LEKSIKA
Jurnal Bahasa, Sastra, dan Pengajarannya
Volume 20, Number 2, Mar 2026 – Jun 2026
https://dx.doi.org/10.30595/lks.v20i2.30564

Eco-lexis in Indonesian children's narratives: A mini-corpus study of frequency, range, and collocations

[*]Nur Melansari [1,2], Khaerun Nisa'a Tayibu [1,3], Raviqa Raviqa [1,4], Madani Madani [1], Andi Muhammad Syafri Idris [1]

[1] *Department of Linguistics, Faculty of Cultural Sciences, Hasanuddin University, Makassar, 90245, South Sulawesi, Indonesia*
[2] *English Language Education Department, Faculty of Teacher Training and Education, Universitas Dayanu Ikhsanuddin, Baubau, 93721, Southeast Sulawesi, Indonesia*
[3] *Department of Elementary School Teacher Education, STKIP Andi Matappa, Pangkajene Kepulauan, 90611, South Sulawesi, Indonesia*
[4] *Department of Marine Transportation Management, Politeknik Ilmu Pelayaran, Makassar, 90225, South Sulawesi, Indonesia*

Abstract

Children's stories are a powerful medium for introducing environmental awareness while simultaneously supporting literacy development. Yet, empirical descriptions of the environment-related vocabulary that children encounter in Indonesian narrative texts remain limited, particularly those drawn from openly accessible digital libraries. This study addresses that gap by conducting a corpus-based descriptive lexical analysis of eco-lexis in a mini-corpus of Indonesian children's narratives sourced from Let's Read (The Asia Foundation). The dataset comprises 10 narratives (N = 10; 4,886 cleaned tokens) spanning Reading Levels 1–5, selected using purposive stratified sampling with inclusion criteria of Indonesian language, narrative genre, environmental focus, and Creative Commons licensing (CC BY or CC BY-NC). The analysis integrates three complementary procedures: (1) lexical profiling through raw and normalized frequency (per 1,000 words), (2) distributional profiling using range (number of texts containing an item) as a pragmatic proxy for dispersion in a small corpus, and (3) phraseological analysis through recurrent bigrams/short phrases around key eco-lexis items. Eco-lexis identification combines a seed list with corpus-driven refinement and light manual lemmatization to group closely related variants. Results show that eco-lexis is organized around high-salience environmental entities (e.g., hutan, pohon, ikan, lumba-lumba, karang, udang, air, sungai, bakau), supported by recurring problem and action vocabulary that frames everyday stewardship (e.g., sampah, plastik, buang, bersih, jaga). Range patterns differentiate a compact core set (widely distributed across texts) from topic-bound clusters concentrated in one or two narratives (e.g., mangrove ecology and forest governance). Phrase-level patterns further reveal how narratives encode environmental responsibility through reusable "chunks," including disposal routines (e.g. tempat sampah, membuang sampah) and protection frames (e.g. menjaga hutan, menjaga kelestarian). Overall, the study offers a transparent, lightweight workflow for eco-lexis profiling and provides an empirically grounded basis for selecting story texts and designing low-burden classroom tasks that link vocabulary development with environmental responsibility discourse.

Keywords: *Children's literature; Corpus linguistics; Ecolinguistics; Environmental education; Environmental vocabulary; Semantic mapping*

Article History: Received 18 Apr 2026, Final revision 5 Jun 2026, Published 10 Jun 2026

*Corresponding author: **Nur Melansari**, Address: Universitas Dayanu Ikhsanuddin, Jl. Sultan Dayanu Ikhsanuddin No.124, Betoambari, Baubau, 93721, Southeast Sulawesi, Indonesia, Email: nurmelansari@unidayan.ac.id; melanniiz@gmail.com

Introduction

Climate change is widely recognized as a pressing global challenge, with widespread impacts and risks that require coordinated mitigation and adaptation efforts across societies (Intergovernmental Panel on Climate Change, 2023). Because responses to climate change depend not only on technology and policy but also on public understanding and participation, education becomes a strategic lever for building long-term climate action. In the UNFCCC framework, Action for Climate Empowerment (ACE) explicitly includes climate change education and public awareness as key elements for empowering society to engage in climate action (United Nations Framework Convention on Climate Change, n.d.). In a similar vein, UNESCO frames climate change education as a means to equip people with the knowledge, skills, values, and attitudes needed to act as "agents of change" (The United Nations Educational Scientific and Cultural Organization, n.d.).

Within this educational agenda, children and young learners are not merely passive recipients of information; they are future decision-makers and present-day participants in shaping norms and everyday practices. Formal schooling and literacy activities provide sustained, developmentally appropriate opportunities to introduce environmental responsibility and to connect it with broader civic values. Global Citizenship Education—linked to SDG Target 4.7—positions education as a pathway for developing the knowledge and skills needed to promote sustainable development and responsible global participation (United Nations, n.d.). Middle childhood (roughly ages 7–12) is marked by rapid vocabulary growth. Research suggests that, during the school years, repeated exposure through reading and written context can support substantial incidental vocabulary learning (Joseph & Nation, 2018).

Children's literature has long been discussed as a meaningful site for environmental learning, spanning ecocritical to ecopedagogical concerns (Gaard, 2009). In these traditions, children's environmental texts are treated not simply as "messages" about nature but as literary works that can foster ethical reflection, environmental justice awareness, and imaginative engagement with the non-human world (Gaard, 2009). More recent scholarship suggests that "green" children's books increasingly foreground action orientations—for instance, by including explicit tasks or recommendations that invite young readers to act (or at least imagine acting) in response to climate and environmental problems (Rybak, 2025). At the same time, ecocritical work on picturebooks highlights that environmental learning is mediated by narrative, focalization, paratext, and multimodal design, which can cultivate wonder and reflection rather than only transmitting factual information (Laliena & Sala, 2023). Complementing these perspectives, research on climate-related children's books notes their educational promise alongside the need for more systematic empirical attention to children's climate change literature (Benevento, 2023). Together, these discussions suggest that stories can carry environmental meanings and responsibilities in ways that remain accessible to young readers—especially when texts are culturally and linguistically close to children's lived worlds.

