

A Corpus-Based Lexical Profiling of English for Tourism Coursebooks: Pedagogical Register and Implications for ESP Materials Development

Anchalee Veerachaisantikul¹, Wara Chansin²

¹English Department, Rajamangala University of Technology Isan, Thailand

²Humanities Department, Naresuan University, Phitsanulok, Thailand

Corresponding author: anchalee.ve@rmuti.ac.th

Received 29 July 2026 | Received in revised form 06 August 2026 | Accepted 10 August 2026

APA Citation:

Veerachaisantikul, A. & Chansin, W. (2026). A Corpus-Based Lexical Profiling of English for Tourism Coursebooks: Pedagogical Register and Implications for ESP Materials Development. *Journal of English Language Teaching and Linguistics*, 11(2), 2026, 341-356. doi: <http://dx.doi.org/10.21462/jeltl.v11i2.2232>

Abstract

The present study investigates the lexical features and pedagogical discourse of English for Tourism (EFT) textbooks widely used in Thai higher education. A specialized corpus of 15 textbooks (1,873 digitized pages) was constructed based on a nationwide examination of tourism and hospitality programs in four geographical regions of Thailand during the 2020–2021 academic year. WordSmith Tools were used to analyze the corpus on the basis of the course specifications of the Thai Qualifications Framework (TQF). The quantitative analysis revealed 181,303 tokens and 9,613-word types with an overall Type-Token Ratio (TTR) of 5.30%. The distributional analysis of the lexical categories showed that POS were distributed unevenly. The most dominant lexical category was the common and proper nouns (41.00%), offering a strong focus on tourism entities, professional roles, and operational concepts. The category of adjectives only represented 1.28% of the corpus, of which descriptive and evaluative adjectives were particularly scarce. The results point to the fact that the selected EFT textbooks mainly dealt with transactional and operational language, with limited exposure to persuasive and evaluative language that are crucial for the promotion of a destination and hospitality marketing. The study offers corpus-based evidence to improve ESP curricula, to evaluate the content of EFT textbooks, and to develop more balanced instructional materials for higher education settings.

Keywords: Corpus Linguistics, Coursebook Evaluation, English for Tourism, Lexical Profiling, Part-of-Speech Distribution

1. Introduction

International tourism has been growing fast, and this has strengthened English as the most dominant lingua franca in the global hospitality and tourism industry. Thus, English for Tourism (EFT) has become an important sub-field of English for Specific Purposes (ESP) aimed at preparing students for actual communication in professional tourism settings. In many countries where English is not the first language, such as Thailand, EFT courses are widely available at the higher education level, with textbooks being the main source to improve learners' communicative competence (Nguyen & Boonsuk, 2023; Tomlinson, 2012; Wong, 2021). Within this framework, pedagogical materials serve as the primary source of input, directly influencing learners' vocabulary development and lexical accessibility (Nation, 2013). The quality and the authenticity of the language used in these resources are therefore essential for preparing students for real communication. Traditionally, the linguistic components of ESP textbooks have been analyzed based on expert judgement, evaluation checklists, and teachers' perceptions (McGrath, 2016). These are approaches that are useful pedagogically but rarely provide solid evidence of the language found in instructional materials. Corpus linguistics is a more systematic and empirical approach to the study of real language patterns in large electronic corpora (Flowerdew, 2015; Römer, 2019). Corpus-based approaches allow us to assess the authenticity of educational texts and differentiate their lexical features, including frequency, keyword, and part-of-speech (POS) profiling, as well as lexical diversity measures (Biber et al., 2016; Friginal et al., 2021).

Traditionally, the linguistic components of ESP textbooks have been analyzed based on expert judgement, evaluation checklists, and teachers' perceptions (McGrath, 2016). While these approaches are useful pedagogically, they rarely provide solid empirical evidence of the actual language found in instructional materials. The development of corpus linguistics has been made possible by the advent of massive electronic corpora, which have enabled more systematic and empirical studies of real language use (Flowerdew, 2015; Römer, 2019). Corpus-based techniques provide academics with a comprehensive framework for investigating the linguistic features of instructional texts and evaluating their representativeness in connection to real-world communication. Corpus analysis permits the investigation of lexico-grammatical aspects like frequency patterns, keywords, part-of-speech (POS) distributions and measures of lexical variety (Biber et al., 2016; Friginal et al., 2021). This methodological shift toward evidence-based.

