



AN ANALYSIS OF SPECIALIZED VOCABULARY FREQUENCY, LEXICAL DENSITY, AND MULTIMODALITY OF STUDENTS' WORK IN AN EMI CLASS: A CORPUS-BASED STUDY

Fitri Aulia Sanni* and Yanty Wirza

English Language Education Department, Universitas Pendidikan Indonesia, Indonesia

*Corresponding Author's email address: fitriaulia@upi.edu

ABSTRACT

Corpus analysis plays pivotal role in investigating language use within EMI contexts. However, there is still a lack of research on identifying specialized vocabulary with its lexical density along with the multimodal analysis in the science education context. This research aims to identify frequently used specialized vocabulary, measure lexical density, and investigate multimodality in the students' texts. Utilizing quantitative and qualitative methods with AntConc and Compleat Lexical Tutor and Kress and Leeweun multimodal analysis framework, the data, totaling 12.587 words, were collected from assignments in the form of mindmap submitted by 62 students enrolled in the International Program in Science Education, specifically in the Biotechnology subject, over a span of four consecutive years. The finding showed the 20 specialized vocabulary in Biotechnology disciplinary-related content words. The study also revealed the interconnection of the visual and verbal elements in the meaning-making of the disciplinary concepts. The study then confirmed that the specialized vocabulary, lexical density, and multimodality are interrelated components of communication in specialist domains which should considered for pedagogical purposes.

ARTICLE INFO

Article History:

Received 23 Jul 2024

Revised 18 Jan 2025

Accepted 27 Feb 2025

Available online 28 Feb 2025

Keywords:

Corpus, EMI, lexical density, multimodality, specialized vocabulary

To cite this paper (in APA style):

Sanni, F. A., & Wirza, Y. (2025). An analysis of specialized vocabulary frequency, lexical density, and multimodality of students' work in an EMI class: A corpus-based study. *International Journal of Education*, 18(1), 23-34.
<https://doi.org/10.17509/ije.v18i1.69721>

1. INTRODUCTION

Examining students' writing performance is a process that involves assessing their proficiency in effectively conveying information through written language. The process is closely connected to vocabulary profiling and lexical density (Altınmakas & Bayyurt, 2019; Lei & Yang, 2020; Han et al., 2021). Furthermore, examining students' writing performance through vocabulary profile and lexical density is significant for multiple reasons. Maamuujav (2021) asserts that vocabulary measures and lexical analysis of students' writing are essential and valuable indicators of students' writing and language development. They provide insight into how one's lexicon influences the quality of writing. It can assist teachers in choosing suitable vocabulary instruction and materials. Furthermore, lexical density can assist teachers in assessing the complexity of students' writing. This aligns with Nasserri & Thompson (2021), who emphasize the vital role of lexical density in evaluating the complexity and informativeness of written language. It offers valuable understanding of their proficiency in utilizing complex language structures in written form. In addition, vocabulary profiling assists teachers in identifying the strengths and weaknesses of students' vocabulary, along with lexical density, teachers may discover areas where students excel and where they need improvement. They can then provide feedback to students based on these results.

While analyzing vocabulary profiling and lexical density offers useful insights into a student's writing skills, multimodality can be used as a next step after assessing students' writing performance, specifically in selecting resources for students, as it accommodates the varied learning styles and preferences of students (Anderson & Kachorsky, 2019; Hernandez et al., 2020; Payaprom & Payaprom, 2020; Ally et al., 2022; Meda & Waghid, 2022). Furthermore, this approach has the potential to be utilized in various educational environments, including English as a Medium Instruction (EMI) settings.

To support teaching and learning activities in the EMI class, comprehending the complexities of language usage has become crucial. Therefore, mastering specialized vocabulary, lexical density, and multimodality can improve the effectiveness of teaching and learning in EMI. Many students encounter challenges in comprehending specialized vocabulary (Dang, 2019; Macaro, 2022), which may significantly hinder their understanding in EMI lessons. Moreover, comprehending specialized vocabulary may be similar to comprehending important concepts in EMI as Liu & Lei (2019) states acquiring specialized vocabulary is indispensable for enhancing subject knowledge. Students may encounter difficulties understanding technical books or discipline-specific materials without a strong grasp of specialized vocabulary. Furthermore, the EMI teacher should consider reading difficulty while choosing a material or book based on students' abilities. Lexical density can be used to achieve this goal. In addition, to comprehensively meet the diverse needs of students, teachers can leverage multimodality in resource selection. By incorporating a variety of modes, teachers can cater to different learning styles and abilities. Moreover, Grapin & Llosa (2022) emphasize that the use of multimodality in assessment also provides a deeper understanding of students' knowledge and skills, particularly for English learners, across multiple subject areas.

Due to the increasing global trend of internationalization in higher education institutions (HEIs), more HEIs are including English medium instruction (EMI) into their programs (Simbolon, 2021; Willans, 2022; Bolton et al., 2023). There is a growing trend in many countries to provide EMI in both undergraduate and postgraduate programs even though EMI is not exclusively implemented in HEIs. According to Macaro et al. (2017), EMI is the method of teaching academic subjects (except English) in countries or regions where the majority of the population does not have English as their first language, using the English language. The emergence of EMI in the Asia Pacific region is primarily driven by the growing importance of English in the economic and geopolitical spheres in Asia, as well as the adoption of local strategies for disciplinary semiotic transposition. This trend has been supported by various studies, including those by Bolton (2008), Kirkpatrick (2014), Kirkpatrick and Lixun (2020), and Walkinshaw et al. (2017). Furthermore, Airey (2020) and Richards & Pun (2021) contend that English in EMI is closely interconnected with the content of the subject, as lessons and class interactions are conducted in English. According to Bolton and Jenks (2022), English as a Medium of Instruction (EMI) overlaps significantly with English for Specific Purposes (ESP) and World Englishes. They show that Asia has twice as many HEIs using EMI as Europe. Lim and Tan-Chia (2022) emphasize the overlap in science education by utilizing non-verbal semiotic resources to enhance meaning-making interactions.

Vocabulary in ESP (English for Specific Purposes) refers to the vocabulary used in a particular field of study or professional context. It can be represented by several terms such as technical, sub-technical, and semi-technical, special purpose, discipline-specific, specialized vocabulary. Moreover, the learning of vocabulary is crucial in ESP since it enables learners to comprehend the language and concepts related to their specific area of expertise (Costeleanu, 2019). The selection of vocabulary should align with the specific requirements of the learners and the academic discipline they are studying. According to Liu & Lei (2019) there are two main categories of existing approaches used to identify specialized vocabulary: The two approaches are (1) judgment-based and (2) corpus-based.

