, Omarova et al. (2025) stated that function words are more frequently found in spoken language compared to written language, while Burke et al. (2025) have concluded that semantic meaning, or thematic information are carried by content words.

2.4 Content Language Integrated Learning (CLIL) and Corpus Linguistics

The Content and Language Integrated Learning (CLIL) approach promotes the simultaneous learning of subject content and a target language, making it highly relevant in multilingual and multicultural settings like Malaysia. CLIL has been increasingly incorporated into English language education in Southeast Asia, including primary classrooms, where science, moral education, and health topics are embedded into English syllabi (Dalton-Puffer & Smit, 2021). A study by Kruawong & Phoocharoensil (2024) found that CLIL interventions have significantly assisted students' vocabulary comprehension and content knowledge improvements.

Corpus linguistics intersects with CLIL in several important ways. First, corpus-informed material design supports CLIL principles by ensuring that the lexical load of content subjects is manageable for second-language learners (Lo & Lin, 2020). Second, thematic keyword analysis, which is central to this study, can identify whether textbooks achieve a balance between content-specific vocabulary and language development goals. For example, in a Year 5 textbook, science-related topics such as “the water cycle” or “body systems” may introduce low-frequency domain-specific vocabulary. Without proper scaffolding, these may exceed CEFR level A1 and A2 expectations (Martínez, 2023).

Furthermore, the CLIL approach often requires teachers to mediate content using simplified or graded English to promote speaking proficiency among local students in Indonesia (Haq et al. (2024). Corpus-based analyses can assist in identifying lexico-grammatical patterns that align with CLIL's pedagogical goals especially if these patterns are recycled consistently to promote both content mastery and language acquisition (Coyle et al., 2021). By applying corpus tools to CLIL-oriented materials, researchers can also map which language functions (e.g., describing, classifying, hypothesizing) are being developed through thematic units.

In conclusion, recent literature strongly supports the use of corpus-based methodologies to evaluate the lexical and thematic adequacy of ESL textbooks, particularly within CEFR-aligned and CLIL-oriented curricula. Through keyword and wordlist analysis, this study seeks to uncover whether the Malaysian Year 5 English

textbook aligns with CEFR level A1 and A2 expectations, and whether its content scaffolds language acquisition in line with contemporary theories of vocabulary development and integrated learning.

3 Methodology

3.1 Research Design

This study adopts a quantitative, descriptive corpus-based design to examine the linguistic features of an English language textbook currently used in Malaysian primary schools. The design is non-experimental and involves the compilation and analysis of a corpus created from authentic textbook material. The research focuses on identifying the lexical composition and word frequency distribution within the textbook to evaluate its linguistic appropriateness and potential effectiveness for language learning.

3.2 Corpus Compilation Procedure

The textbook selected for this study was the Year 5 Primary School English Textbook, published by the Ministry of Education Malaysia. The textbook is the prescribed instructional material in the national Standards-based English language curriculum (SBELC), thereby ensuring its relevance and representativeness for English language instruction at the primary level. Based on Figure 1, data was collected through text extraction of the entire textbook, where all relevant written content from the textbook, including instructional texts, reading passages, dialogues, exercises, and supplementary materials, were systematically extracted. Following that, data cleaning was conducted where non-linguistic elements such as images, page numbers, and decorative text were excluded from analysis. The textbook was obtained through the Malaysian Ministry of Education website and downloaded in a PDF format. The second step is digitization and formatting of the extracted texts, which include conversion of the scanned textbook images into OCR file format and finally, conversion into plain text (.txt) files. Formatting inconsistencies and typographical errors observed during the digitization process were manually corrected to maintain textual accuracy. Crosschecking of the txt files among the researchers was conducted to minimize interference of non-linguistic elements during the analysis phase. The digitized content was then organized thematically to reflect the original structure of the textbook and ensure the contextual integrity of the textbook material. The finalized text files were imported into WordSmith Tools version 8.0 software for lexical and frequency analysis. The corpus was then processed using the software's wordlist and keyword functions to identify recurring lexical items and patterns.