In the Indonesian context, research on children's stories with environmental concerns has often been conducted through qualitative ecocritical readings that foreground themes such as pollution, nature protection, and environmental ethics. For example, an ecocritical study analyzes environmental care in a children's story using Greg Garrard's ecocritical categories (e.g., pollution, shelter, earth) (Hamida & Indrastuti, 2025). Other ecocritical work similarly emphasizes how Indonesian children's fiction encodes values such as "respect for nature" and environmental wisdom/ethics (Lanta et al., 2022). While these studies demonstrate the relevance of children's literature for environmental education, they typically prioritize thematic and moral interpretation over a systematic account of the linguistic resources through which environmental meanings are built and repeated across texts.

However, environmental learning in stories is not only carried by themes and morals; it is also shaped by recurring lexical and phraseological choices. From an ecolinguistic perspective, patterned language contributes to the "stories we live by," influencing how relationships between humans and the non-human world are framed and evaluated (Stibbe, 2015). Ecolinguistics is also framed more broadly as an interdisciplinary field that examines language in its ecological context, including how discourse contributes to life-sustaining (or life-diminishing) interactions (MDPI, n.d.). Within this orientation, mapping eco-lexis (environment-related vocabulary) provides a practical lens for describing how narratives linguistically encode environmental entities, problems, solutions, and responsibilities. Eco-lexis does not only label environmental referents; it can repeatedly pair environmental nouns with evaluative terms and action verbs, thereby cueing responsibility, agency, and everyday stewardship within the story world.

A corpus-informed approach offers tools for making such lexical patterning visible and reportable. Corpus linguistics provides methods for investigating language patterns using machine-readable text collections, commonly framed as computer-aided empirical analysis of text data used to address research questions (McEnery & Hardie, 2012). Although the present study uses a mini-corpus rather than a large balanced corpus, the analytic logic remains similar: frequency profiles and recurring co-occurrence patterns can describe what is salient in a textual domain and support replicable claims about language use. Two corpus-informed concepts are central to this study. The first is distribution/dispersion—the degree to which occurrences of an item are spread evenly across corpus units or clustered in a small subset of texts (Gries, 2021); and the second concept is collocation, which captures characteristic associations words form with other words in recurrent contexts (*Collocations*, n.d.). For children's environmental narratives, collocational patterns are especially informative because they often realize action frames (e.g., verb–noun routines for disposal and care) and evaluative frames (e.g., cleanliness and contamination), which can be interpreted through an ecolinguistic lens as traces of responsibility and agency.

Corpus-based profiling has also been applied to children's texts in relation to pedagogy, for example to identify frequent content vocabulary for learning purposes, demonstrating that corpus methods can meaningfully characterize language in children's storybooks (Maryani, 2011). However, such work has not typically focused on a targeted eco-lexis inventory combined with frequency, range, and collocational patterning in Indonesian children's environmental narratives. This leaves space for a transparent, replicable lexical account that complements existing theme-based ecocritical readings by showing how environmental meanings and responsibilities are linguistically patterned in narrative input.

This study addresses that gap using an openly accessible digital reading ecosystem. To build a transparent and replicable mini-corpus of Indonesian children's narratives, this study uses Let's Read, an initiative of The Asia Foundation that describes itself as a free digital library for children in Asia and the Pacific and prioritizes locally produced stories in children's own languages and culturally familiar settings (The Asia Foundation, n.d.). Beyond accessibility, Let's Read is methodologically useful for corpus work because each title is accompanied by a dedicated book page that reports key metadata (e.g., reading level, available languages, country of origin, and license type), allowing systematic sampling, documentation, and later verification (Let's Read, n.d.-b, n.d.-e). This metadata-rich structure supports consistency in selecting the Indonesian version of each story (when multiple language versions exist) and in tracking corpus composition across levels (Let's Read, n.d.-b). Finally, the study's focus on Creative Commons–licensed titles (CC BY and CC BY-NC) provides an explicit legal-ethical basis for non-commercial research reuse: CC licenses permit sharing and adaptation with attribution, while CC BY-NC additionally restricts commercial reuse (Creative Commons, n.d.-a, n.d.-b).

Building on this corpus construction and documentation strategy, the study investigates how environment-related vocabulary is distributed and patterned in Indonesian children's narratives. Accordingly, the study is guided by three research questions:

1) RQ1: Which environmental vocabulary items (eco-lexis) are most frequent in a mini-corpus of Indonesian children's narratives?

2) RQ2: How are the eco-lexis items distributed across texts (i.e., which items show wider range versus topic-bound concentration)?

3) RQ3: What collocational patterns occur around key eco-lexis items, and how do these patterns contribute to action- and responsibility-oriented environmental meanings in the narratives?

Method

Research design

This study employs a corpus-based descriptive lexical approach using a mini-corpus of Indonesian children's narratives sourced from an open-access digital library. Small, carefully curated specialized corpora are widely used when the research goal is depth and contextual specificity rather than broad statistical generalization, enabling transparent documentation and close qualitative interpretation alongside quantitative summaries (Koester, 2022; Vaughan & Clancy, 2013). The dataset comprises 10 texts (N = 10), enabling close and transparent documentation as well as manual verification of counts and contexts. Accordingly, findings are interpreted descriptively and illustratively rather than as statistically generalizable, with range (the number of texts in which an item occurs) used as a practical proxy for dispersion in a small corpus. The analysis combines (i) lexical profiling (raw and normalized frequency), (ii) distributional profiling via range-based dispersion, and (iii) phraseological analysis through recurrent bigrams/short phrases surrounding key eco-lexis items (collocational patterning). Here, dispersion is understood as the extent to which occurrences of a word are distributed evenly or unevenly across the corpus (Gries, 2021).

Data source and corpus construction

Texts were collected from Let's Read, a program of The Asia Foundation that provides free access to children's books across languages in Asia and the Pacific (Let's Read, n.d.-a; The Asia Foundation, n.d.). Each title in Let's Read is accompanied by a metadata page (hereafter Book_URL) that displays information such as reading level, language, country of origin, and licensing (Let's Read, n.d.-c, n.d.-d). For each selected title, the story content was accessed via a reading interface (hereafter Read_URL) used for text extraction.

To enable systematic tracking, each story was assigned an internal identifier (Text_ID) following a level-coded scheme (e.g., L1_01–L1_02 … L5_01–L5_02). The corpus comprised 10 unique titles—two narratives per reading level (Levels 1–5)—contingent on the availability of eligible Indonesian-language narratives on the platform.