Corpus linguistics has made significant progress in the areas of second language research and education (Lee et al., 2019; Gao et al., 2022; Kuperman et al., 2023). A corpus is a large, systematic collection of naturally occurring texts, including written and spoken language, that helps analyze and characterize a language. A corpus can include books, articles, and conversations. Corpus analysis is a valuable instrument in selecting specialized vocabulary. A frequency-annotated vocabulary list can be generated from a corpus to find the most common specialized words. A domain-specific corpus helps explain language use in a specific context. Furthermore, specialized corpora offer more comprehensive instances of vocabulary and its application in a particular field, therefore responding to the particular needs of students better than general corpora (Flowerdew, 2012).

Lexical density could be one of the parameters to measure the complexity of written or spoken texts. According to Ure (1971), lexical density is the proportion of words carrying lexical values to the total number of

words in a text. Similarly, Johansson (2009) claims lexical density refers to the ratio of content words (nouns, verbs, adjectives, and adverbs) to the overall number of words in the text. Texts with a lower density are more easily understood. A higher lexical density indicates that the majority of words in the text are content words, which signifies a text that is more informative or content-rich. Nasser & Thompson (2021) claim lexical density is considered an indicator of advanced writing and is associated with the informative quality of writings. Lexical density also predicts academic writing skills, making it a useful indicator of language progression. Automatic essay grading systems need a high lexical density to quantify text complexity objectively, especially in large student writing corpora. It is an essential part in analyzing and assessing language.

Multimodality is the use of multiple modes to construct or enhance meaning making in communication. These modes can be linguistic, visual, gestural, spatial, and audio. According to Kress (2010), multimodality is a theoretical framework that examines the diverse range of modes individuals employ to communicate and convey their thoughts and ideas. Multimodality is relevant nowadays due to technical advances and multimedia creation software. According to Kress et al (2001), science teaching and learning is particularly multimodal. The multimodal use of learning resources in EMI's science involves the utilisation of various visual elements such as tables, figures, photographs, and diagrams. These resources are employed to effectively convey scientific knowledge by utilising a diverse range of semiotic tools, including the students' native language, written words, visual materials, body language, and speech. This combination of modalities aims to enhance the students' comprehension of scientific concepts (Blair, et. al, 2018; Morell, 2020). The use of multimodal semiotic resources in EMI instruction complements translanguaging methods, which facilitate a classroom environment characterised by multivoicedness or heteroglossic (Blair, et al., 2018). It serves as a means of combining multilingual and multimodal competences to facilitate students' advanced cognitive skills and development of topic knowledge (Gu et al., 2021).

Some previous research has been conducted regarding selecting specialized vocabulary utilizing corpus in some fields such as engineering course (Nekrasova-Beker et al., 2019), civil engineering (Otto, 2021), digital science resource (Arndt, 2022), marine engineering (Đurović, 2021), lastly, science and engineering (Uehara et al., 2022). Furthermore, several studies also have examined the research on corpus with lexical density. Sainsbury (2020) examined the impact of lexical density on the simultaneous interpretation of slide presentations using corpus. Nasser and Thompson (2021) conducted a study examining the disparities in lexical density and diversity between English as a first language (L1) and English as a second language (L2) in academic settings using corpus. Moreover, Maamujav (2021) conducted a study where they used Coh-Metric and Vocabprofile to assess and forecast the quality of students' writing based on measures of lexical density, lexical diversity, and lexical complexity. Lastly, previous research on corpus with multimodality. Hirata (2019) undertook a project aimed at creating a multimodal corpus tool specifically designed for young English as a foreign language (EFL) learner. Hiippala et al. (2021) created a multimodal corpus of 1000 science graphics from primary school. In addition, Reece et al. (2023) developed a multimodal collection of 1656 spoken English interactions.

Despite the fact that those studies have provided research in selecting specialized vocabulary in various domains, corpus with lexical density, and corpus with multimodality, there is still a lack of studies in selecting specialized vocabulary in the science education context along with its lexical density and multimodal analysis. To fulfill this gap, this study aims to identify the most frequently used specialized vocabulary in the field of science education based on the results of assignments made by students of the international program at one university in Bandung, measuring the lexical density and analyzing its multimodality. In light of the background described above, this study seeks to investigate the following research questions: (1) What are the most frequently used specialized vocabulary words in Biotechnology subject among the students? (2) What is the lexical density obtained from each of the students' assignments? (3) In what ways multimodal analysis indicate the students' meaning-making of the disciplinary concepts?

2. METHOD

Research Design

The research will be conducted utilizing a mixed-method study. This study needs a mixed-method study to better understand the research issues by combining quantitative and qualitative data (Creswell, 2014), which is consistent with its research questions. To address the initial research question on the most frequently used specialized vocabulary, a quantitative approach will be employed. Coxhead (2017) states corpora are often used in quantitative studies to find the specialized vocabulary for ESP in research that is based on corpus analysis. Likewise, a quantitative approach will be employed to obtain the lexical density of the text. Conversely, qualitative methodologies will be employed to analyze the multimodality of the chosen texts.

Research Context

The data was collected from an international program at one university in Bandung. This program runs as an EMI where it uses English to teach the science topics and prepares graduates to work in an international education environment. The data collection process involves reaching out to three lecturers in the study program, Physics, Chemistry, and Biology lecturers. From each of them, a variety of student assignments related to the three science subfields are obtained. The course under investigation was Biotechnology, which generally deals with the use of living organisms or their components to develop or create products and processes that have practical applications in various fields. The students' assignment selected for the study is in the form of diagram, more specifically mindmap. The data will be analyzed to achieve the research objectives related to specialized vocabulary, lexical density, and multimodal analysis. Moreover, the mindmap predominantly focuses on the application of biotechnology in the domains of health, food and agriculture, environment, and industry.

Data Source

The data of this research consists of the assignments in Physics, Biology, and Chemistry. Out of the three subfields acquired, only data from the Biology field was utilized as it satisfied the requirements set by the research question of this study. The dataset comprises four consecutive years of students' assignments in Biotechnology disciplines from 2020 to 2023. It includes 62 student submissions in the form of diagrams, specifically mindmaps, saved as jpeg and pdf files.

