REVISTA DE ESTUDOS DA
**LINGUAGEM**

# Unveiling the Academic Vocabulary of Sustainable Development Goal-Research Papers: A Comprehensive Lexical Multidimensional Analysis

*Desvendando o vocabulário acadêmico de artigos científicos sobre os objetivos de desenvolvimento sustentável: uma análise multidimensional lexical detalhada*

**Paula Tavares Pinto**
Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP) | São José do Rio Preto | SP | BR
paula.pinto@unesp.br
https://orcid.org/0000-0001-9783-2724

**Tony Berber Sardinha**
Pontifícia Universidade Católica de São Paulo (PUC/SP) | São Paulo | SP | BR
tonycorpuslg@gmail.com
https://orcid.org/0000-0001-8815-1521

**Denis Luiz Marcello Owa**
Pontifícia Universidade Católica de São Paulo (PUC/SP) | São Paulo | SP | BR
deniso118@yandex.com
https://orcid.org/0000-0001-9802-3929

**Maria Claudia Nunes Delfino**
Faculdade de Tecnologia de São Paulo (FATEC) Praia Grande | SP | BR
claudia@fatecpg.com.br
https://orcid.org/0000-0003-1976-8725

**Simone Vieira Resende**
Faculdade Cultura Inglesa | São Paulo SP | BR
simonevieiraresende@gmail.com
https://orcid.org/0000-0002-1439-3840

**Abstract:** Since the United Nations Agenda 2030 was set up, many countries have worked to find solutions to interconnected global issues such as hunger, poverty, education and sustainability just to mention a few. After the striking of COVID-19 pandemic, the world witnessed researchers gathering around in an international effort to find solutions to the problems previously pointed out. In order to describe the lexis used in research papers discussing the Sustainable Development Goals (SDGs), we compiled a 2 million-word corpus from the PLOs platform with the AntCorGen program using the query term "Sustainable Development Goal". After that, we part of speech-tagged the corpus using the Tree-tagger of Sketch Engine in order to carry out a lexical multidimensional analysis (LMDA). Our aim was to (i) identify the major dimensions of variation based on the lexico-grammatical characteristics in research papers published in English by international researchers and (ii) observe how the SDG themes stand out in the study corpus. Results showed six dimensions that were named according to the words concentrated in each one: Government Actions, Presenting Results, Data Interpretation, Data Presentation, Data Quality, Research Procedure. The first dimension is the one that best illustrates a theme related to the SDGs whereas the other five dimensions clearly show the lexis that illustrates how researchers describe their own articles.

**Resumo:** Desde a criação da Agenda 2030 das Nações Unidas, muitos países têm trabalhado para encontrar soluções para questões globais interligadas, como a fome, a pobreza, a educação e a sustentabilidade. Após o início da pandemia da COVID-19, o mundo assistiu à dedicação de pesquisadores num esforço internacional para encontrar soluções para os problemas anteriormente apontados. Com o intuito de descrever o léxico utilizado em artigos científicos que discutem os Objetivos de Desenvolvimento Sustentável (ODS), compilamos um *corpus* de 2 milhões de palavras com artigos da plataforma PLOs com o programa AntCorGen usando o termo de busca "Objetivo de Desenvolvimento Sustentável". Em seguida, etiquetamos o *corpus* utilizando o Tree-tagger, do Sketch Engine, para realizar uma análise multidimensional lexical (AMDL). Nosso objetivo foi: (i) identificar as principais dimensões de variação com base nas características léxico-gramaticais em artigos científicos publicados em inglês por pesquisadores internacionais e (ii) observar como os temas dos ODS se destacam no *corpus* do estudo. Os resultados mostraram seis dimensões que foram nomeadas de acordo com as palavras concentradas em cada uma: Ações Governamentais, Apresentação dos Resultados, Interpretação dos Dados, Apresentação dos Dados, Qualidade dos Dados, Procedimento de Pesquisa. A primeira dimensão é a que melhor ilustra um tema relacionado aos ODS, enquanto as outras cinco dimensões mostram claramente o léxico que ilustra como os pesquisadores descrevem os seus próprios artigos.

**Palavras-chave:** análise multidimensional lexical; artigos científicos; inglês com fins acadêmicos; objetivos de desenvolvimento sustentáveis.

# 1 Introduction

It is well acknowledged that English has become the international lingua franca for scientific research (Jenkins, 2014). Scholars who wish to publish their studies in high impact

journals generally are required to write up their research papers in English (Swales, 1990; Coxhead, 2000; Hyland & Hyland, 2006; Gardner & Davies, 2014; Hyland & Shaw, 2016). In order to improve their impact in international rankings, Higher Education institutions have traditionally implemented English for Academic Purposes (EAP) classes of academic writing in different research areas. According to Flowerdew and Peacock (2001, p.8) "English for Academic Purposes is the teaching of English with the specific aim of helping learners to study, conduct research or teach in that language".

Since 2015, the UN Sustainable Development Goals (SDGs) have become central guidelines discussed by researchers of different countries to find solutions for major social issues in local and global levels. Therefore, we have witnessed several scientific publications discussing the COVID-19 pandemic period, for example. The SDGs have been discussed in interconnected areas such as Sociology, Economy, Chemistry, Math, Linguistics among others, and influenced one another. In the aforementioned areas, we have gradually seen the SDGs being used as guidelines by international researchers and funding agencies since they deal with the major problems faced everywhere around the globe, which were pointed out by the UN organization. International journals and agencies, such as Springer Nature, have debated the importance of publishing studies that, somehow, discuss major social problems and for that reason they have published special issues of papers focusing different Sustainable Development Goals.[1] In this context, journal editors and funding agencies are the stakeholders who will select the research studies to be published or financed, therefore knowing the specific lexicon used in a specific area will be crucial to communicate the author's results, without mentioning the importance of the Sustainable Development Goals for the scientific community. In this study, we will discuss the academic lexicon used by international researchers who write about the UN Sustainable Development Goals.

The United Nations (UN) is an international organization whose aim is to maintain international peace and security, develop peaceful relations among nations and achieve international cooperation (Sachs, 2012). It was founded in 1945[2] and, since then, its membership has grown significantly reaching 193 Member States. Its operations have also expanded to other areas such as promoting Economy and Health campaigns. After several conferences as the Agenda 21 in Rio (1992); the South Africa World Summit on Sustainable Development (2002) and the Paris Agreement on Climate Change (2015), a new document was elaborated whose aim was to "address the global challenges human beings face, including those related to poverty, inequality, climate change, environmental degradation, peace and justice".[3] This document tackles seventeen interconnected Sustainable Development Goals (SDGs) that should be achieved by 2030 in order to have a better and more sustainable future for the whole planet. The seventeen goals are: (1) No Poverty, (2) Zero Hunger, (3) Good Health and Well-being, (4) Quality Education, (5) Gender Equality, (6) Clean Water and Sanitation, (7) Affordable and Clean Energy, (8) Decent Work and Economic Growth, (9) Industry, Innovation and Infrastructure, (10) Reduced Inequality, (11) Sustainable Cities and Communities, (12) Responsible Consumption and Production, (13) Climate Action, (14) Life Below Water, (15) Life On Land, (16) Peace, Justice, and Strong Institutions, (17) Partnerships for the Goals.

---

[1] Access at <https://www.springernature.com/gp/researchers/sdg-impact> July 15th, 2023.

[2] Access at <https://www.un.org/en/about-us> July 15th, 2023.