Although corpus linguistics has been increasingly applied to ESP research, few studies have investigated the lexical features of English for Tourism coursebooks, especially in Thai higher education. Most previous work has focused on vocabulary frequency, thematic content, or readability. There are no comprehensive studies of lexical distributions of grammatical categories. Of particular interest is the contribution of different word classes to the pedagogical register of tourism textbooks, as well as whether these materials provide a balanced representation of the transactional and promotional language that is commonly used in the tourism industry. Therefore, the empirical evidence about the lexical composition of EFT textbooks is scarce. To fill this gap, the present study creates a corpus-based lexical profile of fifteen English for Tourism coursebooks widely used in Thai higher education. This paper examines the distribution of the major lexical categories by means of part-of-speech analysis based on the specialized corpus. The specialized corpus was developed from the textbooks selected by the Thai Qualifications Framework (TQF) course specifications. The findings provide empirical evidence for the evaluation of the linguistic features of EFT coursebooks and practical implications for ESP materials development, curriculum design and English language teaching in tourism education.

To address these methodological and empirical gaps, the present study constructs a specialized corpus of fifteen English for Tourism (EFT) coursebooks widely used in Thai higher education. Using part-of-speech (POS) profiling, the study systematically examines the distribution of major lexical categories to characterize the pedagogical register represented in these instructional materials. The corpus was compiled from textbooks identified through official Thai Qualifications Framework (TQF) course specifications. By providing a corpus-informed account of the lexical and grammatical composition of EFT coursebooks, the findings contribute empirical evidence for ESP materials evaluation, curriculum development, and the design of instructional resources that better reflect authentic tourism communication.

2. Literature Review

2.1 Corpus Linguistics in Textbook Evaluation and ESP

The use of corpus linguistics in the evaluation of ESP materials has grown increasingly important in recent decades. Traditional textbook development has often relied on writers' intuition, which may create a gap between the language taught in educational institutions and the language used in professional settings (Misnawati et al., 2024). The creation of specialized digital corpora and computational text analysis has provided researchers with the opportunity to investigate linguistic patterns in specialized domains in an objective and empirical manner (Baker, 2018). Corpus-based approaches have moved beyond lexical frequency to examine the use of language in relation to communicative contexts and professional practices. Lexical patterns and structural characteristics of texts have been quantified by means of computational measures such as Type-Token Ratio (TTR) and Part-of-Speech (POS) distribution (Hunston, 2010). Such analyses provide empirical evidence that can be useful for the design of ESP materials and the evaluation of textbooks in the choice of linguistic input which more closely reflects authentic professional communication (Flowerdew, 2016). Accordingly, corpus-informed approaches have become useful tools to cross the gap between language teaching materials and evidence of actual language use and pedagogical needs (Tono, 2020).

The relevance of corpus-based approaches for English for Tourism (EFT) textbooks is particularly pertinent due to the communicative functions that are involved in tourism communication, ranging from transactional service encounters to persuasive promotional discourse. Previous research has shown that tourism discourse varies from predominantly informative to persuasive, interpersonal registers (Wahyuningsih & Mahsar, 2024). Corpus profiling enables researchers to find out if teaching materials achieve a suitable balance between these two communicative functions. For example, empirical findings may show that a textbook overemphasizes mechanical transactional patterns, offering little exposure to descriptive and evaluative language. Distributional analyses of word classes can therefore be used as diagnostic tools in identifying systematic strengths and lexical gaps in ESP curricula (Römer, 2011).

2.2 Lexical Profiling and Pedagogical Register in Tourism Discourse

Lexical profiling refers to the systematic analysis of the vocabulary structure of a text by examining the concentration and distribution of different grammatical word classes. In linguistic theory, words are commonly classified into two broad categories: content words (e.g., nouns, verbs, and adjectives) and function words (e.g., pronouns, prepositions, and auxiliary verbs) (Biber et al., 1999). The proportional distribution of these word classes contributes to the communicative register of a text by reflecting its dominant communicative purposes. A higher proportion of common nouns is generally associated with informational and transactional registers that emphasize the efficient transmission of factual information. In contrast, a greater use of descriptive adjectives

and interactional markers is typically associated with interpersonal or persuasive registers intended to attract audiences, establish rapport, and influence consumer perceptions.

Within tourism discourse, different communicative contexts require distinct lexical resources. Front-line service encounters, such as hotel reservations and airline ticketing, primarily rely on standardized transactional expressions and explicit informational language characterized by a relatively high frequency of nouns and other content words (Saragih, 2024). In contrast, destination marketing and tourism promotion predominantly employ evaluative and descriptive language, where carefully selected adjectives play an important role in constructing attractive destination images and influencing tourists' perceptions (Sun, 2022). Consequently, effective English for Tourism (EFT) coursebooks should provide learners with balanced exposure to both transactional and promotional registers. Lexical profiling offers an objective means of evaluating whether instructional materials achieve this balance by examining the distribution of different word classes. When substantial lexical imbalances exist—for example, an overrepresentation of informational nouns accompanied by limited descriptive and evaluative language—learners may be well prepared for routine service interactions but receive fewer opportunities to develop the persuasive communication skills required in destination marketing and tourism promotion.