This study will involve utilizing the AntConc software as it is one of the recommended tools in corpus linguistics due to its efficient and precise performance when processing diverse small and mid-sized corpora (Anthony, 2013). Additionally, it provides a variety of functions and features, one of them is keyword frequency generators (Anthony, 2005). Furthermore, this research will utilize Compleat Lexical Tutor website, the feature that is used is Vocabprofile where it can perform lexical text analysis. The program categorizes the words in any given text into four groups based on their frequency: the top 1000 most commonly used words in English (K1), the next 1000 words in terms of frequency (K2), both K1 and K2 are come from General Service List (West, 1953), ranging from 1001 to 2000, third is the 550 academic words of English known as the AWL (Coxhead, 2000), which are frequently used in academic texts across various subjects, and lastly the remaining words that do not appear on any of the previously mentioned lists which is off-list words or specialized vocabulary (Cobb, 2024).

In selecting the mindmaps, three mindmaps that have the highest lexical density value were chosen. The purpose of this option is to explore the notions of lexical density and multimodality, which are both related to the complexity and richness of language and communication. These concepts can be beneficial in helping the EMI class achieve its objectives.

Data Analysis

Initially, all the datasets comprising 62 student assignments in PDF and JPEG format will be converted into plain text using the Google Docs tool. Subsequently, a careful proofreading process was conducted on the 62 transformed data to ensure the fidelity of the text to the original data. Once the proofreading was finished and the text was verified to be identical to the original, all of the data was compiled into one and submitted to the Vocabprofile in Compleat Lexical Tutor. In addition, the results will display four distinct categories of words: K1 and K2, AWL (Academic Word List), and Off-list words that do not belong to any of the previous groupings. Off-list words usually show a higher degree of specificity, technicality, or rarity compared to the words found in the other lists. Off-list words will be inputted into the AntConc program to create a small-sized corpus. Then, the word feature in the AntConc application will be utilized to determine the frequency of the specialized vocabulary that has been established. The next stage involves determining the lexical density of all the texts. This investigation also use the Vocabprofile function to ascertain the lexical density of a text. The three texts with the highest lexical density will be chosen for multimodal analysis consisting verbal and visual elements in order to address the last research question based on Visual Grammar Framework by Kress and Leeweun (2006). It involved a multimodal analysis comprising multiple steps. The initial phase included categorization, as well as the interaction between visual and textual components. Furthermore, this study examined the meaning-making or interpretation of the multimodal resources employed, such as images, text, and layout. The final stage involved the inference of the mindmap. This step entailed making conclusions and discerning patterns in the utilization of multimodal resources.

3. RESULTS AND DISCUSSION

Specialized Vocabulary

Several procedures have already been completed to pick specialized vocabulary based on the small corpus of students' assignments. The findings reveal the order of specialized vocabulary, with the top-ranked words being those with the most frequency. The total tokens or words of the entire text amounted to 12.587 tokens. In corpus linguistics, tokens refer to the total words. While types are distinct words or unique terms inside a text that are counted as one regardless of how often they occur, and families are collections of words that may have diverse forms or endings but a common base or root.

Table 1 shows the result of the analysis of Vocabprofil in Lextutor is shown below. The total tokens of the entire text were 12.587, as shown in Table 1. There were 3384 total types. The type-token ration, which is defined as the measure of lexical diversity within a text, was 0.27. The tokens per type, or the average number of times each unique word appears in the text, was 3.72. Finally, the lexical density for the entire text was 0.74.

Table 1 – Vocabprofile analysis of the students' assignment corpus

Words in text (tokens):	12587
Different words (types):	3384
Type-token ratio:	0.27
Tokens per type:	3.72
Lex density (content words/total)	0.74

Ultimately, Off-List terms, which do not belong to the K1, K2 words, and AWL categories, make up 35.24% of the corpus with a total of 4.436 words. It can be said this small corpus, comprising 62 students' projects, contains 35.24% specialized vocabulary which indicates a strong presence of discipline-specific biotechnology words, suggesting that these assignments are heavily focused on biotechnology concepts. Moreover, comprehending these materials requires a knowledge of discipline-specific concepts in Biotechnology, which may provide challenges for individuals who aren't familiar with the topic. After that, the specialized vocabulary that has been obtained is calculated using the AntConc program to find the frequency of each specialized vocabulary. The 20 words with the highest occurrence in the Biotechnology subject are presented in Table 2. The table includes the highest specialized vocabulary, total word frequency, and example sentences containing the specialized vocabulary.

Table 2 – Twenty specialized vocabulary with the highest occurrence

No.	Vocabulary	Frequency	Example of sentence
1.	Cell(s)	95	Lymphocyte cells possess an enzyme able to synthesise this metabolite and so it is also present in hybridoma cells, allowing their survival.
2.	DNA	77	To produce a GM plant, new DNA is transferred into plant cells.
3.	Enzyme(s)	58	The protease enzyme can be used as a meat softener for clay meat for easy chewing, and helps strip off the skin of fish.
4.	Recombinant	40	Recombinant DNA technology comprises altering genetic material outside an organism to obtain enhanced and desired characteristics in living organisms or as their products.
5.	Genetic	38	Phytoremediation and Plant Resistance Development, Genetic engineering has been with Monoclonal antibody is an antibody of single specificity, generated from the immortalisation of a plasma B cell in vitro. dely used for the detection and absorption of contaminants in drinking water and other samples.
6.	Biotechnology	34	Biotechnology has revolutionized the use of traditional enzymes to be applicable in industries such as food, beverage, personal and household care, agriculture, bioenergy, pharmaceutical, and various other segments.
7.	Gene	33	Corn modified with a bacterial insecticide gene so that it produces insect toxins within its cells, protecting it from pest species
8.	Antibiotics	32	Antibiotics are a group of molecules, some natural or synthetic, which have the effect of inhibit or stopping a biochemical process in the organism, especially in the process of infection by bacteria.
9.	Vaccines	32	The traditional vaccines are either killed microorganism or attenuated one to generate immune response in body after their inoculation
10.	Bacteria	32	The bacteria then rapidly produce large quantities of the monoclonal antibody.
11.	Microorganisms	29	Microorganisms can produce metabolites which inhibits the growth of other microorganisms
12.	Protein	27	This enzim causes the milk protein to coagulate, leading to the precipitation
13.	Immune	25	DNA vaccines have higher immune response inducing potential as compared to conventional live attenuated or killed vaccine
14.	Fermentation	24	Fermentation is the process of changing raw materials into food or beverages by using microorganism such as yeast.
15.	Antigen	23	Dendritic cells present antigen on its surface for other types of cells could produce immune response.
16.	Yeast	23	For example, mutant yeast <i>Saccharomyces cerevisiae</i> can produce thiamine (vitamin B1) and blue-green algae can produce vitamin E
17.	Insemination	22	Artificial insemination can be a beneficial and successful treatment for some couples who have trouble conceiving.
18.	Organism	22	The recombined DNA molecule is inserted into a host organism to produce new genetic combinations that are of value to science, medicine, agriculture, and industry.
19.	Tissue	22	Plant tissue culture also known as micropropagation because it involves rapid multiplication of small amount of plant material to produce more and more off springs.
20.	Antibody	21	Monoclonal antibody is an antibody of single specificity, generated from the immortalisation of a plasma B cell in vitro.

