

Integrating Corpus-based Methods to Determine Grammatical Topics for Teaching English Writing in the Thai Context

Raymund T. Palayon* and Yenyong Chongchit

King Mongkut's Institute of Technology Ladkrabang, Chumphon Campus,
Chumphon, Thailand

Corresponding author: raymund.pa@kmitl.ac.th

Article information	
Abstract	Grammatical knowledge for effective English writing remains a significant need among Thai university students, while identifying the relevant grammatical topics to address this need continues to be a challenge in teaching writing courses. Previous papers on writing in Thai universities have mainly focused on writing performance issues, rather than proposing methods to address these issues, particularly in grammar. This quantitative case study employed corpus analyses to determine the relevant grammatical topics that address the grammatical needs of a specific group of Thai university students in English writing. ChatGPT was utilized to improve the essays of the target students, resulting in two datasets (namely, students' essays and refined essays). Keyword analysis was employed to illustrate their grammatical characteristics. Biber's (1988; 1989) multidimensional analysis and key part-of-speech analysis were applied to verify the keyword findings and determine the target grammatical topics for teaching writing. The analyses revealed 12 significant grammatical features that address the grammatical needs of the target students in English writing (namely, nominalizations, determiners, present participial clauses, prepositional phrases, attributive adjectives, conjuncts, independent clause coordination, phrasal coordination, past

	participial clauses, <i>that</i> relative clauses in the subject position, adverbial subordinators, and sentence relatives). These features can make the students' writing compositions informational, narrative, and formal, reflecting the characteristics of academic language. This paper provides procedures for identifying grammatical topics that are aligned with the current grammatical needs of students in writing.
Keywords	keyword analysis, multidimensional analysis, key part-of-speech analysis, essays, grammatical features
APA citation:	Palayon, R. T., & Chongchit, Y. (2025). Integrating corpus-based methods to determine grammatical topics for teaching English writing in the Thai context. <i>PASAA</i> , <i>70</i> , 273–307.

1. Introduction

Writing is a fundamental aspect of communication that plays a crucial role both in academic and professional contexts (Klimova, 2012). Given its significance in academic success and career advancement, tertiary institutions have integrated writing into English language learning curricula, highlighting its integral role in fostering critical thinking (e.g., Condon & Kelly-Riley, 2004) and effective communication skills (e.g., Holmes et al., 2018). Learners are required to complete writing courses to master writing skills in order to produce well-organized and purpose-driven documents essential for career and business development.

A key requirement for effective writing is a solid understanding of grammar, which provides the foundation for constructing clear and coherent sentences. Without grammatical knowledge, even the most insightful ideas can be lost or misunderstood due to poor sentence structure (Hans & Hans, 2017). Educators believe that grammar serves as the framework for organizing thoughts and ideas logically, ensuring that the writer's message is accurately conveyed to the reader (e.g., Sacal & Potane, 2023). Mastery of grammar allows students to express their ideas with precision and fluency, which is essential for both academic and professional writing (e.g., Etfita, 2019).

In Thailand, particularly at tertiary level, the integration of writing into English courses has consistently been practiced to enhance students' language proficiency and prepare them for professional endeavors in an increasingly globalized society (Chuenchaichon, 2015; Watcharapunyawong & Usaha, 2013). However, teaching English writing in Thai university context remains a challenging task among educators due to the disparity between the expectations of writing courses and the English background of Thai learners (Thuong, 2020). In some cases, the writing teaching approach does not meet the learning expectations of Thai students, resulting in poor performance (Iyer, 2013). Previous studies have highlighted various factors affecting the writing performance of Thai students (e.g., Kareviati et al., 2020; Nguyen & Suwannabubpha, 2021; Rodsawang, 2017), and one of the main problems identified is their low level of grammatical knowledge.

According to Boonyarattanasoontorn (2017), Thai students often do not focus on learning grammar and one reason may be the complexities of grammar, which make the process of learning writing tedious and unengaging. This is reflected in the low grammar test scores of a group of Thai university students, indicating their poor performance in English, as reported by Saengboon (2017). The consistent grammatical problems among Thai students can be considered a major reason why their writing performance remains low. The difficulties with grammar are not merely a matter of academic challenges but they also affect their overall communication skills (Lhorsumeth, 2017). Addressing these persistent grammatical issues is crucial for improving students' writing abilities and enhancing their proficiency in English. Thus, targeted interventions focusing on simplifying grammar instruction and making it more engaging could potentially lead to better writing outcomes for these students.

Therefore, this paper applied corpus analyses (namely, keyword analysis, multidimensional analysis, and key part-of-speech analysis) to identify the grammatical topics that addressed the grammatical needs of a specific group of

Thai university students in English writing. To perform the analyses, we used ChatGPT, AntConc 3.4.4w, and Multidimensional Analysis Tagger as tools. The essays of the target students were compiled and improved using ChatGPT (see Section 4.1 for further details on data preparation). The corpus of students' essays and the corpus of refined essays produced by ChatGPT were analyzed through keywords, identified by keyword analysis, to initially view the characteristics of the two datasets. To verify the results, the multidimensional analysis of Biber (1988; 1989) and key part-of-speech analysis were employed, allowing us to describe the general characteristics of the two datasets, verify the keyword findings, and determine the grammatical needs and topics for teaching writing. This study aimed to offer insights to enhance the pedagogical framework for teaching writing to students who are learning English as a second language, specifically in the Thai university context.

2. Issues with Teaching English Writing in Thai Universities

In Thai universities, teaching writing presents several issues that hinder students' ability to develop strong writing competencies. Numerous studies have been conducted to explore the English language difficulties faced by Thai students, and the findings consistently reveal that writing is the most problematic aspect for Thai undergraduate students (Boonyarattanasoontorn, 2017; Seensangworn & Chaya, 2017; Wahyuni & Ilyas, 2016). These problems include low engagement in learning to write, difficulties with grammar in sentence writing, and a lack of effective writing tasks tailored to students' needs. Rodsawang (2017) identifies various writing problems faced by Thai EFL students, including issues with writing styles, grammar, vocabulary, and insufficient practice hours. Nguyen and Suwannabubpha (2021) highlight the struggles of Thai students with grammar and vocabulary in learning EFL writing due to a lack of knowledge. Additionally, teachers lack the language and educational skills necessary for effective writing instruction. These studies suggest that a low level of grammatical knowledge is a major problem, as it prevents students from writing compositions effectively. As a result, students often rely on translation software or computer programs to express

their thoughts in writing (Chompurach, 2021), indicating that they struggle with understanding and producing comprehensive English texts independently.

These issues with teaching and learning English writing in the Thai context, particularly with regard to grammar, have led scholars to explore this area. However, previous studies have primarily focused on identifying grammatical errors in writing rather than establishing procedures for determining the relevant grammatical topics based on the current needs of the students to improve their compositions (e.g., Iamsiu, 2014; Inpanich et al., 2023; Kampookaew, 2020; Promsupa et al., 2017). One study attempted to conduct a needs analysis of English paragraph writing using a quantitative approach with a questionnaire, which revealed that the students were less proficient in writing (Pongsukvajchakul, 2023). Although the analysis provided useful information, it did not specifically address the identification of students' grammatical needs, which is a major factor affecting their writing performance. Simply identifying the grammatical errors in writing is insufficient because it does not provide students with necessary tools or strategies to correct their mistakes and improve their overall writing skills. The focus should not only be on pointing out errors but also be on teaching students how to recognize and use the appropriate grammatical structures to enhance clarity, coherence, and accuracy in their writing. The reported previous studies have enriched the literature on this topic; however, they have not fully addressed how to systematically identify students' grammatical needs, so as to indicate topics for teaching writing that could significantly enhance their compositions. Consequently, there is a need for further research to develop and implement effective methods for assessing and addressing these challenges in order to provide more targeted support for improving English writing proficiency among Thai students.