2.3 Corpus-Based Studies on English for Tourism and ESP Materials

Corpus-based research has played an increasing role in the evaluation and development of English for Specific Purposes (ESP) materials, providing empirical evidence of the use of specialized language in a variety of professional contexts. In previous studies, it has been shown that corpus-based approaches can reveal vocabulary patterns, lexical frequency and grammatical features that may not be easily identified by traditional textbook evaluation methods (Hyland, 2022; Khojasteh & Mukundan, 2025). Studies in domains such as Business English, Academic English, and professional communication, in particular, have shown that corpus-informed analyses can help researchers evaluate the extent to which instructional materials adequately reflect the language demands of specific communities of practice.

The research on ESP corpora has focused mainly on the identification of frequent vocabulary, specialized expressions and linguistic features of the genre. The studies suggest that teaching materials should be developed on the basis of actual language use, rather than the intuition of textbook writers or assumptions about general English (Lacková, 2025; Laosirattanachai & Ruangjaroon, 2021). In addition, corpus analysis can provide insight into the distribution of lexical items and discourse patterns to determine whether learning materials expose students to the linguistic resources they will be required to have in their future professional contexts. These findings have led to a shift from experience-based textbook development to evidence-based materials design (Misnawati et al., 2025).

Studies on tourism discourse have shown that the industry is characterized by complex communicative features, where language works both transactionally and promotionally. Tourism experts need to handle everyday service interactions and convey attractive visuals, experiences and values to potential clients. Consequently, tourism literature research has examined the impact of linguistic choices on persuasion, evaluation and the way destinations are represented. However, much of the existing research has focused on authentic tourism categories such as brochures, websites, advertisements or promotional materials rather than educational materials designed for tourism students. Despite the application of corpus-based approaches in various ESP domains, there is a glaring absence of studies on English for Tourism (EFT) textbooks, particularly in higher education settings. Previous studies on EFT resources have concentrated on the issues of general vocabulary selection, thematic subjects or pedagogical assessment, with only a few exceptions of

comprehensive analysis of the overall lexical features of different textbooks (Wahyuningsih & Mahsar, 2024). The distribution of the various word classes as an indicator of the balance between informational and persuasive registers in EFT instructional resources has been largely neglected. Accordingly, further corpus-based studies are needed to determine the extent to which EFT textbooks provide learners with a comprehensive exposure to the range of linguistic demands of the tourism industry. The analysis of the lexical profiles of the different textbooks can shed light on the empirical aspects of the pedagogical focus of current resources and help to improve the formulation of ESP curricula and the creation of textbooks.

3. Research Methods

3.1 Data Sources and Sampling Criteria

This study explored the lexical features of English for Tourism (EFT) textbooks used in Thai higher education by using a corpus-based descriptive approach. The corpus comprised 15 purposively selected commercially available EFT textbooks. The selection criteria were the use of these resources as teaching materials for tourism and hospitality programs found in the Thai Qualifications Framework (TQF) course specifications for the academic year 2020-2021.

Textbooks were selected from tourism and hospitality programs in public and private universities in four geographical regions of Thailand: Northern, Central, Northeastern and Southern regions. Textbooks from different institutional contexts were included in order to provide greater representation of EFT materials currently being used in Thai higher education. The selected textbooks were digitized and assembled into an EFT specialized corpus of 1,873 pages. This corpus was the primary source of data for a corpus-based analysis of lexical patterns and vocabulary distributions in EFT teaching materials.

3.2 Analytical Tools and Reference Frameworks

The main computational tools used for the processing of corpora and lexical analysis were WordSmith Tools. The application was used to generate frequency word lists, to estimate the size of the corpus in terms of running words (tokens) and lexical diversity (types), and to calculate the Type-Token Ratio (TTR) of the collected EFT corpus. The metrics provided quantitative insights into the lexical features and distributional patterns of the textbooks.

In order to standardize the lexical identification and categorization, the study employed several well-known lexicographical reference materials, that is, the Oxford Advanced Learner's Dictionary (10th edition), the Oxford Dictionary of Travel and Tourism, the Collins COBUILD Bank of English and West's General Service List (GSL). These reference frameworks were utilized in the examination procedure for validation of the word forms, headwords, lexical categories and grammatical classifications. A systematic investigation of vocabulary trends in EFT teaching materials was possible through the use of corpus-driven computational analysis coupled with authoritative lexical references.

