LLT Journal, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 28, No. 2, October 2025, pp. 587-613

LLT Journal: A Journal on Language and Language Teaching
http://e-journal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

CORPUS-INFORMED AND CRITICAL APPROACHES TO ADDRESSING SUSTAINABLE DEVELOPMENT GOALS (SDGS)

Joseph Ernest Mambu
Universitas Kristen Satya Wacana, Indonesia
joseph.mambu@uksw.edu
correspondence: joseph.mambu@uksw.edu
https://doi.org/10.24071/llt.v28i2.12023
received 28 March 2025; accepted 19 August 2025

Abstract

This paper showcases how a do-it-yourself (DIY) corpus on the United Nations' Sustainable Development Goals (SDGs) can be used to increase corpus users' language awareness (at word, phrase, sentence, and discourse levels) and criticality. To this end, the author justifies the close link between corpus linguistics, various critical approaches, and SDGs in the applied linguistics/ELT literature. Subsequently, the author provides an overview of principles for creating a specialized DIY corpus containing around 700,000 words/tokens based on 882 articles of *The Jakarta Post* from 2012 to mid-2024. In view of critical discourse analysis in corpus linguistics, SDG-related key word(s) in context (KWIC) were analyzed in four stages to examine their (1) frequencies and collocates, (2) concordance lines, (3) larger contexts within or across texts in the corpus (or beyond), and (4) potential to stimulate questions aiming at social transformation. In the findings, the four-stage analyses explore two of the most frequent SDGs in the corpus – "gender" (SDG 5) and "clean water and sanitation" (SDG 6) – to illustrate poststructuralist and Marxist criticality, respectively. The decolonial criticality is demonstrated through a corpus analysis of the word "indigenous" and its collocates. Possible pedagogical applications of the corpus-informed approach are also discussed.

Keywords: corpus linguistics, criticality, Sustainable Development Goals (SDGs), *The Jakarta Post* (JP)

Introduction

The United Nations' Sustainable Development Goals (SDGs) have attracted the attention of applied linguistics scholars and (TESOL) practitioners. They have identified the insufficient attention to language in SDGs (Tonkin, 2023) and the lack of access to SDGs materials in English or other languages (McEntee-Atalianis, 2017). Some others have suggested lesson plans on SDGs-related themes (Maley & Peachey, 2017) and incorporated SDGs into critically oriented lessons at an English language education undergraduate program or beyond (Mambu, 2023a, b). Maijala et al. (2023) explored how pre-service language teachers perceived social, cultural, economic, and environmental dimensions of sustainability in relation to their personal lives and pedagogical practices. A similar study focuses on English

language teachers' identified challenges and support when exploring the ecological dimension of sustainability (Mercer et al., 2023). How English language teachers can develop SDGs-related materials for their EFL students with technology (especially that inspired by corpus linguistics), however, is still largely unknown. Language teachers can use generative AI tools like ChatGPT to produce SDGs-related materials to reinforce language skills and prompt the tool(s) to adapt levels of difficulty tailored for their students. Nonetheless, generative AI tools might not always be suitable for reinforcing learners' language skills. Despite its claims that learners can learn inductively when interacting with the tool, it is the ChatGPT "doing the induction," the generated texts of which could then simply be copied and pasted by the learners. As such, the learners do not really learn the (foreign) language (Crosthwaite & Baisa, 2023, p. 3) nor critical thinking.

Against the above backdrop, the purpose of this paper is to illustrate how English language teachers can employ corpus-informed and critical approaches to addressing SDGs. The former approach is expected to increase students' language awareness (at lexical, phrasal, syntactical, and discourse levels) and the latter criticality (including critical thinking). To justify my stance, I will review (1) applied linguists'/English language educators' growing interest in SDGs, (2) how SDGs are potentially analyzed in light of corpus linguistics, and (3) critical approaches to addressing SDGs. Subsequently, I will exemplify a corpus analysis of 882 *The Jakarta Post* (henceforth JP) articles. The analysis of the JP corpus of SDGs can then be developed as "narrow reading" materials, to borrow Schmitt and Carter's (2000) term, based on corpus-informed and critical approaches to SDGs.

Literature Review

The place of SDGs in applied linguistics/English language education

Not long after 2015 when SDGs were officially launched by the United Nations, applied linguists had noticed the importance of SDGs in language studies. McEntee-Atalianis (2017) was more on the pessimistic side, noting that access to SDGs materials is restricted due to digital and linguistic constraints as they were more readily available in English in areas with decent internet connections. From the optimistic side, Maley and Peachey (2017) collected numerous lesson plans developed by English language teachers who see the relevance of SDGs in their language classrooms. They also believe that English language teaching (ELT) will be impoverished if it focuses on vocabulary and grammar only without any attention to current global issues captured in SDGs. Aligned with Maley and Peachey's (2017) optimism, in my earlier works, I explored how my undergraduate English Language Education students interpreted SDGs, developed critically oriented lesson plans revolving around SDGs (Mambu, 2023a), and sustained an interest in SDGs two years after finishing the undergraduate studies (Mambu, 2023b).

In a larger scale than Mambu (2023a, b), Arikan and Zorba (2024) involved 30 Turkish-based preservice English language teachers to investigate what they understood about SDGs and how the teachers became familiar with SDGs. They also explored the degree to which the three-day workshop on connecting literary works to SDGs transformed the teachers' perceptions and understanding of SDGs in relation to ELT mediated by 16 literary texts. Prior to the workshop, 83% of the teachers had limited understanding of SDGs, with social media being chosen by 21 teachers as the most frequently prevalent source of knowledge of the SDGs (p. 6).

Besides, most of the teachers associated SDGs with environmental and social issues (23 and five teachers, respectively), with only two of them thinking of SDGs as related to economy (p. 7). The three-day project immensely enabled the teachers to understand SDGs better and how they were related to ELT and the usefulness of literary texts in improving English language learning.

Also in Türkiye, Arslan and Curle (2024) addressed how the teachers defined Education for Sustainable Development (ESD), how essential ESD was to the teachers, and whether the teachers thought that ESD was adequately incorporated into the EFL curriculum. First, their qualitative content analysis suggested that the teachers defined ESD in various ways, such as "learning for sustainability" socially, economically, and ecologically (20 teachers) and "lifelong learning," which emphasized nurturing inclusive education and critical thinking (two teachers). Second, the teachers thought of ESD as crucial and English as a lingua franca made it possible for learners to be exposed to issues of sustainability. ELT was also considered an important medium enabling students to learn the English language and content (especially various topics related to sustainability). Third, some teachers thought that ESD was not addressed in-depth and often restricted to education, economy, and ecology, thanks to the perceivably insufficient contact hours of the English course. Arslan and Curle (2024) admitted that their study was limited only to the teachers' perceptions and recommended that other data collection methods (e.g., classroom observations and reflective journals) be used in future studies. In my opinion, stakeholders' perceptions and how SDGs-informed ELT are implemented in classes are not sufficient. ELT materials development, especially that informed by a corpus of SDGs, should also be explored.

Different from some empirical studies reviewed above, Yu et al. (2024) have made a comprehensive literature review of recent ELT educators' attempts to integrate SDGs in their language classrooms. Reviewing 90 peer-reviewed journal articles published from 2013 to 2023 in English, the researchers found that effective pedagogical methodologies include designing interdisciplinary curriculums, among others, particularly those that integrate Education for Sustainable Development (ESD) and Second Language Acquisition (SLA) emphasizing "communicative competence and sociocultural theory" (p. 5). In particular, some pedagogical methodologies explored critical pedagogy (Mambu, 2023a, b), the CLIL (Content and Language Integrated Learning) approach, and Project-Based Learning (PBL) (Yu et al., 2024, p. 6). Despite the recently documented success, Yu et al. (2024) still notice a gap in the impact of the pedagogical methodologies (p. 6), especially in maintaining ecological consciousness to support sustainable life styles. Therefore, they call for more research into teachers' preparedness in delivering interdisciplinary sustainability issues in the ELT classroom and the effectiveness, or lack of it thereof, of curriculum design integrating sustainability issues in the language classroom across cultural and linguistic contexts.

Relating language-learning objectives to sustainability themes and lack of SDG resources were the identified challenges (Yu et al., 2024). However, professional development programs and "open educational resources" seem to have overcome the hurdles (p. 1). Yu et al. (2024) also recommend that future studies should examine "the role of digital tools and online platforms in enhancing sustainability education within ELT, including the effectiveness of virtual and augmented reality, gamification, and social media" (p. 10). Yu et al. (2024) have

not addressed corpus linguistics and the latest development of the generative AI, but I believe these technologies should align with their recommendation, especially in terms of curriculum design and materials development integrating SDGs into ELT.

Corpus linguistics and SDGs

A corpus is "a collection of sampled texts, written or spoken, in machine readable form which may be annotated with various forms of linguistic information" (McEnery et al., 2006, as cited in Toriida, 2016, p. 87). A corpus does not contain a random selection of texts; rather, it has to be "carefully sampled and balanced in order to be representative of a specific variety of language (e.g. … British newspaper articles about poverty, political speeches, … Indian English…)" (Baker & McEnery, 2015, p. 1). "Data-driven learning" or DDL in the corpus linguistics literature is built upon a premise that learners do not depend too much on language specialists "to pre-digest the language" but they "could chew on the language data themselves" (Boulton, 2016, p. 115). "Corpus queries" in DDL are likened to "web searches" in everyday life (p. 116). However, it is unrealistic to assume that learners will keep consulting corpus data after taking an introductory corpus-based course. It is more sensible to expect that learners will refine their "search techniques" (e.g., using a "wildcard" like "play a * role in" when Googling or pressing Ctrl + F for searching a word or a chunk in an electronic document) after being introduced to corpus linguistics (pp. 116-117). Utilizing DDL based on large general corpora like the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA) can be overwhelming for teachers and low-proficient learners. However, Boulton (2016) argues that with a tactful strategy, teachers could design manageable learning activities for learners by selecting concordance lines (see Figure 1 for an example) displaying salient patterns.

Conceding that large general corpora might be limited to address specific aims, Boulton (2016) explores the potential benefits of "purpose-built corpora" which can be designed manually using a free software like AntConc (p. 122). Building a corpus might help teachers and learners to be exposed to naturally occurring language, especially that typically used in a specific field relevant to learners' discipline (Boulton, 2016).

Corpus linguistics has inspired applied linguists/English language educators to investigate the degree to which sustainable development transpires in published academic English (Pinto, 2021) and young English language learners' written texts (Sundh, 2016). Using the Sketch Engine corpus tool, Pinto (2021) investigated the degree to which and how Brazilian researchers raised Sustainable Development Goals (SDGs) themes in their academic articles published in English in open-access journals accessible in the "Scientific Electronic Library Online (SciELO)" between 2017 and 2018 from various disciplines. The articles were then compiled in the 906,035-word "Brazilian Corpus of Academic English [BrACE]" (pp. 1, 5). The Sketch Engine tool generated 117 occurrences of the word "sustainability" and how the word collocates with modifiers, adverbs, prepositions, and nouns (p. 6). The word "sustainability" appeared to be (1) objects (e.g., "sustainability envisioned," "managing sustainability, and "oppose sustainability"); (2) together with or preceded by modifiers (e.g., "sub-regional sustainability," "coastal sustainability," and "social sustainability"); and (3) a notion being discussed (e.g., "sustainability

is a cure," "sustainability is a far broader phenomenon" [p. 6]). Besides the word "sustainability," Pinto (2021) also focuses on the word "poverty." It appeared only 21 times in BrACE, with the word "poverty" as an object (e.g., "fighting poverty," "reducing poverty") and as a concept being modified (e.g., "commodifying poverty," "massive urban poverty," and "human poverty" [p. 10]). The larger contexts of some concordance lines focusing on the word "poverty" provide the authors' critical views. For example, real poverty was hidden under the guise of "tourism and environmentalism" or commodified (p. 7).