Sampling and inclusion criteria

Sampling followed a stratified purposive procedure based on the platform's reading level labels (Levels 1–5), in order to avoid over-representing any single difficulty level. For each level, the target was two titles, yielding a 10-text mini-corpus (2 × 5 levels).

A title was included if it met all criteria below:

a. Language: Indonesian (Bahasa Indonesia) version used for analysis.

b. Genre: Narrative children's story (informational/non-narrative texts excluded).

c. Theme: Environmental focus (e.g., waste, pollution, forests, water, cleanliness), where environmental content is central to the narrative rather than incidental.

d. License: Titles licensed under CC BY-NC 4.0 or CC BY 4.0 were included, verified on the Book_URL metadata page. Titles labelled with any NoDerivatives (ND) license (e.g., CC BY-NC-ND 4.0) or marked as copyright/all rights reserved were excluded.

e. Uniqueness: Duplicate titles surfaced by multiple search terms were included only once, using Book_URL/ID as a unique key, with deduplication performed during screening.

Retrieval procedure and deduplication

Candidate titles were located through platform search using environment-related keywords (e.g., *sampah* 'trash/waste', *plastik* 'plastic', *bersih* 'clean', *lingkungan* 'environment', *alam* 'nature', *hutan* 'forest', *sungai* 'river', *laut* 'sea/ocean', *banjir* 'flood', *polusi* 'pollution', etc.). A predefined set of environment-related keywords was prepared and applied iteratively as needed to fill the target quota per reading level. Eligible titles were screened sequentially within each reading level until the quota for that level was met. Duplicates retrieved through different keywords were removed during screening by matching the Book_URL/ID. Only the final selected titles were recorded in the study log and corpus metadata table; excluded candidates were not retained to keep the workflow lightweight and focused on the finalized corpus.

Text extraction and preprocessing (spreadsheet workflow)

For each included title, the story text was copied from the Read_URL interface and pasted into a spreadsheet (one row per text). Preprocessing was conducted to support consistent counting:

a. Cleaning and normalization: removal of non-story elements (e.g., navigation text or platform UI fragments, if present), normalization of whitespace, conversion to lowercase, and normalization of hyphenated forms (hyphens replaced by spaces). As a result, hyphenated lexical items were counted in their cleaned multiword forms (e.g., *lumba-lumba*, 'dolphin').
b. Tokenization: words were operationalized as tokens separated by whitespace after cleaning.
c. Word count: total tokens per text were computed to support normalization.

After automated cleaning, minor manual corrections (e.g., restoring missing spaces, correcting obvious typos, and normalizing a small number of numeric tokens) were applied where needed and documented in the Text Repository sheet to ensure consistent tokenization and counting. Because the corpus contains texts of varying lengths (as expected across reading levels), frequency results were reported using normalized frequency (per 1,000 words) so that texts of different lengths can be compared more accurately (Biber et al., 1998; Xiao, n.d.).

Operationalizing eco-lexis

Eco-lexis was defined as vocabulary items that explicitly denote (a) environmental entities (e.g., *hutan* 'forest', *sungai* 'river'), (b) environmental problems/impacts (e.g., *polusi* 'pollution', *banjir* 'flood'), (c) actions/solutions (e.g., *daur ulang* 'recycle', *mengurangi* 'reduce'), and (d) value-laden stances relevant to environmental responsibility (e.g., *menjaga* 'keep/maintain', *melestarikan* 'preserve').

Eco-lexis identification proceeded in two steps:

a. Seed list development: an initial eco-lexis list was created based on common Indonesian environmental vocabulary likely to occur in children's narratives.

b. Corpus-driven refinement: after initial frequency checks, additional environment-relevant items attested in the corpus were added, while ambiguous items were reviewed in context.

A light manual lemmatization was applied for counting purposes by grouping closely related morphological variants under a single lemma label (e.g., *bersih* 'clean', *kebersihan* 'cleanliness', *membersihkan* 'cleaning' → *BERSIH* 'CLEAN'), recorded transparently in an eco-lexis inventory table.

Analytical procedures aligned with RQ1–RQ3

RQ1: Frequency of eco-lexis

For each eco-lexis lemma, raw frequency was computed (i) per text and (ii) for the corpus overall. To account for different text lengths, normalized frequency was calculated as occurrences per 1,000 words (Biber et al., 1998; Xiao, n.d.).

RQ2: Range (distribution across texts)

To approximate dispersion in a mini-corpus, range was calculated as the number of texts in which a lemma occurred at least once (frequency > 0). This enables a practical distinction between (a) core eco-lexis that appears widely across narratives and (b) topic-bound eco-lexis concentrated in a small number of stories (Gries, 2021).

Given the small corpus size (10 texts), dispersion was operationalized using range. Items were labeled core (range ≥4), semi-core (range =3), and topic-bound (range ≤2) as a pragmatic, data-driven grouping.

RQ3: Recurrent bigrams/phrases (collocational patterning)

Collocational patterning was analyzed using recurrent bigrams/short phrases around key eco-lexis items. In corpus linguistics, collocation is commonly approached as a phenomenon of repeated co-occurrence in texts (Brezina & Gablasova, n.d.; Lancaster University, n.d.).

Procedure:

1) Select a small set of node words (eco-lexis lemmas) based on salience in RQ1 (e.g., top-frequency items such as *SAMPAH* ('trash/waste'), *BERSIH* ('clean'), *AIR* ('water'), *PLASTIK* ('plastic'), *HUTAN* ('forest')).

2) Extract short contexts from each occurrence (KWIC-style lines) in the spreadsheet.

3) Identify recurring two-word or short multiword sequences (e.g., verb–noun action frames such as "*buang sampah*" (throw away the trash), noun–modifier frames such as "*sampah plastik*" (plastic waste)).

4) Count phrase frequencies across the corpus and verify interpretations through concordance context to avoid false positives.

Reliability, transparency, and ethics

To improve transparency, the study records (i) a full corpus metadata table (Text_ID, title, Book_URL, Read_URL, reading level, word count, license, date accessed) and (ii) an eco-lexis inventory table listing

lemma groupings and category assignments. Only final selected titles were archived in the corpus documentation; screening exclusions were not retained. Because the corpus contains titles licensed under Creative Commons CC BY-NC 4.0 and CC BY 4.0, reuse follows the applicable license requirements—attribution in all cases, and non-commercial restrictions where specified (Creative Commons, n.d.-a, n.d.-b). In dissemination, the study prioritizes sharing metadata and derived results (frequency/range/phrase tables and brief illustrative concordance excerpts) rather than redistributing full story texts, which remain accessible through the original Book_URLs.