[3] Access at <https://www.un.org/en/academic-impact/page/sustainable-development-goals> July 15th, 2023.

Brazil, as one of the UN members, is also involved in dealing with the SDGs issues regarding government actions, businesses and the academia. A great number of research has been conducted by specialists in private and public sectors where we see some of the major funding agencies launching new calls based on the SDGs. Considering the language used in these publications, although Brazilian researchers have been dealing with the SDGs in their work, many of them are still influenced by their native language (Pinto et al., 2021b; Pinto et al. 2021c). At the same time, it is not easy to map the lexicon used in studies that are directly related to the SDGs, that could be presented to Brazilian scholars in order to help them write their own research papers. The relevance of this material goes along with what world specialists, such as the ones from Elsevier´s World Research, have pointed out about Brazilian studies. To them, although Brazil already has a sizeable research output:

> [...] the country has room to improve, as its performance along measures of research impact (citations, highly cited articles) does not keep pace" (Elsevier's World Research, 2016).

Regarding the first issue, i.e., Brazilian researchers writing in English without so much influence of their native language, several studies about Academic Writing and Corpora have been carried out (Charles, 2012; Karpenko-Seccombe, 2020; Nesselhauf, 2005; Pinto et al., 2021a; Pinto et al., 2021b). Several authors have discussed the importance of internationalization of Higher Education institutions in Brazil (Abreu-e Lima *et al.* 2021; Baumvol & Sarmento, 2016; Dutra *et al.* 2019; Finardi & Guimarães, 2021; Gimenez et al. 2018; Sarmento *et al.* 2016.) and some actions taken in a university in the state of São Paulo to support academic writing among graduate students, professors and researchers are discussed in Pinto et al. (2021a) and Pinto et al. (2021b).

In order to deal with the second issue regarding the mapping of SDGs, we believe the methodology of Corpus Linguistics can be very useful in pointing out the SDGs themes which are present in a corpus of research papers. There have been several studies using Corpus Linguistics as its main methodology (Berber Sardinha, 2003, 2004, 2019; Péres-Paredes, 2020; Sinclair, 2001; Viana & Tagnin, 2011), however, there are still few studies that use Corpus Linguistics to analyze the SDGs, some of them are the studies carried out by Cordeiro (2019), Jodoin & Singer (2020), Törö (2020), Valvason, 2021 and Pinto (2021a).

Cordeiro (2019) uses a corpus-driven approach to carry out a systematic review of literature to analyze the gaps between the academic subject discourses and business sector practices in fisheries, aquaculture and seafood. The study indicates a conceptual gap between knowledge and business practices in how market access is conceived, studied and managed in the pharmaceutical and healthcare marketing industry in comparison to fisheries and aquaculture. This systematic review of literature is directly aimed at in the 2030 Agenda for Sustainable Development and Zero Hunger. Cordeiro (2019) suggests a conceptual model in understanding the challenges, potential barriers and solutions to developing market access for fisheries and aquaculture international business.

Jodoin & Singer (2020) analyze a corpus of texts with themes about the environment taken from English as a Foreign Language textbooks used in Japan. By employing codification and corpus analysis techniques the study suggests that the image-text interplay found in the corpus has little effect on the attitudes, values, and practices considered important to an effective Education for Sustainable Development. Besides that, there

seems to be little thematic interconnectedness among topics, images, and texts. The results also suggest that textbook creators are paying little regard to the inherent aims of including environmental content, both in terms of promoting Sustainable Development Goals and in promoting sustainability competencies.

Törö (2020) used tools from Corpus Linguistics to conduct a critical discourse analysis of the Finnish development policy (2016) that is based on the United Nations 2030 Agenda and its 17 Sustainable Development Goals. The study focuses on the definitions of development and its implementation as put by the policy and how its text relates to its production and consumption as well as to the social and political context in which it is situated. Results show that the policy reinforces the image of top-down aid dictated by institutions of the Global North in which Finland seems to be a giver and savior with countries in the Global South as vulnerable receivers of aid, granting little agency for the institutions, let alone the people from Global South countries.

Valvason (2021) carried out a corpus-based study of the SDG keywords in three languages (British, Hungarian and Italian). Results showed that the British, Hungarian and Italian cultures, as they are published in the 2030 Agenda and in the press, are based on concepts such as "development", "technology", "economy", "climate", and in values like "inclusivity" and "sustainability". Nevertheless, frequency lists revealed a different perspective. They do not show words related to economy, animals or plants as salient, so such words are not in use.

In a recent pilot study, Pinto (2021a) analyzed a corpus of research papers written in English by Brazilian researchers with the aim of detecting whether the SDG themes had been discussed in their research papers. Results showed that the keywords "sustainability" and "poverty" were present in the Brazilian research papers; however, the first one was more evident and had a more positive semantic meaning than the second one, where authors emphasized the importance of fighting against low conditions for human beings.

In this study, we used multidimensional analysis (MDA) to analyze the discourse which is present in the texts of the corpus. We aimed at identifying the major dimensions of variation based on the  lexical characteristics in the research papers published by international  researchers through the underlying themes present in a corpus of scientific texts. After having presented the main characteristics of the Lexical MDA, we present the aims of this paper which are two-fold: 1. Finding which lexical dimensions are present in the SDGPlos corpus[4]; and 2. Analyzing how the SDGs are present in research papers written by international research teams. The results of this research can be used in the future to help Brazilian researchers to write their papers according to the international scientific community.

## 2 Lexical Multidimensional Analysis

Multidimensional Analysis (MDA) is a multivariate statistical analysis used to identify correlations among linguistic features across the texts in a corpus which was developed by Biber (1988, 2001) to compare written and spoken registers in English and it has since been applied extensively in synchronic and diachronic analysis of registers in English as well as Western and non-Western languages (Asención-Delaney, 2011; Biber et al., 1992; Berber Sardinha et

---

4  From now on this is how we will call this corpus.

al., 2014). This methodology takes into consideration a collection of texts that will be used for a quantitative study of Linguistics by applying statistical procedures, which identify co-occurrent patterns of language, which are grouped into factors and then linguistically interpreted, to receive the name of dimensions (Delfino, 2021). Biber's MDA model has been dedicated to analyzing lexical-grammatical categories to reveal the function of the language, that is why his approach to MDA has been known as functional MDA.

In Brazil, Tony Berber Sardinha, who also developed MDA studies in other languages such as Portuguese, German and Spanish, has most recently elaborated a Lexical Multidimensional Analysis (LMDA) in which only content words such as nouns, adjectives, verbs and adverbs are used as variables in the analysis. In this study we use this same approach of the entirely lexical variables to perform the LMDA.

In our study, a corpus of 400 research papers in English with 2,044,674 million words was collected from the open source platform PLOS ONE[5]. The corpus was normalized for 1,000 words to avoid that the size of the text would interfere in the analysis. After that, it was part-of-speech tagged for hundreds of linguistic features. The dimensions were identified using factor analysis, which is "a statistical technique to identify factors and sets of correlated linguistic features that correspond to latent (unobserved) variables in the corpus" (Berber Sardinha, 2022, p. 664). With the purpose of mapping the SDGs and observing how international authors use academic language to discuss these goals, we have adopted a Lexical Multidimensional Analysis (Berber Sardinha, 2022; Berber Sardinha; Pinto, 2017; Delfino *et al.*, 2018). Summarizing, the main focuses of Lexical MDA are shown in Table 1:

**Table 1** - Main characteristics of the Lexical MD Analysis

| Goal | Underlying parameters of lexical variation |
|---|---|
| Method | Multivariate statistics (Factor Analysis) |
| Linguistic features | Lexical |
| Primary interpretation | Discourse-based |

Source: Adapted from Berber Sardinha (2017)

It is important to point out that all research papers discuss the SDGs, therefore, if the vocabulary related to them is central in those studies, they will emerge as dimensions in the Lexical MDA and will represent the discourses that underlie such texts. On the other hand, if the vocabulary related to the SDGs is peripheral, it will not be salient on the dimensions, and, in that case, we are interested to know which discourses such vocabulary represent.