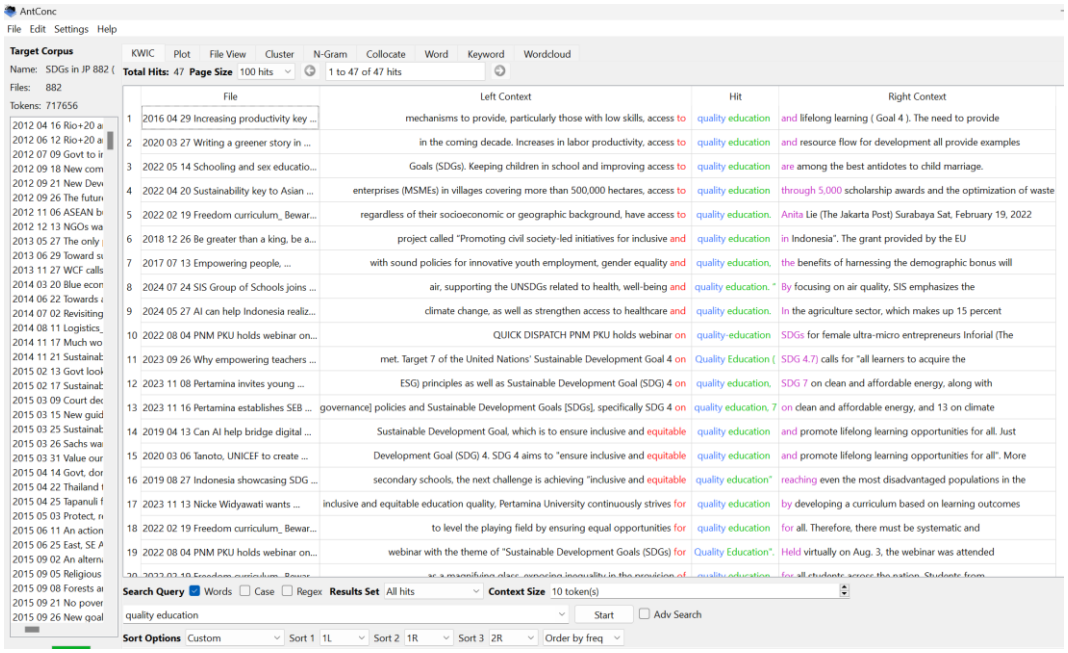


Figure 1. An example of concordance lines from my JP-based corpus of SDGs, with the phrase "quality education" as the focus

Sundh (2016) compiled a corpus of 2,200 SDG-related written texts produced by Baltic early-teen English language learners from Sweden, Russia, Lithuania, Latvia, and Estonia. Sundh focused on key words in nouns from the Baltic Young Learners of English Corpus (BYLEC). The study was motivated by the researcher's belief that young learners of English found solutions to economic, social, and ecological problems by engaging them in "future-facing activit[ies]" (Sundh, 2016, p. 93). As he further argues, "To learn a foreign language is thus to learn for a sustainable future" (p. 95). Consistent with the CLIL (Content and Language Integrated Learning) approach, Sundh believes that language learning should not only be "form-focused" but also "content-based," so "engaging topics" like those related to SDGs are pivotal. Central to learning contents in the language classroom are "a rich and advanced vocabulary and an awareness of challenges on the global sense" (Sundh, 2016, p. 94). Against the backdrop of emphasizing vocabulary learning for a sustainable future, Sundh (2016) investigated how issues related to sustainability and the future surfaced in the Baltic learners' texts. To make the analysis manageable, Sundh (2016) focused on nouns (as key words) which "occurred more than once in the [BYLEC] data," were used across learners, and were "used in an abstract sense and in a context that can be associated with issues of a sustainable life style" (p. 97). The analysis revealed that nominal key words

fitting into the criteria included "charity," "communication," "culture," "energy," "invention," "peace," "technology," and "war" (p. 100). These key words were used to discuss abstract themes in three prominent areas: "(1) sharing resources and communicating in different cultures, (2) finding new inventions for energy and technology, and (3) finding ways to cope with and possibly solve international conflicts" (pp. 100-101).

Though insightful, the two studies summarized above are situated in relatively well-resourced contexts – academics privileged enough to get their work published in English (Pinto, 2021) and young English language learners in eastern Europe. How English language teachers can facilitate their students to learn English and SDGs critically using appropriate technologies is still understudied. Corpus analysis has been supported by the concordancing technology (e.g., AntConc [Anthony, 2023]). Corpus linguistics, however, has recently encountered a fierce contender: generative AI (GenAI) tools.

The technicalities of operating corpus tools to analyze linguistic data can be really challenging for learners and teachers alike. People can simply make prompts that generate instant responses to language-related queries like "can you list of a few example sentences including 'phrase X'"? (Crosthwaite & Baisa, 2023, p. 3). GenAI tools are way better than corpus tools in helping teachers to design differentiated learning as teachers could make prompts (e.g., "re-write this text for a 9th grader" [p. 3]) to the former in such a way that the results can be used for a specific group of learners. Currently, no corpus tools, unlike GenAI tools, can "[train] data token counts in the billions or trillions" very quickly online (p. 3). To top it all, GenAI is a lot more phenomenal than corpus tools for teachers in transforming responses to an earlier prompt (e.g., "what are some nouns that fill slot X in this sentence") into a pedagogical assignment or evaluation (e.g., "take the list of nouns you generated and create a sequence of multiple-choice questions based on them") or a whole lesson plan (e.g., "build a lesson plan for 2nd graders based on acquiring these norms" [p. 3]).

Despite the apparently shortcomings of corpus tools, Crosthwaite and Baisa (2023) contend that the advantages of corpora still outweigh GenAI. First, general and specialized, do-it-yourself (DIY) corpora are based on real, naturally occurring data whose sources can be precisely identified or cited. As Crosthwaite and Baisa (2023, p. 2) put it, "we know, exactly, the domain of texts from which the corpus is derived," including JP articles on SDGs as a DIY corpus in the current paper. This corpora's superiority is lacking in GenAI. Second, corpora provide authentic use of language as it is based on language produced naturally by human in various contexts. GenAI language is produced "through a statistical procedure," which is typically grammatical but is not necessarily appropriate in terms of register or context. Third, corpus-based queries are replicable. If a person or many people make repeated queries based on the same corpus or corpora, they can expect to generate the same or similar results because the language in electronically stored texts is static, although the corpus or corpora may be updated from time to time. GenAI, on the other hand, relies on complex statistical procedures when generating responses, such that "the answers are randomly sampled leading to a unique answer for each subsequent identical query" (Crosthwaite & Baisa, 2023, p. 2). Most importantly, when learners use corpora for linguistic analysis, they can do it inductively, thus ensuring active data-driven learning (DDL) on their (not on GenAI's) part.

Teachers with limited technical skills can still implement the corpus-informed approach in that the use of concordancing software in corpus linguistics (CL) can be complemented by GenAI tools. The superiority of concordancing software in CL can compensate for the shortcomings of GenAI tools, or vice versa. GenAI tools are typically better at helping users (especially teachers and students) to provide immediate answers and produce simplified, paraphrased, summarized, or synthesized texts. It is even possible for GenAI tools to simplify naturally occurring texts (including texts containing SDG themes from *The Jakarta Post*). The AI-simplified texts are then built as a DIY corpus specifically designed for beginner learners of English (e.g., levels A1 and A2 on the Common European Framework of Reference for Languages [CEFR]). Naturally occurring language and replicability (Crosthwaite & Baisa, 2023) might be compromised as AI tools can generate relatively improper register and make countless simplified versions of the same texts, although the sources can still be traced. However, once a version of AI-simplified texts is stored in a DIY corpus, one can expect to generate the same results for any corpus-based queries. However, concordancing software is better at immersing learners in their own DDL utilizing unabridged texts (like those published in *The Jakarta Post*) or their (AI-)simplified counterparts. The former, which is more suitable for upper-intermediate and advanced learners (i.e., language users at B2, C1, and C2 levels on CEFR), becomes the focus of the current paper.

Critical approaches to viewing SDGs

Before demonstrating the capability of JP-based SDG corpus to help English language teachers design DDL for their students, in this section, I will review critical approaches to examining SDGs.

Criticality in language studies and SDGs

In language studies, Kubota and Miller (2017) have summarized that criticality focuses on scrutinizing relations of power, the nuances of which can be viewed through postmodern/poststructuralist, Marxist, and decolonial theoretical lenses separately or in combination. The poststructuralist perspective explores how identity is discursively formed in everyday life. Social constructions can be problematic or transformative. Poststructuralist criticality, as it were, is appropriate for inquiring into issues related to gender (SDG 5) and inequalities (SDG 10), among others. The Marxist perspective in language studies pays substantive attention to physical or tangible realities that can be sensed or seen (e.g., poverty and its ramifications; see Kubota & Miller, 2017). It also investigates the relationship between a person's or a community's social class and language teaching, learning, or use, especially in terms of access to standard academic English (Block, 2015). The missions of Marxist, Freirean critical pedagogies in English language learning include engaging teachers and learners to reflect on their own historical, political, and economic conditions, especially those that are oppressive; aspire to have a better and just future; and take concrete transformative actions (Abednia & Crookes, 2019). The Marxist lens is useful to shed light on socioeconomic and class gaps accounting for the abundance, the scarcity, or even the absence of materials for human beings to fulfill their basic needs (Block, 2015). Therefore, concerning the SDGs, the Marxist lens is suitable to address issues related to poverty (SDG 1), hunger (SDG 2), health care system (SDG 3), quality

of education (SDG 4), access to quality water and sanitation (SDG 6), and various forms of materially based inequalities (SDG 10), among others. Recall Pinto (2021) who focuses not only on the word "sustainability," but also "poverty" when analyzing the Brazilian Corpus of Academic English. Viewed through a Marxist lens, Pinto's (2021) study also disclosed Brazilian scholars' criticality in that "economic liberalism and how industrial capitalism was 'characterized by massive urban poverty and social dislocation'" (pp. 11-12).

The post-/decolonial perspective grew out of scholars' dissatisfaction with the Eurocentric Marxism and poststructuralism. At the crux of decolonial research is that studies are conducted together with Indigenous people, not *about* or *on* them, to address issues that matter for Indigenous communities, such as claiming Indigenous people's rights, maintaining oral cultures through telling local stories, revitalizing Indigenous languages, connecting with nature in one's own land, rediscovering Indigenous women's spiritual prominent role in society, establishing partnerships with fellow Indigenous communities, renaming (of people, events, and places) with their earlier Indigenous names, and appreciating Indigenous knowledge frequently suppressed by colonial mindsets, among others (Smith, 2021). Ideally, the decolonial perspective is relevant to achieving sustainable cities and communities (SDG 11) and other SDGs related to nature (e.g., climate action [SDG 13], life below water [SDG 14], and life on land [SDG 15]).

Intersectionally, Mambu (2023a) discussed how Marxism and post-/decoloniality can be used together to interpret the long-lasting, rampant poverty (SDG 1) and poor quality of education (SDG 4) in Papua, as implied from Pamela's story in which some of her fellow Papuans were not creative in processing natural resources to make their traditional cuisines from sago sustainable. The lack of creativity seems to be deeply rooted in perpetuated poverty and ignorance, thanks to the Dutch's colonization up to the early 1960s and the U.S. de facto occupation (Saltford, 2003) in Grasberg gold mining operated by Freeport-McMoran (Firdaus, 2017; Sulaiman, 2018) in Indonesia's Papua.