Results and Discussion

This section presents the Results and Discussion as separate subsections. The Results are reported in a research-question-driven sequence: the Results section begins with a brief corpus overview and key metrics to contextualize the dataset, followed by findings for RQ1 (frequency and normalized frequency), RQ2 (range as a proxy for dispersion in the mini-corpus), and RQ3 (recurrent bigrams/short phrase patterns around selected node words). The subsequent Discussion interprets these patterns in relation to prior work and highlights pedagogical implications for story-based environmental literacy.

Results

Corpus overview and metrics

A 10-text mini-corpus of Indonesian children's narratives (total 4,886 cleaned words) was analyzed using raw frequency, corpus-level normalized frequency (per 1,000 words), and range as a practical proxy for dispersion. Reading levels (1–5) follow the platform's difficulty labels, where Level 1 texts tend to be short and highly repetitive, while higher levels gradually increase in length and lexical/syntactic complexity; this is reflected in the corpus word counts, which range from 95 to 1,655 cleaned words (Table 1). The corpus includes titles licensed under Creative Commons CC BY 4.0 and CC BY-NC 4.0. Both licenses permit reuse with attribution, while CC BY-NC additionally restricts commercial reuse; accordingly, the study reports derived results and short illustrative excerpts rather than redistributing full texts.

Tables 1–4 summarize the corpus composition, the most salient eco-lexis items, their distribution across texts, and recurrent phraseological patterns.

Table 1. Mini-corpus composition and metadata (N=10 Indonesian children's narratives from Let's Read; total words=4,886).

No.	Text ID	Title	Reading Level	Word Count (cleaned)	License	Country of Origin	Publisher	Theme Tag
1	L1_01	Berjalan-jalan di Taman	1	106	CC-BY-4.0	Cambodia	The Asia Foundation - Let's Read	plastic litter
2	L1_02	Moah Sahabat yang Baik	1	95	CC-BY-NC-4.0	Indonesia	UNICEF	deforestation
3	L2_01	Kumpulan Awan Sampah	2	491	CC-BY-4.0	India	Pratham Books	littering and waste management
4	L2_02	Batik Tanah Liek	2	330	CC-BY-NC-4.0	Indonesia	The Asia Foundation - Let's Read	organic waste reuse
5	L3_01	Selamatkan Udang dan Ikan Kami!	3	615	CC-BY-NC-4.0	Not specified	Plan Cambodia	mangrove conservation

6	L3_02	Tas Plastik Milik Tok	3	271	CC-BY-NC-4.0	Cambodia	The Asia Foundation - Let's Read	plastic litter
7	L4_01	Bora dan Teman Lumba-lumbanya	4	505	CC-BY-NC-4.0	Cambodia	The Asia Foundation - Let's Read	river pollution
8	L4_02	Petualangan Menemui Matahari	4	308	CC-BY-NC-4.0	Cambodia	The Asia Foundation - Let's Read	air pollution
9	L5_01	Terumbu Karang yang Memanas	5	510	CC-BY-NC-4.0	Not specified	The Asia Foundation - Let's Read	ocean warming
10	L5_02	Luh Ayu Manik Mas, Sayang Hutan	5	1655	CC-BY-NC-4.0	Indonesia	Not specified	deforestation

Table 1 indicates that the mini-corpus spans a broad thematic spectrum of environmental narratives, ranging from everyday waste and cleanliness practices to ecosystem- and policy-oriented topics such as mangrove conservation, marine ecology, and deforestation. The distribution of reading levels also yields heterogeneous text lengths, which motivates the use of normalized frequency measures in subsequent analyses. In addition, the corpus draws on titles from multiple origins and publishers within the same Indonesian-language dataset, allowing the analysis to focus on lexical patterning in the target language while treating origin and publisher information as contextual metadata.

RQ1: Most frequent eco-lexis items (frequency and normalized frequency)

To address RQ1, eco-lexis items were ranked using corpus-level normalized frequency (per 1,000 words) to reduce the influence of text-length variation across reading levels. Table 2 therefore summarizes the most salient items in the mini-corpus by combining raw frequency (overall occurrence counts), normalized frequency (comparability across texts), and range (how many stories contain the item), providing an initial profile of which environmental meanings are most prominently foregrounded in the narratives.

Table 2. Top eco-lexis items by normalized frequency (per 1,000 words) in the mini-corpus (N=10; total words=4,886).

No.	Lemma	Category	Total Raw	Norm per 1000 (corpus)	Range
1	hutan	entity	32	6.55	2
2	pohon	entity	31	6.34	6
3	ikan	entity	28	5.73	3
4	sampah	problem	24	4.91	4
5	lumba-lumba	entity	21	4.30	1
6	sungai	entity	19	3.89	3
7	air	entity	18	3.68	5
8	kulit	entity	16	3.27	2
9	karang	entity	15	3.07	1
10	tas plastik	entity	15	3.07	3
11	bakau	entity	12	2.46	1
12	buang	action	10	2.05	4
13	udang	entity	9	1.84	1
14	tanah	entity	9	1.84	2

15	terumbu karang	entity	9	1.84	1
16	bersih	value	8	1.64	4
17	tanam	action	8	1.64	4
18	jaga	action	8	1.64	3
19	pantai	entity	8	1.64	3
20	tebang	action	7	1.43	3

Overall, eco-lexis was dominated by entities (e.g., nature and habitat terms) alongside a smaller set of problem and action vocabulary. The most frequent items by normalized frequency included *hutan* (32 tokens; 6.55/1,000), *pohon* (31; 6.34/1,000), *ikan* (28; 5.73/1,000), and *sampah* (24; 4.91/1,000). Water-related and place-based vocabulary was also prominent, such as *air* (18; 3.68/1,000) and *sungai* (19; 3.89/1,000). In addition to these entities, a clear behavioral/action strand appeared among the higher-frequency items, including *buang* (10; 2.05/1,000), *tanam* (8; 1.64/1,000), and *jaga* (8; 1.64/1,000), as well as evaluative/behavioral framing through *bersih* (8; 1.64/1,000). Taken together, the high-frequency profile suggests that the narratives foreground environmental entities as story worlds while repeatedly cueing everyday behavioral routines (e.g., disposal, cleanliness, care).