Our research questions are:
1.      To what extent do the SDGs come through as dimensions?
2.      Which themes are mostly being discussed in each lexical dimension?
3.      Do contexts related to the SDGs show positive or negative discussions?

The following sections are organized to present the following topics: Methodology; Thematic Dimensions; Discussion & Results and Final Considerations.

---

[5]  Available at <https://journalsearches.com/journal.php?title=P> November, 2023.

# 3 Methodology

The first step was the compilation of a corpus using AntCorGen (Anthony, 2019), which is a tool that quickly compiles specialized corpora with research papers from PLOS, which is a nonprofit, open access multi-disciplinary publisher. For this study, we included the term "Sustainable Development Goal" as a query word in AntCorGen and set the tool to extract 400 articles from Plos. Since we wanted to have mostly written material we selected the articles' abstracts, introduction, materials & methods, results & discussion and conclusions, as we can see in the figure below:

Figure 1 - Screenshot of AntCorGen
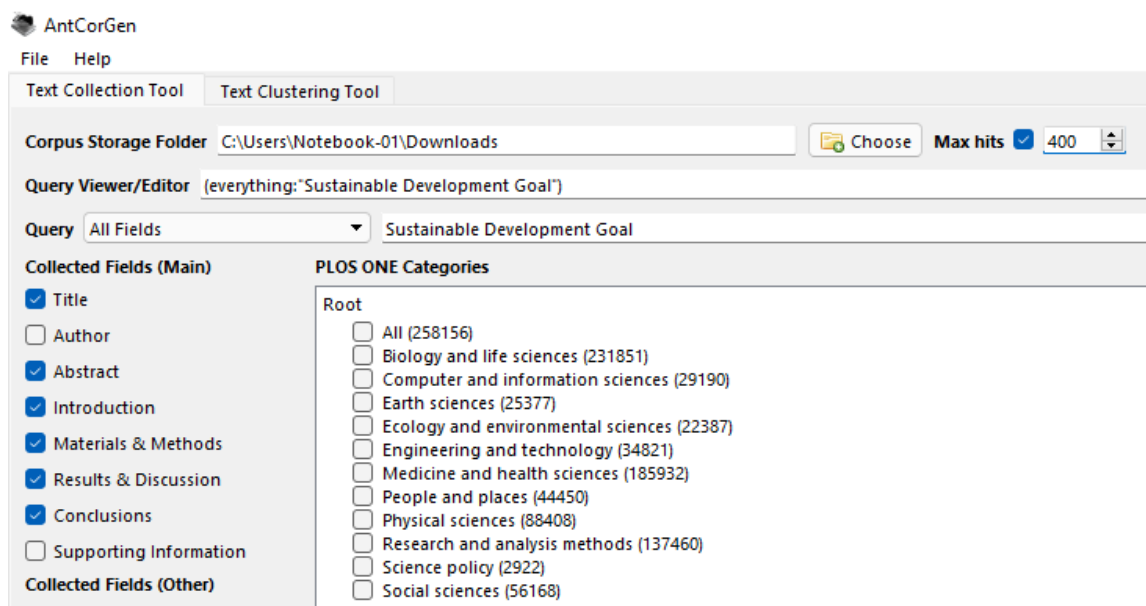


Source: the authors.

As a result, we compiled a corpus of 2,044,674 words, 2,640,307 tokens which was uploaded to the Sketch Engine tool (Kilgariff, 2014), a platform with a set of tools for text analysis and text mining applications. Since the intention was analyzing only content words, the Tree Tagger was used in order to tag and lemmatize the words, so that words such as 'boy' and 'boys' are counted as the same one and not two different ones. The TreeTagger is a tool for annotating text with part-of-speech and lemma information and has been successfully used to tag English, among other languages. It tags each word of the text with the corresponding part of speech, as we can see in the tagged sentence "Implementation of ecosystem-based fisheries management requires a clear conceptual and quantitative framework for assessing how different harvest options can modify benefits to ecosystem and human beneficiaries", on Table 2.

**Table 2 -** Tagged sentence by Sketch Engine from SDG Plos corpus

| Implementation | NN | implementation-n | for | IN | for-i |
|---|---|---|---|---|---|
| of | IN | of-i | assessing | VVG | assess-v |
| ecosystem-based | JJ | ecosystem-based-j | how | WRB | how-x |
| fisheries | NNS | fishery-n | different | JJ | different-j |
| management | NN | management-n | harvest | NN | harvest-n |
| ( | ( | (-x | options | NNS | option-n |
| <g/> | | | can | MD | can-x |
| EBFM | NP | EBFM-n | modify | VV | modify-v |
| <g/> | | | benefits | NNS | benefit-n |
| ) | ) | )-x | to | IN | to-i |
| requires | VVZ | require-v | ecosystem | NN | ecosystem-n |
| a | DT | a-x | and | CC | and-c |
| clear | JJ | clear-j | human | JJ | human-j |
| conceptual | JJ | conceptual-j | beneficiaries | NNS | beneficiary-n |
| and | CC | and-c | <g/> | | |
| quantitative | JJ | quantitative-j | . | SENT | .-x |
| framework | NN | framework-n | </s> | | |
| | | | <s> | | |

Source:the authors.

# 4 Thematic Dimensions

After the variables were tagged in the sketch engine, they were counted by a script developed for this research originating a spreadsheet which was processed by SAS program[6] in order to run two Factor Analyses. Factor analysis is a statistical method used to identify underlying relationships among a large set of variables. It reduces the dimensionality of data by grouping variables that are highly correlated into factors, which represent the shared variance among the variables (Biber, 1988). In MDA studies we perform two factor analysis with different but complementary objectives. The first one (unrotated) is necessary to determine the number of factors to retain and it generates a scree plot graph (below), with the number of eigen-values which leads us to the number of factors to be used for the second factor analysis, the rotated one, that minimizes the number of variables that have high loadings on each factor. This method simplifies the interpretation of the factors.

---

[6]   SAS platform at <https://portalsas.com.br/login> February, 2022.