The most frequently occurring term is "cell(s)" with a total of 95 occurrences, while the term "antibody" has the lowest occurrence rate, appearing only 21 times. In addition of frequency, the example of the sentence

containing specialized vocabulary included. Nevertheless, certain sentences include small inaccuracy. Regarding the sample sentence about antibiotics, there is a slight mistake where the term "inhibit" should be replaced with "inhibiting" to maintain parallelism with "stopping". Furthermore, in the given sentence "to produce more and more off springs" the term "off springs" should be corrected to "offsprings". In the last example, the word "enzim" which is the Indonesian term for "enzyme" should be written as "enzyme". In the context of EMI, the example of sentence which is not standardized shows trivial inaccuracies, this also shows the students are learners of English as a foreign language in Indonesia.

Lexical Density

This study utilizes the Vocabprofile feature to assess the lexical density of each text, specifically targeting materials used in EMI contexts. It should be noted that lexical density in Vocabprofiles has formula content words per total. The higher the value, the more content words there are, and vice versa. The outcomes are displayed in the table provided below.

Table 3 – Lexical density found in the students' assignments

No	Year and Topic	Lexical Density (content words/total)
1.	2021 Tape	0.97
2.	2021 Tempe	0.97
3.	2020 Biogas	0.93
4.	2021 Microbial Pesticides	0.93
5.	2022 Cloning	0.92
6.	2022 Water Purification	0.92
7.	2021 Biotechnology	0.90
8.	2020 The Baking Industry	0.89
9.	2021 Methane Production	0.89
10.	2021 Cheese Manufactures	0.89
11.	2021 Oil Recovery	0.89
12.	2021 Vaccines	0.88
13.	2022 Methane Production	0.88
14.	2022 Vaccine	0.87
15.	2021 Artificial Insemination	0.86
16.	2021 Yoghurt	0.86
17.	2021 Plant Tissue Culture	0.86
18.	2021 Food Additives	0.85
19.	2021 Bread	0.84
20.	2020 Plant Tissue Culture	0.83
21.	2022 Gene Engineering in Animals	0.83
22.	2022 Genetic Engineering in Plants	0.83
23.	2020 Recombinant DNA Technology	0.81
24.	2020 Food Additives	0.81
25.	2020 Methane Production	0.79
26.	2021 Biogas	0.79
27.	2021 Genetic Engineering in Animals	0.79
28.	2020 Antibiotics	0.78
29.	2020 Microbial Pesticides	0.78
30.	2021 Recombinant DNA Technology	0.78
31.	2022 Microbial Pesticides	0.77
32.	2022 Waste Management Mind Map	0.77
33.	2021 Cloning	0.76
34.	2022 Recombinant DNA Technology	0.75
35.	2020 Genetic Engineering in Animals	0.75
36.	2021 Waste Management	0.75
37.	2021 Antibiotics	0.75
38.	2021 The Baking Industry	0.75
39.	2022 Plant Tissue Culture	0.75
40.	2020 Yoghurt	0.74
41.	2020 Hydrogen Production	0.72
42.	2022 Antibiotics	0.72
43.	2022 Cheese Manufacture	0.72

44.	2022 Oil Recovery	0.71
45.	2020 Waste Management	0.70
46.	2020 Vaccines	0.69
47.	2020 Tempe	0.69
48.	2020 Genetic Engineering in Plants	0.69
49.	2021 Genetics Engineering in Plants	0.69
50.	2023 Enzymes	0.69
51.	2023 Food Additives	0.69
52.	2020 Bread	0.68
53.	2022 Artificial Insemination	0.68
54.	2023 Antibiotics	0.68
55.	2020 Tapai	0.67
56.	2023 Antibodies	0.67
57.	2023 Vaccines	0.67
58.	2020 Artificial Insemination	0.64
59.	2020 Cheese Manufacture	0.63
60.	2020 Water Purification	0.63
61.	2020 Oil Recovery	0.62
62.	2020 Cloning Dolly	0.61

Table 3 displays in descending order based on lexical density. The table above contains the lexical density of each text and also contains the year when the text was created and the topic, which serves to differentiate each text. Based on the collected data, the highest lexical density is 0.97 (97%) from the 2021 Tape text. On the other hand, the lowest lexical density is 0.61 (61%) achieved by 2020 Cloning Dolly text. In other words, the average is 0.78 (78%), meaning that the data contains a significant number of content words. This is a common in science classes, where content words are frequently encountered during teaching and learning activities.



Fig. 1 – Student mindmap 1 2021 Tape

Multimodal Analysis

This study also concerns with the multimodal analysis to see the meaning-making of the data by analyzing its visual and verbal element, this research selected three students' assignment to be analyzed with the highest lexical density which consists of 2021 Tape, 2021 Tempe, and 2020 Biogas. The analysis focuses on the concepts in Visual Grammar Framework by Kress and Leeweun (2006).

Figure 2 above displays the first mindmap employing the topic "Tape" as a part of Biotechnology. The title "Tape" is located at the center of Figure 2, it is the salience of this mindmap and is accompanied by four distinct images of tape forms serving as the background. Each sub-division and branch is accompanied with a rectangular colored background to enhance readability. Moreover, there are four sub-divisions of it represented by arrows, indicating the processes that relate with Tape, namely example, process, definition, and type, indicating classification since it involves the act of grouping things together based on shared features, particularly to

categorize or gain understanding about something connected to Tape. Furthermore, Kress et al. (1996) claim texts should be read in the manner they are intended to be read, which is from left to right. In the case of Gunther's statement, when the reader read starting from the left side which is example and process then moves to the right side which is definition and type, the reader may experience confusion, as they will encounter the example section before the definition section, which is often the first thing one expects when encountering a new knowledge. Each subdivision will be further divided into multiple branches represented by arrows which show connections or relationships between the subdivision and its respective branches. The definition part also includes traditional food and conventional biotechnology without defining them or Tape. Following the definition, the text explores the two types of biotechnology which is leaf and jar. The leaf section includes visual representations of its examples, such as guava leaves, hibiscus leaves, banana leaves, and teak leaves. These images serve to enhance the meaning making of Tape type to better understand the examples' physical characteristics of Tape products. The second type is a jar, which also has a supporting photograph. The following section covers aerobic fermentation and yeast. Anaerobic fermentation begins with enzymatic breakdown of starch into glucose and represented by arrows. *Pediococcus* and *acetobacter* bacteria metabolise glucose into alcohol and organic acid. Alcohol oxidation produces esters. In the second phase, *saccharomyces cerevisiae* germs and *rhizopus oryzae*, *endomyopsis burtonii*, *mucor* sp., *candida utilis*, *sacromycopsis fibuligera*, and *pediococcus* sp. are used. Finally, cassava tape, banana tape, and sticky rice are illustrated to show their appearance.