RQ2: Distribution of eco-lexis across stories (range-based dispersion)

To address RQ2, the analysis moves beyond overall frequency to examine how eco-lexis is distributed across stories. Because a mini-corpus is sensitive to story-level thematic effects, distribution is operationalized using range—the number of texts in which an item occurs at least once—so that widely recurring vocabulary can be distinguished from items concentrated in only a few narratives. Tables 3A–3B present this range-based view alongside normalized frequency to support interpretation of both cross-text recurrence and thematic clustering.

Table 3A. Core and semi-core eco-lexis items by range and normalized frequency.

No.	Lemma	Category	Total Raw	Range	Norm per 1000 (corpus)	Label
1	pohon	entity	31	6	6.34	Core
2	air	entity	18	5	3.68	Core
3	sampah	problem	24	4	4.91	Core
4	buang	action	10	4	2.05	Core
5	bersih	value	8	4	1.64	Core
6	tanam	action	8	4	1.64	Core
7	ikan	entity	28	3	5.73	Semi-core
8	sungai	entity	19	3	3.89	Semi-core
9	tas plastik	entity	15	3	3.07	Semi-core
10	jaga	action	8	3	1.64	Semi-core
11	pantai	entity	8	3	1.64	Semi-core
12	tebang	action	7	3	1.43	Semi-core
13	laut	entity	6	3	1.23	Semi-core

Given the small corpus size (10 texts), items were grouped pragmatically by range: core (range ≥ 4), semi-core (range = 3), and topic-bound (range ≤ 2). Range-based dispersion patterns indicated a small "core" set that recurred across multiple narratives (Table 3A). The most widely distributed item was pohon (range=6), followed by air (range=5). A cluster of everyday environmental behavior vocabulary also showed broad coverage: *sampah, buang, bersih,* and *tanam* each occurred in four texts (range=4),

suggesting shared cross-story emphasis on cleanliness and waste-related actions. Several items were moderately distributed (range=3), including *ikan, sungai, tas plastik, jaga, pantai, tebang,* and *laut*, which can be treated as semi-core vocabulary. In contrast, many high-frequency items were concentrated in only one or two texts (topic-bound; Table 3B). Notably, *hutan* had the highest overall normalized frequency yet a low range (range=2), indicating strong thematic concentration in a small subset of stories. Similar topic-focused clustering appeared for coastal and marine ecology terms such as *lumba-lumba, karang, bakau,* and *terumbu karang* (range=1), as well as specific issue vocabulary (e.g., *banjir, polisi hutan, balak liar*) that likely reflects story-level thematic framing rather than cross-corpus generality. While Table 3A highlights items with wider cross-text recurrence (core and semi-core), Table 3B lists topic-bound items (range ≤ 2) whose low distribution suggests greater dependence on specific story themes, settings, or plot events.

Table 3B. Topic-bound eco-lexis items (range ≤2) in the mini-corpus.

No.	Lemma	Category	Total Raw	Range	Norm per 1000 (corpus)	Interpretation
1	hutan	entity	32	2	6.55	Topic-focused
2	kulit	entity	16	2	3.27	Topic-focused
3	tanah	entity	9	2	1.84	Topic-focused
4	kayu	entity	7	2	1.43	Topic-focused
5	tempat sampah	entity	7	2	1.43	Topic-focused
6	panas	impact	6	2	1.23	Topic-focused
7	nelayan	entity	5	2	1.02	Topic-focused
8	sembarangan	descriptor	5	2	1.02	Topic-focused
9	burung	entity	4	2	0.82	Supporting/rare
10	lindung	action	4	2	0.82	Supporting/rare
11	lingkungan	entity	4	2	0.82	Supporting/rare
12	busuk	problem	3	2	0.61	Supporting/rare
13	hanyut	impact	3	2	0.61	Supporting/rare
14	lestari	action	3	2	0.61	Supporting/rare
15	rusak	impact	3	2	0.61	Supporting/rare
16	air laut	entity	2	2	0.41	Supporting/rare
17	gundul	impact	2	2	0.41	Supporting/rare
18	kantong	entity	2	2	0.41	Supporting/rare
19	pungut	action	2	2	0.41	Supporting/rare
20	rumput laut	entity	2	2	0.41	Supporting/rare
21	lumba-lumba	entity	21	1	4.30	Topic-focused
22	karang	entity	15	1	3.07	Topic-focused
23	bakau	entity	12	1	2.46	Topic-focused
24	terumbu karang	entity	9	1	1.84	Topic-focused
25	udang	entity	9	1	1.84	Topic-focused
26	hutan bakau	entity	7	1	1.43	Topic-focused
27	banjir	impact	5	1	1.02	Topic-focused
28	polisi hutan	entity	5	1	1.02	Topic-focused
29	pabrik	entity	4	1	0.82	Supporting/rare
30	balak liar	problem	3	1	0.61	Supporting/rare
31	bungkus	entity	3	1	0.61	Supporting/rare
32	bungkus biskuit	entity	3	1	0.61	Supporting/rare
33	gelondongan	entity	3	1	0.61	Supporting/rare
34	lumpur	entity	3	1	0.61	Supporting/rare