**Image 2** - Scree Plot
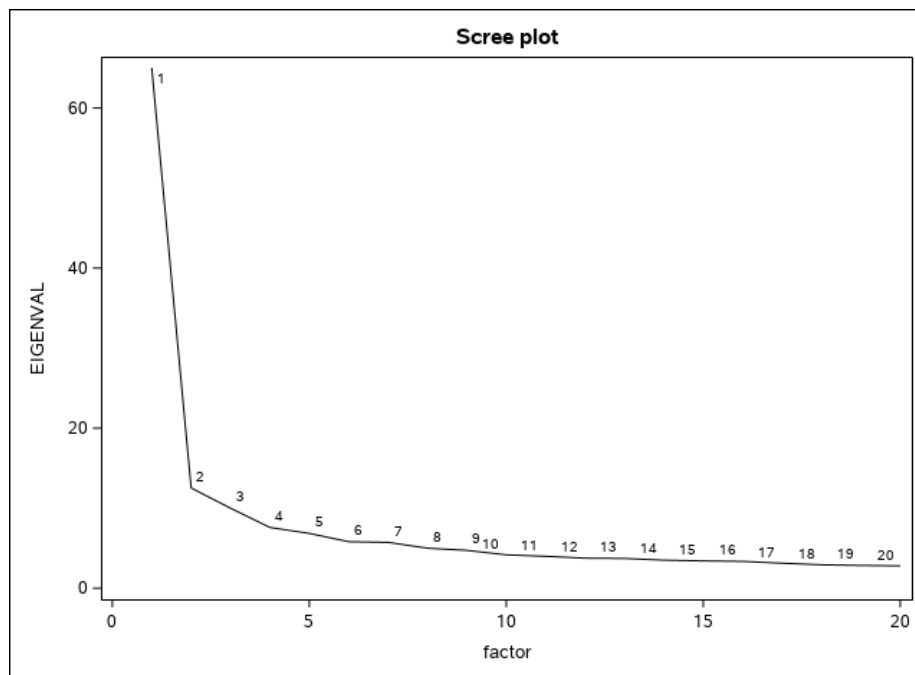


Scree plot

The graph we see in Figure 2 is called Scree Plot, where the curve shows that there are more variations explained  from factor 1 to 2. After that, we have the following factors from 2 to 3, from 3 to 4, from 4 to 5 and from 5 to 6. Following those, we almost have a flat line where there is almost no variation, therefore, we only considered there were six factors generated from SDGplos corpus.

We tested the interpretation with 3, 4 and 5 factors, but the best fit for interpretability was with 6 factors. Therefore, the six factors obtained on the first factor analysis were used as the input for the second one, which is called rotated through the Promax method in the SAS statistical program. The words (variables) were distributed in the six factors according to their co-occurrence, i.e., the words that occur together in the texts were assembled in different factors. For example, the words 'address, focus, many, provide, important' were grouped together in factor 1 along 86 that also occurred together in the texts of the corpus. This happened in all the other five factors.

In the following section, we show the words that are loaded in each factor, and, based on them, we label each dimension according to these words and how they appear in the text of the corpus. It is important to mention that words with weight smaller than .30 were not computed in the factors. Overall, results from the factor analysis indicated that the lexis used by the authors in the corpus is strictly related to the language used in academic research papers. The first factor, however, differed from the others, as we can see in the Appendix.

As we will discuss in the next section, Dimension 1 showed a vocabulary related to decisions taken by stakeholders and governments, therefore, it was labeled *Government Actions*. The other five dimensions clearly showed the lexis related to academic writing and how writers express the results of their studies, consequently, Dimensions 2, 3, 4, 5 and 6 were called

*Presenting Results, Data Interpretation, Data Presentation, Data Quality* and *Research Procedure*, respectively. In the next subsection we will discuss each dimension separately.

## 4.1 Dimension 1: Government Actions

Dimension 1, which concentrates most of the words, a total of 86, indicates that the lexis that stands out is the one related to government actions. The content words in Dimension 1 illustrate very clearly the idea of actions taken by governments and stakeholders. As an example, some of these content words subdivided into categories are the **verbs** "address" (.58) and "focus" (.56); the **nouns** "government" (.55) and "stakeholder" (.49); the **adjectives** "important" (.55) and "effective" (.48) and the **adverbs** "particularly" (0.40) and "especially" (.36). The theme that emerged related to stakeholders' decisions showed a very important outcome which is that actions are really being taken by governments and stakeholders in different countries, which is what all members of the United Nations have been looking forward to seeing regarding the SDGs.

After reading the texts that loaded on Dimension 1, we selected the excerpt below to illustrate it as authors discuss how the government should deal with farmers that misreport information about their production. In the article *Strategies for poverty alleviation supply chain with government subsidies and misreporting behavior in China*, whose excerpt is shown in example 1, Ye & Deng (2021) highlight the importance of farmers informing the government about their production and that, by failing to do so, they will impair the projects that support low-income people to have access to those products. The article also points out that the government should "establish punitive measures to curtail misreporting". In this text we see content words from Dimension 1 such as "ensure", "policy" and "effective". The words in capital were loaded in Dimension 1. After all words from a corpus are loaded and distributed in different factors they are separated in different dimensions. Each dimension will be named according to the themes they relate to. So, from now on, the terms factor and dimension are used interchangeably.

---

**Example 1**

Our findings demonstrate that both the enterprise and the farmer can obtain maximum profit from the misreporting behavior. Unfortunately, this behavior always damages the profit of OTHER participants and weakens the efficiency of subsidy POLICY. Moreover, to mitigate the negative implication of misreporting behavior, the GOVERNMENT can ESTABLISH punitive measures to curtail misreporting. Our WORK PROVIDEs IMPORTANT POLICY implications for GOVERNMENTs and enterprises. To ENSURE that more consumers have access to poverty alleviation products, GOVERNMENT organizations should prioritize such projects. In addition, the provision of PUBLIC facilities and technical guidance should be more EFFECTIVE and prompt to share enterprises' and farmers' costs. We further recommend that subsidy policies be formulated according to recipients' performance in poverty alleviation projects, with corresponding supervision and punitive measures.

---

This text is related to the SDGs 12 (Responsible Consumption and Production), 16 (Peace, Justice and Strong Institutions) and 17 (Partnerships for the Goals).

## 4.2 Dimension 2: Presenting Results

Dimension 2 was labeled Presenting Results since the lexis in this dimension seems to draw readers attention over the results of the studies being described. Some of the content words are **verbs** such as "show " (.62) and "indicate" (.48); **nouns** as "model" (.52) and "pattern" (.39); **adjectives** such as "possible" (.37)  and "large" (.36) and the **adverbs**  such as  "even" (.36) and "here" (.35).

In the article *Patterns of Vertebrate Diversity and Protection in Brazil*, in example 2, Jenkins et al. (2015) used the lexis to describe their results by using verbs such as "occur" and "indicate" and nouns such as "contrast", "patterns" and "results". By doing so, they show their results on patterns of biodiversity in Brazil contrasting distinct characteristics in different regions of the country. Again, the capitalized words had been loaded in Dimension 2.

> **Example 2**
> Data deficient mammals OCCUR across the country, concentrating in the Amazon and southeast Atlantic Forest, and there are no data deficient birds in Brazil. In CONTRAST, nearly a third of amphibians are data deficient, widespread across the country, but with a high concentration in the far southeast. Spatial biases in species locality data, however, possibly influence the perceived PATTERNs of biodiversity. Regions with low sampling density need more biological studies, as do the many data deficient species. All biomes except the Amazon have less than 3% of their area under full protection.  Reassuringly though, rates of protection do correlate with higher biodiversity, including higher levels of threatened and small-ranged species. Our RESULTs INDICATE a need for expanded formal protection in Brazil, especially in the Atlantic forest, and with an emphasis on fully protected areas. Previous observational work revealed that transient populations in a sustainable building disposed of waste more accurately when compared to patrons in a non-sustainable building.

The article discusses conservation decisions taken in Brazil and how they have a huge impact on the survival of global biodiversity. This article is related to the SDG 15 (Life on Land).