Critical discourse analysis (CDA) and corpus linguistics

While important, criticality summed up by Kubota and Miller (2017) has yet to provide a framework for a corpus analysis. I find Baker and McEnery's (2015) work useful as it discusses how CDA and corpus linguistics can be done in four stages: quantitative analysis (or "description"), "interpretation," "explanation," and "critical evaluation" (p. 3). In the first stage of description, Corpus linguistics allows researchers to have "a quantitative flavour" (Baker & McEnery, 2015, p. 2) because it provides information of how frequent a word, a phrase, or a chunk occurs in a corpus. Frequency in a corpus is also related to the notions of collocates and keywords. Collocates refer to words that are closely associated with, and are located adjacent to, or somewhat near, other words, "more than would be expected if all the words in a corpus were ordered in a random jumble" (p. 2). A keyword is a word prominent in a smaller, specialized corpus, but might not be as frequently occurring as that in a language, as indicated by the large general, "reference" corpus (p. 2) of spoken and written language representing various genres (e.g., the over one-billion-word Corpus of Contemporary American English [COCA]).

Once the quantitative parts (i.e., frequency, collocates, and keywords) have been identified, corpus linguistics proceeds to qualitative analyses where each word

under scrutiny has to be interpreted in its context(s) – the second stage. The focused word is displayed in the center of a few or even thousands of "concordance lines," with the contexts on the left or right of the word (Baker & McEnery, 2015, p. 3).

At the third stage of corpus-based discourse analysis, researchers attempt to explain "descriptive and interpretative findings within a wider social context." For example, Baker and McEnery (2015) notice that there have been more people using second person pronouns in various English written genres in the 20th century, especially in "informational and official texts" (p. 3). This phenomenon seems to signify that writers want to have a closer relationship with their readers. To account for this, the findings have to be situated in the wider social context. Baker and McEnery pose a question to find an explanation: "Can phenomena like a move towards relaxed and more informal social conventions… or even increased use of persuasive language due to the capitalist imperative to position everyone as a consumer help to explain our finding about pronouns?" (p. 3). Baker and McEnery (2015) even push corpus-based discourse analysts to take another step – i.e., being "critical or inspir[ing] social change." Therefore, in the fourth stage, analysts should be "more evaluative, pointing out the consequences of such uses of language (asking 'who benefits?' or who is potentially disempowered), perhaps making recommendations for good practice" (p. 3).

In this paper, Kubota and Miller's (2017) summary of criticality in language studies and Baker and McEnery's (2015) CDA in relation to corpus linguistics will serve as the theoretical framework delineating corpus-informed and critical approaches to addressing SDGs.

Method

Constructing a specialized corpus of SDGs from JP

A corpus-informed approach to materials development emphasizes language that occurs naturally in actual texts, including in a specialized corpus compiled by a language teacher/teacher educator/researcher him-/herself (McCarthy, 2004). One of possible specialized corpora compiled by a language teacher is a corpus from a newspaper. SDGs-related texts circulating in mass media based in Indonesia are still largely unexplored and hence potentially analyzed in light of corpus linguistics. In Indonesia, *The Jakarta Post* (JP) seems to be the most representative, especially because it is in English. A 42-year-old newspaper like JP is still read quite widely (mainly in Jakarta, but also in other cities in Indonesia) and both EFL teachers and students can benefit from JP.

The SDGs have become more frequently used worldwide, but how and for what purpose it is used in the Indonesian context, especially in mainstream mass media like *The Jakarta Post* (JP), is still absent in the literature. In view of Reppen (2022, p. 31), SDGs as a "[type] of language" seems "not adequately represented by existing corpora," although JP is a de facto corpus. As an applied linguist and an English language teacher educator, I am interested in understanding how SDGs have permeated the mass media (particularly JP) published in English in the Indonesian context and how the understanding could inform my pedagogy, especially when introducing my student-teachers to SDGs-related contents and language features in JP to help them read and respond to SDGs academically (e.g., in an op-ed article as a genre). Principles in constructing a corpus, including a JP-based corpus of SDGs, are as follows.

Corpus size

Reppen (2022) argues, "There is not a specific number of words… Corpus size is certainly not a case of one size fits all" (p. 31). That said, two factors are necessary to consider: "representativeness (have I collected enough texts (words) to accurately represent the type of language under investigation?) and practicality (time constraints)" (p. 32). A relatively small, specialized corpus (e.g., 40,000-word humor in English language teacher faculty meetings in two institutions) could be regarded as sufficient, depending on research questions raised by researchers. However, Reppen (2022) reminds that if the purpose of research is to investigate scarce grammatical features or rare vocabulary words, a corpus containing millions of words is necessary.

In the current paper, I collected over 700,000-word/token texts containing SDGs from 882 JP articles since the first occurrence of SDG in the newspaper on April 16, 2012 (see Figure 2). Older and more recent texts can be useful. The former is especially important when learners or researchers want to know how a certain concept like SDGs evolves. As Bowker (1998, p. 638) puts it, "concepts are better explained when they first come out." The recent texts are also valuable "to reflect the state of the field [or a theme] at present" (p. 638). Due to time constraints, I did not collect JP articles after July 24, 2024. As I collected the articles as exhaustively as possible from the mentioned date in 2012 until mid-2024, I attempted to ensure that my DIY was representative, especially in mapping out which SDGs were well- or under-represented in the JP-based corpus of SDGs and what salient themes could be exploited from the corpus for critical English language teaching and learning.

Permission

Permission should be secured for copyrighted materials, even when they are available online (Reppen, 2022). Accordingly, I asked permission for the 882 articles from JP and the Chief Editor signed it (Ref. 176/Srt-TJP/Ed/X/2024).

Corpus design

When I designed the DIY corpus, several criteria were considered. First, I determined what kind(s) of texts and from which genres were collected (Reppen, 2022). According to Bowker (1998), there should a variety of genres "to help [learners] achieve an understanding of [a] subject field and allow them to see different registers of usage" (p. 638). It is also desirable that a specialized corpus contains "texts by a variety of authors (to determine if usage is widespread or idiosyncratic)" (p. 638). In building the corpus, I collected as many genres written by various authors and Englishes as possible (e.g., op-ed, feature, advertorial, and news articles) from JP. Second, how the files were named (Reppen, 2022) was important to consider. My strategy for file naming was YYYY MM DD Title (e.g., 2012 11 06 ASEAN business must focus on sustainability). Third, it was necessary to decide on what file formats the texts were stored (Reppen, 2022). Therefore, I saved the articles in pdf and txt formats, the latter particularly suitable to be used in AntConc, a corpus tool employed here. Fourth, I also took the text size into account. "Complete texts" are desirable "so examples of usage or explanations of concepts are not cut short" (Bowker, 1998, p. 638).

Based on the criteria, the following procedure was followed (and could be replicated) when building a JP-based corpus of SDGs. First, I opened

https://www.thejakartapost.com/. Second, I subscribed to *The Jakarta Post* monthly to ensure that I could access paid contents. Third, I clicked the search button on JP and typed "sustainable development goal," "sustainable development goals," "sustainable development goal 2012," "sustainable development goals 2012," and so forth. Fourth, I used a similar search template when repeating the process from 2013 to mid-2024. Fifth, I opened the Notepad tool on my Windows. Sixth, I copied and pasted any texts containing "sustainable development goal(s)" to the Notepad, storing the texts in a .txt format. As I collected texts from JP, I also downloaded and installed the AntConc software (Anthony, 2023). Next, I fed the txt files into AntConc compositely (882 articles) and on a yearly basis (2012-2024) so I was able to present the total word/token count in over 12 years (see Figure 2) and frequencies of SDGs-related phrases on an annual basis (Table 1).

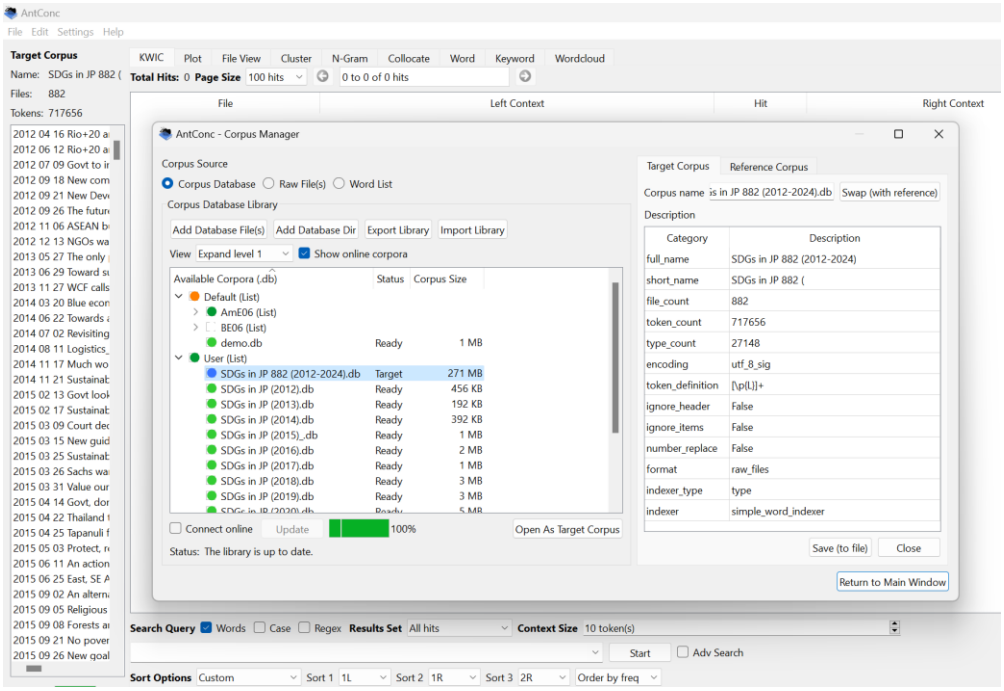


Figure 2. The size of JP-based specialized corpus of SDGs

Limitations

As an English newspaper published and distributed in Indonesia, JP's audience is quite limited (e.g., expatriates working in Indonesia and highly educated advanced English users living in big cities). Therefore, JP articles, especially those containing SDG themes, do not necessarily represent SDG discourses circulating nationwide, much less across the globe. Besides, as an exploratory study in which I constructed and delved into the DIY corpus on SDGs from JP myself, teacher and learner perspectives on engaging with the corpus data are absent.

Findings and Discussion

Findings

The corpus analyses here focused on (1) the frequency of each SDG-related key word(s) in context (KWIC) as a phrase (e.g., "gender equality" as SDG 5), (2) a specific KWIC strongly collocating with the SDG (e.g., "discrimination" as in

"gender discrimination"), (3) sample word clouds, highlighting some KWICs, (4) sample concordance lines containing an SDG-related KWICs and their collocates, and (5) the larger contexts of a KWIC to illustrate poststructuralist, Marxist, and decolonial criticality (Kubota & Miller, 2017) in light of the four stages of critical discourse analysis (CDA) in corpus linguistics (Baker & McEnery, 2015).

Based on the current DIY corpus (Table 1), gender equality (SDG 5), climate action (SDG 13), quality education (SDG 4), zero hunger (SDG 2), and clean water and sanitation (SDG 6) were the top five KWICs in terms of frequency, with SDG 5 the most consistently occurring from 2012 to mid-2024, except in 2013. Due to space restrictions, in the current paper, a poststructural analysis focused on SDG 5 and a Marxist analysis on SDG 6. There was no mention of "indigenous" in any of the 17 SDG labels, so for a decolonial analysis, I scrutinized how the word "indigenous" collocated with other words in my DIY corpus of SDGs.