3. Corpus-based Methods for Identifying Significant Grammatical Topics

Despite the increasing volume of research on the difficulties of Thai students in writing using the English language and teaching English writing to Thai

students, there are still significant challenges and gaps in understanding how to effectively address their specific needs and improve their writing skills. Therefore, we conducted this case study to find a specific framework aimed at filling these gaps and reducing these challenges for the improvement of teaching English writing courses within the Thai context. This framework focuses on grammar, given that addressing issues in this area is highly important for enhancing the overall writing proficiency of Thai students. Having a specific focus for investigation may enable us to establish a clear purpose that can be applied to a broader context. In this study, we employed corpus analyses to find reliable evidence, describing the students' grammatical needs and determining relevant grammatical topics to address their needs. To the best of our knowledge, this approach based on the methods employed in this study has never been used before, which implies that this approach offers new insights that shed light on the identification of grammatical topics.

The corpus approach refers to a methodology in linguistic research with the aim to analyze large datasets, known as corpora, based on the linguistic patterns (McEnery, 2019; Rayson, 2008). It uses quantitative analysis through the aid of specialized computer programs to illustrate how language is actually used in various contexts (Reppen & Simpson-Vlach, 2019). The use of the corpus approach is common within applied linguistics research. However, integrating different corpus-based methods is a unique methodological strategy that allows for a more comprehensive analysis. This approach provides a deeper understanding of the specific linguistic features and their patterns that are helpful to characterize the students' grammatical needs and determine grammatical topics for teaching writing based on the datasets gathered.

Three corpus-based methods were employed to serve the goals and purposes of this study. Keyword analysis, which aims to identify the important lexical items that are more frequent in a target corpus than in a benchmark corpus (Pojanapunya, 2017; Rayson, 2019; Rayson & Potts, 2021), was used to initially

check the linguistic sophistication of each dataset. To strengthen the first set of findings, multidimensional analysis and key part-of-speech analysis were applied. Multidimensional analysis was used to display the general characteristics (or dominant text dimensions) of a target corpus based on the six text dimensions in the theoretical framework of Biber (1988; 1989). These six dimensions are involved and informational discourse, narrative and non-narrative concerns, context-independent discourse and context-dependent discourse, overt expression of persuasion, abstract and non-abstract information, and on-line informational elaboration, demonstrated by sets of co-occurring patterns of syntactic features. Key part-of-speech analysis which identifies the important grammatical groups of words that are more frequent in a target corpus than in a benchmark corpus (Culpeper, 2009; Palayon et al., 2022; Rayson, 2008) was used to allow us to verify the syntactic findings in the multidimensional analysis and provide further insights into the characterization of each dataset. The use of these two methods guided us in identifying significant grammatical topics for teaching writing empirically. Triangulating various corpus techniques offered a robust framework for addressing the complex grammatical challenges faced by Thai students, which should contribute to more effective strategies for teaching grammar in English writing courses.

To identify the target linguistic features, we found ChatGPT, AntConc (Anthony, 2005), and Multidimensional Analysis Tagger or MAT (Nini, 2019) to be the most suitable tools. ChatGPT is an Artificial Intelligence (AI) language model that applies Natural Language Processing (NLP). This tool is widely used because it is capable of generating human-like text responses based on the information or text it receives (Wu et al., 2023), and it was considered relevant in this study which aimed to identify relevant grammatical topics that could address the grammatical needs of a specific group of students. ChatGPT was deemed effective to assist this process. Although ChatGPT has undergone several developments over time (e.g., Sahib et al., 2023), its core functionality—particularly its ability to generate text with appropriate grammatical structures—remains consistent. This core

function directly aligns with the goal of this study, which was to identify grammatical structures that could enhance the writing quality of the target students. Therefore, despite the tool's ongoing development, its ability to assist in addressing grammatical needs and improving writing remains unaffected, ensuring that the findings of this study would still be relevant and valid. Furthermore, AntConc and MAT are powerful and multipurpose corpus tools designed for linguistic analysis, displaying a range of features that enabled us to identify linguistic patterns in texts. AntConc facilitates the identification of keywords and key part-of-speech tags within the target datasets (Palayon et al., 2022). MAT performs the part-of-speech tagging and automatically calculates the dimension scores of each text dimension and z-score of each syntactic feature used for characterizing the dominant text dimensions (Nini, 2019). These tools helped us quantify the target linguistic features in the datasets that served as empirical evidence for this study.

4. Methodology

This study aimed to provide a pedagogical framework for teaching grammar in English writing courses at the tertiary level. This framework began by using corpus analysis to identify the grammatical topics for teaching English writing that addressed the grammatical needs of the students. To achieve this goal, we needed datasets that served as the target corpus and benchmark corpus to identify the target linguistic features with the help of corpus tools. In this section, we present the procedures for conducting the study, including methods for data selection and preparation, units of linguistic features for data analysis, and steps for determining the relevant grammatical topics.

4.1 Data Preparation

The first phase of this methodology involved sourcing datasets. To ensure the reliability and relevance of the datasets, a structured approach was applied (e.g., Sarma et al., 2012). We collected narrative essays from third-year Thai

students from the English Language Department of a university in Thailand. These essays focused on life experiences, such as family, friendship, and schoolwork, and served as their writing output from one semester in the essay writing course. They were written in the classroom to ensure that the students did not use computer applications to produce their essays. The students were, however, instructed to use a dictionary to assist them in writing. We aimed to minimize variations in academic background and educational context by narrowing our scope to a single university and department to enhance the internal validity of our study. Additionally, focusing on third-year students provided a balance between linguistic proficiency and academic experience, offering insights into both foundational language skills and advanced writing abilities. The texts were compiled as one dataset and termed *students' essays*. Ethical concerns were minimal, as the texts did not contain sensitive personal information or any details that could identify the students. However, permission was obtained to use the essays for research purposes.

After compiling the students' essays, ChatGPT was instructed to refine the students' essays. The given prompt instructed the tool to improve the quality of the texts, particularly their grammatical structures. This was sufficient, as this study only aimed to identify grammatical structures for teaching writing, specifically those that aligned with the students' needs and the teaching context. After refining the essays, we reviewed each text for accuracy. We focused on reviewing the introductions, bodies, conclusions, words, and sentence structures to ensure the coherence and improvement of the texts. The texts were compiled as one dataset and termed *refined essays*.

Each dataset was then grammatically tagged using MAT, and the tags produced were used for key part-of-speech analysis. The two datasets served as the target and benchmark corpora. Each target corpus contained specific linguistic features that described the grammatical needs of the students (based on the *students' essays*) and determined the grammatical topics for teaching writing

(based on the *refined essays*) to address the needs of the students. Each benchmark corpus acted as a point of comparison to reveal the distinctive features of the target corpus and provided a baseline for determining whether a particular feature in the target corpus was statistically significant or unusual. Table 1 shows the sizes of the datasets, sufficient to find linguistic patterns.