## 4.3 Dimension 3: Data Interpretation

In Dimension 3, the lexis mostly used points out to the interpretation of data. This result can be seen by some words that show higher scores such as the **verbs** "examine" (.77) and "investigate" (.68); the **nouns** "study" (.71) and "effectiveness" (.70); the **adjectives** " individual" (.56) and "common" (.55); and the **adverbs**  "furthermore" (.52) and "however"(.49). Although there are some words that are found in Dimension 2 as well, here they have more statistical relevance. This result may also be true to the fact that the vocabulary used to present results may also be used to interpret data, therefore, we can see a correlation between these two dimensions. Here the authors are describing what results have shown, for example, by comparing results based on the respondents' answers.

In the article *Happiness and Social Exclusion of Indigenous Peoples in Taiwan-A Social Sustainability Perspective* (Wang, 2015), in example 3, the content words that illustrate Dimension 3 in the text are "study", "regression", "examine" and "effect", which are used to describe a binary logistic regression used to examine peoples' answers. The capitalized words were identified in Dimension 3.

> **Example 3**
> This STUDY employed binary logistic REGRESSION to EXAMINE the EFFECTs of different domains of social exclusion on the likelihood of perceiving happiness; other exogenous factors were controlled. The results show that among the respondents, mountain indigenous peoples, females, the elderly and those who are healthier, wealthier, highly educated, possessing western beliefs, and are more likely to be happy, compared to their counterparts. As expected, the results reveal that the likelihood of being happy is higher for those who have received medical benefits, as well as those persons without housing problems or financial difficulties, compared to their excluded counterparts. However, no significant association is found between happiness and some social exclusion domains, such as child and youth benefits, and unemployment benefits. The disengagement of the indigenous peoples in mainstream society, with respect to the accessibility of welfare provisions, is a crucial element in regard to social exclusion and happiness. Several policy implications for the social sustainability of indigenous peoples can be inferred from these findings.

The excerpt shows a discussion over how socially excluded groups perceive happiness in society. This article is related to the SDGs 1 (No Poverty), 3 (Good Health and Well Being) and 10 (Reduced Inequalities).

## 4.4 Dimension 4: Data Presentation

Dimension 4 was labeled Data Presentation since we observe that the lexis here is being used to show how data was collected and presented. Some content words related to that are the **nouns** "datum" (.57) and "number" (.48); the **verbs** "estimate" (.51) and "calculate" (.46); the **adjectives** "high" (.50) and "available" (.49); the **adverbs** "less" (.39) and "respectively" (.38). The excerpt in example 4 was taken from the article *Understanding the Sustainability of Retail Food Recovery*, by Phillips et al. (2013) . Some of the content words in example 4 that clearly represent how researchers present their study data are "data", "rate", "number" and "increased". The capitalized words in this excerpt were part of Dimension 4.

> **Example 4**
> In this paper we study the simultaneous problems of food waste and hunger in the context of food (waste) rescue and redistribution as a means for mitigating hunger. To this end, we develop an empirical model that can be used in Monte Carlo simulations to study the dynamics of the underlying problem. Our model's parameters are derived from a DATA set provided by a large food bank and food rescue organization in north central Colorado. We find that food supply is a non-parametric heavy-tailed process that is well modeled with an extreme value peaks over threshold model. Although the underlying process is stochastic, the basic approach of food rescue and redistribution to meet hunger demand appears to be feasible. The ultimate sustainability of this model is intimately tied to the RATE at which food expires and hence the ability to preserve and quickly transport and redistribute food. The cost of the redistribution is related to the NUMBER and density of participating suppliers. The results show that costs can be reduced (and supply INCREASEd) simply by recruiting additional donors to participate. With sufficient funding and manpower, a significant amount of food can be rescued from the waste stream and used to feed the hungry.

The text will show how the authors present their data about problems of food waste and hunger in the context of food waste, rescue and redistribution as a means for mitigating hunger. This article is related to the SDGs 1 (No Poverty), 2 (Zero Hunger), 3 (Good Health and Well Being) and 10 (Reduced Inequalities).

## 4.5 Dimension 5: Data Quality

Dimension 5 is somehow related to Dimension 4, however, the difference is that in this case the lexis is not only being presented, but the authors are highlighting the quality of their data. Here we have **nouns**, such as "response" (.58) and "fact" (.50); **verbs** such as "reveal"(.57) and "observe" (.50); **adjectives**, as "significant" (.62) and "consistent" (.51) and **adverbs** such as "significantly" (.72) and "more" (.52). Excerpt 5 was taken from the article *Combining Aesthetic with Ecological Values for Landscape Sustainability* by Yang et al. (2014). Some of the content words in example 5 that show how the authors point out the quality of their data are "survey", "more", "significantly", "associated" and "significant" which are frequently used in the sections of research papers where results and data are being presented and emphasized. The words in capital were part of Dimension 5.

---

**Example 5**
In this study, a hybrid framework was proposed to evaluate ecological and aesthetic values of five landscape types in Houguanhu Region of central China. Data from the public aesthetic SURVEY and professional ecological assessment were converted into a two-dimensional coordinate system and distribution maps of landscape values. Results showed that natural landscapes (i.e. water body and forest) contributed positively MORE to both aesthetic and ecological values than semi-natural and human-dominated landscapes (i.e. farmland and non-ecological land). The distribution maps of landscape values indicated that the aesthetic, ecological and integrated landscape values were SIGNIFICANTLY ASSOCIATEd with landscape attributes and human activity intensity. To combine aesthetic preferences with ecological services, the methods (i.e. field SURVEY, landscape value coefficients, normalized method, a two-dimensional coordinate system, and landscape value distribution maps) were employed in landscape assessment. Our results could facilitate to identify the underlying structure-function-value chain, and also improve the understanding of multiple functions in landscape planning. The situation context could also be emphasized to bring ecological and aesthetic goals into better alignment. The growth of human populations and their resource needs have stressed the conservation of natural land resources. Many policies and programs have been implemented to address the pressures on land resources and notwithstanding this pressure, SIGNIFICANT acquisition of land for conservation has occurred throughout history in the U.S., and internationally.

---

Example 5 proposes a hybrid framework to evaluate ecological and aesthetic values which are beneficial to humans in various landscapes. The SDGs whose themes are related to this article are the SDGs 3 (Good Health and Well Being) , 11 (Sustainable Cities and Communities) and 15 (Life on Land).

## 4.6 Dimension 6: Research Procedure

Dimension 6 shows content words related to research methodology and procedures. In this dimension we only have two categories of content words which are nouns and verbs since this is the last dimension, where we usually find less variables. For **nouns** we see, for example, "information" (.43) and "procedure" (.39); and some of the **verbs** are "consent" (.50) and "approve" (.44).

The paper in example 6 discusses the mobilization and collaboration among diverse partners as vital components to reduce and eliminate cancer disparities in the United States. In order to do so, the authors observed the impact of intersectoral connections among the members of the Massachusetts Community Network for Cancer Education, Research, and

Training (MassCONECT). The  excerpt in example 6 was taken from the article *Addressing cancer disparities via community network mobilization and intersectoral partnerships: a social network analysis,* by Ramanadhan et. al. (2012). As we see, some of the content words that clearly describe research procedures in different studies are "consent", "procedure", and "conduct". In this case, researchers are pointing out that the ethics committee approved the interviews conducted in the study that is being presented, which is a common section we find in research papers. Once again, the words in capital letters were part of Dimension 6.