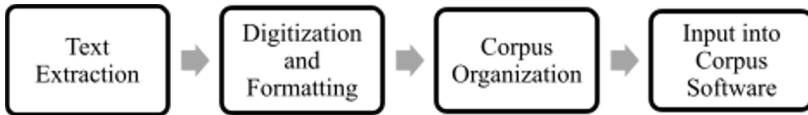


Fig. 1. Corpus compilation procedure.

3.3 Tool for Analysis

The compiled text corpus was processed using WordSmith Tools (version 8), a software that is widely used in corpus linguistics research. The tool facilitated the generation of word frequency lists, key word analysis, concordance lines and type-token ratio.

3.4 Analytical Procedure

The analysis began with a frequency analysis using WordSmith Tools to generate a wordlist and identify the most commonly used words in the textbook. This helped determine the frequency and repetition of lexical items. The next step involved comparing the textbook vocabulary with CEFR level A1 and A2 wordlists to assess whether the content aligns with the language proficiency levels expected of beginner ESL learners. This comparison highlighted the textbook's lexical adequacy and revealed potentially unfamiliar words that may require additional instructional support.

3.5 Reliability and Validity

To ensure the reliability of the corpus analysis, only textual data directly taken from the textbook was included. The use of established corpus software (WordSmith Tools) and recognized wordlists CEFR LEVEL A1 and A2 enhanced the study's validity. Manual checks were conducted during text cleaning to prevent OCR-related errors from skewing the results.

4 Findings and Discussion

4.1 Word Frequency List

Table 1 shows the list of words which are most frequently found in the Year 5 English language textbook. Out of the 11,572-word tokens from the textbook corpus, more than 50% of the words belong to the CEFR level A1 category. The wordlist analysis reveals a high frequency of CEFR level A1 function words. This includes words such as "the" (5.4%), "and" (4.3%), "in" (4.1%), "to" (2.1%), and "is" (1.9%). These findings meet typical English usage pattern expectations and confirm that the lexical content of the textbook is linguistically authentic and suitable for learner-level texts designed specifically for learners in the CEFR A1 or A2 categories. These function words are foundational to sentence construction, supporting learners' acquisition of

grammatical competence. Additionally, high-frequency pronouns like “you” (3.7%), “I” (2.3%), “he” (2.0%), and “it” (1.7%) point to a discourse that includes personal communication or classroom dialogues, likely reinforcing the interpersonal mode of communication emphasized in beginner to intermediate-level ESL instruction.

The prominence of task-related verbs such as “listen” (1.2%), “check” (1.1%), and “complete” (1.0%) suggests that the corpus may have originated from classroom materials or instruction-based texts. These verbs indicate procedural language that supports task-based learning approaches, emphasizing learner autonomy and interaction. Similarly, the presence of prepositions such as “on” (2.5%), “with” (2.2%), “for” (3.3%), and “at” (1.6%) shows a strong emphasis on locative and directional expressions, which are crucial for both everyday communication and classroom instructions.

Notably, the appearance of terms like “pairs” (1.4%), “exercise” (1.3%), and “good” (1.6%) underscores a pedagogical focus on collaborative activities, positive reinforcement, and structured tasks just like in hallmarks of communicative language teaching (CLT). These keywords indicate that learners are engaged in paired or group work, likely to be completing exercises that are scaffolded and reinforced through encouragement. While the list predominantly consists of CEFR level A1 instead of A2 level vocabulary, the range of high-utility words suggests a balanced input that supports both receptive (listening, reading) and productive (speaking, writing) skills development. This result shows that there is a need to have a strong foundation of basic high-frequency and essential words to be proficient English language users.

Table 1. Word frequency analysis.