Table 1

Sizes of the Datasets

<i>Approach of comparison</i>	<i>Target corpora</i>	<i>Benchmark corpora</i>
Comparison 1 (Students' essays vs. Refined essays)	Students' essays *150 text files *30,659 word tokens	Refined essays *131 text files *31,393 word tokens
Comparison 2 (Refined essays vs. Students' essays)	Refined essays *131 text files *31,393 word tokens	Students' essays *165 text files *34,206 word tokens

4.2 Linguistic Units for Data Analysis

An essay is comprised of words, phrases, and sentences, which together form the essential components of the text. Thus, analyzing both words and grammatical structures in the target corpora was crucial, as it allowed us to describe the grammatical needs of the target students and determine the grammatical topics for teaching writing that would address these needs. To analyze the target corpora at the lexical and grammatical levels, we needed to identify the words and syntactic tags within the texts that exhibit high relative frequencies or significant statistical scores. Generally, keywords play a crucial role in characterizing the prevalent concepts in the texts, while syntactic tags highlight the relevant grammatical topics for teaching writing (Culpeper, 2009; Rayson, 2008). Both linguistic units served as indicators to achieve the purpose of this study.

4.3 Identifying Keywords, Dominant Text Dimensions, and Significant Syntactic Tags

The methodology outlined by Pojanapunya and Watson Todd (2018; 2021) for identifying keywords was employed in this study and also applied to identify key part-of-speech tags. To conduct these analyses, we compared the datasets reciprocally using log-likelihood (LL), a probability statistic that assigns scores to linguistic items based on their importance. This statistical approach was utilized to discern the general characteristics of linguistic features rather than focusing on specific attributes of each feature, as described by Pojanapunya and Watson Todd (2018) in their previous work on LL utilization. This step enabled us to identify linguistic items that were significantly more frequent in the target corpus compared to the benchmark corpus. After the data comparison, an extensive list of linguistic items emerged, so a cutoff point was needed. As mentioned by Palayon et al. (2022), three possible methods can be used for determining the cutoff point: (1) a cutoff LL value or its associated probability value (e.g., Esimaje, 2012), (2) Top N (e.g., Palayon et al., 2024; Pojanapunya & Watson Todd, 2018), and (3) a proportion of the range of LL values identified by a z-score or percentage (e.g., Palayon et al., 2022; Pojanapunya & Watson Todd, 2021). However, given that log-likelihood scores are influenced by corpus size (Pojanapunya, 2017) and the identification of the N value is highly subjective, which makes these methods problematic, a proportion of the range of LL values identified by a z-score was employed. This statistical measure provides a standardized approach to identifying the most important features within a dataset and interpreting them within the context of the overall distribution. A z-score threshold of 3 was applied for keywords, while a z-score of 1 was used for key part-of-speech tags based on the variation in the range of each feature (Palayon et al., 2022). These scores ensured the identification of sufficient items for data analysis.

To examine the characteristics of the essays, keywords were selected randomly to see how these words were used in the texts by analyzing their collocations, as specified by Mutual Information (MI) scores. This statistical

measure was appropriate for this analysis based on the sizes of the corpora. It allowed us to determine the degree of association or strength of the relationship between words within the target corpus. In this way, we could gain a clearer understanding of the linguistic sophistication in the texts, building on the initial analysis of the datasets. According to statistical principles, collocates with MI scores greater than 3 indicate strong associations with the main words, suggesting that they are significant features for interpretation. Therefore, we generally selected collocates with MI scores of 3 or higher and a high frequency for the analysis (see Baker, 2020; Kang, 2018; Palayon et al., 2024, for further use of MI scores).

After identifying the keywords, we examined the dominant text dimensions of the datasets using multidimensional analysis as per Biber (1988; 1989). Understanding the general characteristics of the datasets from the six text dimensions is crucial, as it validates the language level of each dataset. After the tagger of this method revealed the sets of dimension scores and syntactic items, we identified the score patterns, indicating the dominant text dimensions. This process helped us determine the prevalent syntactic features in each dataset based on their z-scores, which in turn allowed us to describe the students' grammatical needs and determine the grammatical topics for teaching writing. To strengthen these findings, we used the key part-of-speech tags measured by log-likelihood. These items confirmed the multidimensional analysis findings, particularly the syntactic features, and offered additional linguistic insights. Using two statistical measures allowed us to find more important grammatical features that may not be detected by a single statistical measure.

5. Results/Findings

The first set of results included the list of keywords of each dataset and the collocates of the selected keywords in each dataset. As reported in Table 2, the two lists of keywords demonstrated the differing language characteristics of the two datasets. The keywords in the refined essays displayed more linguistic sophistication when compared to the students' essays. This characteristic could

be found in the keywords such as *unwavering*, *crucial*, *superheroes*, *encouragement*, and *environment*. These words suggested that the refined essays used a wide variety of words including unfamiliar words, to convey complex ideas, illustrating the language level of intermediate or advanced texts. Additionally, the use of prepositions, articles, and nouns in the refined essays was frequent, which indicated that these texts tended to provide details specifying relationships between different elements in a sentence. On the other hand, the students' essays contained more basic and common words such as *good*, *think*, *is*, *be*, *have*, and *know*, which were fundamental to everyday communication but may be perceived as less sophisticated in certain contexts. This implied that the students' essays, while clear and functional, may lack the complexity and depth found in more advanced texts. The reliance on basic and common words suggested that the target students were still developing their linguistic range and may not yet be comfortable incorporating more varied language. As a result, the students' essays tended to be more straightforward and easier to understand, but they may not fully capture or express comprehensive or sophisticated ideas. These results highlighted the progression from basic language use to a more sophisticated level of writing, which was essential for advanced communication.

Table 2*Keywords in Each Dataset*

Rank	<i>Students' essays</i>		<i>Refined essays</i>	
	<i>Keywords</i>	<i>LL</i>	<i>Keywords</i>	<i>LL</i>
1	I	205.1	of	137.6
2	Very	190.3	despite	88.4
3	Because	169.7	journey	86.7
4	So	141.0	challenges	72.3
5	Like	140.6	into	70.3
6	Will	132.4	their	69.1
7	Good	112.7	unwavering	63.4
8	think	112.4	a	63.1
9	is	109.8	the	60.6
10	be	107.1	conclusion	59.4
11	have	103.3	however	59.0
12	do	99.9	our	58.4
13	don	80.4	an	58.1
14	school	74.3	as	57.6
15	study	74.0	beyond	57.5
16	teach	72.4	valuable	56.6
17	t	72.2	joy	55.1
18	teacher	72.1	providing	51.6
19	want	70.1	insights	48.6
20	friend	66.2	during	48.2
21	you	60.3	friendships	48.0
22	always	60.1	contribute	44.2
23	when	56.4	crucial	44.2
24	she	55.3	through	42.4
25	lot	52.3	guidance	41.3
26	give	51.0	while	39.9
27	can	49.8	diverse	39.8
28	know	49.1	currently	39.8
29	about	46.2	sense	39.5
30	many	46.2	individuals	38.3
31	happy	45.7	unique	38.3