---

**Example 6**
The ethics committee specifically APPROVEd this CONSENT PROCEDURE and interviewers documented the CONSENT process as part of the study protocol. Study Design: We CONDUCTed a cross-sectional study at the end of Year 4 of the MassConect initiative to DESCRIBE the social network that developed over the time since the network's founding. We CONDUCTed a sociometric network analysis, meaning that we had a pre-defined network and attempted to COLLECT data from each member about relationships to all other members on the LIST. This type of network analysis supports evaluation of network growth and resource exchange [33], [38]. General study results and community-specific INFORMATION were presented to each of the community coalitions after the analysis was completed. This study was conceptualized, planned, implemented, and evaluated using CBPR principles [39], by a dedicated working group which was a subset of the Community Advisory Group, which included community partners from each community as well as investigators, dissemination partners and study staff. To limit potential conflicts and biases, most working group participants ensured that colleagues would take the survey on behalf of their group; however, two working group participants answered the survey in collaboration with colleagues.

---

In this article we have SDGs 3 (Good Health and Well Being), 10 (Reduced Inequalities) and 17 (Partnerships for the Goals).

## 5 Discussion

We start this discussion by reviewing the first research question of this study which was 1.) *To what extent do the SDGs come through as dimensions?* The results showed that researchers worldwide are discussing the Sustainable Development Goals and how their studies can, in different contexts, describe and support better solutions to globalized societies. An important finding was that the themes related to the Sustainable Development Goals are varied and based on different groups of society such as researchers, privileged and lower income groups and the way people's behaviors directly affect their lives. By reading the excerpts presented in this study we observed, for example, discussions over how the misreporting of farmers' production will interfere in social projects to reduce hunger (SDGs 1 and 2) and how different age and background groups perceive happiness (SDGs 3 and 11). Although most of the authors do not openly refer to which SDG they are dealing with, the readers can identify them in the articles because of the topics discussed and how they are interconnected. In Example 1, the same text has three themes, 'Responsible Consumption and Production' (SDG 12), 'Peace, Justice and Strong Institutions' (SDG 16) and 'Partnerships for the Goals' (SDG 17) which appear through the words "Policy", "Other", "Government", "Establish", "Work", "Provide", "Important", "Ensure", "Public" and "Effective".

The second question proposed in the beginning of this study was 2.) *Which themes are mostly being discussed in each lexical dimension?* This procedure showed which words presented

higher co-occurrences in the research papers and detected 6 themes in the corpus which were labeled *Government Actions*; *Presenting Results*; *Data Interpretation*; *Data Presentation*; *Data Quality* and *Research Procedure*. As previously presented, the first dimension is the one where we have most of the content words loaded and it was the only one that was strictly related to the SDGs since it revealed words that described how governments and stakeholders act regarding decisions and sanctions taken depending on different issues. As for the other dimensions, we see that the academic register is very salient and that might be the reason why we have the co-occurent lexis dedicated to describing different sections of research papers. Different from Delfino et. al (2018), where the Lexical MDA was used with a corpus of research papers about Applied Linguistics, in our study, we have a broad variety of topics from the seventeen SDGs in different research areas. In Delfino et. al's study (2018), specific themes related to Applied Linguistics were identified such as Teacher Training Programs, Language and Technology, Academic Genres and Text Types and Social Practices in Classes. However, in our study, because of the variety of themes, the academic register is more salient and co-occurent, resulting in topics such as data presentation, research procedures and results presentation. Nevertheless, as we look closely at each excerpt and read the texts with the lexis present in the dimensions, the SDGs emerge. Finding out more about the academic register does meet one of the main purposes of this research which is to know what lexis international researchers use in order to discuss their findings on the SDGs. We believe this data can be used in teaching materials for graduate students and less experienced researchers.

The last question of our study was 3.) *Do contexts related to the SDGs show positive or negative discussions?* If we consider the excerpts shown in this paper, we see that there has been a general concern over which methods and actions should be taken in order to have better results in different societies. Also, distinct groups are being focused such as women, indigenous tribes, lower and higher income families, so there has been special concern over all individuals in society, which we consider a positive aspect in the articles. Another fact that called our attention is how much concern has been addressed to sanctions and punitive regulations to institutions, groups and individuals that do not contribute to projects that are thought for general society. This can be seen by the content words loaded in Dimension 1, which are strictly related to this issue.

## 6 Final Considerations

This study set out to investigate how international researchers have been discussing the Agenda 2030 and its seventeen Sustainable Development Goals (SDGs) in their studies. Since its publication in 2015, countries organized themselves to meet those goals as civil society, governments and researchers. Our aim in our study was to observe by a bottom-up procedure how the goals were being discussed through the lexis used by researchers. By applying the lexical multidimensional analysis we found six dimensions of content words that showed that in the first dimension, where we have a higher load of words, the theme being discussed was Government Decisions. On the other hand, the other five dimensions were dedicated to academic lexis, each one of them, in their own way, which were named as Presenting Results; Data Interpretation; Data Presentation; Data Quality and Research Procedure. These results

may become valuable information to be used in EAP classes in order to help international researchers to better communicate their studies with previously approved academic lexis.

At the same time, further work needs to be carried out to analyze how the same SDGs are being dealt with in different countries and by different research groups. In order to do so, we intend to observe the metadata from each article to better map the SDGs internationally. The challenge now is to follow publications and their results as the world and governments have been going through major changes such as climate change, which was discussed at the UN Climate Change Conference UK 2021 (SDG 13), and world peace (SDG 16), also recently discussed during the G7 summit in Hiroshima 2023. Continued efforts are needed to make the SDGs implemented and more accessible to general society. We believe that, by analyzing the discourse used in research papers on these themes, we can contribute to EAP instruction and research paper writing that discuss the SDG in a comprehensive way. Academic classes and terminological databases can be shared by research groups and international institutions so that these topics are discussed and solutions to major challenges can reach governments and stakeholders.

## Acknowledgements

## Authors' contributions

Paula Tavares Pinto collected and analyzed the data, wrote the text and added references. Tony Berber Sardinha tagged the corpus, ran it in the SAS program and reviewed the text. Denis Luiz Marcello Owa, Maria Claudia Nunes Delfino and Simone Vieira Resende analyzed the data, reviewed the text and added references.

## References

ABREU-E-LIMA, D. P. M.; MORAES FILHO, W. B. ; NICOLAIDES, C.; QUEVEDO-CAMARGO, G. ; SANTOS, E. M. *Idiomas sem Fronteiras - Internacionalização da Educação Superior e formação de professores de língua estrangeira*. 1. ed. Belo Horizonte: Editora UFMG, 2021. v. 1. 287p.

ANTHONY, L. *AntCorGen* (Version 1.1.2) [Computer Software]. Tokyo, Japan: Waseda University, 2019. Available at https://www.laurenceanthony.net/software/antcorgen/

ASENCIÓN-DELANEY, Y.; COLLENTINE, J. A multidimensional analysis of a written L2 Spanish corpus. *Applied linguistics*, Oxford, v. 32, n.3, p. 299-322, 2011. DOI: https://doi.org/10.1093/applin/amq053

BAUMVOL, L. K. ; SARMENTO, S. . A internacionalização em Casa e o uso de inglês como meio de instrução. *In*: BECK, M. S.; MORITZ, M. E.; MARTINS, M. L. M. ; HEBERLE, V. (orgs.). *Echoes: Further Reflections on Language and Literature*. 1ed. Florianópilis: UFSC, 2016, v. 1. p. 65-82.

BERBER SARDINHA, A. P. Uso de corpora na formação de tradutores. *Delta*: documentação de estudos em lingüística teórica e aplicada, São Paulo, v. 19, n. 2, p. 43-70, 2003. DOI: https://doi.org/10.1590/S0102-44502003000300005

BERBER SARDINHA, T.. *Lingüística de corpus*. Barueri: Editora Manole Ltda, 2004.