Word	Frequency	% of total tokens	Word	Frequency	% of total tokens
THE	690	5.67%	THEN	58	0.48%
AND	367	3.02%	WRITE	56	0.46%
IN	364	2.99%	PHRASES	54	0.44%
YOU	198	1.63%	WHEN	54	0.44%
TO	182	1.50%	LISTEN	53	0.44%
I	166	1.36%	CAN	52	0.43%
ARE	141	1.16%	SHE	51	0.42%
YOUR	135	1.11%	COMPLETE	47	0.39%
IS	131	1.08%	FROM	47	0.39%
OF	129	1.06%	IT'S	45	0.37%
IT	104	0.85%	LIKE	45	0.37%
WITH	103	0.85%	WORDS	44	0.36%
AT	94	0.77%	DID	43	0.35%
HE	93	0.76%	KEY	43	0.35%
THERE	87	0.71%	SENTENCES	43	0.35%
ON	86	0.71%	HOW	42	0.35%
ABOUT	79	0.65%	SCHOOL	42	0.35%

USE	79	0.65%	PAST	41	0.34%
WHAT	75	0.62%	SPORTS	41	0.34%
OR	73	0.60%	WORK	41	0.34%
DO	72	0.59%	GOOD	40	0.33%
FOR	72	0.59%	LOOK	40	0.33%
WAS	64	0.53%	THEY	40	0.33%
MY	62	0.51%	THINK	40	0.33%
QUESTION	58	0.48%	GOT	39	0.32%
S					

*(Assume total word tokens = 11,572 in the textbook corpus)

4.2 Keyword List

The keyword analysis was conducted to identify the thematic and lexical patterns of the textbook. WordSmith Tools is used to compare the Year 5 English textbook corpus with the CEFR level A1 and A2 reference corpus. The comparison aimed to determine which lexical items are significantly more frequent in the textbook when compared to reference CEFR A1 and A2 levels, thus indicating possible lexical challenges for primary ESL learners. For example, function words such as “the”, “and” and “in” are CEFR level A1 while content words such as “look”, “school” and “work” are CEFR level A2 (Capel, 2010), which several words are also found in the Year 5 English Textbook. Based on Table 2, the keyword analysis of the Year 5 English textbook corpus, when compared against the CEFR level A1 and A2 reference corpus, has revealed a pronounced thematic pattern on environmental awareness. This thematic orientation aligns with broader educational aims to integrate 21st-century global competencies such as environmental stewardship and sustainability into English language instruction. Such integration reflects current curricular trends that seek to foster both linguistic proficiency and global citizenship among young learners. This is also evidenced in the Standard-Based Curriculum and Assessment Document for Primary School and Year 5 Scheme of Work that includes environment as one of the themes that the students are expected to learn.

A significant proportion of the keywords identified (pollution, recycle, protect, damage) fall outside the scope of the CEFR level A1 core vocabulary. These keywords were categorized as CEFR level A2 (Capel, 2010) when computed into the Oxford Learners Dictionaries website (Oxford University Press, n.d.). These lexical items, while thematically relevant, may exceed the expected proficiency level of beginner ESL learners. These deliberate lexical inclusions suggest an underlying assumption by the textbook developers that learners can navigate content-rich vocabulary independently or with minimal support. However, for learners at the A1 level, this assumption may not hold true. Therefore, lexical scaffolding strategies such as pre-teaching key vocabulary, providing multimodal input, or using simplified definitions are crucial to ensure better reading comprehension and higher learning effectiveness.

Furthermore, some of the extracted keywords are not only present within reading passages but also embedded in task instructions and activity prompts. This dual presence increases the cognitive load for learners, as they must decode both the content and the instructions without necessarily having prior exposure to such lexis. This calls for deliberate attention in instructional design, particularly in the sequencing of vocabulary introduction and recycling.

The presence of content-specific terminology (e.g., plastic, litter, recycle) is also indicative of a Content and Language Integrated Learning (CLIL) approach, wherein language is used as a medium to access subject-specific knowledge. While this pedagogical approach has been recognized for promoting deeper engagement and meaningful language use, it necessitates a careful balance between content complexity and linguistic accessibility. In the absence of such balance, learners may experience frustration or disengagement due to lexical overload.