<i>Rank</i>	<i>Students' essays</i>		<i>Refined essays</i>	
	<i>Keywords</i>	<i>LL</i>	<i>Keywords</i>	<i>LL</i>
32	me	45.5	dedicated	38.1
33	superhero	45.2	personal	37.1
34	finally	45.1	holds	36.8
35	are	44.3	shaping	36.8
36	m	43.0	superheroes	36.7
37	tell	42.9	growth	36.7
38	help	42.0	importance	36.7
39	they	40.0	experiences	36.1
40	autobiography	38.1	effective	35.8
41	things	37.8	lives	35.3
42	talk	37.6	encouragement	33.9
43	take	37.5	environment	33.7
44	people	36.4		

To gain further insight into the characteristics of the words in each list, the collocates of the selected keywords were examined, as they were words frequently appearing near a given keyword, offering additional context and understanding of its usage. As reported in Table 3, the pattern of pronoun collocates in the students' essays, such as *I* and *my*, revealed that these texts heavily focused on the students' personal information. However, the concordance lines showed that most sentence structures in the students' essays tended to be simple, primarily presenting their own thoughts and experiences without engaging them with broader or more complex concepts. These findings suggested that there was a gap in the grammatical knowledge of the students, which affected their ability to produce comprehensive compositions. In the refined essays, the pattern of pronoun collocates was frequent, such as *my* and *me*, characterizing the involved characteristic of texts. The way these features were used in the texts reflected a more reflective and sophisticated approach to writing because they were integrated into a broader narrative (e.g., *guiding me through life's challenges*). Additionally, the concordance lines displayed grammatical complexity and registered appropriateness, demonstrating the comprehensiveness of the texts.

From these results, while the characteristics observed in the students' essays highlighted areas for improvement such as vocabulary expansion, sentence complexity, and engagement with broader topics, the development of more sophisticated writing abilities in the refined essays indicated progress in narrative construction and argumentation. This entailed that addressing these areas of improvement could help students enhance their writing proficiency and effectively communicate their ideas in various contexts.

Table 3*Collocates of the Selected Keywords*

<i>Students' essays</i>			
<i>Collocates</i>	<i>Freq</i>	<i>MI</i>	<i>Concordance lines</i>
<u>because</u>			
i	99	3.5	<u>because my parents think I will be rich</u>
my	72	3.5	<u>because my superhero is my parents</u>
is	69	3.9	<u>because some of my wish is so hard</u>
to	59	3.2	<u>because we need to do more tasks</u>
<u>will</u>			
i	129	3.9	<u>I will do my best every day</u>
and	55	3.5	<u>I will always love and take care</u>
be	47	4.9	<u>I will be rich in the future.</u>
the	46	3.3	First, I <u>will talk about the purpose.</u>
<u>have</u>			
i	272	3.9	<u>I want to have money and healthy life</u>
to	198	3.9	I would like <u>to have</u> a band.
a	141	4.0	We <u>have to manage a</u> problem
friends	56	4.9	<u>I have a lot of friends.</u>
<u>friend</u>			
my	163	4.3	They good but I miss <u>my old friend.</u>
best	159	6.8	Patter is my <u>best friend.</u>
and	59	3.2	My <u>friend and</u> I cleaned the school

Students' essays

<i>Collocates</i>	<i>Freq</i>	<i>MI</i>	<i>Concordance lines</i>
have	52	4.1	But I <u>have a best friend</u> .
<u>when</u>			
i	195	4.6	<u>When</u> I was young, I want to have
and	43	3.2	I feel happy <u>and funny when</u> I talk
school	37	3.3	<u>When I was in high school</u> , I was in the
we	35	5.1	<u>When she comeback we</u> will go to

Refined essays

<i>Collocates</i>	<i>Freq</i>	<i>MI</i>	<i>Concordance lines</i>
<u>of</u>			
the	584	4.3	In <u>the course of</u> my life, ...
a	326	3.7	<u>a natural part of</u> childhood
and	252	3.2	<u>and offering words of</u> encouragement
my	229	3.1	finding acceptance <u>of my</u> identity
<u>unwavering</u>			
and	26	4.1	<u>and provide unwavering</u> support
support	25	7.9	They offer <u>unwavering support</u>
my	23	4.0	<u>unwavering support of my</u> parents
in	11	3.5	<u>in their unwavering</u> efforts
<u>valuable</u>			
and	26	4.1	<u>and valuable</u> knowledge
insights	12	8.0	I gained <u>valuable insights</u> into
lessons	10	7.8	offering <u>valuable lessons</u> that
for	10	4.3	<u>for teachers as valuable</u> mentors
<u>through</u>			
me	12	4.8	guiding <u>me through</u> life's challenges
of	11	3.4	<u>through the challenges of</u> living
with	7	4.3	<u>with you through</u> both the sad and...
life	7	5.0	making my journey <u>through life</u>

<i>Students' essays</i>			
<i>Collocates</i>	<i>Freq</i>	<i>MI</i>	<i>Concordance lines</i>
<u>while</u>			
the	18	3.2	<u>While I cherish the friends I have, ...</u>
i	18	3.8	<u>While I don't have many friends, ...</u>
of	12	3.2	<u>While their occasional lack of respect</u>
not	11	5.5	<u>While my family may not possess</u>

To verify the characteristics of the two datasets, the dominant text dimensions based on the patterns of dimension scores are presented in Table 4. Patterns emerged in the first, second, and fifth dimensions. In the first dimension, the positive score of the students' essays indicated an involved characteristic, whereas the negative score of the refined essays indicated an informational characteristic. This result aligned with Dimension 1 in Biber's (1988; 1989) framework, indicating that a low score reflected informationally dense texts (e.g., a high frequency of nouns and adjectives, as shown in Table 5), while a high score corresponded to affective and interactive texts (e.g., a greater use of pronouns and verbs, as shown in Table 5). In the second dimension, the negative score of the students' essays indicated a non-narrative characteristic, as reflected in features such as the high frequency of causative adverbial subordinators and amplifiers (see Table 5). In contrast, the positive score of the refined essays suggested a narrative characteristic, marked by frequent features such as the use of present and past participial clauses (see Table 5). In the fifth dimension, the negative score of the students' essays indicated that these texts provided information in a non-abstract way (see the tags and examples in Table 5), whereas the positive score of the refined essays indicated that these texts presented information in a technical, abstract, and formal manner. Based on Biber's (1988; 1989) framework, specifically in Dimension 5, a high score indicated that the texts contained a high frequency of clauses and conjuncts, as shown in Table 5.

In summary, the students' essays exhibited characteristics associated with involved discourse, non-narrative texts, and non-abstract information, whereas the

refined essays displayed characteristics associated with informational discourse, narrative texts, and abstract information. The findings implied that the refined essays were more aligned with academic writing conventions, whereas the students' essays tended to be more conversational and less structured, a style that was more typical of everyday communication. The differences in these dimensions highlighted the need for further instruction in academic writing skills to help students transition from a more informal, involved style to a more formal, abstract, and technically precise style of writing.