BERBER SARDINHA, T. *A corpus-based history of Applied Linguistics*. Apresentação de Trabalho. Associação Internacional de Linguística Aplicada (AILA), Rio de Janeiro, 2017.

BERBER SARDINHA, T. Corpus linguistics and the study of social media. *In:* O'Keefe, Anne; McCarthy, Mike. (orgs.). *The Routledge Handbook of Corpus Linguistics*. V.1. 2ed.Londres: Routledge, 2022. p. 656-674.

BERBER SARDINHA, T. Corpus linguistics and historiography: Finding the major discourses in the first 50 years of TESOL Quarterly. *Journal of Research Design and Statistics in Linguistics and Communication Science*, v.7, n.1, p.S73–S94, 2022. DOI: https://doi.org/10.1558/jrds.18538

BERBER SARDINHA, T.; KAUFFMANN, C.H.; ACUNZO, C. A multidimensional analysis of register variation in Brazilian Portuguese. *Corpora*, v.9, n.2, p. 239-271, 2014.DOI: https://doi.org/10.3366/cor.2014.0059

BERBER SARDINHA, T. B.; PINTO, M. V. (eds.). *Multi-dimensional analysis*: Research methods and current issues. London: Bloomsbury Publishing, 2019.

BIBER, D. *Variation across speech and writing*. Cambridge: Cambridge University Press, 1988.

BIBER, D.; HARED, M. Dimensions of register variation in Somali. *Language Variation and Change*, Cambridge, v.4, n.1, p. 41-75, 1992. DOI: https://doi.org/10.1017/S095439450000065X

BIBER, D. *Variation in English*: Multi-dimensional studies. London: Pearson, 2001.

BURSTEIN, R., WANG, H., R. Jr, R. C.;HAY, S. I. Development and validation of a new method for indirect estimation of neonatal, infant, and child mortality trends using summary birth histories. *PLoS medicine*, San Francisco, v.15, n.10, e1002687, 2018. DOI: https://doi.org/10.1371/journal.pmed.1002687.

CHARLES, M. 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, Amsterdam, v.31, n.2, p. 93-102, 2012. DOI: https://doi.org/10.1016/j.esp.2011.12.003

CORDEIRO, C. M. A corpus-based approach to understanding market access in fisheries and aquaculture international business research: A systematic literature review. *Aquaculture and Fisheries*, Amsterdam, v. 4, n. 6, p. 219-230, 2019. DOI: https://doi.org/10.1016/j.aaf.2019.06.001.

COXHEAD, A. A new academic word list. TESOL quarterly, Alexandria, v.34, n.2, p.213-238, 2000. DOI: https://doi.org/10.2307/3587951.

DELFINO, M. C. N. Análise multidimensional: os números na Linguística. *Cadernos de Linguística*, Campinas v. 2, n. 4, e474, 2021. https://orcid.org/0000-0003-1976-8725

DELFINO, M. C.; ARAÚJO, R. F.; BERBER SARDINHA, T. Revista brasileira de Linguística Aplicada: multidimensões temáticas. *In*: FINATTO, M. J.B.; REBECHI, R. R.; SARMENTO, S.; BOCORNY, A. E. P. (orgs.). Linguística de Corpus: Perspectivas. 01ed.Porto Alegre: Instituto de Letras da UFRGS, 2018. p. 93-124.

DUTRA, D. P.; NUNES, L. P.; ORFANÒ, B. M. ; ARRUDA, C. F. B. Institutional internationalisation through academic literacies in English: teaching and learning written genres in the Brazilian higher education context. *The Especialist* (PUCSP), São Paulo, v. 40, p. 1-14, 2019.

FINARDI, K. R.; GUIMARÃES, F. F.. Internacionalização, rankings e publicações em inglês: a situação do Brasil na atualidade. *Estudos em Avaliação Educacional*, São Paulo, v. 28, n. 68, p. 600-626, 2017. DOI: https://doi.org/10.18222/eae.v28i68.4564

FLOWERDEW, J.; PEACOCK, M. *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press, 2001.

GARDNER, D.; DAVIES, M. A new academic vocabulary list. *Applied linguistics*, Oxford, v.35, n.3, 305-327, 2014. DOI: https://doi.org/10.1093/applin/amt015

GIMENEZ, T. et al. *Guide to English as a medium of instruction in Brazilian Higher Education institutions 2018-2019*. [S. l.]: FAUBAI, 2018. Disponível em: http://faubai.org.br/britishcouncilfaubaiguide2018.pdf. Access on Dec. 2024.

HYLAND, K.; HYLAND, F. Feedback on second language students' writing. *Language teaching*, Cambridge, v.39, n.2, p.83-101. 2006. DOI: https://doi.org/10.1017/S0261444806003399

HYLAND, K.; SHAW, P. (eds.). *The Routledge handbook of English for academic purposes*. London: Routledge, 2016

*International Comparative Performance of the UK Research Base*. Elsevier UK´s Department for Business, Energy & Industrial Strategy (BEIS), 2026.

JENKINS, J. *Global Englishes*: A resource book for students. Routledge, 2014.

JENKINS, J.; LEUNG, C. English as a lingua franca. *The companion to language assessment*, Hoboken, v,4, p.1607-1616, 2014.

JENKINS, C. N., ALVES, M. A. S.; UEZU, A.; VALE, M. M. Patterns of vertebrate diversity and protection in Brazil. *PloS one*, San Francisco, v.10, n.12, e0145064, 2015. DOI: https://doi.org/10.1371/journal.pone.0145064

JODOIN, J.; SINGER, J. Mainstreaming Education for Sustainable Development in English as a Foreign Language: An Analysis of the Image-Text Interplay Found in EFL Textbooks in Japanese Higher Education. In: LEAL FILHO, W. et al. (org.)*Universities as Living Labs for Sustainable Development*. Springer, Cham p. 545-565, 2020. DOI: 10.1007/978-3-030-15604-6_34

KARPENKO-SECCOMBE, T. *Academic writing with corpora*: A resource book for data-driven learning. Routledge, 2020. DOI: https://doi.org/10.4324/9780429059926

KILGARRIFF,, A. et al. The Sketch Engine: ten years on. *Lexicography*, Berlin, v. 1, n. 1, p. 7–36, 1 jul. 2014. DOI: https://doi.org/10.1007/s40607-014-0009-9.

KLAES, C.; SCHNEEGANS, S.; SCHÖNER, G.; GAIL, A. Sensorimotor learning biases choice behavior: a learning neural field model for decision making. *PLoS computational biology*, San Francisco, v.8, n.11,e1002774, 2012. DOI: https://doi.org/10.1371/journal.pcbi.1002774.

NESSELHAUF, N. *Collocations in a learner corpus*. Amsterdam: John Benjamins Publishing, 2005.

PÉREZ-PAREDES, P. *Corpus Linguistics for Education*: A Guide for Research. Abingdon: Routledge, 2020.