35	udara	entity	3	1	0.61	Supporting/rare
36	bahan kimia	entity	2	1	0.41	Supporting/rare
37	botol plastik	entity	2	1	0.41	Supporting/rare
38	hilir	entity	2	1	0.41	Supporting/rare
39	lalat	entity	2	1	0.41	Supporting/rare
40	lingkungan hidup	entity	2	1	0.41	Supporting/rare
41	air minum	entity	1	1	0.20	Supporting/rare
42	air pasang	entity	1	1	0.20	Supporting/rare
43	air sungai	entity	1	1	0.20	Supporting/rare
44	asap	problem	1	1	0.20	Supporting/rare
45	babat	action	1	1	0.20	Supporting/rare
46	cangkang	entity	1	1	0.20	Supporting/rare
47	cemar	action	1	1	0.20	Supporting/rare
48	erosi	impact	1	1	0.20	Supporting/rare
49	gotong royong	action	1	1	0.20	Supporting/rare
50	hutan gunung	entity	1	1	0.20	Supporting/rare
51	hutan lindung	entity	1	1	0.20	Supporting/rare
52	kantong plastik	entity	1	1	0.20	Supporting/rare
53	kerang	entity	1	1	0.20	Supporting/rare
54	kering	impact	1	1	0.20	Supporting/rare
55	kotor	value	1	1	0.20	Supporting/rare
56	ladang	entity	1	1	0.20	Supporting/rare
57	mainan	entity	1	1	0.20	Supporting/rare
58	mati	impact	1	1	0.20	Supporting/rare
59	minyak	entity	1	1	0.20	Supporting/rare
60	pabrik karet	entity	1	1	0.20	Supporting/rare
61	pelihara	action	1	1	0.20	Supporting/rare
62	pijah	action	1	1	0.20	Supporting/rare
63	polip karang	entity	1	1	0.20	Supporting/rare
64	polusi	problem	1	1	0.20	Supporting/rare
65	pulih	impact	1	1	0.20	Supporting/rare
66	rabasan lahan	problem	1	1	0.20	Supporting/rare
67	rawa	entity	1	1	0.20	Supporting/rare

In Table 3B, topic-bound items (range ≤ 2) are further divided into two pragmatic subtypes for interpretive clarity. "Topic-focused" items have low range but comparatively higher normalized frequency within the topic-bound set (here, ≥ 1.0 per 1,000 words), indicating dense concentration in one or two stories. By contrast, "supporting/rare" items show both low range and lower normalized frequency (here, < 1.0 per 1,000 words), reflecting peripheral or episodic mentions that mainly support story-specific contexts.

RQ3: Recurrent bigrams/phrases (collocational patterning)

To address RQ3, the analysis shifts from single-word salience to phraseological patterning by examining recurrent two-word sequences (bigrams) surrounding selected high-salience eco-lexis nodes. For each node, adjacent co-occurrences were recorded as **left** (one token preceding the node) and **right** (one token following the node) bigrams, and then aggregated across texts to obtain total counts and range. Table 4 reports the resulting patterns, distinguishing bigrams that recur across multiple stories from those that are concentrated within a single narrative, and provides short contextual examples to support interpretation.

Table 4. Recurrent and topic-focused bigrams around selected eco-lexis nodes

No.	Node	Bigram	Total Count	Range (texts)	Pattern type	Example (short context)
1	plastik	tas plastik	15	3	Recurrent	"… kalau kita memungut semua **tas plastik** yang kita lihat …"
2	sampah	tempat sampah	7	2	Recurrent	"… memasukkan tas tas plastik ke dalam **tempat sampah** …"
3	sampah	membuang sampah	4	2	Recurrent	"… kamu tidak boleh **membuang sampah** sembarangan …"
4	sungai	di sungai	4	2	Recurrent	"… jembatan tersebut hancur oleh gelondongan gelondongan kayu yang tertambat **di sungai** …."
5	sungai	ke sungai	3	2	Recurrent	"… ketika hujan turun tas plastik itu bisa hanyut **ke sungai** …"
6	jaga	menjaga hutan	3	2	Recurrent	"… mengajak orang orang untuk **menjaga hutan** bakau …"
7	jaga	menjaga kelestarian	2	2	Recurrent	"… dan **menjaga kelestarian** lumba lumba …"
8	hutan	hutan bakau	7	1	Topic-focused	"… inilah **hutan bakau** yang bermanfaat sebagai tempat pemijahan ikan dan udang …"
9	hutan	polisi hutan	5	1	Topic-focused	"… bukankah ada **polisi hutan** yang bertugas menjaga hutan …"
10	pohon	anakan pohon	3	1	Topic-focused	"… datang untuk bertanam **anakan pohon** …"
11	pohon	pohon bakau	2	1	Topic-focused	"… **pohon bakau** adalah tanaman yang dilindungi …"
12	pohon	penebangan pepohonan	1	1	Topic-focused	"… jika **penebangan pepohonan** terus berlanjut semua warga yang hidup di hilir akan menghadapi masalah …"
13	buang	jangan buang	1	1	Topic-focused	"… **jangan buang** kulitnya di jalan …"
14	bersih	kebersihan sungai	1	1	Topic-focused	"… penduduk desa terus bergotong royong menjaga **kebersihan sungai** …"
15	hutan	kelestarian hutan	1	1	Topic-focused	"… mereka semua sepakat untuk menjaga **kelestarian hutan** …"

The phraseological analysis around selected high-salience nodes identified a small set of recurrent bigrams that foreground everyday environmental behavior (Table 4). Waste-management patterns were particularly salient: *tas plastik* showed the highest total count (15) across three texts, alongside recurrent co-occurrences such as *tempat sampah* (7; range=2) and *membuang sampah* (4; range=2). Together, these phrases form a coherent behavior–solution script in which children are prompted to recognize plastic litter and perform concrete actions (collecting, disposing, and placing waste in designated bins). Stewardship-oriented action frames also appeared through *menjaga hutan* (3; range=2) and *menjaga kelestarian* (2; range=2), indicating that narratives not only name environmental objects and problems but also encode responsibility and care as actionable routines.

Beyond behavior scripts, recurrent river-related prepositional bigrams functioned as place/trajectory frames rather than content-light fillers. Specifically, *di sungai* (4; range=2) marks key locations where environmental events unfold (e.g., debris obstruction or contamination contexts), whereas *ke sungai* (3;

range=2) encodes directionality and consequence pathways (e.g., rainfall/runoff carrying items into the river). Topic-focused phrase frames further highlight story-specific environmental knowledge and governance terms that occur intensively within single narratives, including *hutan bakau* (7; range=1) and *polisi hutan* (5; range=1), alongside conservation-oriented wording such as *kelestarian hutan* (1; range=1). Overall, the phrase patterns suggest a dual function of children's narratives: reinforcing shared everyday behaviors while also introducing specialized ecological and institutional vocabulary within focused story themes.