Table 4

Dimension Scores of Each Dataset

<i>Dimensions (D)</i>	<i>Students' essays (D scores)</i>	<i>Refined essays (D scores)</i>
<i>D1 Involved and informational discourse</i>	<i>12.78</i>	<i>-9.01</i>
<i>D2 Narrative and non-narrative concerns</i>	<i>-3.20</i>	<i>1.37</i>
<i>D3 Context-independent and context-dependent discourse</i>	<i>3.96</i>	<i>7.35</i>
<i>D4 Overt expression of persuasion</i>	<i>0.59</i>	<i>-3.21</i>
<i>D5 Abstract and non-abstract information</i>	<i>-1.69</i>	<i>1.55</i>
<i>D6 On-line informational elaboration</i>	<i>-1.02</i>	<i>-1.12</i>

To confirm syntactically the dominant text dimensions identified by the patterns of dimensions scores, the significant grammatical tags are presented in Table 5. The two sets of grammatical features, identified by log-likelihood and z-scores, helped us better understand the text dimensions of each dataset. The students' essays exhibited first person pronouns (e.g., *I* and *my*) and second person pronouns (e.g., *you* and *your*), indicating that these texts contained involved characteristics. This signified that the students were highly focused on their personal aspect, suggesting their limited knowledge of the topics. Additionally, only a few grammatical features were found in these essays, reflecting their low level of sophistication, which aligned with their non-narrative and non-abstract characteristics. These features were generally limited to causative adverbial subordinators, phrasal coordination, *that* relative clauses in the object position,

and *that* relative clauses in the subject position (see <https://tinyurl.com/samples-student-essays> for samples of the students' essays).

In the refined essays, the frequent use of sophisticated syntactic features contributed to the informational, narrative, and formal elements of the text, which confirmed the findings in the previous analyses. The presence of prepositional phrases, attributive adjectives, and determiners suggested that the texts were informational, indicating their broad characteristics. The presence of present participial clauses, past participial clauses, *that* relative clauses in the subject position, independent clause coordination, nominalizations, sentence relatives, conjuncts, and adverbial subordinators implied that the texts were narrative and formal, indicating their descriptive, factual, structured, and organized characteristics. These features indicated that these were relevant grammatical topics that needed to be taught to the target students to improve the quality of their compositions (see <https://tinyurl.com/samples-refined-essays> for samples of the refined essays).

Table 5

Grammatical Features in Each Dataset

Rank	Students' essays			
	Key part-of-speech tags	LL	Syntactic features	z
1	CAUS Causative adverbial subordinators (e.g., because my home is...)	171.0	CAUS Causative adverbial subordinators (e.g., because my home is...)	2.9
2	AMP Amplifiers (e.g., so bad)	159.8	PHC Phrasal coordination (e.g., happiness and safety)	1.9
3	FPP1 First person pronouns (e.g., I, me, my)	133.4	TOBJ That relative clauses on object position	1.2
4	BEMA Be as main verb (e.g., is, are, am)	115.7	TSUB That relative clauses on subject position	1.1
5	VPRT Present tense (e.g., talk, watch, finish)	112.4	AMP Amplifiers (e.g., so bad)	1.1

<i>Rank</i>	<i>Students' essays</i>			
	<i>Key part-of-speech tags</i>	<i>LL</i>	<i>Syntactic features</i>	<i>z</i>
6	PRMD Predictive modals (e.g., will, can)	105.1		
7	SPP2 Second person pronouns (e.g., you)	77.2		
8	PROD Pro-verb do (e.g., ...those who don't)	74.0		
9	QUAN Quantifiers (e.g., many, some)	72.2		
10	EMPH Emphatics (e.g., really, so)	57.4		
11	CONT Contractions (e.g., don't)	53.1		
12	TO Infinitives (e.g., to help)	51.6		
<i>Rank</i>	<i>Refined essays</i>			
	<i>Key part-of-speech tags</i>	<i>LL</i>	<i>Syntactic features</i>	<i>z</i>
1	NOMZ Nominalizations (e.g., encouragement)	390.8	PRESP Present participial clauses (e.g., Coming home feels like...)	4.8
2	PIN Total prepositional phrases (e.g., in the face of diverse challenges)	289.8	PHP Phrasal coordination (e.g., and)	3.8
3	PRESP Present participial clauses (e.g., Coming home feels like...)	217.6	PASTP Past participial clauses (e.g., characterized by cleanliness)	3.0
4	DT Determiners (e.g., a, the, this)	189.6	TSUB That relative clauses on subject position (e.g., The unwavering support *** my family provides)	2.6
5	JJ Attributive adjectives (e.g., fictional, unwavering, valuable, strict)	122.6	CONJ Conjuncts (e.g., while, because, nor)	1.7
6	CONJ Conjuncts (e.g., while, because, nor)	70.6	ANDC Independent clause coordination (e.g., ... and for	1.5

<i>Rank</i>	<i>Students' essays</i>			
	<i>Key part-of-speech tags</i>	<i>LL</i>	<i>Syntactic features</i>	<i>z</i>
7	ANDC Independent clause coordination (e.g., ... and for me, my superhero is my family)	67.8	me, my superhero is my family)	1.2
8			NOMZ Nominalizations (e.g., encouragement)	
9			JJ Attributive adjectives (e.g., fictional, unwavering, valuable, strict)	
10			OSUB Other adverbial subordinators (e.g., without them)	
			SERE Sentence relatives (e.g., ... which necessitates finding effective ways...)	1.0

(Note: The examples were taken from the essays available in the links provided.)

6. Discussion

The analyses revealed a set of significant grammatical features found in the refined essays, which were infrequent in the students' essays. These features were nominalizations, total prepositional phrases, present participial clauses, determiners, attributive adjectives, conjuncts, independent clause coordination, phrasal coordination, past participial clauses, *that* relative clauses in the subject position, adverbial subordinators, and sentence relatives. These features served as grammatical topics for teaching English writing to address the grammatical needs of the target students.

The frequent use of determiners, prepositional phrases, and attributive adjectives illustrated the informational characteristic of the refined essays by contributing to the clarity, precision, and density of these texts. Determiners (e.g., *the, a, some*) specify and define nouns, clarifying what is being referred to in a

sentence (Parisse & Maillart, 2019). In informational texts, such as academic writing or reports, precision and clarity are considered to be particularly important. These words help define the exact quantity or specificity of the subject (Davidse, 2004). The presence of determiners indicates a high level of explicitness and referential clarity, which is typical of informational discourse (Epstein, 2002; Gray, 2010). Prepositional phrases are composed of prepositions followed by noun phrases, which provide additional information about time, place, direction, and relationships between elements in a sentence (Schweikert, 2005). Informational texts are dense in detail, providing specific contextual or relational information, and prepositional phrases contribute to this by linking nouns and verbs to other parts of the sentence (Swarniti, 2021). These features pack more details into sentences and they are prevalent in academic writing, where providing precise and contextual information is essential. Attributive adjectives are modifying words placed before nouns to describe them, providing additional information about the nouns (Rind & Tillinghast, 2008). These words contribute to informational precision by adding descriptive detail (Hollmann, 2021). In informational texts, these features often convey descriptive qualities, and they are crucial in academic writing where describing qualities in a concise and informative way is key. The frequent use of determiners, prepositional phrases, and attributive adjectives ensures that the information conveyed is detailed and organized, which is typical of texts aiming to inform or explain complex topics in a clear and concise manner.