PHILLIPS, C.; HOENIGMAN, R.; HIGBEE, B.; REED, T. Understanding the sustainability of retail food recovery. PLOS one, San Francisco, v.8, n.10, e75530, 2013. DOI: https://doi.org/10.1371/journal.pone.0075530

PINTO, P. T.; GARCIA, D. N. M.; SANTOS, D. C. Ações de internacionalização para o ensino e a pesquisa na área de línguas. *Fórum linguístico*, Florianópolis. v. 18, n. 1, p. 5618-5630, 2021a. DOI: https://doi.org/10.5007/1984-8412.2021.e72639

PINTO, P. T.; DE CAMARGO, D. C.; SERPA, T.; DA SILVA, L. F. Analysing the behaviour of academic collocations in a corpus of research-papers: a data-driven study. *Revista de Estudos da Linguagem*, Belo Horizonte [S. l.], v. 29, n. 2, p. 1229–1252, 2021b. DOI: https://doi.org/10.17851/2237-2083.29.2.1229-1252

PINTO, P. T.; REES, G. P.; FRANKENBERG-GARCIA, A. Identifying collocation issues in English L2 research article writing. In: CHARLES, M.; FRANKENBERG-GARCIA, A. (eds) *Corpora in ESP/EAP writing instruction*. New York: Routledge, 2021c. p. 147-170. DOI: https://doi.org/10.4324/9781003001966

PINTO, P. T. The Sustainable Development Goals (SDGs) words 'poverty' and 'sustainability' in the brazilian research: a preliminary thematic corpus-based analysis. *Cadernos de Linguística*, [S. l.], v. 2, n. 4, p. e440, 2021a. DOI: http://dx.doi.org/10.25189/2675-4916.2021.v2.n4.id440

PINTO, P. T. A multidimensional analysis of research papers on the sustainable development goals (SDGs) (2021b). *22 InPLA - Linguagem e Interfaces - aproximações e distanciamentos*. Pontifícia Universidade Católica de São Paulo, São Paulo, 2021b.

RAMANADHAN, S.; SALHI, C.; ACHILLE, E.; BARIL, N., D'ENTREMONT, K.; GRULLON, M.; VISWANATH, K. Addressing cancer disparities via community network mobilization and intersectoral partnerships: a social network analysis. *PloS one*, San Francisco, v.7, n.2, e32130, 2012. DOI: https://doi.org/10.1371/journal.pone.0032130.

SACHS, J. D. From millennium development goals to sustainable development goals. *The Lancet*, London, v.379, n.9832, p.2206-2211, 2012.

SARMENTO, S.; THIAGO, E. M. C. P.; ANDREOTTI, V. O. Science without Borders – An alternative framework for evaluation. *Interfaces Brasil/Canadá*, São Paulo, v.16, n. 1, 2016. p. 40-71. DOI: https://doi.org/10.15210/interfaces.v16i1.7760

SINCLAIR, J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

SWALES, J. M. *Genre analysis*: English in academic and research settings. Cambridge: Cambridge University Press, 1990.

TÖRÖ, T. *Cooperation or Aid*? A Corpus-Assisted Critical Discourse Analysis of Finland's Development Policy, 2020. Dissertation (Master's in English Studies) — Faculty of Culture and Society, Malmö University, Malmö, 2020. Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-22715. Accessed on: [April 9th, 2025].

VALVASON, E. *The semantics of sustainable development*: A corpus-assisted, ecological analysis of discourse across languages. 228 f. Thesis (PhD in Linguistic Sciences) — Department of Humanistic Studies, University of Pavia, Pavia, 2021. Available at: http://hdl.handle.net/11571/1438355. Accessed on: [April 9th, 2025].

VIANA, V.; TAGNIN, S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo Cidade: Hub Editorial, 2011.

YANG, D.; LUO, T.; LIN, T., QIU, Q.; LUO, Y.. Combining aesthetic with ecological values for landscape sustainability. *Plos one*, San Francisco, v.9, n.7, e102437, 2014. DOI: https://doi.org/10.1371/journal.pone.0102437

YE, F.; DENG, Q. Strategies for poverty alleviation supply chain with government subsidies and misreporting behavior in China. *Plos one*, San Francisco, v.16, n.7, e0253761, 2021. DOI: https://doi.org/10.1371/journal.pone.0253761

WANG, J. H. Happiness and Social Exclusion of Indigenous Peoples in Taiwan-A Social Sustainability Perspective. *PLoS One*, San Francisco, v.10, n.2, e0118305, 2015. DOI: https://doi.org/10.1371/journal.pone.0118305

# Appendix – Lexical Multidimensional Analysis Factors

| | |
|---|---|
| Factor 1 | address (.58); focus (.56); many (.56); provide (.55); important (.55); government (.55); also (.54); help (.54); need (.54); lack (.53); develop (.53); health (.53); make (.52); support (.52); local (.52); broad (.52); such (.52); challenge (.52); example (.50); community (.49); role (.49); identify (.49); stakeholder (.49); build (.48); other (.48); effective (.48); effort (.47); understanding (.47); involve (.47); approach (.47); understand (.47); most (.46); opportunity (.46); good (.45); several (.45); strengthen (.45); implement (.45); often (.44); ensure (.44); national (.44); create (.44); include (.44); exist (.43); public (.42); establish (.42); essential (.42); work (.41); level (.41); financial (.41); framework (.40); specific (.40); particularly (.40); resource (.40); program (.40); difficult (.40); importance (.40); able (.40); take (.40); development (.40); problem (.40); initiative (.39); international (.39); long (.39); critical (.39); inform (.39); policy (.39); promote (.38); necessary (.38); part (.38); lead (.38); well (.38); practice (.38); implementation (.38); contribute (.38); improve (.37); system (.37); strong (.36); especially (.36); impact (.36); very, (.35) influence (.35) sector (.35) major (.35); knowledge (.35) process (0.35) capacity (.35). |
| Factor 2 | show (.62); value (.61); represent (.59); assume (.55); same (.53); model (.52); give (.50); time (.50); indicate (.48); contrast (.47); state (.46); then (.45); line (.45); result (.45); action (.44); explain (.44); thus (.44); different (.43); become (.42); particular (.42); rather (.41); consider (.41); goal (.41); allow (.40); generate (.40); pattern (.39); occur (.38); start (.38); present (.37); possible (.37); large (.36); condition (.36); choose (.36); even (.36); expect (.36); account (.35); here (.35). |
| Factor 3 | examine (.77); study (.71); effectiveness (.70); outcome (.68); find (.68); investigate (.68); individual (.56); intervention (.55); common (.55); review (.54); furthermore (.52); report (.52); group (.51); explore (.50); however (.49); effect (.49); size (.48); small (.48); finding (.47); measure (.47); relationship (.46); design (.46); sample (.45); positive (.45); highlight (.44); improvement (.42); relate (.42); only (.40); risk (.39); setting (.38); performance (.38); multiple (.38); characteristic (.38); background (.37); regression (.35). |
| Factor 4 | datum (.57); data (.56); year (.56); estimate (.51); high (.50); available (.49); total (.48); number (.48); score (.47); base (.46); journal (.46); calculate (.46); region (.43); recent; (.42); increase (.42); country (.41); overall (.41); compare (.40); pone (.40); area (.40); reference (.40); status (.40); point (.39); average (.39); less (.39); rate (.39); respectively (.38); regional (.37); scenario (.37); index (.36); decrease (.36); likely (.35); vary (.35); relative (.35); proportion (.35). |
| Factor 5 | significantly (.72); significant (.62); response (.58); reveal (.57); more (.52); consistent (.51); participate (.51); observe (.50); fact (.50); difference (.49); reduction (.49); suggest (.44); interest (.39); prior (.38); associate (.38); survey (.36). |
| Factor 6 | consent (.50), approve (.44), information (.43), assign (.42), informed (.41), collect (.41), conduct (.41), select (.40) collection (.39) procedure (.39) list (.37) description (.37) obtain (.37) follow (.35) describe (.35). |