3.3 Data Processing and Cleaning

Before computational examination the collected textbooks were subjected to a methodical digitization and corpus creation process. The hard copies of the textbooks were converted to machine-readable formats using Optical Character Recognition (OCR). The retrieved texts were first stored in Microsoft Word (.docx) files to allow for manual validation, proofreading and correction of OCR errors. This quality assurance procedure was employed in order to ensure that the digitized texts were faithful representations of the original textbook material and to reduce the possibility of potential inaccuracies in subsequent corpus analysis. After the validation process, the

texts underwent a systematic purification procedure in order to make sure that only relevant linguistic material was left in the corpus. Non-linguistic and non-analyzable elements have been removed, e.g. table labels, image captions, ornamental text, page numbers, copyright statements and other formatting related features. Moreover, to ensure the validity and reliability of the specialized EFT corpus, components that did not reflect the intended linguistic input of the research were omitted. After cleaning, the created text files were converted to plain text (.txt) files with UTF-8 encoding and a consistent format. The files were then imported into WordSmith Tools. The preparation process provided consistency in computational processing, which allowed for a reliable investigation of the lexical frequency, vocabulary distribution and lexical patterns in the selected EFT textbooks.

3.4 Analytical Procedures

To ensure systematic, transparent and replicable procedures for corpus construction and lexical profiling, the implementation of the corpus analysis was carried out through three sequential operational phases, as shown in Figure 1. The three stages involved (1) corpus building and data collection, (2) digitization, normalization and corpus preparation, and (3) lexical profiling and quantitative analysis. These phases were designed to maintain consistency throughout the research process and to ensure that the collected textbook data were accurately prepared before computational analysis.

The overall procedure implemented in this study is presented in Figure 1. The first step is the selection and collection of EFT textbooks in accordance with the Thai Qualifications Framework (TQF). The next step is the preparation of the digitized corpus data, and the last step is the computational analysis of lexical features using corpus-based approaches. The procedure involves a systematic extension from data collection to linguistic interpretation to provide a clear methodological framework for investigating lexical patterns of EFT coursebooks.

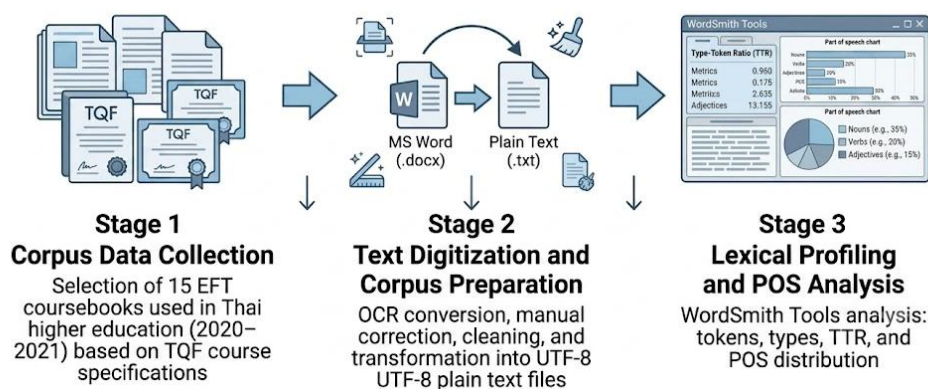


Figure 1. Three-Stage Procedure for Corpus Construction and Lexical Profiling of EFT Coursebooks

Stage 1: Corpus Construction and Data Collection

In order to construct the corpus, the course specifications of the Thai Qualifications Framework (TQF) were examined to identify the eligible tertiary-level tourism and hospitality programs in four geographical regions of Thailand. Based on the defined selection criteria, 15 EFT textbooks used in the academic year 2020–2021 were selected. The collected materials were compiled into a specialized digital corpus comprising 1,873 pages.

Stage 2: Digitization, Normalization and Corpus Preparation

The texts from the selected textbooks were digitized, cleaned and normalized before computational processing. The OCR-produced texts were manually checked for recognition errors and textual accuracy. The cleaned documents were exported as Microsoft Word (.docx) files and then converted to plain text (.txt) files with UTF-8 encoding. The WordList function of WordSmith Tools was used to generate preliminary corpus statistics such as token and type counts.