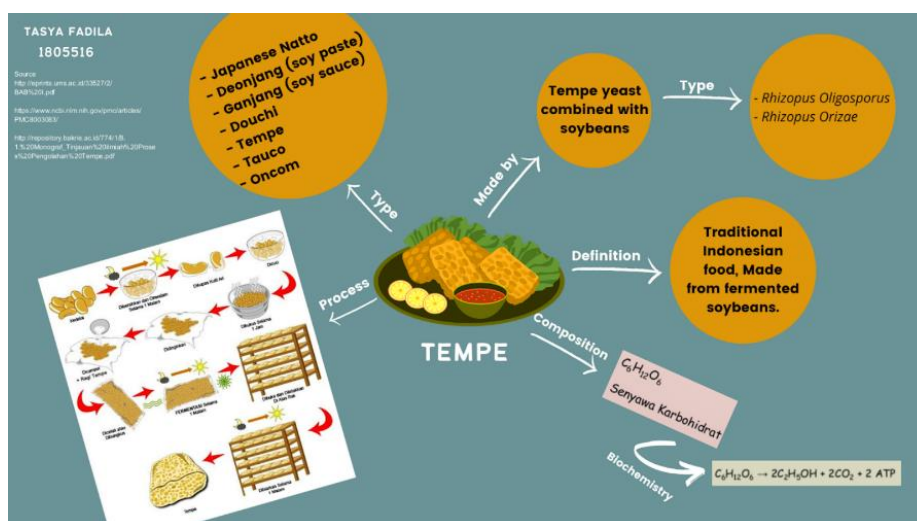


Fig. 2 – Student mindmap 2 2021 Tempe

Figure 3 is the second mindmap. This mindmap covers tempeh. Like figure 2, the word “tempe” is in the middle of the mindmap and has high salience, suggesting its importance. Tempe uses all-capitalization and boldface. This font is meant to make a big impression and grab readers' attention as in line with Bezemer & Kress (2015), the use of bold font or by capitalizing shows the intensity. Furthermore, the middle section includes illustrative images of cooked tempeh. Additionally, the images feature vegetables and sambal, which are commonly served alongside tempeh in Indonesian culture. Overall, these visuals portray the final outcome of processed tempeh. This mindmap is divided into five sections: type, process, made by, definition, and composition. It is represented by arrows indicating the connection or relationship between the subdivision and its respective branches. Each section is represented by a circle colored in Goldenrod, which is a shade of yellow. The initial categorization is the type of tempe, comprises Japanese natto, deonjang (fermented soybean paste), ganjang (fermented soy sauce), douchi, tempeh, tauco, and oncom. The font color employed to depict the various forms of tempeh is black. There is also the process of manufacturing tempeh in the second sub-division. However, the graphics depicting the process have low resolution, making it difficult to discern the process details due to the small print size. Additionally, there is a sub-division called “made by”. This sub-division explains that tempeh is made from yeast and soybeans, then an arrow points to the specific yeasts used, *Rhizopus Oligosporus* and *Rhizopus Orizae*. The following section of the mindmap defines tempe, but it puts it in the right side rather than the left, which may confuse the reader. It describes tempe as an Indonesian fermented soybean dish. Images of tempeh, vegetables, and sambal, representing Indonesia, support this concept in the central area of mindmap. This is the salience of this mindmap, indicating the element that stands out and to attract reader attention. Lastly, the composition subdivision of tempeh consists of carbohydrate substances or $C_6H_{12}O_6$. This process is represented by the chemical formula $C_6H_{12}O_6 \rightarrow 2C_2H_5OH + 2CO_2 + 2ATP$. This part is further categorized under biochemistry, which demonstrates the process of carbohydrate molecules conversion into energy.

The last mindmap is figure 4 “Biogas”. The title is displayed in all capital letters and bold type to emphasize its importance and the salience of the mindmap. This area is represented by a rectangular shape with a green background. The mindmap is entirely made by hand. Nevertheless, the author of this mindmap missed the description of the categories that created. The mindmap consists of six sections, each shown by a dark green line indicating the process or relationship with the sub-branches. Beginning from the upper left corner, this subdivision consists of four branches, each of which is accompanied with a supporting image to depict the description. The

first branch belongs to electricity, shown by red lightning and some blue resistor sign. In Indonesia context, it is a symbol of PLN (Perusahaan Listrik Negara) or Indonesian power company which is the company that distribute electricity to all over Indonesia, the use of this symbol demonstrates the localization regarding electricity and made by L2 student in Indonesia. The second branch relates to heat, depicted as an orange flame. The third branch represents fuel, symbolized by a red car, indicating fuel is the energy source for the red car. Lastly, the fourth branch represents a gas grid, depicted by a blue flame. The second subdivision showcases the construction of biogas plant with illustrative image which consists of the mixing tank, gas storage tank, gas control valve, and overflow tank. Furthermore, the student also produces subterranean sketches, which include the inlet pipe, inlet tank, digester tank, partition wall, and outflow pipe. Next is oxygen-free anaerobic procedure. This section is mostly text. The absence of oxygen is symbolised by a slanted O₂. This section has multiple branches and a blue background. The first branch covers pH 6.5–7.8.