The frequent use of present participial clauses, independent clause coordination, phrasal coordination, past participial clauses, *that* relative clauses in the subject position, adverbial subordinators, and sentence relatives illustrates the narrative characteristic of essays. Present and past participial clauses add detail and dynamism to descriptions and actions. These clauses often show ongoing actions, creating vivid imagery, while past participial clauses emphasize completed actions or states. Both can enhance the flow of narrative events and descriptions (Biber et al., 1998; Sleeman, 2017). Independent clause coordination helps convey a sequence of events, which is central to narrative writing. This structure can show

cause and effect or actions happening simultaneously, driving the plot forward (Verstraete, 2005). Phrasal coordination allows for detailed descriptions, which enrich narrative settings and character development, giving more details to help the readers imagine the flow of the text (Kytö & Smitterberg, 2023). *That* relative clauses in the subject position provide additional information about subjects, allowing for deeper characterization and context within narrative, offering background or defining details important for narrating (Hoffmann, 2008; Mak et al., 2008; Sharvit, 1999). Adverbial subordinators (e.g., *because, although, since*) help explain reasons, causes, and conditions, providing context for actions or events in the narrative (Dryer, 2021). These subordinators often clarify why something happens, which is crucial for building a coherent storyline (Łęcki & Nykiel, 2017). Sentence relatives link sentences or ideas, creating coherence across a narrative (Caponigro et al., 2013; Patterson & Caponigro, 2016). They help maintain the flow of information and ensure the reader follows the connections between events or descriptions.

The frequent use of nominalizations, conjuncts, and past participial clauses illustrates the abstract characteristic of the refined essays. Nominalization is the process of converting action verbs or adjectives into nouns (Shibatani, 2019). Nominalizations often create a more impersonal and formal tone by avoiding conversational language and emphasizing the concepts or processes (Baratta, 2010). These words help in abstracting actions and qualities into concepts, making complex ideas easier to discuss at both a conceptual and technical level. Conjuncts are linking words used to connect clauses, sentences, or ideas within the texts (Nordquist, 2019). These words help in structuring abstract ideas by linking related concepts, making it easier to understand abstract relationships between different pieces of information (Povolná, 2016). Conjuncts contribute to a formal and technical tone by providing sophisticated ways to link ideas, which is essential for creating coherent and polished academic or professional writing. Past participial clauses are grammatical features that use the past participle of a verb and often describe a state or condition resulting from an action, allowing for abstract

representations of processes or results. These features provide detailed information about the outcome or state of a noun, focusing on concepts rather than actions. They help in presenting information in a structured and formal way, avoiding conversational language. These clauses describe specific conditions, resulting from technical processes, contributing to operational processes (Biber et al., 1998).

The 12 significant grammatical features offer teachers key insights into shaping a pedagogical framework for writing instruction. These features help students move from basic sentence structures to more complex, formal writing. For instance, nominalizations and conjuncts are essential for teaching academic writing, promoting clarity and formality, while participial clauses and independent clause coordination enable students to create more dynamic narratives. Additionally, features like determiners and attributive adjectives help students achieve greater precision in their descriptions, enhancing both narrative and academic styles. Integrating these elements into lessons allows teachers to guide their students in constructing coherent, grammatically sound, and stylistically rich sentences, which are crucial for effective communication in different contexts. However, these findings may not be broadly applicable, as they are based on a specific group of students. Every group of students, especially in Thailand, may have distinct grammatical needs. These findings would only be relevant to other groups if their linguistic backgrounds and learning requirements closely align with those of the students in this study.

7. Methodological Contributions

The integration of the three corpus-based methods (namely, keyword analysis, multidimensional analysis, and key part-of-speech analysis) makes significant contributions to both research and teaching. For research, integrating these methods provides systematic procedures that have not been widely practiced in previous corpus studies for analyzing the language characteristics of the target corpora. It allows language researchers to choose an appropriate

statistical measure and a cutoff method to produce the final lists of linguistic features. These methods establish a robust foundation for identifying linguistic patterns, such as frequency and distribution of grammatical foundation. In contrast to intuition-based methods, these methods offer quantitative evidence, ensuring that the findings produced are representative of actual language use. Keyword analysis and key part-of-speech analysis use an inductive approach to examine the target texts, while multidimensional analysis employs a deductive approach to examine the target texts. These methods triangulate the results, which increases the reliability of the findings. For teaching, especially in English writing instruction, the three methods may allow teachers to identify the grammatical needs of students from a corpus perspective, and this approach can contribute to the development of the traditional methods of conducting needs analysis (e.g., Pongsukvajchakul, 2023). These methods provide evidence-based insights into how students actually use the target language. Rather than relying on generalized teaching materials, teachers can use corpus findings to design relevant writing activities that focus on addressing the specific grammatical and lexical needs of their students. The sets of grammatical features associated with different genres (e.g., narrative texts) increase teachers' awareness of the need to tailor their instruction to the specific grammatical demands of each genre, preparing students to write effectively across a range of contexts.

8. Conclusion

This paper employed keyword analysis, multidimensional analysis, and key part-of-speech analysis to highlight the significant grammatical features in the target corpora. The essays in English of a specific group of Thai university students were collected and refined using ChatGPT, and the corpora produced (namely, students' essays and refined essays) served as the target corpora and benchmark corpora. The features in the students' essays describe the grammatical needs of the students in English writing, while the features in the refined essays serve as grammatical topics that can be used to teach writing and improve the target students' compositions. The students' essays contain features associated with

involved, non-narrative, and non-abstract characteristics, indicating a lack of sophistication in their writing and the need for further development of their grammatical skills. The refined essays contain features associated with informational, narrative, and abstract characteristics, indicating complexity as a quality typical of advanced-level compositions. In the refined essays, 12 grammatical features were found, which served as relevant grammatical topics for teaching the writing course. This paper sheds light on specific grammatical areas that need attention, providing insights for targeted instruction to improve students' writing skills. It also outlines steps for conducting grammatical needs analysis from a quantitative perspective, enhancing traditional methods of needs analysis. To maximize the methods used in this study, future research could further examine different types of essays (e.g., persuasive, expository, argumentative, etc.) to find linguistic patterns, characterizing each text type, that can be used to improve students' compositions.

9. About the Authors

Raymund T. Palayon is a lecturer at King Mongkut's Institute of Technology Ladkrabang, Chumphon Campus. He completed his Ph.D. in Applied Linguistics at King Mongkut's University of Technology Thonburi, Bangkok, Thailand. His research interests include corpus linguistics, discourse analysis, religious discourse, classroom discourse, and English language teaching and learning.

Yenyong Chongchit is a lecturer at King Mongkut's Institute of Technology Ladkrabang, Chumphon Campus. She completed her M.A. in English at Thammasat University, Bangkok, Thailand. Her research interests include English for specific purposes, English for careers, and ASEAN cultures.