Stage 3: Lexical profiling and quantitative analysis

The total number of running words (tokens: 181,303), different lexical items (types: 9,613) and the overall Type-Token Ratio (TTR: 5.30%) were calculated to determine the overall lexical profile of the final corpus. In addition, the Part-of-Speech (POS) distribution was analyzed to determine the relative frequency of major lexical categories such as common nouns (41.00%), auxiliary verbs (4.85%), pronouns (4.09%) and adjectives (1.28%). These quantitative measures were subsequently used to describe the pedagogical register represented in EFT textbooks.

3.4 Data Analysis Procedures

A corpus-based approach was used to analyze the corpus data to investigate the lexical input of English for Tourism (EFT) coursebooks at the tertiary level. The analysis included corpus preparation, lexical frequency analysis, keyword analysis and contextual analysis. The texts of the chosen EFT coursebooks were transcribed into electronic form and compiled and processed into a specialized corpus. In the process of corpus preparation, irrelevant elements such as page numbers, headings, publisher information, exercise instructions and duplicated texts were discarded to make the data accurate and consistent. The finalized corpus comprised 15 English for Tourism (EFT) coursebooks containing 181,303 running words (tokens). After compiling the corpus, a lexical frequency analysis was performed with WordSmith Tools (Scott, 2024). Word frequency lists were generated using the WordList function, and the most frequent lexical items in the corpus were identified. Frequency data were analyzed to identify common vocabulary items and patterns of lexical distribution across the selected EFT coursebooks, thus providing an overview of the lexical characteristics of the compiled corpus. In order to identify vocabulary items especially characteristic of the EFT corpus, a keyword analysis was carried out using the Keywords function in WordSmith Tools. Keyness values were obtained through the comparison of the specialized EFT corpus with a reference corpus. Keywords were then selected and classified according to their relevance to the contexts of tourism communication and professional contexts and according to their statistical significance in the corpus. Concordance and collocation analyses were also performed to examine the contextual use of frequent words and keywords. The concordance lines were examined via the Concordance function, and recurrent word combinations and lexical patterns were identified through the Clusters function. These analyses provided a more nuanced understanding of how vocabulary related to tourism is presented and used in tertiary EFT coursebooks. Finally, the findings of the corpus analysis were discussed from the view of English for Specific Purposes (ESP) to evaluate the adequacy and appropriateness of the lexical input in EFT coursebooks and to provide implications for tourism English teaching, materials development and vocabulary selection for ESP learners.

4. Results

This section presents the results of the corpus-based analysis of the English for Tourism (EFT) coursebook corpus. The results are presented in four sections. Section 4.1 describes the structural features of the compiled corpus, and Section 4.2 describes its quantitative lexical profile. Sections 4.3 and 4.4 then illustrate the distribution of major word classes and the specific lexical items that are found in the corpus.

4.1 Structural Characteristics of Tourism Corpus (TC)

The compiled Tourism Corpus (TC) contains 15 selected English for Tourism (EFT) coursebooks from Thai higher education institutions. The corpus comprises 1,873 pages and has a digital size of 3,423,973 bytes. Table 1 shows the structural characteristics of the compiled coursebooks, such as the number of pages, the size of the digital file, and the running words of each volume.

Table 1: Structural Characteristics of the Tourism Corpus (TC)

Coursebook	Pages	File Size	Running Word
1	114	130,841	14,200
2	138	301,763	28,400
3	135	275,546	25,100
4	142	338,596	52,395
5	124	163,489	15,800
6	110	130,060	13,900
7	140	338,434	31,200
8	135	261,020	24,800
9	85	98,409	9,125
10	125	178,333	16,500
11	130	259,334	23,900
12	120	227,370	20,100
13	105	124,974	12,800
14	130	258,770	23,700
15	140	337,034	30,800
Total	1,873	3,423,973	181,303

The compiled corpus is highly diverse in terms of textbook length and lexical coverage. Individual course books range in length from 85 to 142 pages and contain between 9,125 and 52,395 running words. Course book 4 is the largest volume in the corpus, whereas course book 9 is the smallest. Overall, the corpus is composed of 181,303 running words distributed across 15 course books. This variation suggests that the selected course books differ significantly in terms of the amount of lexical input offered to learners. The variety of such structures emphasizes the value of a corpus-based analysis to investigate the lexical characteristics of English for Tourism teaching materials and to evaluate how well the lexical content is reflected in the commercially available course books.

4.2 Quantitative Lexical Profile of the Tourism Corpus

The quantitative characteristics of the Tourism Corpus were established using WordSmith Tools. Basic corpus statistics, including the number of running words (tokens), unique word forms (types), and the type-token ratio (TTR), provide an overview of the lexical composition of the compiled corpus. These measures serve as the quantitative baseline for the subsequent lexical analyses.