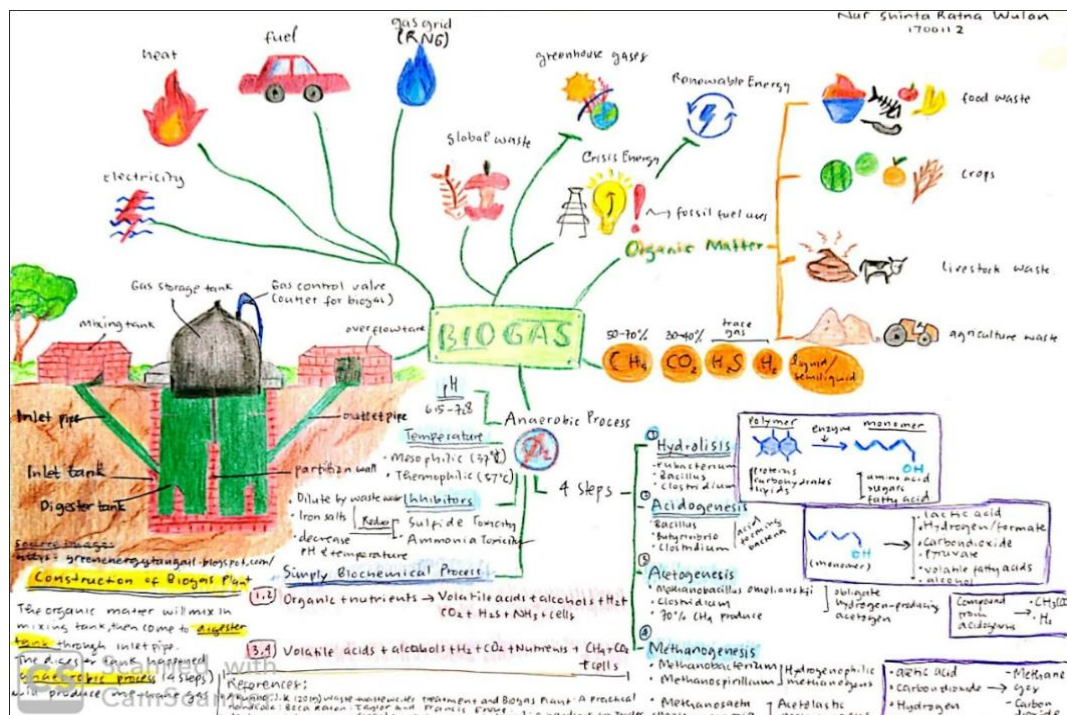


Fig. 3 – Student mindmap 3 2020 Biogas

The second branch covers mesophilic and thermophilic temperature. The next branch covers inhibitors and waste water dilution. Sulphide and ammonia toxicity reduction are subbranches of this branch. There are also branches for iron salts and pH and temperature drop. Two simple biochemical steps follow the inhibitor. The anaerobic process includes hydrolysis, acidogenesis, asetogenesis, and methanogenesis. Polymer—molecule components—is visualised in hydrolysis. Each of these steps involves the participation of microorganisms and entails the transformation from the initial form to a new form within a rectangular container.

The following top-right subdivision has numerous branches. Global waste is symbolised by fish bones and devoured apples in the first branch. Both fish bones and devoured apples are recognizable signs of consumption and waste generation. These symbols highlight the fact that much of our waste comes from organic materials.

The second branch shows greenhouse gases as the light shining into the earth and a blue curving line trapping them. The following branch depicts the energy crisis with a lit electric tower and red exclamation. This section covers renewable energy with visuals of blue lightning and two arrows making a circle. The blue lightning's energy has two circular arrows signify circularity and continual flow. In the following section, organic matter is the next sub-division. This section is divided into four branches: food waste, which includes representations of garbage, fish bones, and fruit skins; crops represented by images of watermelons, melons, oranges, and rice; livestock waste, which includes representations of cows and manure; agriculture waste, which includes images of tractors and grain. The last subdivision is a description of numerous chemical formulae in orange circles such as CH₄ with a percentage of 50-70%, the second chemical formula is CO₂ or carbon dioxide with a percentage of 30-40%, then there are trace gases specifically H₂S and H₂, and finally there is liquid/semiliquid.

Based on the multimodality analysis approach on Figure 2 (Tape), 3 (Tempe), and 4 (Biogas), the study shows the important points of meaning-making these mindmaps to convey the disciplinary concepts. Firstly, each mind map starts with a title, image, and bold writing to emphasise a main idea. Secondly, all three employ branches with arrows from the centre to categorise main topic information including definition, type, procedure, and example. Arrows and lines show the process and link between categories and subcategories. Furthermore, all mindmaps use images, icons, and illustrations to improve meaning and express concepts. In Figures 2 and 3, the definition section is on the right, which may confuse readers because when they read from left to right, they should know the definition before jumping to the type, process, etc. Finally, low visual resolution makes mindmap analysis harder.

Discussion

The mixed-method approach employed in this study provided valuable insights into the specialized vocabulary, lexical density, and multimodal analysis of student assignments in EMI class. While previous research focused on select specialized vocabulary utilizing corpus in engineering course (Nekrasova-Beker et al., 2019), civil engineering (Otto, 2021), digital science resource (Arndt, 2022), marine engineering (Đurović, 2021), and science and engineering (Uehara et al., 2022), this research focus on finding specialized vocabulary in biotechnology class as EMI program. The findings highlighted a significant presence of specialized vocabulary related to biotechnology, indicating a strong emphasis on subject-specific language in the assignments. Identification of frequently used specialized vocabulary like "cell", "DNA", and "enzyme" underscored their crucial role in conveying disciplinary concepts in biotechnology. It is important to understand these specialized vocabularies because it is equal to mastering the content of Biotechnology field, the student need to familiar these discipline-specific words. As Costeleanu (2019) claims that the learning of vocabulary is crucial in ESP since it enables learners to comprehend the language and concepts related to their specific area of expertise. Furthermore, Liu & Lei (2019) also asserts that understanding discipline-specific vocabulary is essential for acquiring subject knowledge, in this case Biotechnology field. This finding suggest that it is important to teach ESP seriously, these discipline-specific words should be encouraged to learn in order to acquire the discipline-specific content and knowledge.

In addition, the analysis found that the data had a relatively high lexical density, ranging from 61% - 98%, with an average of 78%. This aligns with Ure's (1971, as stated in Johansson, 2009) claim that text with a lexical density above 40% can be categorized as high-lexical density text. The presence of a significant number of disciplinary words in the text indicates that the data is highly informative and content-rich. Moreover, This study emphasizes the significance of lexical density in student work for teachers to assess whether students are employing suitable vocabulary in their writing. According to Nasser & Thompson (2021), lexical density serves as an indicator of advanced writing and is associated with the informative quality of writing. Examining lexical density can also indicate the progress of the student in acquiring and employing new vocabulary accurately. By looking at the specialized vocabulary and lexical density, this Biotechnology class in the EMI context and ESP context is a class that is designed for students to achieve a deep understanding of the subject matter. It brings the implication that the program is preparing students for careers as scientists and science teachers properly.