Across the mini-corpus, the frequency profile (RQ1) indicates that environmental entities (e.g., forests/trees, water/river terms, and marine life) dominate the lexical landscape, while problem and action vocabulary adds explicit behavioral cues. Range-based analysis (RQ2) shows a compact core set—especially *pohon, air, sampah, buang, bersih,* and *tanam*—suggesting cross-story emphasis on cleanliness and everyday pro-environmental routines, alongside semi-core items that recur moderately across texts. At the same time, several high-frequency items are topic-bound, reflecting dense thematic concentration in specific narratives (e.g., *hutan*, mangrove and marine ecology terms). Phraseological patterns (RQ3) reinforce this duality: recurrent bigrams foreground practical behavior scripts (plastic disposal and cleanliness) and place/trajectory framing (*di/ke sungai*), while topic-focused bigrams introduce more specialized conservation and governance vocabulary (e.g., *polisi hutan*). Taken together, the results set up the Discussion by showing how children's stories may blend actionable daily behaviors with selective, theme-driven ecological knowledge.

Discussion

Environmental vocabulary as story-world building and behavioral cueing

Across Tables 2–3, eco-lexis is anchored in concrete entities (e.g., forests/trees, rivers, marine life) while also including a smaller but pedagogically consequential set of problem and action terms (e.g., *sampah, buang, bersih, jaga*). This distribution is consistent with observations that children's environmental/climate texts often communicate through tangible referents and relatable situations, rather than technical exposition, making the environmental message more accessible to young readers (Gaard, 2009).

From an ecolinguistic perspective, these recurrent lexical and phraseological choices can contribute to circulating "stories we live by" that normalize care, stewardship, and responsibility, particularly when action verbs repeatedly frame what characters should do (Stibbe, 2015).

Core vs. topic-bound eco-lexis as a curriculum design map

The range-based grouping (Table 3A–3B) provides a practical bridge from corpus description to instructional design. In a small corpus, using range (the number of texts containing an item) as a proxy for dispersion is an established and defensible corpus-linguistic strategy for differentiating widely shared vocabulary from story-specific clusters (Gries, 2021). Pedagogically, the core set (range ≥ 4) is well suited for iterative revisiting of key items across lessons, while topic-bound items (range ≤ 2) are better treated as theme modules (e.g., mangroves, forest governance, coastal ecology) that support deeper contextual learning within a single unit. This logic is strengthened by evidence that children can acquire new vocabulary incidentally through reading when words recur in meaningful contexts (Joseph & Nation, 2018).

Phraseology as teachable "chunks" and environmental action scripts

RQ3 (Table 4) highlights a compact set of recurrent bigrams that foreground ready-made action scripts. Waste-management routines (e.g., *tas plastik, tempat sampah, membuang sampah*) encode concrete

behavioral sequences—recognizing waste, collecting it, and disposing of it appropriately. These action frames align with climate and environmental education agendas that emphasize not only awareness but also skills, values, and dispositions that support action (The United Nations Educational Scientific and Cultural Organization, n.d.).

Additionally, river-related prepositional bigrams (*di sungai* vs. *ke sungai*) function as place vs. trajectory frames, supporting developmentally appropriate causal-systems reasoning about where impacts occur and how waste moves (e.g., rainfall/runoff carrying items into waterways). Furthermore, topic-focused phrases (e.g., *hutan bakau, polisi hutan, kelestarian hutan*) indicate that some narratives introduce specialized ecological and governance vocabulary densely within a single story theme, creating opportunities for project-based learning tied to local contexts (Stibbe, 2015).

Educational implications: turning corpus findings into teachable resources

The combined frequency–range profile and phrase patterns can be translated into concrete classroom supports for elementary learners. Core vocabulary items (e.g., *sampah–buang–tempat sampah–bersih*) can be organized as a semantic network (e.g., via semantic mapping/concept mapping) and recycled across multiple stories to strengthen retention, while semi-core items (e.g., *sungai, tas plastik, jaga*) can anchor thematic reading cycles. Research on concept/knowledge maps supports the use of node–link representations to strengthen retention, and semantic mapping has also been shown to support vocabulary learning more effectively than traditional approaches in instructional settings (Dilek & Yürük, 2013; Nesbit & Adesope, 2006).

Phrase-level patterns from Table 4 can be treated as formulaic "chunks" or lexical phrases for speaking and writing frames (e.g., imperatives and pledges built from *jangan buang, membuang sampah, menjaga kelestarian*), supporting both language development and environmental responsibility discourse (Hinkel, 2018; Nattinger & DeCarrico, 1992; The United Nations Educational Scientific and Cultural Organization, n.d.). Because incidental vocabulary growth is supported by repeated meaningful exposure, recurrent eco-lexis and recurrent phrases provide a principled basis for selecting texts and designing low-burden noticing and rewriting tasks (Joseph & Nation, 2018; Nation, 2015; Schmidt, 1990).

Limitations and future directions

This study is intentionally lightweight (mini-corpus, manual workflow), so findings should be treated as illustrative rather than generalizable. Range remains an operational proxy for dispersion; expanding the corpus (more titles per level and broader coverage of themes) would allow more stable dispersion statistics and more robust phrase frequency patterns (Gries, 2021). Future work could (a) scale the corpus to test whether currently topic-focused items become semi-core with broader coverage, and (b) triangulate textual patterns with classroom uptake through short interventions using Table 4 "chunks" to evaluate how readily learners reuse action frames in speaking and writing tasks aligned with climate education objectives (The United Nations Educational Scientific and Cultural Organization, n.d.).

Conclusion

This mini-corpus study examined eco-lexis in 10 Indonesian children's narratives using frequency, range as a proxy for dispersion, and recurrent bigrams/short phrases. The findings show that environmental meaning is constructed through a combination of high-salience entities (e.g., forests/trees, rivers, marine life) and a compact set of problem and action vocabulary that cues everyday stewardship (e.g., waste disposal, cleanliness, care). Range-based patterns distinguish a small core eco-lexis shared across texts from topic-bound clusters that concentrate specialized ecological and governance terms within particular story themes (e.g., mangroves and forest protection). Phraseological results further demonstrate that

narratives encode environmental responsibility through recurring action and place/trajectory frames (e.g., disposal routines and river-direction pathways), offering practical "chunks" that can be leveraged for elementary literacy and environmental learning. Overall, the study provides a lightweight, replicable workflow and an empirically grounded basis for selecting vocabulary targets and designing story-based activities that integrate language development with environmental awareness.