Table 2: Descriptive Statistics of the Tourism Corpus (TC)

Corpus Metric	Value
Running Words (Tokens)	181,303
Unique Word Forms (Types)	9,613
Type-Token Ratio (TTR)	5.30%

The Tourism Corpus contains 181,303 running words (tokens) and 9,613 different word-forms (types), with an overall type-token ratio (TTR) of 5.30%. The relatively low TTR signals a high degree of repetition of lexical items in the corpus, which suggests that the selected coursebooks cover a relatively stable set of core tourism-related vocabulary. Such repetition is a common feature of ESP instructional materials, in which key domain-specific vocabulary is systematically introduced and consolidated across several units and learning contexts. These descriptive statistics provide a quantitative overview of the corpus and represent a baseline for the further analyses of word classes and specialized tourism vocabulary presented in the following sections.

4.3 Part-of-Speech (POS) Distribution of the Tourism Corpus

To further examine the lexical characteristics of the Tourism Corpus, the distribution of major word classes was analyzed using the POS tagging function in WordSmith Tools. The analysis focused on three broad lexical categories: content words, function words, and non-word artefacts. Table 3 summarizes the overall distribution of these categories within the compiled corpus.

Table 3: Macro-Level Distribution of Word Classes in the Tourism Corpus

Major Class	Constituent Categories	Frequency	Share (%)
Content Words	Nouns, Verbs, Adjectives, Adverbs	102,817	56.72
Function Words	Determiners, Prepositions, Auxiliaries, Pronouns	69,290	38.22
Non-Words / Artefacts	Punctuation, Digitization Symbols	9,196	5.07
Total		181,303	100.00

As shown in Table 3, content words make up the highest percentage of the corpus (56.72%), followed by function words (38.22%). Non-word artefacts account for 5.07% of the corpus after the data processing. The high frequency of content words demonstrates that the lexical items related to tourism are a significant part of the selected EFT coursebooks. This distribution reflects the lexical features of EFT materials where there is a need for a significant number of content words in relation to specialized concepts, professional activities, and tourism-related processes. The relatively high percentage of content words also provides evidence that the selected coursebooks are rich in lexical input that is related to the tourism domain, while the high percentage of function words supports the grammatical organization and communicative structure of the texts.

4.3.1 Key Part-of-Speech Categories

Table 4: Key Part-of-Speech (POS) Distribution in the Tourism Corpus

Part-of-Speech Category	Representative Examples	Frequency (N)	Percentage (%)
Common/ Proper Nouns	hotel, guest, service, flight	74,335	41.00%
Auxiliary Verbs	is, are, can, will, would	8,787	4.85%
Pronouns	you, I, your, we	7,411	4.09%
Qualifying Adjectives	scenic, comfortable, deluxe, available	2,312	1.28%

Table 4 revealed the most frequent grammatical category in the corpus is common and proper nouns, which comprise 41.00% of all tokens (n=74,335). This recommends that the EFT coursebooks contain a large number of lexical items that refer to entities, roles and concepts related to tourism. The next most frequent categories are auxiliary verbs (4.85%) and pronouns (4.09%), which are smaller but important categories in terms of interactional and service-oriented language. A further interesting finding is the relatively low proportion of adjectives, which are only 1.28% (n=2,312) of the corpus. This distribution advocates that adjectival resources are less prominent than noun-based lexical resources in the selected EFT coursebooks.

4.3.2 Functional Distribution of Adjectives in the Tourism Corpus

In order to explore the adjectival patterns in the Tourism Corpus in more depth, the adjectives were categorized according to their functional meanings. The analysis was conducted on two main categories: functional/operational adjectives and descriptive/evaluative adjectives. Table 4.1 shows the frequency and proportional distribution of these adjectival sub-categories.

Table 4.1: Distribution of Adjectival Sub-Categories in the Tourism Corpus

Adjectival Category	Representative Tokens	Occurrence (n)	Percentage within Adjectives (%)	Share of Total Corpus (%)
Functional / Operational	available, single, double, local, total, main, direct, daily, international	1,891	81.79%	1.04%
Descriptive / Evaluative	comfortable, scenic, deluxe, traditional, famous, delicious, beautiful, breathtaking	421	18.21%	0.23%
Total Adjectives		2,312	100%	

Table 4.1 displayed the distribution of adjectives in the Tourism Corpus. Functional/operational adjectives are the most frequent category, representing 81.79% (n=1,891) of all adjective tokens in the Tourism Corpus. They are mainly related to practical descriptions, service specifications and operational information such as room types, availability and procedural conditions. Descriptive/evaluative adjectives represent 18.21% (n=421) of adjective tokens and only 0.23% of the total corpus. The distribution of adjectives in the Tourism Corpus proposes that the use of adjectives in the selected EFT coursebooks is more frequently associated with functional descriptions than with evaluative or promotional expressions.