Moreover, this research also demonstrated the meaning-making of EMI students' mindmaps in three different biotechnology concepts which consists of Tape (figure 2), Tempe (figure 3), and Biogas (figure 4). Each map highlights a central topic as the salience element with visuals, categorizes information using arrows, and incorporates visuals. However, the organization of categories is unclear, potentially confusing readers. Additionally, the low graphic resolution makes the mindmap more challenging to analyze. Overall, the study emphasizes the importance of verbal and visual aids in conveying disciplinary concepts. This result contributes to the field in several key ways. Firstly, mindmap can facilitate student in reorganizing how they think and made students abstract the main topic or idea they wanted to communicate and build around that topic (Cendros Araujo & Gadaniadis, 2020). Furthermore, it shows the potential of mind maps as a tool for EMI students to analyse and depict intricate scientific concepts. Furthermore, the study emphasises the importance of using multimodal analysis to comprehend student learning. By integrating mind maps into classes, students can actively organise their thoughts and comprehend intricate subjects, particularly in EMI disciplines. It also promotes critical thinking (Sari et al., 2021). Furthermore, mind maps empower students to self-organize knowledge, fostering ownership and engagement with the learning process.

4. CONCLUSION

The study has identified 20 specialized words that are most frequently used in a small corpus of assignments by students utilizing Vocabprofile and AntConc software. The other finding revealed the text submitted by students has high lexical density which is more challenging to read, and has demonstrated the interconnection of the visual and verbal elements in the meaning-making of the disciplinary concepts. Based on these finding, specialized vocabulary, lexical density, and multimodality are interconnected elements of communication in specialized domains. Specialized vocabulary contributes to the higher lexical density of the texts, while multimodal communication can enhance the meaning-making and accessibility to the texts by integrating diverse modes of expression alongside textual contents.

For future research, it is suggested to investigate specialized vocabulary, lexical density, and multimodality on the other specific disciplinary. This investigation can assist researchers in uncovering language patterns and obstacles unique to each discipline, as well as its impact on communication. It can be employed by corpus analysis methodologies, such as Vocabprofile and AntConc, in combination with multimodal analysis and qualitative case studies. Furthermore, it can contribute significantly to the scholarship of Applied Linguistics in EMI contexts, providing information to create language learning materials, teaching methods, and assessment tools that are tailored to meet the specific needs of various fields of study.

6. REFERENCES

- Airey, J. (2020). The content lecturer and English-medium instruction (EMI): Epilogue to the special issue on EMI in higher education. *International Journal of Bilingual Education and Bilingualism*, 23(3), 340–346. <https://doi.org/10.1080/13670050.2020.1732290>

- Ally, F., Pillay, D. J., & Govender, N. (2022). Teaching and learning considerations during the COVID-19 pandemic: Supporting multimodal student learning preferences. *African Journal of Health Professions Education*, 14(1), 13–16. <https://doi.org/10.7196/ajhpe.2022.v14i1.1468>
- Altınmakas, D., & Bayyurt, Y. (2019). An exploratory study on factors influencing undergraduate students' academic writing practices in Turkey. *Journal of English for Academic Purposes*, 37, 88–103. <https://doi.org/10.1016/j.jeap.2018.11.006>
- Anderson, K. T., & Kachorsky, D. (2019). Assessing students' multimodal compositions: An analysis of the literature. *English Teaching: Practice & Critique*, 18(3), 312–334. <https://doi.org/10.1108/etpc-11-2018-0092>
- Anthony, L. (2005). AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. IPCC 2005. Proceedings. International Professional Communication Conference, 2005. Retrieved May 15, 2024, from <http://dx.doi.org/10.1109/ipcc.2005.1494244>
- Anthony, L. (2013). Developing AntConc for a new generation of corpus linguists. *Proceedings of the Corpus Linguistics Conference (CL 2013)*, 14–16.
- Arndt, R. (2022). A specialized vocabulary list from an original corpus of digital science resources for middle school learners. *Journal of English for Academic Purposes*, 60, 101187. <https://doi.org/10.1016/j.jeap.2022.101187>
- Bezemer, J., & Kress, G. (2015). *Multimodality, Learning and Communication: A social semiotic frame*. Routledge.
- Blair, A., Haneda, M., & Bose, F. N. (2018). Reimagining english-medium instructional settings as sites of multilingual and multimodal meaning making. *TESOL Quarterly*, 52(3), 516–539. <https://doi.org/10.1002/tesq.449>
- Bolton, K. (2008). English in Asia, Asian Englishes, and the issue of proficiency. *English Today*, 24(2), 3–12. <https://doi.org/10.1017/s026607840800014x>
- Bolton, K., & Jenks, C. (2022). World Englishes and English for specific purposes (ESP). *World Englishes*, 41(4), 495–511. <https://doi.org/10.1111/weng.12604>
- Bolton, K., Hill, C., Bacon-Shone, J., & Peyronnin, K. (2023). EMI (English-medium instruction) in Indonesian higher education. *World Englishes*, 42(3), 424–446. <https://doi.org/10.1111/weng.12622>
- Cendros Araujo, R., & Gadanidis, G. (2020). Online collaborative mind mapping in a mathematics teacher education program: A study on student interaction and knowledge construction. *ZDM*, 52(5), 943–958. <https://doi.org/10.1007/s11858-019-01125-w>
- Cobb, T. (2024). Vocabprofile English. Retrieved May 15, 2024, from <https://www.lexutor.ca/vp/eng/>
- Costeleanu, M. (2019). The role of vocabulary in esp teaching. *The European Proceedings of Social and Behavioural Sciences*. <http://dx.doi.org/10.15405/epsbs.2019.08.03.120>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213. <https://doi.org/10.2307/3587951>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.
- Dang, T. N. Y. (2019). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related TV programs: Insights from medical corpora. *TESOL Quarterly*, 54(2), 436–459. <https://doi.org/10.1002/tesq.552>
- Đurović, Z. (2021). Corpus linguistics methods for building ESP word lists, glossaries and dictionaries on the example of a marine engineering word list. *Lexikos*, 31(1). <https://doi.org/10.5788/31-1-1647>
- Flowerdew, L. (2012). *Corpora and language education*. Springer.
- Gao, J., Pham, Q. H. P., & Polio, C. (2022). The role of theory in quantitative and qualitative second language learning research: A corpus-based analysis. *Research Methods in Applied Linguistics*, 1(2), 100006. <https://doi.org/10.1016/j.rmal.2022.100006>
- Grabin, S. E., & Llosa, L. (2022). Multimodal tasks to assess English learners and their peers in science. *Educational Assessment*, 27(1), 46–70. <https://doi.org/10.1080/10627197.2022.2028139>
- Gu, M. M., Li, Z. J., & Jiang, L. (2021). Navigating the instructional settings of EMI: A spatial perspective on university teachers' experiences. *Journal of Multilingual and Multicultural Development*, 45(2), 564–578. <https://doi.org/10.1080/01434632.2021.1914064>
- Han, Y., Zhao, S., & Ng, L.-L. (2021). How technology tools impact writing performance, lexical complexity, and perceived self-regulated learning strategies in EFL academic writing: A comparative study. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.752793>
- Hernandez, J. E., Vasan, N., Huff, S., & Melovitz-Vasan, C. (2020). Learning styles/preferences among medical students: Kinesthetic learner's multimodal approach to learning anatomy. *Medical Science Educator*, 30(4), 1633–1638. <https://doi.org/10.1007/s40670-020-01049-1>
- Hirata, E. (2019). The development of a multimodal corpus tool for young EFL learners. In *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. Routledge.
- Hiippala, T., Alikhani, M., Haverinen, J., Kalliokoski, T., Logacheva, E., Orekhova, S., Tuomainen, A., Stone, M., & Bateman, J. A. (2021). AI2D-RST: A multimodal corpus of 1000 primary school science diagrams. *Language Resources and Evaluation*, 55(3), 661–688. <https://doi.org/10.1007/s10579-020-09517-1>
- Johansson, V. (2009). Developmental aspects of text production in writing and speech.
- Kirkpatrick, A. (2014). The language (s) of HE: EMI and/or ELF and/or multilingualism? *The Asian Journal of Applied Linguistics*, 1(1), 4–15.
- Kirkpatrick, A., & Lixun, W. (2020). *Is English an Asian language?* Cambridge University Press.
- Kress, G., Charalampous, T., Jewitt, C., & Ogborn, J. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom*. Continuum.
- Kress, G. R. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Taylor & Francis.