In summary, while the inclusion of environmental topics and domain-specific vocabulary enriches the learning experience and reflects progressive pedagogical priorities, it is imperative that language input remains lexically appropriate, explicitly supported, and contextually reinforced, especially for learners operating within the CEFR level A1 and A2 proficiency bands.

Table 2. Keyword analysis.

Keyword	Observed frequency	CEFR A1/A2 expected frequency	Notes
recycle	High	Low	Topic-specific; environmental education focus
ocean	High	Low	Reflects thematic content (science/environment)
pollution	High	Low	Content-rich word; may challenge lower-level learners
harmful	High	Rare	Low-frequency adjective at A2/B1 level
protect	High	Moderate	Common verb in environmental topics
litter	High	Low	Culturally relevant term; possibly unfamiliar to some learners
environment	High	Low	Central theme of the unit
save	Moderate	Common	Polysemous verb; requires context clarification
plastic	High	Rare	Domain-specific vocabulary
damage	Moderate	Low	Higher-level verb for A1 learners

4.3 Keyword Analysis Word Cloud

Further findings for Research Question 2 can be found in the keyword analysis word cloud shown in Figure 2. The word cloud reveals a high frequency of function words such as the, and, in, to, is, and of, which are foundational to sentence structure and common in beginner-level English materials. Their presence aligns with CEFR

level A1 and A2 texts, indicating that the material is linguistically accessible for second language learners. In addition, several instructional and task-related verbs (listen, check, complete, do, can) suggest that the materials are activity-driven and likely adopt a task-based or communicative approach to language learning. This focus on functional classroom language supports learner autonomy and facilitates understanding of instructions, especially when paired with visual aids or modelled demonstrations.

Other recurring words such as good, pairs, very, exercise point to a classroom context that emphasizes collaborative learning and positive reinforcement. However, the presence of a misspelled word (that for that) may reflect common learner errors, highlighting the need for spelling awareness activities. Collectively, the keyword patterns suggest that the materials are well-structured for beginner ESL learners, blending grammar exposure with instructional clarity. Pedagogically, it is recommended that teachers explicitly teach task-related verbs, reinforce core function words through contextualized practice, and implement spelling correction strategies. Such approaches ensure both linguistic development and comprehension support within a low-proficiency ESL learners context.



Fig. 2. Keyword analysis word cloud.

5 Conclusion

This study highlights the significant role of keyword analysis in identifying high-frequency and topic-specific vocabulary that can support ESL learners in a CLIL-based classroom. The findings outline the importance of incorporating task-focused vocabulary lists that are derived from keyword analysis as it provides targeted lexical support for ESL learners. This research also advocates for a structured, corpus-based approach to vocabulary development within the textbook syllabus.

Future studies should also consider findings from past studies that support the use of form-focused instruction and task-based learning as effective methods to teach lexical items, especially in CLIL settings. Tailoring these approaches for primary-level ESL learners requires attention to developmental stages, cognitive load, and engagement strategies. To further enhance vocabulary acquisition, it is recommended that educators and stakeholders refine word lists based on text analysis and learner's needs. As vocabulary remains a cornerstone for both language proficiency and content mastery especially in CLIL classrooms, such integrative strategies are vital for fostering long-term lexical growth and academic success among young ESL learners.

Acknowledgments. We would like to thank the University of Poly-Tech Malaysia for providing us with the UPTM Internal Research Grant (URG). It has provided indispensable financial, technical, and logistical assistance which enabled the research project to be conducted smoothly.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

- Alali, F., & Schmitt, N. (2012). A corpus-based study of the most frequent vocabulary in the English Vocabulary Profile. *English Profile Journal*, 3(1), 1–21. <https://doi.org/10.1017/S2041536212000045>
- Beglar, D., & Hunt, A. (2005). Thirty years of vocabulary teaching in English language teaching: A review and some recommendations. *Language Teaching*, 38(4), 279–302. <https://doi.org/10.1017/S0261444805003130>
- Burke, E., Patrick, K., Hamrick, P., & Gunstad, J. (2025). Effects of normal cognitive aging on spoken word frequency: Older adults exhibit higher function word frequency and lower content word frequency than young adults. *The Open Psychology Journal*, 18, Article e18743501354793. <https://doi.org/10.2174/0118743501416115251129091125>
- Capel, A. (2010). A1–B2 vocabulary: Insights and issues arising from the English Vocabulary Profile. *English Profile Journal*, 1(1), e3. <https://doi.org/10.1017/S2041536210000048>
- Council of Europe. (n.d.). *English Vocabulary Profile*. <https://www.englishprofile.org/>
- Coyle, D., Hood, P., & Marsh, D. (2021). *CLIL: Content and language integrated learning*. Cambridge University Press.