References

Benevento, S. V. (2023). Communicating climate change risk to children: A thematic analysis of children's literature. *Early Childhood Education Journal*, *51*(2), 201–210. https://doi.org/10.1007/s10643-021-01294-y

Biber, D., Conrad, S., & Reppen, R. (1998). Norming frequency counts. In *Corpus Linguistics: Investigating Language Structure and Use* (pp. 263–264). Cambridge University Press. https://doi.org/10.1017/cbo9780511804489.017

Brezina, V., & Gablasova, D. (n.d.). *Corpus approaches to collocation: Challenges and opportunities*.

Collocations. (n.d.). Corpus Linguistics for Education: A Guide for Research. Retrieved March 11, 2026, from https://ebrary.net/310140/language_literature/collocations

Creative Commons. (n.d.-a). *Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)*. Creative Commons. Retrieved March 27, 2026, from https://creativecommons.org/licenses/by-nc/4.0/deed.en

Creative Commons. (n.d.-b). *Attribution 4.0 International (CC BY 4.0)*. Creative Commons. Retrieved March 27, 2026, from https://creativecommons.org/licenses/by/4.0/deed.en

Dilek, Y., & Yürük, N. (2013). Using semantic mapping technique in vocabulary teaching at pre-intermediate level. *Procedia - Social and Behavioral Sciences*, *70*, 1531–1544. https://doi.org/10.1016/j.sbspro.2013.01.221

Gaard, G. (2009). Children's environmental literature: From ecocriticism to ecopedagogy. *Neohelicon*, *36*(2), 321–334. https://doi.org/10.1007/s11059-009-0003-7

Gries, S. T. (2021). Analyzing Dispersion. In M. Paquot & S. T. Gries (Eds.), *A Practical Handbook of Corpus Linguistics* (pp. 99–118). Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_5

Hamida, N., & Indrastuti, N. S. K. (2025). Kepedulian terhadap lingkungan dalam cerita anak Nanti Saja karya Fransisca Emilia: Kajian Ekokritik Greg Gerrard [Concern for the environment in children's stories Later by Fransisca Emilia: Greg Gerrard's Ecocritical study]. *Widyaparwa*, *53*(2), 132–141. https://doi.org/10.26499/wdprw.v53i2.1600

Hinkel, E. (2018). Teaching and learning formulaic sequences and prefabs. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0925

Intergovernmental Panel on Climate Change. (2023). Summary for Policymakers. In H. Lee & J. Romero (Eds.), *Climate change 2023: Synthesis report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1–34). https://doi.org/10.59327/IPCC/AR6-9789291691647.001

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*, 190–211. https://doi.org/10.1016/j.jecp.2017.08.010

Koester, A. (2022). Building small specialised corpora. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (2nd Edition, pp. 45–58). Routledge. https://doi.org/10.4324/9780367076399

Laliena, D., & Sala, R. T. (2023). Picturebooks and reader training in the 21st century. An ecocritical reading of canonical works of children's literature. *Frontiers in Education*, *8*(1304027), 1–8. https://doi.org/10.3389/feduc.2023.1304027

Lancaster University. (n.d.). *Collocations*. Corpus Linguistics (Lancaster University). Retrieved April 18, 2026, from https://www.lancaster.ac.uk/fss/courses/ling/corpus/Corpus3/3COLL.HTM

Lanta, J., Rahman, F., Lewa, I., & Akhmar, A. M. (2022). Respect for nature in Indonesian children's fiction: Ecocriticism perspective. *Webology*, *19*(1), 6010–6021. https://www.webology.org/data-cms/articles/20220206020817pm19402.pdf

Let's Read. (n.d.-a). *About us*. Let's Read. Retrieved April 18, 2026, from https://www.letsreadasia.org/about

Let's Read. (n.d.-b). *Lari*. Let's Read. Retrieved March 11, 2026, from https://www.letsreadasia.org/book/lari

Let's Read. (n.d.-c). *Luh ayu manik mas, pahlawan lingkungan*. Let's Read. Retrieved April 18, 2026, from https://www.letsreadasia.org/book/87c92403-22b4-462d-a95d-c98f27f65035?bookLang=6260074016145408

Let's Read. (n.d.-d). *Main air, yuk*. Let's Read. Retrieved April 18, 2026, from https://www.letsreadasia.org/book/main-air-yuk?bookLang=6260074016145408

Let's Read. (n.d.-e). *Siapakah yang bisa membantuku?* Let's Read. Retrieved April 17, 2026, from https://www.letsreadasia.org/book/6691895d-a56e-47bd-8dcf-a006bc952aab?bookLang=6260074016145408

Maryani, M. (2011). Identifying Indonesian-core vocabulary for teaching English to Indonesian preschool children: a corpus-based research. *K@ta*, *13*(2), 147–161. https://doi.org/10.9744/kata.13.2.147-161

McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

MDPI. (n.d.). *Ecolinguistics*. Encyclopedia. Retrieved March 11, 2026, from https://encyclopedia.pub/entry/35470

Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, *27*(1), 136–145. http://nflrc.hawaii.edu/rfl

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*(3), 413–448.

Rybak, K. (2025). Towards a literature of actions: Green informational picturebooks and critical engagement with fighting climate change. *Children's Literature in Education*, *56*, 210–224. https://doi.org/10.1007/s10583-023-09549-5

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Stibbe, A. (2015). *Ecolinguistics: Language, ecology and the stories we live by*. Routledge.

The Asia Foundation. (n.d.). *Let's Read*. Retrieved March 11, 2026, from https://asiafoundation.org/programs/education-and-leadership/lets-read/

The United Nations Educational Scientific and Cultural Organization. (n.d.). *Climate change education*. Retrieved March 23, 2026, from https://www.unesco.org/en/climate-change/education

United Nations. (n.d.). *Global citizenship education*. Retrieved March 23, 2026, from https://www.un.org/en/academic-impact/page/global-citizenship-education

United Nations Framework Convention on Climate Change. (n.d.). *Action for climate empowerment*. Retrieved March 11, 2026, from https://unfccc.int/topics/education-and-youth/big-picture/ACE

Vaughan, E., & Clancy, B. (2013). Small corpora and pragmatics. In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2013: New Domains and Methodologies* (pp. 53–73). Springer. https://doi.org/10.1007/978-94-007-6250-3_4

Xiao, R. (n.d.). *Unit 6: Making statistical claims*. Retrieved April 18, 2026, from https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xCBLS/chapters/A06.pdf