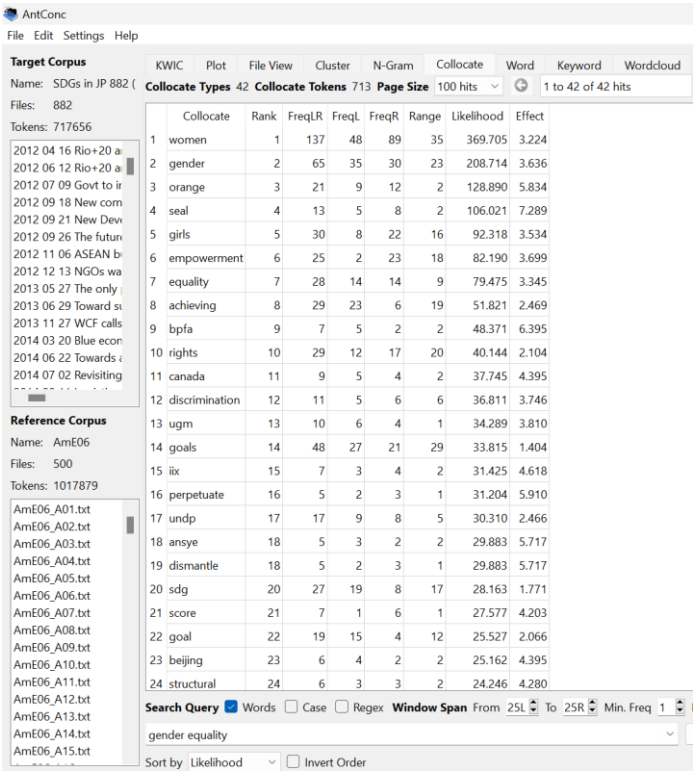
Table 1. Year-by-year frequencies of SDGs-related key words in the JP-based specialized DIY corpus

No	SDG	Key words	2012-2024 (882 articles)	Rank (2012-2024)	2012 (8 articles)	2013 (3 articles)	2014 (6 articles)	2015 (45 articles)	2016 (60 articles)	2017 (35 articles)	2018 (75 articles)	2019 (78 articles)	2020 (104 articles)	2021 (97 articles)	2022 (119 articles)	2023 (206 articles)	2024 (up to early July - 46 articles)
1	5	gender equality	179	1	2	0	6	8	16	18	4	40	25	25	7	11	17
2	13	climate action	95	2	0	0	0	1	0	1	6	8	9	14	16	33	7
3	4	quality education	47	3	0	0	0	1	5	1	2	5	5	1	12	12	3
4	2	zero hunger	35	4	0	1	0	0	3	1	18	1	5	2	2	2	0
5	6	clean water and sanitation	23	5	1	0	0	1	1	0	1	0	7	2	0	8	2
6	8	decent work and economic growth	19	6	0	0	0	0	0	0	0	3	4	2	0	10	0
7	14	life below water	16	7	0	0	0	0	0	0	2	1	3	4	3	1	2
8	1	no poverty	15	8	0	0	0	1	2	0	0	2	4	1	1	4	0
9	3	good health and well being	12	9	0	0	0	0	1	0	2	0	2	3	0	3	1
10	7	affordable and clean energy	9	10	0	0	0	0	0	0	0	0	6	0	0	3	0
11	11	sustainable cities and communities	8	11	0	0	0	0	0	0	0	1	1	3	1	1	1
12	12	responsible consumption and production	8	11	0	0	0	0	0	0	0	0	2	1	0	4	1
13	17	partnerships for the goals	8	11	0	0	0	0	0	0	0	2	1	0	1	4	0
14	16	peace, justice and strong institutions	6	12	0	0	0	0	0	0	1	2	0	0	0	3	0
15	9	industry, innovation and infrastructure	4	13	0	0	0	0	0	0	0	2	1	0	0	1	0
16	10	reduced inequalities	4	13	0	0	0	0	0	1	0	1	2	0	0	0	0
17	15	life on land	4	13	0	0	0	0	0	0	0	0	1	2	0	0	1
		Total	492														

A sample of poststructuralist analysis: SDG 5

While it was still unclear why gender equality was the most frequently occurring in 2019 (Table 1), the corpus suggested that JP, viewed through a poststructuralist lens (Kubota & Miller, 2017), seemed to be a strong advocate for women or at least provides a forum for women or men to raise their voices, fighting against gender-based discrimination. The word "discrimination" was one of the 42 collocates within a range of 25 words to the left and 25 words to the right of "gender equality" (see Figure 3). The collocates of "gender equality" was visualized in a Wordcloud (Figure 4), with bigger fonts typically signifying more salient collocates. Teachers and students can zoom in on the word "discrimination" by clicking it on the AntConc tool (see the collocate ranked #12 in Figure 3) generating the collocates of "gender equality." The reference corpus set for the current analysis of collocates was the one million word "AmE06 Corpus." The corpus was built in

2006 (Goldsmith, 2022) from 15 sub-genres of written American English in 500 files containing around 2,000 word samples (The EdUHK Corpus Team, 2024).



	Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1	women	1	137	48	89	35	369.705	3.224
2	gender	2	65	35	30	23	208.714	3.636
3	orange	3	21	9	12	2	128.890	5.834
4	seal	4	13	5	8	2	106.021	7.289
5	girls	5	30	8	22	16	92.318	3.534
6	empowerment	6	25	2	23	18	82.190	3.699
7	equality	7	28	14	14	9	79.475	3.345
8	achieving	8	29	23	6	19	51.821	2.469
9	bpfa	9	7	5	2	2	48.371	6.395
10	rights	10	29	12	17	20	40.144	2.104
11	canada	11	9	5	4	2	37.745	4.395
12	discrimination	12	11	5	6	6	36.811	3.746
13	ugm	13	10	6	4	1	34.289	3.810
14	goals	14	48	27	21	29	33.815	1.404
15	iix	15	7	3	4	2	31.425	4.618
16	perpetuate	16	5	2	3	1	31.204	5.910
17	undp	17	17	9	8	5	30.310	2.466
18	ansye	18	5	3	2	2	29.883	5.717
19	dismantle	18	5	2	3	1	29.883	5.717
20	sdg	20	27	19	8	17	28.163	1.771
21	score	21	7	1	6	1	27.577	4.203
22	goal	22	19	15	4	12	25.527	2.066
23	beijing	23	6	4	2	2	25.162	4.395
24	structural	24	6	3	3	2	24.246	4.280

Figure 3. A screenshot of collocates linked to "gender equality" in the corpus

According to Laurence Anthony (2022), FreqLR denotes "the frequency of the collocate to the left and right of the search term," with FreqL only to the left and FreqR only to the right. Range means that a search term, like "women," the collocate of "gender equality" in Figure 3, occurs in 35 files. Likelihood is "a statistic measure of saliency (keyness)." Effect is "an effect size measure" Likelihood is useful to identify keywords or words that occur remarkably more frequently in a target corpus (the DIY JP corpus here) compared to the AmE06 Corpus as a reference corpus here.

When clicked, the AntConc generated the concordance lines highlighting "discrimination" as a KWIC and its collocates, focusing on one word on the left ("Sort 11L") and two words on right ("Sort 2 1R;" "Sort 3 2R") of the KWIC, ordered by value. Order by value allows AntConc users (i.e., language teachers and students) to learn alphabetically ordered collocates of the KWIC (e.g., "discrimination") on the left ("sort 1:1L"), such as "**e**nding discrimination," **g**ender discrimination," "**o**f discrimination," and so forth (Figure 5).

One might wonder why the word "orange" (see Figure 3) is a prominent collocate of "gender equality." Quantitative analysis to describe frequency and collocates (stage 1 in Baker & McEnery, 2015) was not enough. Proceeding to the second stage, I interpreted the word "orange." Based on concordance lines (see Figure 6), it turned out that "orange" here had nothing to do with a fruit but a color associated with a movement supporting women. Different from Figure 5 where the order by value was set at "Sort 1: 1L," in Figure 6, AntConc users can also analyze

what comes after the "orange" as KWIC in an alphabetical order by setting the "Sort 1" with "1R." This setting generates "Orange **B**ond…," "Orange **M**ovement™, and "Orange **P**ledge," among others.

Figure 4. A wordcloud of "gender equality"

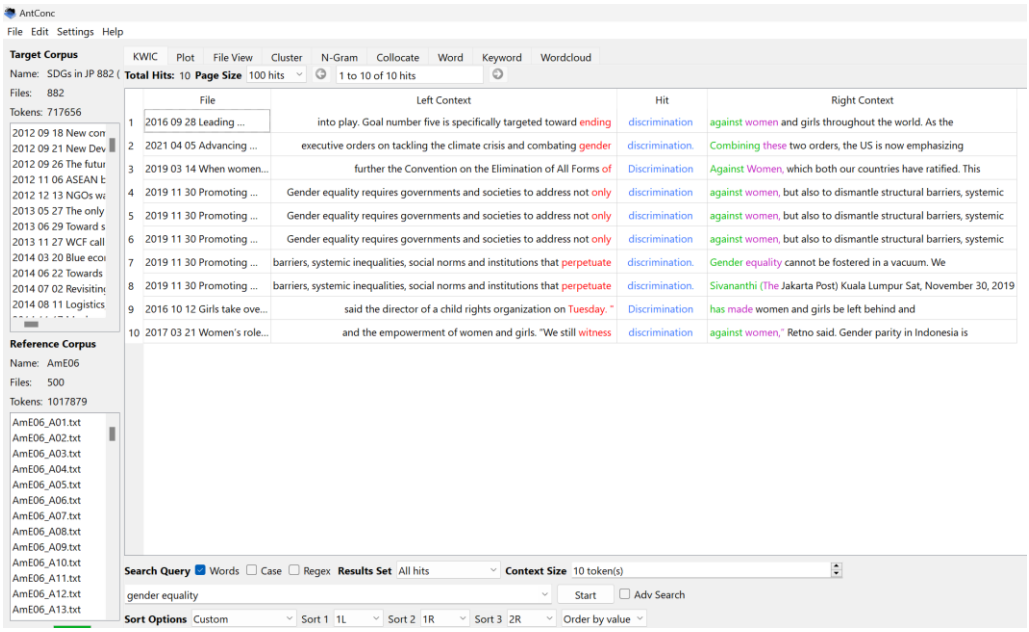


Figure 5. A screenshot of concordance lines highlighting "discrimination" as the KWIC

In the third stage, an analyst should explain the word within "a wider social context" (Baker & McEnery, 2015, p. 3). Regarding when the Orange Movement gains prominence in Indonesia, file names in Figure 6 suggested that all occurrences of the Orange Movement were in 2024. Based on one of the files dated July 4, 2024, the movement has three main objectives: "Establishing a gender-empowered financing system with transparent standards; mobilizing new capital through gender-lens investment products; [and] innovating for gender equality and economic empowerment" (Sagir, 2024).

Pushing to the fourth stage, Baker and McEnery (2015) suggest that analysts should be critical and initiate social transformation. Women should benefit from the Orange Movement ™, but the critical questions are which women, and how can

they be facilitated to achieve the movement's missions? How can (English language) education be used to disseminate the Orange Movement™ discourse?