- Dalton-Puffer, C., Nikula, T., & Smit, U. (2010). *Language use and language learning in CLIL classrooms*. John Benjamins Publishing Company
- Dalton-Puffer, C., & Smit, U. (2021). Content and language integrated learning: A research agenda for language education. *Language Teaching*, 54(1), 97–114.
- Dang, T. N. Y., & Webb, S. (2020). Making an essential word list for learners of English. *Language Teaching Research*, 24(5), 645–666
- Fromkin, V., Rodman, R., & Hyams, N. M. (2003). *An introduction to language* (7th ed.). Thomson/Heinle.
- Hamid, B. A., Yamat, H., & Subramaniam, G. (2021). Lexical alignment in Malaysian primary ESL textbooks: A CEFR perspective. *Malaysian Journal of ELT Research*, 17(2), 1–18.
- Haq, Z., Purnawan, A., Thamrin, A. H., & Anggrainy, F. P. N. (2025). The implementation of CLIL (Content and Language Integrated Learning) to promote students' speaking at SMA Tumbuh Yogyakarta. *International Journal of Multicultural and Multireligious Understanding*, 12(3), 435–447. <http://dx.doi.org/10.18415/ijmmu.v12i3.6598>
- Kementerian Pendidikan Malaysia. (2020). *English Language Education Reform in Malaysia: The Roadmap 2015–2025*. Ministry of Education Malaysia.
- Kruawong, T., & Phoocharoensil, S. (2024). Enhancing science vocabulary and content knowledge of Thai EFL students through Content and Language Integrated Learning (CLIL) and English science textbook word lists (ESTWL). *rEFLections*, 31(1), 275–294. <https://doi.org/10.61508/refl.v31i1.272797>
- Lo, Y. Y., & Lin, A. M. Y. (2020). Language across the curriculum and CLIL in English as an additional language (EAL) contexts: Theory and practice. *Language, Culture and Curriculum*, 33(3), 224–236.
- Loewen, S., & Sato, M. (2018). *The Routledge handbook of instructed second language acquisition*. 1st edn. Routledge, New York.
- Martínez, R. (2023). Evaluating the CEFR-aligned lexical load of ELT textbooks using corpus tools. *International Journal of Applied Linguistics*, 33(1), 112–130 (2023)
- Nation, I. S. P. (2022). *Learning vocabulary in another language*. 3rd edn. Cambridge University Press.
- Omarova, S., Ospanova, D., Aitova, N., Tokenkyzy, G., Ormanova, A., & Alshynbekova, M. (2025). A corpus approach in language discovery: A word frequency analysis based on the corpus outcomes in Kazakh. *Forum for Linguistic Studies*, 7(2), 869–881. <https://doi.org/10.30564/fls.v7i2.8317>
- Oxford University Press. (n.d.). *Oxford Learner's Dictionaries*. <https://www.oxfordlearnersdictionaries.com>
- Schmitt, N., Jiang, X., & Grabe, W. (2022). The percentage of words known in a text and reading comprehension: A meta-analysis. *The Modern Language Journal*, 95(1), 26–43.
- Webb, S., & Nation, P. (2017) *How vocabulary is learned* (2nd ed.). Oxford University Press.
- Zou, D., Wang, D., & Xing, M. (2021). A review of corpus-based textbook evaluation studies: Implications for CEFR-aligned material development. *Language Teaching Research*, 25(2), 267–289.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