- Kress, G. R., Leeuwen, T. V., & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. Psychology Press.
- Kuperman, V., Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Lõo, K., Marelli, M., ... Usal, K. A. (2023). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition*, 45(1), 3–37. <https://doi.org/10.1017/s0272263121000954>
- Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5), 721–753. <https://doi.org/10.1093/applin/amy012>
- Lei, S., & Yang, R. (2020). Lexical richness in research articles: Corpus-based comparative study among advanced Chinese learners of English, English native beginner students and experts. *Journal of English for Academic Purposes*, 47, 100894. <https://doi.org/10.1016/j.jeap.2020.100894>
- Lim, F. V., & Tan-Chia, L. (2022). Designing for multimodal literacy learning. In *Designing Learning for Multimodal Literacy* (pp. 18–36). Routledge. <http://dx.doi.org/10.4324/9781003258513-2>
- Liu, D., & Lei, L. (2019). Technical Vocabulary. In *The Routledge Handbook of Vocabulary Studies*. Routledge.
- Maamujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, 49, 100540. <https://doi.org/10.1016/j.asw.2021.100540>
- Macaro, E. (2022). English Medium Instruction: What do we know so far and what do we still need to find out?. *Language Teaching*, 55(4), 533–546.
- Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2017). A systematic review of English medium instruction in higher education. *Language Teaching*, 51(1), 36–76. <https://doi.org/10.1017/s0261444817000350>
- Meda, L., & Waghid, Z. (2022). Exploring special need students' perceptions of remote learning using the multimodal model of online education. *Education and Information Technologies*, 27(6), 8111–8128. <https://doi.org/10.1007/s10639-022-10962-4>
- Morell, T. (2020). EMI teacher training with a multimodal and interactive approach: A new horizon for LSP specialists. *Language Value*, 12(1). <https://doi.org/10.6035/languagev.2020.12.4>
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, 100511. <https://doi.org/10.1016/j.asw.2020.100511>
- Nekrasova-Beker, T., Becker, A., & Sharpe, A. (2019). Identifying and teaching target vocabulary in an ESP course. *TESOL Journal*, 10(1). <https://doi.org/10.1002/tesj.365>
- Otto, P. (2021). Choosing specialized vocabulary to teach with data-driven learning: An example from civil engineering. *English for Specific Purposes*, 61, 32–46. <https://doi.org/10.1016/j.esp.2020.08.003>
- Payaprom, S., & Payaprom, Y. (2020). Identifying learning styles of language learners: A useful step in moving towards the learner-centred approach. *Dil ve Dilbilimi Çalışmaları Dergisi*, 16(1), 59–72. <https://doi.org/10.17263/jlls.712646>
- Reece, A., Cooney, G., Bull, P., Chung, C., Dawson, B., Fitzpatrick, C., Glazer, T., Knox, D., Liebscher, A., & Marin, S. (2023). The CANDOR corpus: Insights from a large multimodal dataset of naturalistic conversation. *Science Advances*, 9(13). <https://doi.org/10.1126/sciadv.adf3197>
- Richards, J. C., & Pun, J. (2021). A typology of english-medium instruction. *RELJ Journal*, 54(1), 216–240. <https://doi.org/10.1177/0033688220968584>
- Sainsbury, C. (2020). The impact of lexical density in simultaneous interpreting of slide presentations: A replication. *Master: Univ. Genève*.
- Sari, R. M., Sumarmi, S., Astina, I. K., Utomo, D. H., & Ridhwan, R. (2021). Increasing students critical thinking skills and learning motivation using inquiry mind map. *International Journal of Emerging Technologies in Learning (IJET)*, 16(03), 4. <https://doi.org/10.3991/ijet.v16i03.16515>
- Simbolon, N. E. (2021). English Medium Instruction (EMI) practice: Higher education internationalization in Indonesia. *Englisia: Journal of Language, Education, and Humanities*, 8(2), 72. <https://doi.org/10.22373/ej.v8i2.8961>
- Uehara, S., Haraki, H., & McLean, S. (2022). Developing a discipline-specific corpus and high-frequency word list for science and engineering students in graduate school. *Vocabulary Learning and Instruction*, 11(2), 57–68. <https://doi.org/10.7820/vli.v11.2.uehara>
- Ure, J. (1971). Lexical density and register differentiation. In *Applications of Linguistics* (pp. 443–452). Cambridge University Press.
- Walkinshaw, I., Fenton-Smith, B., & Humphreys, P. (2017). EMI issues and challenges in asia-pacific higher education: An introduction. In *English Medium Instruction in Higher Education in Asia-Pacific* (pp. 1–18). Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-51976-0_1
- Wang, S. P., & Chen, Y. L. (2018). Effects of Multimodal Learning Analytics with Concept Maps on College Students' Vocabulary and Reading Performance. *Educational Technology & Society*, 21(4), 12–25.
- West, M. (1953). A general service list of English words: With semantic frequencies and a supplementary word-list for the writing of popular science and technology.
- Willans, F. (2022). The scramble for EMI: Lessons from postcolonial 'old EMI' universities. *Teaching in Higher Education*, 27(4), 546–557. <https://doi.org/10.1080/13562517.2022.2029395>