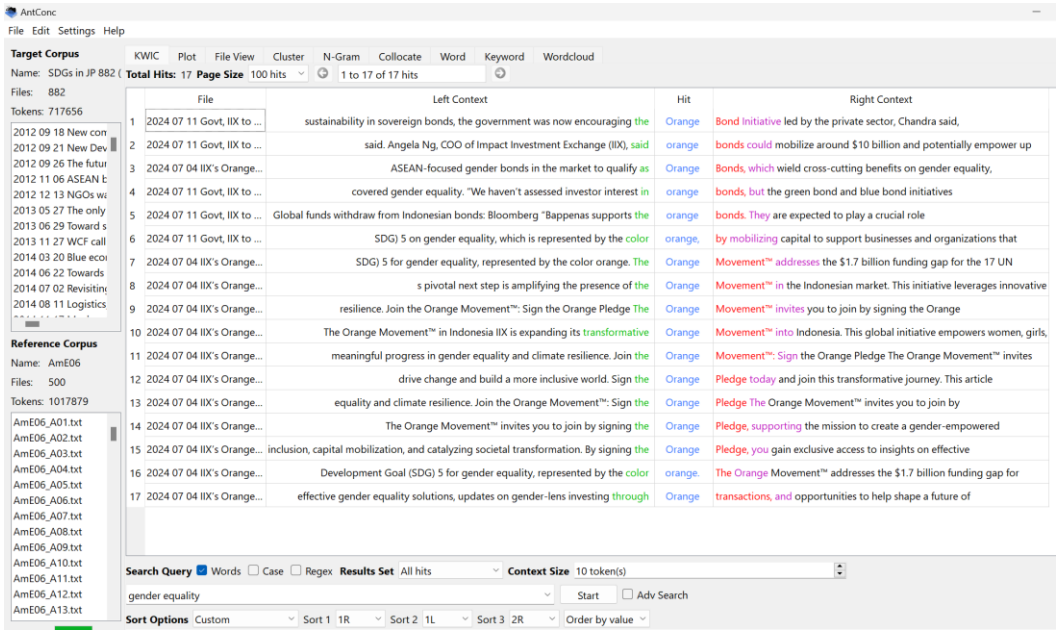


Figure 6. The concordance lines of "orange" as the KWIC

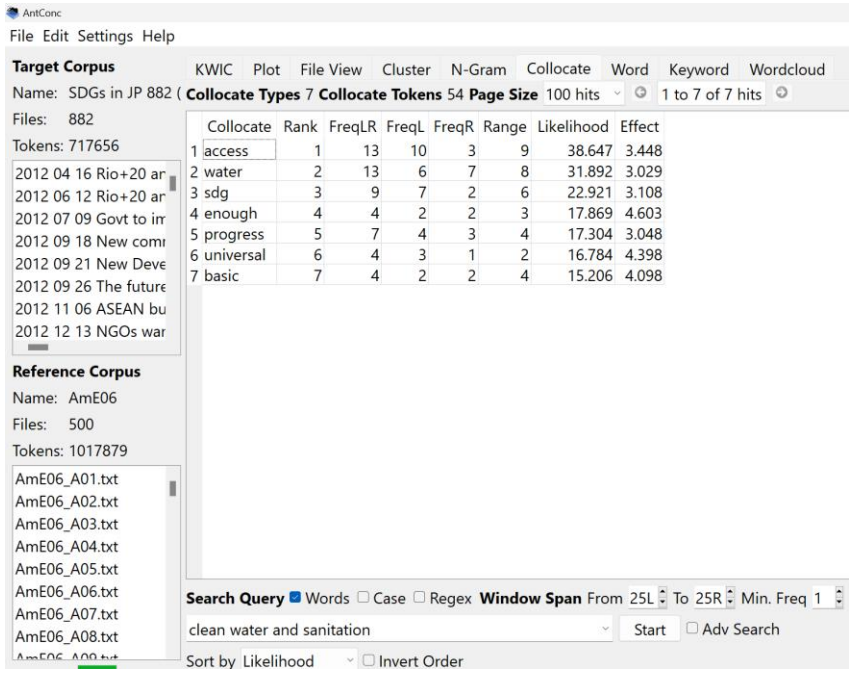


Figure 7. A screenshot of collocates linked to "clean water and sanitation" in the corpus

A sample of Marxist analysis: SDG 6

The key words "clean water and sanitation" is in the fourth rank of the most frequently occurring phrase in the JP corpus (see Table 1). The word "access" ranks the highest as the collocate of "clean water and sanitation" in the corpus (see Figure 7). When clicked for further analysis, AntConc generated concordance lines highlighting the word "access" as a KWIC preceeding by either "universal" or

"poor." Between the universal-poor continuum, there existed "increasing" and "improved access to clean water and sanitation" (see Figure 8). Interpreting the corpus data (the second stage in Baker & McEnery's [2015] framework) in light of the Marxist perspective (Block, 2015), lack of access to clean water and sanitation deprives people of their basic human rights. Situating SDG 6 in Indonesia (Baker & McEnery's [2015] third stage), one can learn that the idealized "universal access to clean water and sanitation" is enormously more difficult for children with disabilities (see concordance line 12 in Figure 8), particularly in areas commonly impacted by heat and drought like East Nusa Tenggara (Ekaristi, 2023). The critical question (stage 4 in Baker & McEnery [2015]) then becomes how we can help children with disabilities in drought impacted areas to increase access to clean water and sanitation. Ekaristi (2023), the author of concordance lines 12-13 in Figure 8, recommended that "the private sector should contribute to ensuring inclusive water access through initiatives like corporate social responsibility (CSR) programs." Accordingly, the civil society, including English language teachers and students, can practice writing a proposal to convince corporations to have meaningful and relevant CSR programs supporting children with disabilities attempting to access decent water and sanitation.

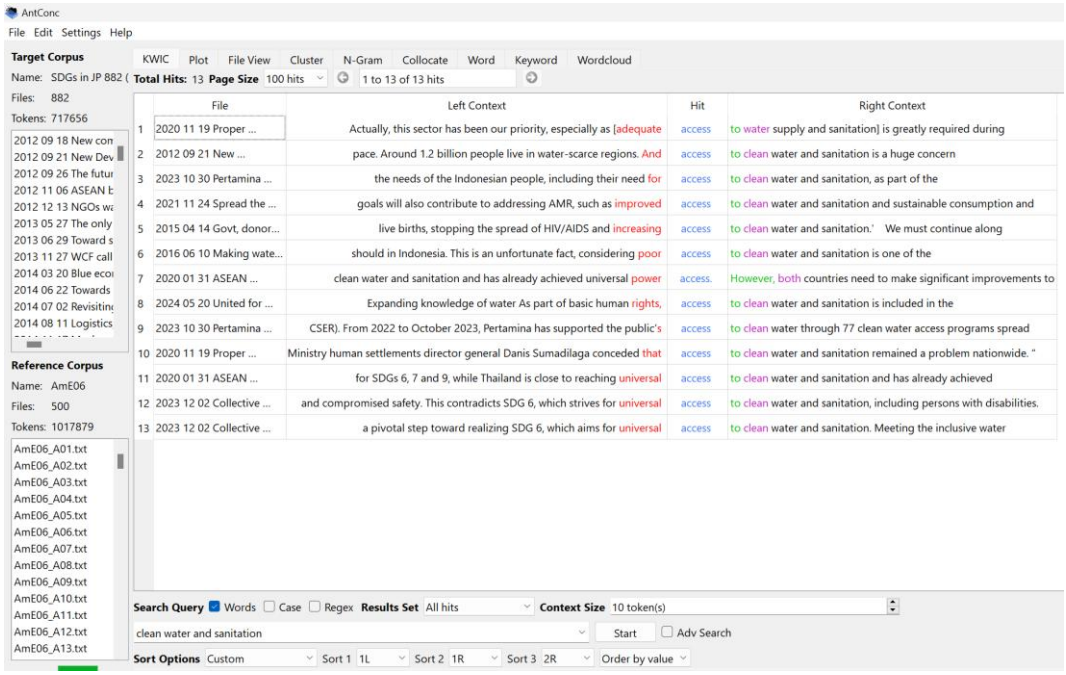


Figure 8. The concordance lines of "access" as the KWIC

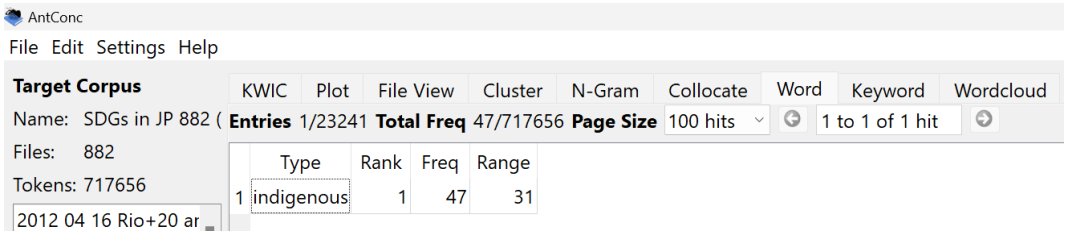


Figure 9. The word "indigenous" in the DIY corpus based on the JP

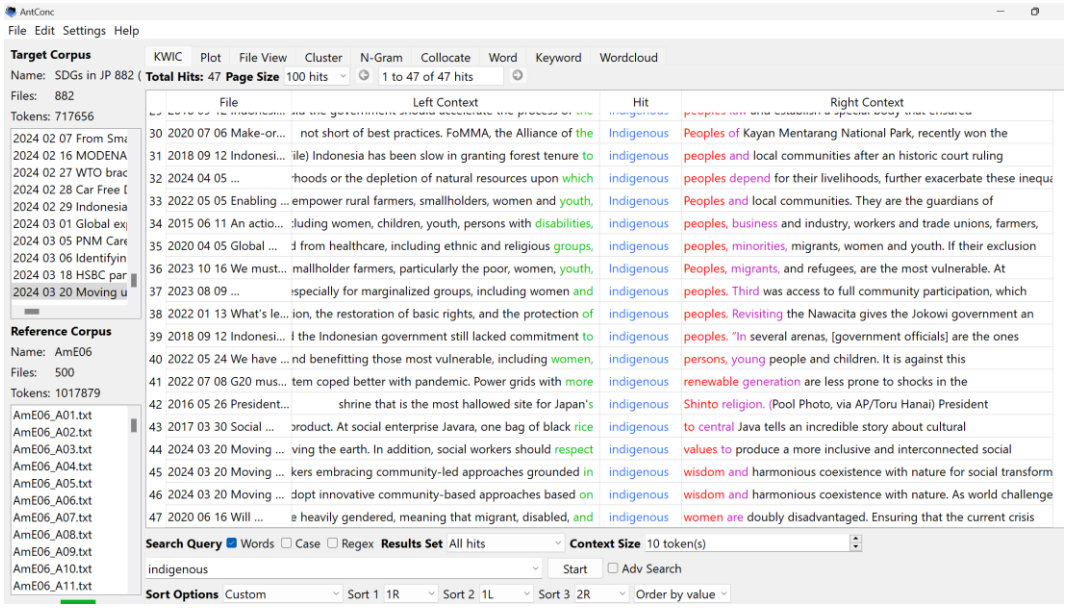


Figure 9 (continued). The word "indigenous" in the DIY corpus based on the JP

A sample of decolonial analysis on the "Indigenous" key word

The word "indigenous" appeared rather scarcely (i.e., 47 times in 31 texts published by JP from 2012 to mid-2024; see Figure 9). Furthermore, "peoples" and "people" (ranked first and ninth respectively) mattered in the "Indigenous" context (see Figures 10 and 11). Visualized in a Wordcloud, the words "peoples" and "people" were indeed prominent, followed by words like "nature," "communities," and "community" (see Figure 12), ranked fifteenth, twenty-first, and twenty-fourth, respectively (Figure 10). Of a particular interest is the word "peoples" (Figure 11), which indicates the diversity of Indigenous communities.

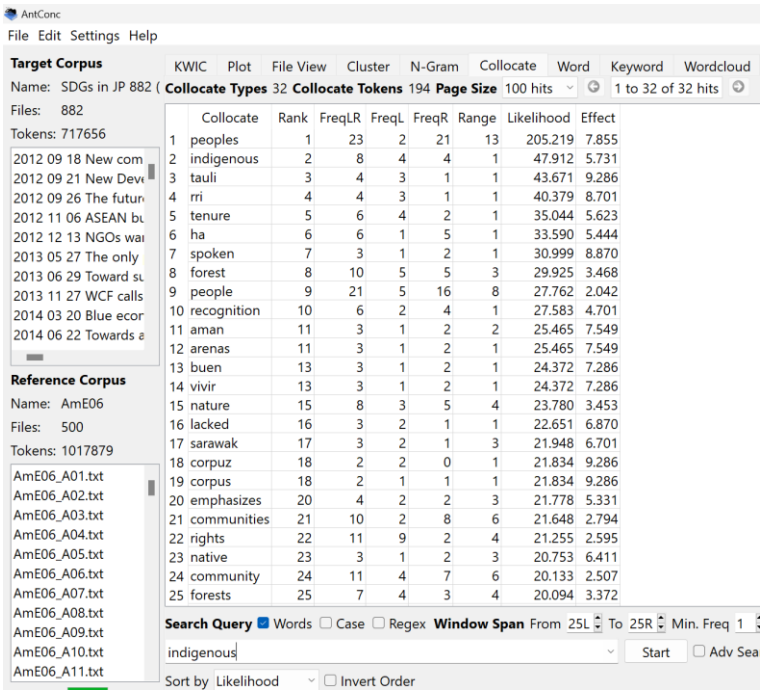


Figure 10. A screenshot of collocates linked to "indigenous" in the corpus

Going beyond the quantification of word occurrences (stage 1 in Baker & McEnery [2015]), in stage 2, corpus analysts examine concordance lines to interpret the immediate context of a text. For instance, AntConc users curious about how "people" and "peoples" are used in a larger context can click the "indigenous" hit, which happens to originate from the same text file (see Figure 13). Examining "people" and "peoples" in larger contexts within and across texts (Figures 11 and 13) constitutes "narrow reading" where learners "[read] on the same topic over the course of a number of texts" (Schmitt & Carter, 2000, p. 5; see also Chang & Renandya, 2021). AntConc users can also explore the place of "nature" in Indigenous peoples' lives for "social transformation" and "future generations" (Figure 14).