4.4 Specialized Lexicon: High-Frequency Domain Analysis

The POS analysis was followed by an analysis of high-frequency specialized lexical items in the Tourism Corpus to establish the dominant lexical domains represented in the selected EFT coursebooks. The analysis was based on the most frequent tourism-related lexical items and their functional categories. The top 20 specialized lexical items identified in the corpus are presented in Table 5.

Table 5: Top 20 Specialized Lexical Items in the Tourism Corpus

Rank	Specialized Word	Frequency (n)	POS Category	Functional Category
1	HOTEL	738	Noun	Institutional Entity
2	GUEST	728	Noun	Primary Role / Actor
3	TRAVEL	518	Noun	Industry Domain
4	TOURISM	514	Noun	Industry Domain
5	ROOM	486	Noun	Facility / Amenity
6	QUESTIONS	411	Noun/Verb	Transactional Task
7	SERVICES	329	Noun	Operational Output
8	EMPLOYEE	313	Noun	Primary Role / Actor
9	TOUR	276	Noun	Product / Package
10	TIME	270	Noun	Operational Parameter
11	TRANSPORT	260	Noun	Operational Sector
12	CUSTOMER	258	Noun	Primary Role / Actor
13	BUSINESS	256	Noun	Commercial Domain
14	AIRPORT	251	Noun	Facility / Infrastructure
15	ATTRACTION	227	Noun	Promotional Target
16	BOOKING	215	Noun	Operational Task
17	INFORMATION	208	Noun	Communication Content
18	FLIGHT	194	Noun	Service Sector
19	RESERVATION	186	Noun	Operational Task
20	CHECK-IN	179	Noun	Operational Task

The lexicon of the Tourism Corpus is mainly composed of nouns which present tourism-related entities, service providers, amenities and operational activities as presented in Table 5. The identified lexical units are grouped into three major functional groups. Initially, roles and institutions are illustrated by frequently occurring terms such as GUEST, CUSTOMER, EMPLOYEE, HOTEL and AIRPORT, highlighting important participants and environments in tourism and hospitality. Second, TRAVEL, TOURISM, TOUR, FLIGHT and SERVICES are terms for offerings and amenities that are basic elements of the tourism industry. Third, procedural operations are composed of such items as BOOKING, CHECK-IN, INFORMATION, RESERVATION, which advises typical actions and communication modes in the transactions of tourism service.

5. Discussion

This section discusses the implications of the findings for academics and pedagogy. The findings are discussed in the context of English for Specific Purposes (ESP) materials design and communicative competence, exploring the ways in which commercial English for Tourism (EFT) coursebooks present tourism-related language input in Thai higher education settings.

The findings reveal a strongly nominal and operational orientation in the English for Tourism (EFT) coursebooks, with common and proper nouns constituting 41.00% of the corpus, whereas adjectives account for only 1.28%; more strikingly, descriptive/evaluative adjectives represent merely 0.23% of the total corpus. This lexical configuration indicates that the materials prioritize entities, occupational roles, facilities, services, and procedural activities—such as hotel, guest, travel, tourism, room, booking, and reservation—over the evaluative and persuasive resources required for destination description and tourism promotion. The finding converges with Laosirattanachai and Ruangjaroon (2021), whose large specialized corpora of tourism, hotel, and airline discourse demonstrated the importance of domain-specific and technical vocabulary for ESP learners, supporting the present study's observation that specialized lexical items constitute a central

component of tourism English. At the same time, the present results extend their work by showing that the pedagogical issue is not merely whether specialized vocabulary is represented, but which communicative functions that vocabulary enables: extensive operational vocabulary does not necessarily provide adequate resources for promotional and evaluative discourse. This interpretation is consistent with Hyland's (2022) conception of ESP, which emphasizes that instruction should respond to the particular discourse practices learners will encounter in their future academic or occupational communities. It also accords with Biber and Conrad's (2019) register framework, according to which recurrent linguistic features need to be interpreted functionally in relation to the situational purposes of a register rather than treated merely as decontextualized frequency counts. From this theoretical perspective, the predominance of nouns and operational adjectives can be interpreted as a linguistic realization of the coursebooks' informational and transactional orientation, while the scarcity of descriptive/evaluative adjectives signals restricted representation of the persuasive register needed for destination marketing. This imbalance is particularly significant because professional tourism communication encompasses multiple communicative purposes; consequently, a coursebook may possess substantial specialized lexical coverage while still exhibiting inadequate register diversity. The findings therefore refine the notion of lexical adequacy in ESP materials by suggesting that adequacy should be evaluated not only through vocabulary quantity and domain specificity but also through the correspondence between lexical distributions, communicative functions, and the target professional register.

