CLUSTERS OF SDG KEYWORDS

Angvarrah Lieungnapar

*Faculty of Humanities and Social Sciences, Suan Sunandha Rajabhat University,
Bangkok, Thailand*
E-Mail: angvarrah.li@ssru.ac.th,

ABSTRACT

Since the UN adopted the Sustainable Development Goals (SDGs) in 2015, SDG concepts have been integrated in several fields including education. For EFL lecturers, a simple way to allow their students adopt SDG concepts is through reading documents related to SDGs. However, identifying reading topics related to SDG concepts is problematic since the concept is too broad and the available SDG keywords are limited. Under the assumption that keywords and their clusters (multi-word terms) based on co-occurrence can reflect the topics of the content, this study aims to develop a list of clusters of SDG keywords on the basis of Log-likelihood (LL), a statistical calculation which generate keyness values based on frequency of occurrence. Reports, explanations and descriptions related to the 17 SDG goals were downloaded from UN Sustainable Development websites. The keywords were identified by Key-BNC, which is a online application providing a simple interface for calculating comparative keyword statistics against a word list from the British National Corpus (BNC). The AntConc program was used to identify clusters of keywords. Most of the keywords found were obviously related to each SDG goal: poverty (SDG1), hunger (SDG2), health (SDG3), and education (SDG4). Some of them were indirectly but relatively related to the goals: pandemic (SDG1), wasting (SDG2), global (SDG3) and closures (SDG4). The results of clusters of keywords can be applied to identify more specific topics related to each SDG goal.

Keywords: Key-BNC, Keyword analysis, Topic identification

INTRODUCTION

Sustainable development goals (SDGs) have been incorporated into several disciplines, due to the growing public interest in social concerns. Concerning a wide range of disciplines, research in this area has evolved into different implementations and directions involving diverse practices and applications (e.g. [1], [2], [3]).

UNESCO pinpoints the significance of the implementation of the concepts of sustainable development into education, which is formally called Education for Sustainable Development (ESD). ESD allows teaching and learning methods to motivate learners to acquire the knowledge, skills, attitudes and values that are necessary to change their behavior and take action for sustainable development [4].

In the area of English as a Foreign Language (EFL), many EFL studies apply the Sustainable Development Goals (SDGs) as a framework for the EFL classroom to promote competencies like building learners' knowledge and awareness of global issues, critical

thinking and analytical skills; and action for positive social and political change. The example studies are an application of the United Nations' Sustainable Development Goals (SDGs) as reading content (Content-Based Instruction) to enhance and discover affective factors related to learners as individuals and citizens of the world [5], embedding Sustainable Development Goals into critical English language teaching and learning [6], and the analysis of a corpus built from sample texts with environmental themes found in English as a Foreign Language textbooks [7].

Although there seems to be a large number of studies inducing the concepts of SDGs, there has been little research identifying topics related to SDGs on the basis of keyword analysis. With this lack of SDG keywords availability, it can be difficult to pinpoint reading topics that are relevant to SDGs key concepts. Under the assumption that keywords and their clusters (multi-word terms) based on co-occurrence can reflect topics in a given corpus, this study aims to develop a list of clusters of SDG keywords so that potential topics related to SDGs can be identified for further study.

Keyword analysis is a corpus-based method, using automatically-generated analytical techniques, to identify words having greater frequency in one corpus when compared to another [8]. In other words, keyword analysis is a corpus-based technique which identifies important words, or keywords, in a given corpus as compared against a benchmark corpus [9]. The term keywords used in this paper refers to words with high keyness scores in a target corpus when compared to a comparative corpus.

OBJECTIVE

This study uses keyword analysis as a theoretical framework for identifying topics as a goal. This paper argues that topics related to SDGs can be revealed through clusters of keywords. The purpose is then to examine the keywords of the SDGs corpus created and explore the clusters of keywords related to SDGs.

METHODOLOGY

Corpora were collected from UN websites and created as the target corpus for the analysis compared to the reference corpus, which were BNC. The corpora consisted of 144 texts, including reports, explanations and descriptions from the 17 SDG goals. The total size of the corpus was 347,753 tokens, 11,353 types. They were purposefully collected from 7 UN websites showing reports and giving information about SDGs.

Key-BNC, an online tool with a user-friendly interface for computing comparative keyword statistics against a word list from the British National Corpus, was initially used to reveal keywords compared with the BNC as the reference corpus and SDGs corpus as the target corpus. Keywords in this study were analyzed based on Log-likelihood (LL), a statistical calculation which generates keyness values based on frequency of occurrence.