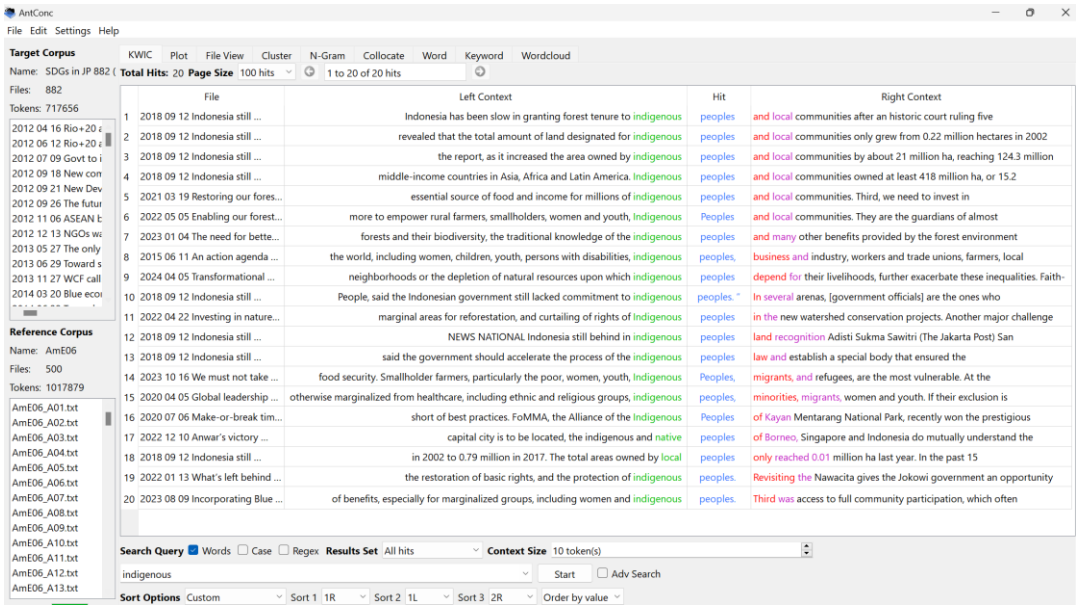


Figure 11. The concordance lines of "peoples" collocating with "indigenous," among others

Figure 12. A wordcloud of "indigenous"

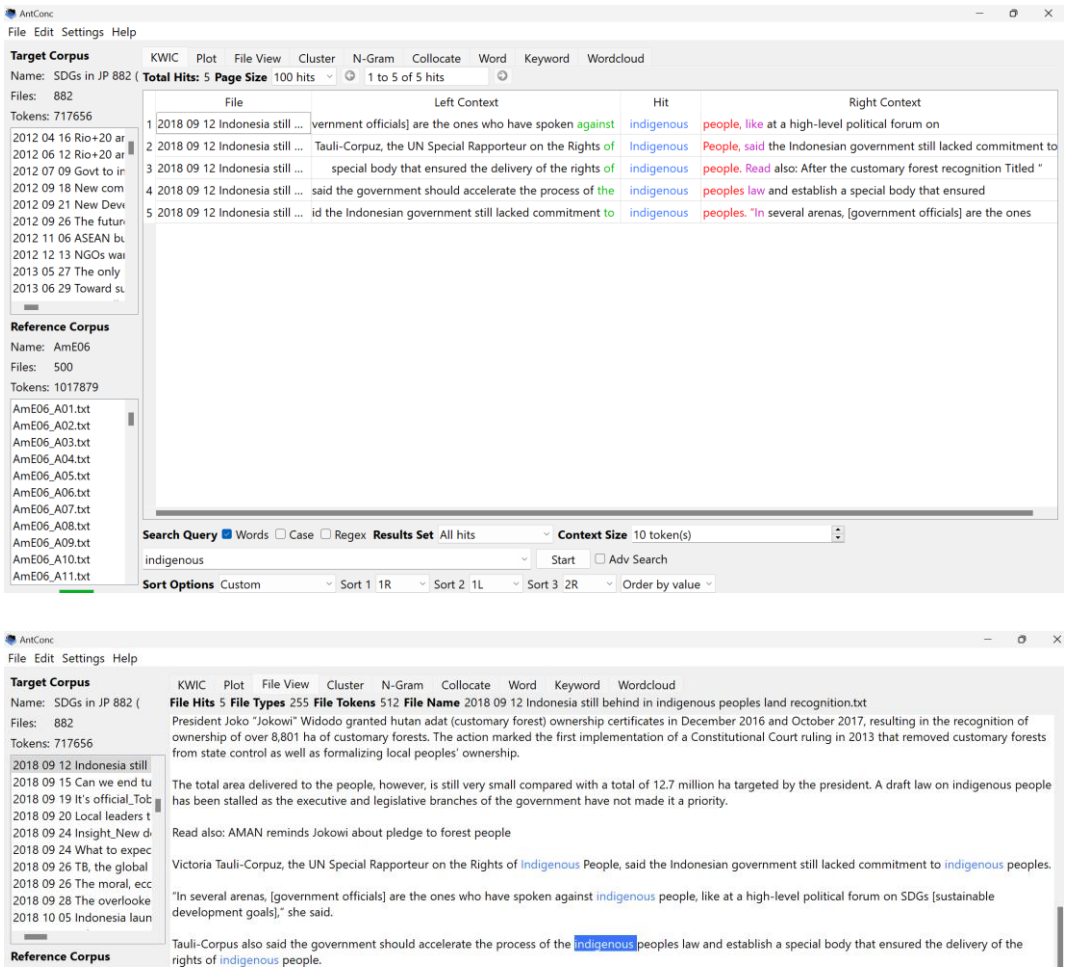


Figure 13. The concordance lines and a larger context of "indigenous" collocating with "people*"

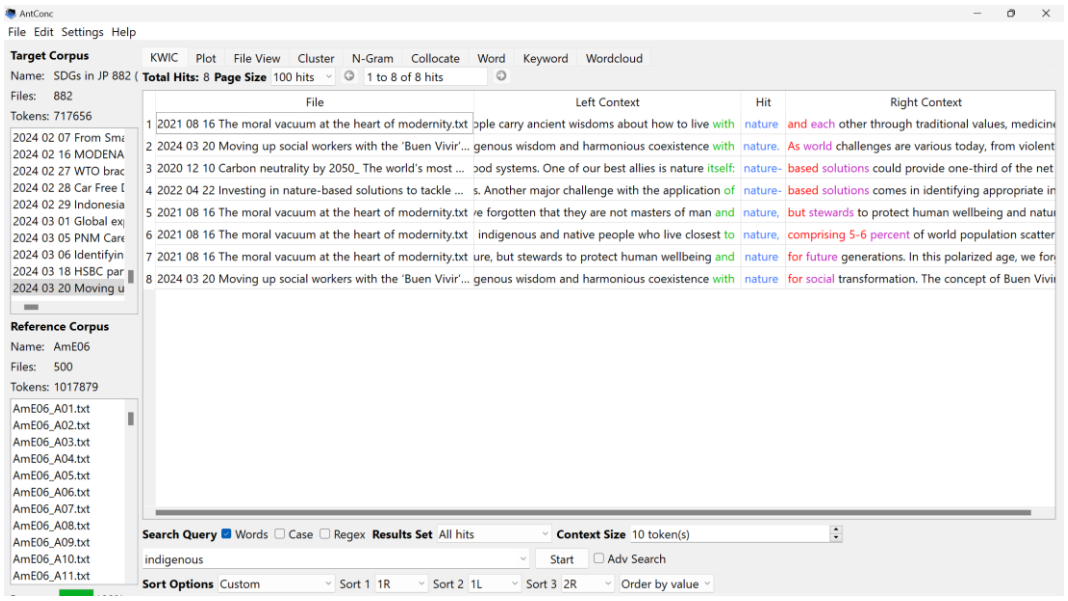


Figure 14. The concordance lines of "nature" as a collocate of "indigenous"