10. References

Anthony, L. (2005). AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. *In Proceedings of the*

- IEEE International Professional Communication Conference (IPCC 2005)* (pp. 729–737). IEEE. <https://doi.org/10.1109/IPCC.2005.1494244>
- Baker, P. (2020). Corpus-assisted discourse analysis. In C. Hart (Ed.), *Researching discourse* (pp. 124–142). Routledge. <https://doi.org/10.4324/9780367815042>
- Baratta, A. M. (2010). Nominalization development across an undergraduate academic degree program. *Journal of Pragmatics*, 42(4), 1017–1036. <https://doi.org/10.1016/j.pragma.2009.08.007>
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1), 3–44. <https://doi.org/10.1515/ling.1989.27.1.3>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Boonyarattanasoontorn, P. (2017). An investigation of Thai students' English language writing difficulties and their use of writing strategies. *Journal of Advanced Research in Social Sciences and Humanities*, 2(2), 111–118. <https://dx.doi.org/10.26500/JARSSH-02-2017-0205>
- Caponigro, I., Torrence, H., & Cisneros, C. (2013). Free relative clauses in two Mixtec languages. *International Journal of American Linguistics*, 79(1), 61–96. <https://doi.org/10.1086/668608>
- Chompurach, W. (2021). “Please let me use Google Translate”: Thai EFL students' behavior and attitudes toward Google Translate use in English writing. *English Language Teaching*, 14(12), 23–35. <https://doi.org/10.5539/elt.v14n12p23>
- Chuenchaichon, Y. (2015). A review of EFL writing research studies in Thailand in the past 10 years. *Journal of Humanities Naresuan University*, 11(1), 13–30.
- Condon, W., & Kelly-Riley, D. (2004). Assessing and teaching what we value: The relationship between college-level writing and critical thinking abilities. *Assessing Writing*, 9(1), 56–75. <https://doi.org/10.1016/j.asw.2004.01.003>

- Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics*, 14(1), 29–59.
<https://doi.org/10.1075/ijcl.14.1.03cul>
- Davidse, K. (2004). The interaction of quantification and identification in English determiners. In M. Achard & S. Kemmer (Eds.), *Language, culture and mind* (pp. 507–533). CSLI Publications.
- Dryer, M. S. (2021). Adverbial subordinator prefixes. *Linguistic Discovery*, 17(1), 84–92.
- Epstein, R. (2002). The definite article, accessibility, and the construction of discourse referents. *Cognitive Linguistics*, 12(4), 333–378. <https://doi.org/10.1515/cogl.2002.007>
- Esimaje, A.U. (2012). A corpus-based lexical study of sermons in Nigeria. *English Language Teaching*, 5(9), 24–32. <https://doi.org/10.5539/elt.v5n9p24>
- Etfita, F. (2019). The correlation between students' grammar mastery and news writing ability. *Al-Ishlah: Jurnal Pendidikan*, 11(1), 58–70.
<https://doi.org/10.35445/alishlah.v11i1.113>
- Gray, B. (2010). On the use of demonstrative pronouns and determiners as cohesive devices: A focus on sentence-initial this/these in academic prose. *Journal of English for Academic Purposes*, 9(3), 167–183.
<https://doi.org/10.1016/j.jeap.2009.11.003>
- Hans, A., & Hans, E. (2017). Role of grammar in communication–writing skills. *International Journal of English Language, Literature and Humanities*, 5(1), 39–50.
- Hoffmann, T. (2008). English relative clauses and construction grammar: A topic which preposition placement can shed light on? In G. Trousdale & N. Gisborne (Eds.), *Constructional approaches to English grammar* (pp. 77–112). De Gruyter Mouton. <https://doi.org/10.1515/9783110199178.2.77>
- Hollmann, W. B. (2021). The 'nouniness' of attributive adjectives and 'verbiness' of predicative adjectives: Evidence from phonology. *English Language & Linguistics*, 25(2), 257–279. <https://doi.org/10.1017/S1360674320000015>

- Holmes, A. F., Zhang, S., & Harris, B. (2018). An analysis of teaching strategies designed to improve written communication skills. *Accounting Education, 28*(1), 25–48. <https://doi.org/10.1080/09639284.2018.1477055>
- Iamsiu, C. (2014). *An analysis of grammatical errors in Srinakharinwirot University students' writing* [Master's thesis, Srinakharinwirot University]. <https://ir.swu.ac.th/handle/123456789/3562>
- Inpanich, P., Chaya, P., Wongsittikan, N., & Evans, M. D. (2023). A comparison of grammatical errors made by Thai EFL undergraduate students in different types of essays. *Manusat Paritat: Journal of Humanities, 45*(2), 165–191. <https://so06.tci-thaijo.org/index.php/mpjh/article/view/268155>
- Iyer, P. (2013). Effects of collaborative blogging on communicative skills in writing of Thai university EFL students. *English for Specific Purposes World, 14*(39), 1–14. https://utr.spb.ru/ESP-World/Articles_39/39_Purnima.pdf
- Kampookaew, P. (2020). An analysis of grammatical errors made by Thai EFL university students in an EAP writing class: Issues and recommendations. *rEFLECTIONS, 27*(2), 246–273. <https://doi.org/10.61508/refl.v27i2.248862>
- Kang, B. -m. (2018). Collocation and word association: Comparing collocation measuring methods. *International Journal of Corpus Linguistics, 23*(1), 85–113. <https://doi.org/10.1075/ijcl.15116.kan>
- Kareviati, E., Yugafiati, R., & Resmini, S. (2020). Need analysis of English Department students as the basic of material development of writing for general communication. *EL TIN Journal, 8*(1), 1–8. <https://www.e-journal.stkipsiliwangi.ac.id/index.php/eltin/article/view/1307>
- Klimova, B. F. (2012). The importance of writing. *Paripex-Indian Journal of Research, 2*(1), 9–11. https://www.researchgate.net/profile/Blanka-Klimova/publication/274925223_The_Importance_of_Writing/links/58b71c9992851c471d47a5f0/The-Importance-of-Writing.pdf

- Kytö, M., & Smitterberg, E. (2023). Clausal and phrasal coordination in recent American English. *Corpus Linguistics and Linguistic Theory*, 19(1), 23–46. <https://doi.org/10.1515/cllt-2022-0035>
- Łęcki, A. & Nykiel, J. (2017). Grammaticalization of the English adverbial subordinator in order that. In D. Olmen, H. Cuyckens & L. Ghesquière (Eds.), *Aspects of grammaticalization: (Inter)Subjectification and directionality* (pp. 237–256). De Gruyter Mouton. <https://doi.org/10.1515/9783110492347-009>
- Lhorsumeth, M. P. (2017). *A survey study of attitudes towards English grammar learning of Thai private university students* [Master's thesis, Thammasat University]. https://ethesisarchive.library.tu.ac.th/thesis/2017/TU_2017_5921040332_9125_7297.pdf
- Mak, W. M., Vonk, W., & Schriefers, H. (2008). Discourse structure and relative clause processing. *Memory & Cognition*, 36(1), 170–181. <https://doi.org/10.3758/MC.36.1.170>
- McEnery, T. (2019). *Corpus linguistics*. Edinburgh University Press. <https://web.stanford.edu/~eckert/PDF/McEneryWilson.pdf>
- Nguyen, T. T. L., and Suwannabubpha, S. (2021). EFL writing at Thai secondary schools: Teachers and students' views, difficulties and expectations. *Language Related Research*, 12(3), 187–214.
- Nini, A. (2019). The multi-dimensional analysis tagger. In T. Berber Sardinha & M. V. Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 67-94). Bloomsbury. <https://andreanini.com/wp-content/uploads/2019/06/pre-print-the-multidimensional-analysis-tagger.pdf>
- Nordquist, R. (2019, January 21). The definition of 'conjunct' in grammar. *ThoughtCo*. <https://www.thoughtco.com/what-is-conjunct-grammar-1689910>
- Palayon, R. T., Todd, R. W., & Vungthong, S. (2022). From the temple of life to the temple of death: Keyness analyses of the transitions of a