Typically, distinctive characteristics provided by keyword analysis can be divided into three types: open-class lexical words or aboutness keywords indicating what topic the corpus is mainly about, grammatical words indicating the style of the texts, and proper nouns [10]. In

this study, only the former types of keywords will be examined. The AntConc program was used to identify key clusters of keywords.

The corpus of SDGs was analysed through keyword analysis [11]. The keyword list of each of the SDGs categories was identified and then ranked by log-likelihood statistic [12], [13] as identified by Key-BNC. Only top 100 keywords of each category were focused for further analysis. The next step is to identify keywords functioning as nouns under the assumption that nouns can reflect topics. CLAWS was used to identify part of speech of keywords of each SDG. Only nouns keywords that were found in each specific goal were focused to identify keyword clusters. Two-word Clusters of nouns as keywords of each SDG and their concordance were identified by AntConc program, Cluster toolbar.

RESULTS

Table 1 shows keywords (noun part of speech) that were found in all SDGs. These seven words reflect the main goal of SDG concept that is to promote global sustainable development in many countries.

Table 1 Keywords found in every SDG goal

No.	Keywords	Frequency	No.	Keywords	Frequency
1	countries	2262	5	sustainable	2257
2	development	2133	6	targets	596
3	global	1391	7	UN	1027
4	SDG	2104			

Table 2 shows that most of the keywords with high frequency of LL were directly related to the title of each SDG goal such as poverty (SDG1), hunger (SDG2), health (SDG3), and education (SDG4). Some of them were indirectly but relatively related to the title of each goals such as living (SDG1), agriculture (SDG2), births (SDG3) and access (SDG4). Within the keywords with high frequency of LL identified, keywords of each specific goal can be identified. These groups of keywords are likely to reflect a variety of more specific topics of each goal. However, a cluster of each of these keywords can be reflect a wide range of more varied but specific topics of each goal. Interestingly, many acronyms can be found.

Table 2 Keywords of each specific goal

SDGs Goals	Top 5 keywords with high LL	Keywords of each specific goal
1. No poverty	Poverty(273), Disasters(s)(91), Extreme(89), Living(70), Countries(63)	alleviation(14), catastrophes(8), coronavirus(4), dimensions(21), eradication(16), exposure(13), line(37), protection(44), situations(13), vulnerability(13)
2. Zero hungry	Food(341), Hunger(108), Agricultural(134),	farmers(21), prevalence(43), price(39), prices(39), producers(40), security(35),

SDGs Goals	Top 5 keywords with high LL	Keywords of each specific goal
	Agriculture(91), Sustainable(75)	seeds(15), stunting(33), undernourishment(28), volatility(20)
3. Good health and well-being	Health(365), Mortality(188), Deaths(163), Diseases(100), Births(76),	malaria(64), medicines(49), mortality(188), substance(33), suicide(23), tb(34), tobacco(30), tuberculosis(32), uhc(14), vaccine(56)
4. Quality education	Education(377), School(137), Primary(123), Access(122), Children(120)	adults(31), boys(31), childhood(23), citizenship(14), closures(23), enrollment(10), enrolment(20), ict(12), learning(101), literacy(69), numeracy(31), outcomes(25), proficiency(21), scholarships(22), school(137), schooling(15), schools(64), skills(56), students(32), teachers(45)
5. Gender equality	Women(382), Gender(139), Violence(104), Sexual(93), Rights(93)	contraceptive(19), decision(31), sex(22), mutilation(34), empowerment(42), leadership(26), marriage(29), ownership(36), participation(25), partner(8), positions(27),
6. Clean water and sanitation	Water(770), Drinking(110), Access(82), Hygiene(78), Freshwater(67)	hygiene(78), lack(26), lakes(17), monitoring(19), rivers(18), scarcity(20), supplies(20), water(770), wetlands(21), toxicology(8), withdrawals(15)
7. Affordable and clean energy	Energy(3195), Access(839), Renewable(593), Electricity(400), Clean(355),	efforts(106), grid(190), intensity(88), investments(126), region(187), sids(114), solutions(174), stakeholders(51), tracking(67)
8. Decent work and economic growth	Employment(192), Economic(178), Growth(160), Labour(141), Energy(108)	labour(141), creation(24), decouple(10), economy(32), entrepreneurship(13), ilo(25), job(43), jobs(91), slavery(14), unemployment(28)
9. Industry, innovation and infrastructure	Development(161), Infrastructure(134), Manufacturing(91), Developing(77), Innovation(74)	capabilities(15), chains(19), diversification(15), mobile(30), industrialization(28), industries(49), value(56), mobile(30), industry(64), infrastructures(13), network(23), credit(21), passenger(18), retrofit(8), tech(56), upgrade(28)