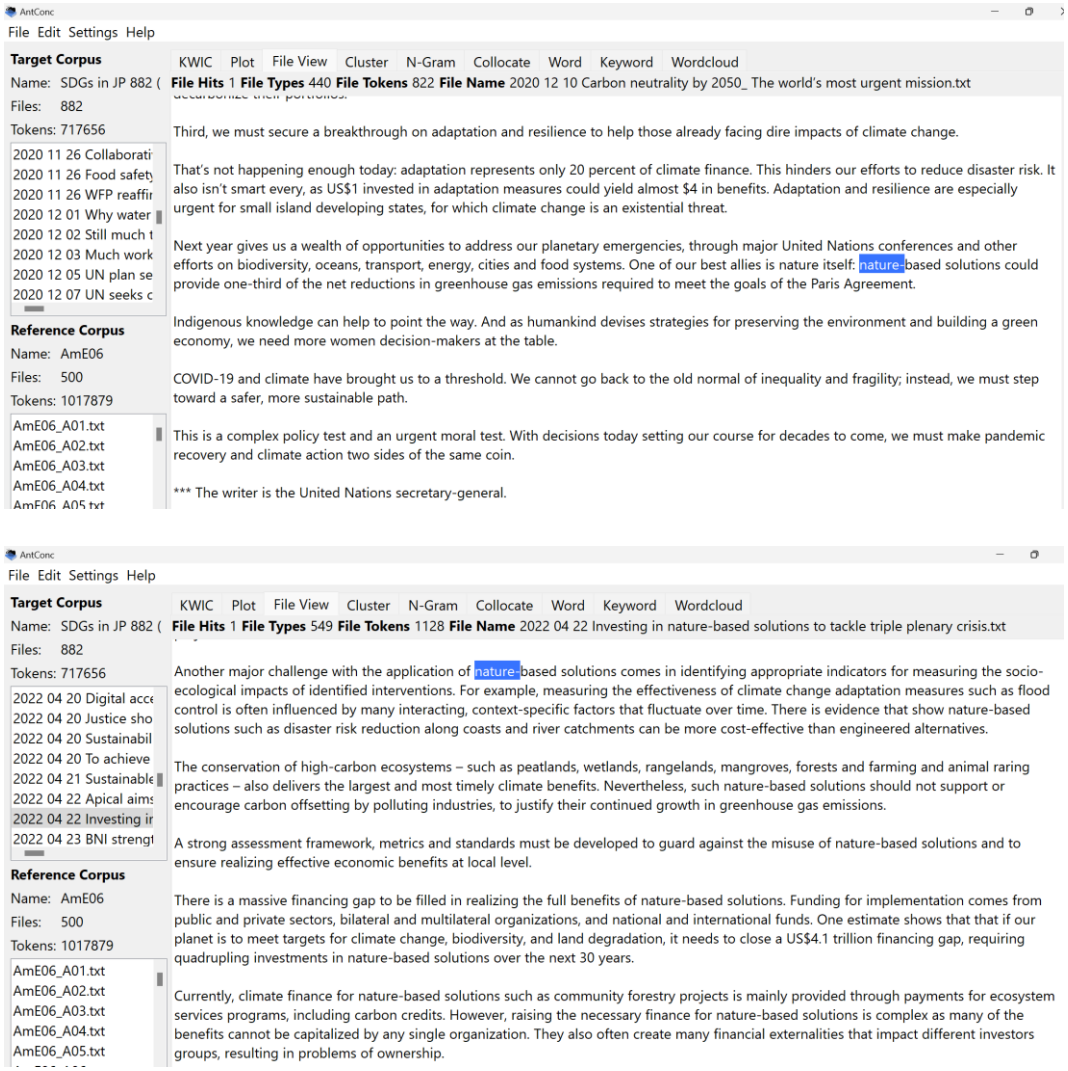


Figure 15. Exploring "nature-based solutions" as a narrow-reading activity

Narrow reading is also facilitated by the DIY JP-based corpus when AntConc users come across a phrase like "nature-based solutions" (Figure 14). In the corpus, the phrase came from two sources and their larger contexts are hence explorable. One can critically ask *what* "nature-based solutions" entail, *who* promote(s) them, *who* can/should be involved in them, *why* they matter, and *how* various stakeholders (especially English language teachers/learners) can contribute to them. Some of the answers can be found in two texts published by JP. In the text published on April 22, 2022, "nature-based solutions" include "disaster risk reduction along coasts and river catchments" and "community forestry projects." Language students can learn how to scan a textual signpost like "such as" to identify what "nature-based solutions" involve (Anbumozhi, 2022; Figure 15). The United Nations secretary-general endorsed nature-based solutions, arguing that they "could provide one-third of the net reductions in greenhouse gas emissions…" (Guterres, 2020; Figure 15). Guterres also believed that Indigenous people and "women-decision makers" should be involved in protecting the natural world, making our earth more sustainable. Although not captured in Figure 15, in the remainder of the text, Anbumozhi (2022) suggested that the costly nature-based solutions are only

possible with "effective governance system," requiring solid coordination and collaboration between stakeholders.

In the third stage (Baker & McEnery, 2015), the sentiment of Sawitri's (2018) report that "the Indonesian government still lacked commitment to indigenous peoples"—the UN Special Rapporteur on the Rights of Indigenous People was quoted as saying—seemed to account for the scarcity of the key word "indigenous" collocating with "people*" (see Figure 13). From a decolonial perspective (Smith, 2021), the Indigenous peoples are still marginalized, particularly in the wider context of "modernity" (see the JP article title under "File" dated August 16, 2021 in Figure 14) glorifying profits and excessive consumptions leading to the destruction of the Earth (Sheng, 2021).

Moving to the fourth stage (Baker & McEnery, 2015), the phrase "indigenous-community empowerment programs" found in Inforial (2022) sounds promising to be a good practice, but what are they? Who determine(s) that the programs are indeed relevant and useful to Indigenous communities? How can Indonesian civil society push the government to be more committed to Indigenous peoples? Furthermore, how language learners of various levels of proficiency can contribute to "nature-based solutions" (Figure 15) have yet to be specifically addressed. Revisiting the idea that multiple stakeholders should synergize in protecting the environment (Anbumozhi, 2022; Guterres, 2020; see Figure 15), language teachers can pose difficult questions to be addressed by their (relatively advanced) students. The questions may include: How can Indigenous knowledge in your localities – be they in forest, coastal, mountainous, or arid areas – guide you to be more environmentally conscious and actively engaged in preserving nature? Do female Indigenous people from a low social class have a place in contributing to ensuring a more sustainable earth? If yes, why, and how? To address these questions, teachers can unpack the key words as prompts (e.g., "female Indigenous people," "social class," and "more sustainable earth") that students can use when consulting the DIY JP corpus, Google, or generative AI tools. By addressing questions like those suggested above, teachers and students explore intersectionality (Kayi-Aydar et al., 2022; Kubota & Miller, 2017): indigeneity (decoloniality), gender identity (poststructuralism), and social class (Marxism), with the SDG discourse as the main focus.

Discussion

Building upon Yu et al.'s (2024) call for more inquiry into English teachers' readiness in providing SDG-related materials, the current study has explored the DIY JP corpus as a data-driven learning resource (Boulton, 2016) on SDGs for language teacher educators (like myself), teachers, and students. Furthermore, the present study is wider in scope than Pinto's (2021) corpus analysis of the Brazilian Corpus of Academic English (BrACE) on poverty (SDG 1); Brun-Mercer's (2021) analysis of gender (in)equality (SDG 5) in the corpus of the United Nations' General Debate addresses; and Liu et al.'s (2024) critical discourse analysis of environmental (sustainability) themes (covered at personal, local, national, and international levels) in a corpus of English language textbooks from 48 volumes used nationwide in Chinese universities, with no specific SDGs being addressed. My study provides a more comprehensive coverage of key phrases of 17 SDGs, despite being limited to keyword frequencies from 2012 to mid-2024 in JP (Table

1), viewed through critical lenses of CDA (Baker & McEnery, 2015) and criticality (Kubota & Miller, 2017; see Table 2).

The current findings can be synthesized in Table 2. The first row is based on Kubota and Miller (2017) and the left column under the theme row on Baker and McEnery (2015). Although English language teachers do not need to mention Poststructuralism, Marxism, and Decoloniality in their language classrooms, these perspectives of criticality can be used as their guidelines for grouping SDGs based on identity politics (e.g., gender [SDG 5]), access to materials/opportunities (e.g., access to clean water and sanitation [SDG 6], and the voices or silences of Indigenous peoples (especially in former colonized nation states in the Global South) encompassing various SDGs (e.g., climate change [SDG 13]), respectively. From Baker and McEnery's (2015) viewpoint, corpus-informed analyses on language can be done in the description and interpretation stages. That is, teachers and students can focus their attention to highly frequent SDGs-related words or phrases and prominently salient collocates of key words (e.g., "gender equality," "indigenous," "nature") under scrutiny.

Table 2. The synthesis of Baker and McEnery (2015) and Kubota and Miller (2017)

	Poststructuralism	**Marxism**	**Decoloniality**
Theme	Gender equality (SDG 5)	Access to clean water and sanitation (SDG 6)	"Indigenous"
Stage			
1. Description	SDGs-related words/phrases: Their frequencies and collocates		
2. Interpretation	Concordance lines of "orange" → "Orange Movement™"	Concordance lines of "access": "poor" vs. "universal"	Concordance lines of "Indigenous," "nature," and "marginalized"
3. Explanation	The "Orange Movement™" is trending in 2024	The difficulty of "universal access to clean water and sanitation" especially for children with disabilities	"The Indonesian government still lacked commitment to Indigenous peoples…" (Sawitri, 2018)
4. Critical evaluation	How can we achieve the Orange Movement™'s mission?	How can we "strive for universal access to clean water and sanitation"?	How can Indonesian civil society push the government to be more committed to Indigenous peoples?

At first, especially for younger and beginner learners of English, teachers can direct students' attention to saliency of (key)words by displaying wordclouds (Figures 4 and 12). Then, they can discuss whether certain words can go together with other words in the wordclouds (e.g., *Can "recognition" be used together with "peoples"? Which words can go together with "peoples"?* etc.). To check whether their guesses are justifiable, the teachers can demonstrate the AntConc/corpus search and explain why certain words can or cannot go with particular words (e.g., "'*Marginalized' can be followed by'communities' because the former is an adjective and the latter is a plural noun*). For more challenging activities, teachers

can use the Collocate feature in AntConc (Figures 3, 7, and 10). For analytical or advanced learners, they can also be guided to inquire into the concordance lines (e.g., "nature" in Figure 14). It is expected that teachers and learners alike can serendipitously find interesting (or even previously unknown) collocates like "the Orange Movement™" (Figure 6), "universal access to clean water" (Figure 8), and "nature-based solutions" (Figure 14).

Beyond the concordance lines are the larger contexts of where certain collocates occur in one (e.g., "indigenous people*" [Figure 13]) or more texts (e.g., "nature-based solutions" [Figure 15]). In the larger contexts, language learners can do narrow reading to find explanations or what accounts for specific collocates (e.g., why the word "indigenous" did not occur very frequently in the JP corpus; why the term "Orange Movement™" became salient in the corpus) and raise critical evaluations (or questions) at a content/discourse level (Baker & McEnery, 2015) that push a conversation forward, aiming at social transformation (e.g., *Which women can benefit from the "Orange Movement™" and how? What about women from Indigenous communities and a lower social class?*).

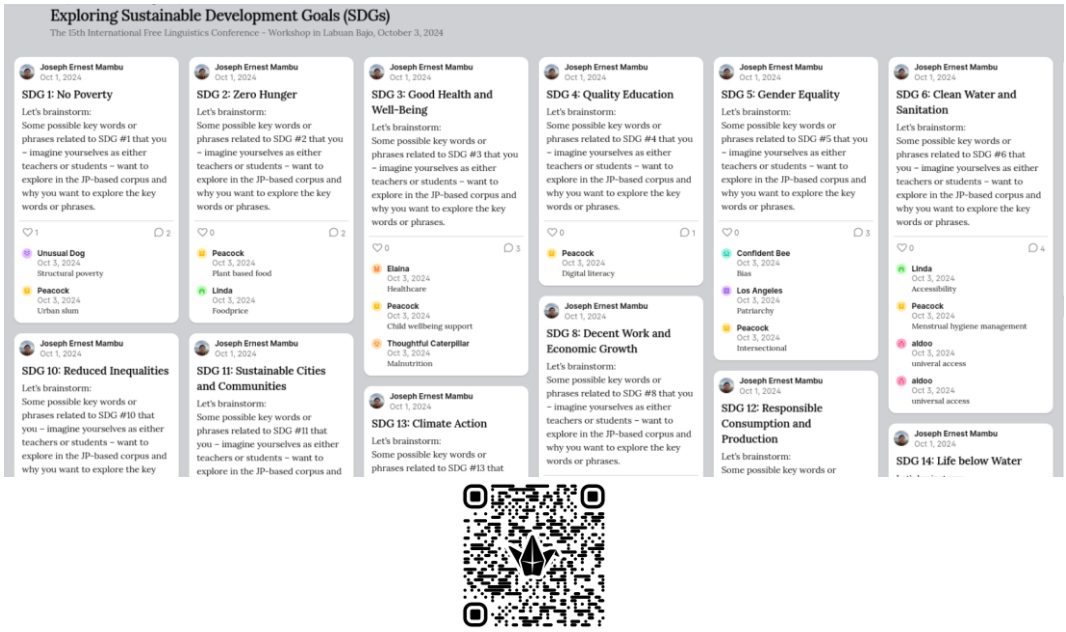


Figure 16. A screenshot of a Padlet-facilitated activity on exploring SDGs

Providing concrete classroom examples or sample lesson plans based the DIY JP corpus and discussing potential challenges in utilizing the corpus go beyond the scope of the present study. That said, I simulated a corpus-informed activity at a workshop in the 15th International Free Linguistics Conference in Labuan Bajo on October 3, 2024 (see Figure 16). Some responses documented in the activity included "universal access" (SDG 6), which aligned with what I prepared for the next day's plenary session (displayed as Figure 8 here). Then, I checked whether their suggested phrases occurred in the JP corpus. Interestingly, there is no occurrence of "structural poverty" (SDG 1) in the DIY JP corpus. The nonexistent phrase might indicate that JP was biased towards the middle-/upper social class. The workshop activity can certainly be applied in classroom practices to explore more SDG-related themes, such as whether the JP corpus, following up on Brun-

Mercer's (2021) corpus analysis of the UN General Assembly General Debate addresses, subtly perpetuates patriarchy by portraying men in more prominent roles than women.