- cult. *Corpora*, 17(3), 331–361.
<https://www.eupublishing.com/doi/10.3366/cor.2022.0262>
- Palayon, R. T., Todd, R. W., & Vungthong, S. (2024). Language characteristics illustrating mass suicide in the destructive cult sermons. *Humanities, Arts and Social Sciences Studies*, 24(3), 796–814.
<https://doi.org/10.69598/hasss.24.3.268800>
- Parisse, C., & Maillart, C. (2019). Determiners. In *The SAGE Encyclopedia of Human Communication Sciences and Disorders* (pp. 578–579). SAGE Publications. <https://doi.org/10.4135/9781483380810.n188>
- Patterson, G., & Caponigro, I. (2016). The puzzling degraded status of *who* free relative clauses in English. *English Language & Linguistics*, 20(2), 341–352.
<https://doi.org/10.1017/S1360674315000325>
- Pojanapunya, P. (2017). A theory of keywords [Doctoral dissertation, King Mongkut's University of Technology Thonburi].
- Pojanapunya, P., & Watson Todd, R. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1), 133–167.
<https://doi.org/10.1515/cllt-2015-0030>
- Pojanapunya, P., & Watson Todd, R. (2021). The influence of the benchmark corpus on keyword analysis. *Register Studies*, 3(1), 88–114.
<https://doi.org/10.1075/rs.19017.poj>
- Pongsukvajchakul, P. (2023). Analysis of Thai EFL university students' needs in learning English paragraph writing. *Journal of Management Sciences*, 2(2), 44–59. <https://kuojs.lib.ku.ac.th/index.php/jmsku/article/view/5681/2711>
- Povolná, R. (2016). A cross cultural analysis of conjuncts as indicators of the interaction and negotiation of meaning in research articles. *Topics in Linguistics*, 17(1), 45–63. <https://doi.org/10.1515/topling-2016-0004>
- Promsupa, P., Varasarin, P., & Brudhiprabha, P. (2017). An analysis of grammatical errors in English writing of Thai university students. *HRD Journal*, 8(1), 93–104.
<https://ojs.lib.buu.ac.th/index.php/hrd/article/view/5587>

- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519–549.
<https://doi.org/10.1075/ijcl.13.4.06ray>
- Rayson, P. (2019). Corpus analysis of key words. In C. Chapelle (Ed.), *The concise encyclopedia of applied linguistics* (pp. 320–326).
<https://core.ac.uk/download/pdf/287592924.pdf>
- Rayson, P., & Potts, A. (2021). Analysing keyword lists. In M. Paquot & S. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 119–139). Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_6
- Reppen, R., & Simpson-Vlach, R. (2019). Corpus linguistics. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 91–108). Hodder Education.
https://pbadoktoral.uin-suka.ac.id/media/dokumen_akademik/130113_20210611_norbert_schmitt_ed-_an_introduction_to_appliedb-ok-org.pdf#page=100
- Rind, M., & Tillinghast, L. (2008). What is an attributive adjective? *Philosophy*, 83(1), 77–88.
<https://doi.org/10.1017/S0031819108000314>
- Rodsawang, S. S. (2017). Writing problems of EFL learners in higher education: A case study of the Far Eastern University. *FEU Academic Review*, 11(1), 268–284. <https://so01.tci-thaijo.org/index.php/FEU/article/view/77107/65454>
- Sacal, P. G. Y., & Potane, J. D. (2023). Students' mastery of English grammar towards effective writing and speaking competence. *International Journal of Multidisciplinary: Applied Business and Education Research*, 4(8), 2894–2904. <http://dx.doi.org/10.11594/ijmaber04.08.27>
- Saengboon, S. (2017). English grammar and Thai university students: An insurmountable linguistic battle? *English Language Teaching*, 10(11), 22–36. <http://doi.org/10.5539/elt.v10n11p22>
- Sahib, T. M., Alyasiri, O. M., Younis, H. A., Akhtom, D., Hayder, I. M., Salisu, S., & Besse, D. M. (2023). A comparison between ChatGPT-3.5 and ChatGPT-4.0 as a tool for paraphrasing English paragraphs. In *Proceedings of the*

- International Applied Social Sciences (C-IASOS-2023) Congress* (pp. 471–480). <https://portal.arid.my/Publications/6faa5026-cc63-445e-8ef7-ce90c89e0bf4.pdf>
- Sarma, S. K., Bharali, H., Gogoi, A., Deka, R., & Barman, A. (2012). A structured approach for building Assamese corpus: Insights, applications, and challenges. In *Proceedings of the 10th Workshop on Asian Language Resources* (pp. 21–28). <https://aclanthology.org/W12-5203.pdf>
- Schweikert, W. (2005). The order of prepositional phrases. In W. Schweikert (Ed.), *Linguistik Aktuell/Linguistics Today* (pp. 195–216). John Benjamins. <https://doi.org/10.1075/la.83>
- Seensangworn, P., & Chaya W. (2017). Writing problems and writing strategies of English major and non-English major students in a Thai university. *Manusat Paritat: Journal of Humanities*, 39(1), 113–136. <https://ejournals.swu.ac.th/index.php/hm/article/view/9327>
- Sharvit, Y. (1999). Functional relative clauses. *Linguistics and Philosophy*, 22, 447–478. <https://doi.org/10.1023/A:1005411720444>
- Shibatani, M. (2019). What is nominalization? Towards the theoretical foundations of nominalization. In R. Zariquiey, M. Shibatani, & D. W. Fleck (Eds.), *Nominalization in languages of the Americas* (pp. 15–167). John Benjamins. https://www.academia.edu/40264880/What_is_nominalization_Towards_the_theoretical_foundations_of_nominalization
- Sleeman, P. (2017). Participial relative clauses. In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics* (pp. 1–30). Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.185>
- Swarniti, N. W. (2021). A corpus based approach to the analysis of structures in prepositional phrase. *Yavana Bhasha: Journal of English Language Education*, 4(1), 18–22. <https://ojs.uhnsugriwa.ac.id/index.php/YB/article/view/649>
- Thuong, T. T. (2020). The integration of process writing and genre-based approach in enhancing Thai students' narrative writing ability at university

level [Master's thesis, Mahasarakham University].

<http://202.28.34.124/dspace/bitstream/123456789/1268/1/59010181012.pdf>

Verstraete, J. C. (2005). Two types of coordination in clause combining. *Lingua*, *115*(4), 611–626.

<https://doi.org/10.1016/j.lingua.2003.09.018>

Wahyuni, S., & Ilyas, M. (2016). An investigation of Thai students' English language problems and their learning strategies at English study program of Teacher Training and Education Faculty of Riau Islamic University. *J-SHMIC: Journal of English for Academic*, *3*(2), 81–91.

[https://doi.org/10.25299/jshmic.2016.vol3\(2\).530](https://doi.org/10.25299/jshmic.2016.vol3(2).530)

Watcharapunyawong, S., & Usaha, S. (2013). Thai EFL students' writing errors in different text types: The interference of the first language. *English Language Teaching*, *6*(1), 67–78. <http://dx.doi.org/10.5539/elt.v6n1p67>

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122–1136.

<https://doi.org/10.1109/JAS.2023.123618>