SDGs Goals	Top 5 keywords with high LL	Keywords of each specific goal
10. Reduced inequalities	Energy(195), Income(148), Inequality(121), Policies(75), Poverty(67)	costs(38), differential(13), energia(8), ethnicity(12), inclusion(18), inequalities(46), inequality(121), migrant(21), migration(54), mobility(19), tariff(19), tier(12)
11. Sustainable cities and communities	Urban(178), Cities(151), Disasters(94), Waste(72), Pollution(61),	air(91), cities(151), heritage(32), transport(57), slum(14), particulate(18), spaces(43), settlements(28), urbanization(35), transportation(22)
12. Responsible consumption and production	Consumption(157), Waste(150), Food(130), Production(121), Material(81)	accordance(24), cycle(14), generation(24), globalization(12), lifestyles(16), metadata(11), patterns(35), procurement(31), recycling(30), tons(21), tool(18), unep(13), wastes(14)
13. Climate action	Climate(422), Change(209), Global(108), Emissions(91), Nations(85),	adaptation(36), dioxide(17), droughts(11), gas(38), greenhouse(47), hazard(16), mechanism(16), mitigation(28), paris(24), raising(17), rise(24), september(34), temperature(38), ipc(30)
14. Life below water	Marine(294), Ocean(144), Fisheries(90), Oceans(80), Coastal(80)	acidification(42), aquaculture(16), coral(24), debris(17), eutrophication(12), fish(90), fisheries(90), fishers(31), fishing(68), law(49), livelihoods(14), mpas(9), nutrient(12), seas(15), oceans(80), plastic(58), plastics(17), seabed(13), stock(59)
15. Life on Land	Biodiversity(192), Forest(s)(187), Species(123), Land(145), Ecosystems(92)	afforestation(10), aichi(14), alien(23), area(59), benefits(30), deforestation(25), degradation(41), desertification(41), drylands(12), extinction(24), forest(100), habitat(19), hectare(21), mountain(52), oceania(8), reforestation(13), restoration(18), rli(7), species(123), use(56), utilization(12), wildlife(47)
16. Peace, justice and strong institutions	Violence(159), Energy(131), Institutions(96), UN(74), Conflict(64)	bribery(36), conflict(64), crime(49), corruption(40), victims(22) economies(23), freedoms(17), governance(29), homicides(19), homicides(19), justice(65), peace(58), societies(27), terrorism(25),

SDGs Goals	Top 5 keywords with high LL	Keywords of each specific goal
17. Partnerships for the goals	Development(239), Global(89), Debt(85), Enhance(81), Partnerships(67)	building(39), diffusion(14), funding(21), oda(60), oecd(13), official(51), partnership(23), statistics(27) partnerships(67), revenue(28), subscriptions(16),

The two-words clusters of keywords of each goal were identified by AntConc program. The clusters are the combination between keywords and co-occurrence words (Nouns). The examples are shown in Table 3.

Table 3
Examples of Clusters of Keywords

SDGs Goals (No. of clusters)	Co-occurrence	Keywords	Co-occurrence
1. No poverty		protection	transfers, measures, access
2. Zero hungry	energy, fertilizer, market, oil, fuel	price	
3. Good health and well-being	tb	deaths	
		malaria	incidence
4. Quality education		school	closures, life, infrastructure electrification
	proficiency	level	
		literacy	rates
		education	scholarships
5. Gender equality		sex	disability, goal
	land, property	ownership	
6. Clean water and sanitation	desalination, toxicology	water	supplies, sanitation, life, pumping, sector, advisories
7. Affordable and clean energy		grid	solutions, infrastructure systems, electrification connection, interconnection
		energy	intensity, solutions, intensity investments
	ECE	region	