Overall, the present study corroborates the spirit of Crosthwaite et al's (2023) study in that language teacher education programs should incorporate technological, pedagogical, and content knowledge (TPACK) into language teachers' teaching preparation, especially that which equips teachers with corpora and DDL. Although I did not explore corpus use for lesson planning, asking teachers to focus on what key words were related to SDGs, why they wanted to focus on them (Figure 16), and whether their chosen key words/phrases occurred in a corpus could be part of a larger corpus-informed lesson. In other words, familiarity with corpus tools like AntConc and DDL (technological knowledge) potentially supports the art of asking questions (pedagogical knowledge) and the exploration of SDGs-related themes (content knowledge).

Conclusion

Language teachers/teacher educators and learners can utilize the JP corpus beyond concordance lines, especially for narrow-reading materials (Schmitt & Carter, 2000). Reading narrowly should be with a purpose, too. Teachers can create critical questions informed by a specialized corpus (like the JP corpus here; see also the last row of Table 2). Then, students explore the corpus or other reliable sources on (SDGs-related) themes (e.g., the Orange Movement™ [Sagir, 2024]; water access to people with disabilities [Ekaristi, 2023]; the criticism toward modernity [Sheng, 2021]) they like/choose/decide to work on.

If teachers decide to use a specialized corpus like the JP corpus, they can expose the students to a freeware like AntConc. Teachers may share the corpus only for private/classroom learning purpose only, or ask them to access available corpora like BNC and COCA, or build their own DIY corpus – with technical guidance from the teachers. Using corpus-informed and critical approaches to analyzing SDGs in JP is only one way of addressing SDGs. Also important to address is how to synergize corpora and GenAI tools. It will be useful, for example, to compare and contrast findings from the JP-based specialized corpus of SDGs with what GenAI tools can say about the dissemination of SDGs in the Indonesian context. Finally, teacher education programs should incorporate the corpus-informed approach to exploring multimodal corpora (Staples & Anthony, 2023), which combine textual and visual SDG-related discourses.

Acknowledgements

I thank the Chief Editor of *The Jakarta Post* for securing permission for me to use and do a corpus analysis of the SDGs-related texts accessed from its website. I also thank the two anonymous reviewers for greatly helping me to improve my paper.

References

Abednia, A., & Crookes, G. V. (2019). Critical literacy as a pedagogical goal in English language teaching. In X. Gao (Ed.), *Second handbook of English Language Teaching* (pp. 1–21). Springer. https://doi.org/10.1007/978-3-319-58542-0_14-2

Anbumozhi, V. (2022, April 22). Investing in nature-based solutions to tackle triple plenary crisis. *The Jakarta Post*. https://www.thejakartapost.com/paper/2022/04/21/investing-in-nature-based-solutions-to-tackle-triple-plenary-crisis.html

Anthony, L. (2022, May 24). Explanation of tabs/headers under each tool? Google Groups. https://groups.google.com/g/antconc/c/xpCD7moyXvM

Anthony, L. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/AntConc

Arikan, A., & Zorba, M. G. (2024). Raising pre-service English language teachers' awareness of sustainable development goals through literary texts. *International Journal of Sustainability in Higher Education, 25*(8), 1681-1695. https://doi.org/10.1108/IJSHE-06-2023-0256

Arslan, S., & Curle, S. (2024). Institutionalising English as a foreign language teachers for global sustainability: Perceptions of education for sustainable development in Turkey. *International Journal of Educational Research, 125*, Article 102353. https://doi.org/10.1016/j.ijer.2024.102353

Baker, P., & McEnery, T. (2015). Introduction. In P. Baker & T. McEnery (Eds.), *Corpora and discourse studies: Integrating discourse and corpora* (pp. 1-19). Palgrave Macmillan.

Block, D. (2015). Social class in applied linguistics. *Annual Review of Applied Linguistics, 35*, 1-19. https://doi.org/10.1017/S0267190514000221

Boulton, A. (2016). Integrating corpus tools and techniques in ESP courses. *ASp: La revue du GERAS, 69*, 113-137. https://doi.org/10.4000/asp.4826

Bowker, L. (1998). Using specialized monolingual native-language corpora as a translation resource: A pilot study. *Meta: Journal des Traducteurs (Translators' Journal), 43*(4), 631-651. https://doi.org/10.7202/002134ar

Brun-Mercer, N. (2021). Women and men in the United Nations: A corpus analysis of General Debate addresses. *Discourse & Society*, *32*(4), 443–462. https://doi.org/10.1177/0957926521992145

Chang, A. C-S., & Renandya, W. A. (2021). The effect of narrow reading on L2 learners' vocabulary acquisition. *RELC Journal, 52*(3), 493-508. https://doi.org/10.1177/0033688219871387

Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics, 3*(3), 1-4. https://doi.org/10.1016/j.acorp.2023.100066

Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning, 36*(7), 1392-1420. https://doi.org/10.1080/09588221.2021.1995001

Ekaristi, H. T. (2023, December 2). Collective action for an equal world. *The Jakarta Post*. https://www.thejakartapost.com/opinion/2023/12/02/collective-action-for-an-equal-world.html

Firdaus, F. (2017, August 3). 'A tragic, forgotten place.' Poverty and death in Indonesia's land of gold. *TIME*. https://time.com/4880190/papua-poverty-shootings-justice-paniai/

Goldsmith, E. (2022, February 28). AntConc version 4: First impressions. *Signs & Symptoms of Translation: A Blog by a Spanish to English Medical Translator*. https://signsandsymptomsoftranslation.com/2022/02/28/antconc-4/

Guterres, A. (2020, December 10). *Carbon neutrality by 2050: The world's most urgent mission*. The Jakarta Post. https://www.thejakartapost.com/academia/2020/12/10/carbon-neutrality-by-2050-the-worlds-most-urgent-mission.html

Inforial. (2022, December 19). PetroChina Jabung confident the best is yet to come as development programs go on. *The Jakarta Post*. https://www.thejakartapost.com/adv/2022/12/19/petrochina-jabung-confident-the-best-is-yet-to-come-as-development-programs-go-on.html

Kayi-Aydar, H., Varghese, M., & Vitanova, G. (2022). Intersectionality for TESOL education: Connecting theory and justice pedagogy. *The CATESOL Journal*, *33*(1), 1–10. https://doi.org/10.5070/b5.35903

Kubota, R., & Miller, E. R. (2017). Re-examining and re-envisioning criticality in language studies: Theories and praxis. *Critical Inquiry in Language Studies*, *14*(2–3), 129–157. https://doi.org/10.1080/15427587.2017.1290500

Liu, Y., An, L., & Chen, S. (2024). Environmental sustainability education in twelve series of Chinese university English language textbooks. *SAGE Open*, *14*(3), 1–12. https://doi.org/10.1177/21582440241266829

Maijala, M., Heikkola, L. M., Kuusalu, S.-R., Laine, P., Mutta, M., & Mäntylä, K. (2023). Pre-service language teachers' perceptions of sustainability and its implementation in language teaching. *Language Teaching Research*. https://doi.org/10.1177/13621688231170682

Maley, A., & Peachey, N. (Eds.). (2017). *Integrating global issues in the creative English language classroom: With reference to the United Nations Sustainable Development Goals*. British Council. https://www.teachingenglish.org.uk/sites/teacheng/files/PUB_29200_Creativity_UN_SDG_v4S_WEB.pdf

Mambu, J. E. (2023a). Embedding Sustainable Development Goals into critical English language teaching and learning. *Critical Inquiry in Language Studies, 20*(1), 46-76. https://doi.org/10.1080/15427587.2022.2099863

Mambu, J. E. (2023b). Embracing Sustainable Development Goals critically to explore life purposes in English language teacher education. *TEFLIN Journal, 34*(2), 264-282. http://dx.doi.org/10.15639/teflinjournal.v34i2/264-282

McCarthy, M. (2004). *Touchstone: From corpus to course book.* Cambridge University Press.

McEnery, T., Xiao, Z., & Tono, Y. (2006). *Corpus based language studies: An advanced resource book*. Routledge. https://doi.org/10.1017/s0047404508080615

McEntee-Atalianis, L. J. (2017). "Leave no one behind": Linguistic and digital barriers to the dissemination and implementation of the United Nations Sustainable Development Goals. *Language Problems & Language Planning, 41*(3), 217-244. https://doi.org/10.1075/lplp.00001.mce

Mercer, S., Correia Ibrahim, N., Bilsborough, K., Jones, C., & Potzinger, C. (2023). Teacher perspectives on addressing environmental issues in ELT. *ELT Journal*, *77*(4), 393–406. https://doi.org/10.1093/elt/ccac039

Pinto, P. T. (2021). The Sustainable Development Goals (SDGs) words 'poverty' and 'sustainability' in the Brazilian research. *Cadenos de Linguística, 2*(4), 1-13. https://doi.org/10.25189/2675-4916.2021.v2.n4.id440

Reppen, R. (2022). Building a corpus: What are key considerations? In A. O'Keefe & M. J. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed, pp. 31-37). Routledge.

Sagir, A. (2024, July 3). IIX's Orange Movement™ leads the charge in Indonesia's impact investment landscape. *The Jakarta Post*. https://www.thejakartapost.com/business/2024/07/04/iixs-orange-movement-leads-the-charge-in-indonesias-impact-investment-landscape.html

Saltford, J. (2003). *The United Nations and the Indonesian takeover of West Papua: The anatomy of betrayal*. Routledge.

Sawitri, A. S. (2018, September 12). Indonesia still behind in indigenous peoples land recognition. *The Jakarta Post*. https://www.thejakartapost.com/news/2018/09/11/indonesia-still-behind-in-indigenous-peoples-land-recognition.html

Sheng, A. (2021, August 16). The moral vacuum at the heart of modernity. *The Jakarta Post*. https://www.thejakartapost.com/paper/2021/08/15/the-moral-vacuum-at-the-heart-of-modernity.html

Schmitt, N., & Carter, R. (2000). The lexical advantages of narrow reading for second language learners. *TESOL Journal, 9*(1), 4-9. https://doi.org/10.1002/j.1949-3533.2000.tb00220.x

Smith, L. T. (2021). *Decolonizing methodologies: Research and indigenous peoples* (3rd ed.). Zed Books.

Staples, S., & Anthony, L. (2023). Corpus linguistics in EFL language teaching: Insights from research and practice. *JALT Journal*, *45*(2), 246–263. https://doi.org/10.37546/jaltjj45.2-4

Sulaiman, S. R. (2018, July 16). A guide to understanding the Freeport divestment deal. *The Jakarta Post*. https://www.thejakartapost.com/news/2018/07/13/a-guide-to-understandingthe-freeport-divestment-deal.html

Sundh, S. (2016). A corpus of young learners' English in the Baltic region – Texts for studies on sustainable development. *Discourse and Communication for Sustainable Education, 7*(2), 92-114. https://doi.org/10.1515/dcse-2016-0018

The EdUHK Corpus Team. (2024). Module 4: AntConc. In the EdUHK Corpus Team, *Enhancing corpus literacy and competence for corpus-based language pedagogy for in- and pre-service English teachers*. Pressbooks. https://pressbooks.pub/cmtry1/chapter/antconc/

Tonkin, H. (2023). Introduction: Diversity of language, unity of purpose. In L. J. McEntee-Atalianis & H. Tonkin (Eds.), *Language and sustainable development* (pp. 1–10). Springer. https://doi.org/10.1007/978-3-031-24918-1_1

Toriida, M-C. (2016). Steps for creating a specialized corpus and developing an annotated frequency-based vocabulary list. *TESL Canada Journal, 34*(1), 87-105. https://doi.org/10.18806/tesl.v34i1.1257

Yu, B., Guo, W. Y., & Fu, H. (2024). Sustainability in English language teaching: Strategies for empowering students to achieve the Sustainable Development Goals. *Sustainability, 16*, Article 3325. https://doi.org/10.3390/su16083325