SDGs Goals (No. of clusters)	Co-occurrence	Keywords	Co-occurrence
8. Decent work and economic growth	child	labour	productivity, income, unit
		unemployment	benefits
	fuel	economy	
9. Industry, innovation and infrastructure	tech	labs	
		network	coverage
10. Reduced inequalities	energy	costs	
	race	ethnicity	origin
	tariff	lines	
11. Sustainable cities and communities		settlements	programme
		slum	dwellers
12. Responsible consumption and production		procurement	practices, policies
		recycling	rates
13. Climate action		mitigation	adaptation, measures, actions
		temperature	rise
14. Life below water		fisheries	aquaculture
		seabed	minerals, mining
		ocean	acidification
		scale	fishers
15. Life on Land		wildlife	trade
		species	extinction
	land, combat	desertification	
16. Peace, justice and strong institutions		producer	economies
		bribery	incidence
17. Partnerships for the goals		building	capacity, sector, mechanism
		funding	transparency
		revenue	service, collection
		broadband	subscriptions

IMPLICATION AND CONCLUSION

This study only included texts from 2015 to 2022, the period during which the SDGs were established by the UN. This leads to a limitation in understanding an overview of sustainability concepts earlier than 2015, before the SDGs were established. However, it seems that keyword analysis can be efficiently used to reflect a wide range of topics. For pedagogical implication, the results of clusters of keywords are likely to be beneficial to EFL teachers, especially to the ones who felt reluctant to incorporate sustainable development in their teaching due to a lack

of scope of SDGs topics for selecting teaching materials [14]. Research on social awareness related to SDGs such as life skill basic necessity of rural learners [15] should be called for.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Suan Sunandha Rajabhat University for invaluable help throughout this research.

REFERENCES

[1]   J.S. Damtoft, J. Lukasik, D. Herfort, D. Sorrentino & E.M. Gartner (2008) -Sustainable Development and Climate Change Initiatives, *Cement and Concrete Research*, Vol. 38, Pp. 115–127.

[2]   C. Wang, Y.K. Cho & C. Kim (2015) -Automatic BIM Component Extraction from Point Clouds of Existing Buildings for Sustainability Applications, *Automation in Construction*, Vol. 56, Pp. 1–13, 2015.

[3]   K. Glyptou, A. Amore & B. Ashton Adie (2022) -In Planning and Managing Sustainability in Tourism, From Aspirations to Applications: The SDGs and the Role of Indicators in the Measurement of Sustainable Tourism, *Springer*, Pp. 13-25.

[4]   United Nations Educational, Scientific and Cultural Organization (UNESCO) (2023), ESD—Building a Better, Fairer World for the 21st Century, URL: http://u4614432.fsdata.se/wp-content/uploads/2013/09/esd.pdf.

[5]   Yastrebova, Elena (2021) -Designing an EFL University Course Around Sustainable Development Goals

[6]   J.E. Mambu (2022) -Embedding Sustainable Development Goals into Critical English Language Teaching and Learning, *Critical Inquiry in Language Studies*

[7]   H.M., Alshuwaikait and I. Abubakar (2008) -An Integrated Approach to Achieving Campus Sustainability: Assessment of the Current Campus Environmental Management Practices, *J Clean Prod*, Vol. 16, Pp.1777–1785.

[8]   M. Scott (1997) -PC Analysis of Key words—and Key Key Words, *System*, Vol. 25(2), Pp. 233-245.

[9]   P. Pojanapunya, and A. Lieungnapar (2017) -The Quality of Choices Determines the Quantity of Key Words, *Proceedings of the International Conference: DRAL 3/19th ESEA*.

[10]   M. Scott, and C. Tribble (2006) -Textual Patterns: Key Words and Corpus Analysis in Language Education. *Amsterdam, John Benjamins*.

[11]   M. Bondi, and M. Scott, M. (Eds.) (2010) -Keyness in Texts, Vol. 41, *John Benjamins Publishing*.

[12]   P. Rayson (2008) -Log-likelihood and Effect Size Calculator.

[13]   P. Rayson (2013) -Corpus Analysis of Key Words, In Carol A. Chapelle (ed.), *The Encyclopaedia of Applied Linguistics*, Pp. 1–7. Oxford: Wiley-Blackwell.

[14] K. Ching Ting Tany (2021) -I Want to Teach Sustainable Development in My English Classroom: A Case Study of Incorporating Sustainable Development Goals in English Teaching, *Sustainability*, Vol. 13(8)8, Pp. 4195.

[15] Achara Wongsothorn, & Suwaree Yordchim (2021) -Reading Enhancement Activities of Students in the Rural Areas in Thailand, *THAITESOL Journal*, Vol. 33(1), Pp.1-19.