The findings also provide empirical support for a corpus-informed approach to materials evaluation, while exposing the limitations of evaluations based primarily on pedagogical judgment. Traditional coursebook evaluation, as represented by McGrath (2016), recognizes materials evaluation and design as central components of curriculum development and emphasizes teachers' informed judgment in coursebook selection and adaptation. The present study does not contradict this approach; rather, it demonstrates how corpus evidence can complement it by revealing distributional imbalances that may be difficult to detect through impressionistic inspection alone. This interpretation is consistent with Tomlinson (2012), who situates evaluation, adaptation, production, and exploitation of materials within an interconnected materials-development process and stresses the continuing need for research-based development of language-learning materials. More specifically, the present evidence strongly supports the conclusion of Khojasteh & Mukundan's (2025) systematic review of 29 corpus-based textbook studies, which found a methodological movement from basic frequency analysis toward increasingly sophisticated lexical, semantic, and syntactic analyses and concluded that corpus-informed evaluation can improve the accuracy and relevance of textbook content. The current findings illustrate precisely why such empirical evaluation matters: although the corpus contains 181,303 tokens and 9,613 types, its TTR of 5.30% indicates substantial lexical repetition, suggesting systematic recycling of a relatively stable domain-specific vocabulary. Such repetition may be pedagogically beneficial because repeated exposure can consolidate essential terminology; however, repetition becomes potentially restrictive when it repeatedly privileges the same transactional register while providing limited exposure to alternative communicative functions. Thus, the results partly support the pedagogical value of controlled lexical input while simultaneously challenging any assumption that frequent recycling alone constitutes sufficient lexical preparation for professional communication. Taken together, the findings, McGrath's materials-evaluation framework, Tomlinson's materials-development perspective, Biber and Conrad's register theory, Hyland's needs-oriented conception of ESP, and contemporary corpus-based textbook research point toward an integrated principle: effective ESP materials should be evaluated simultaneously for lexical coverage, authenticity, functional diversity, and correspondence with target professional discourse.

5.1 Implications for theory

Revisiting the representation of ESP registers These findings contribute to the body of knowledge concerning the presentation of ESP registers in commercially available instructional materials. The ESP materials should ideally mirror the communicative needs and features of a particular professional field. However, the EFT coursebooks that have been studied in this research focus primarily on the operational and transactional dimensions of tourism communication. The high frequency of nouns related to tourism entities, services and operational procedures suggests that the selected materials expose the learner to a considerable amount of core domain-specific vocabulary. However, the lack of descriptive and evaluative lexical resources means that other important aspects of tourism discourse are under-represented, especially the language of experience description, destination promotion and the construction of persuasive meanings.

These findings highlight the importance of register diversity in the design of ESP materials. Good EFT coursebooks should offer a balance of communicative functions, including transactional language, as well as descriptive, evaluative and promotional discourse typical of professional contexts in tourism. Such a balance will better prepare learners to face the varied linguistic demands of the tourism industry.

5.1.1 Lexical Diversity and Controlled Lexical Input (TTR)

The Type-Token Ratio (TTR) of 5.30% shows a relatively high level of lexical repetition in the Tourism Corpus. Repetition of domain-specific vocabulary in ESP materials can support vocabulary acquisition by allowing learners to encounter important lexical items in different units and contexts of learning. The recurrence of the core tourism-related vocabulary prefers the selection of EFT coursebooks is aimed at the reiteration of key terminology, rather than wide lexical variation.

There is also a considerable variation in length of coursebooks (from about 9,125 to 52,395 running words) which indicates variation in the amount of lexical input that commercially available EFT materials provide. This variation indicates that coursebook developers might have different attitudes towards vocabulary coverage and lexical diversity in teaching tourism English. The results imply that corpus-based lexical profiling can be a helpful guideline for vocabulary balance and coverage in EFT materials. The profiling provides empirical evidence to support understanding of how the choice of vocabulary, the repetition of lexical items and the coverage of the language of the domain are structured within the tertiary level EFT textbooks and to facilitate more informed ESP materials evaluation and development.

5.1.2 The Primacy of Nominal-Operational Language and the Restriction of Evaluative Resources

This study has a significant theoretical implication for the differential distribution of lexical categories in the Tourism Corpus. The high percentage of common and proper nouns (41.00%) indicates that the coursebooks chosen for EFT are mainly focused on the entities, professions and operational concepts of tourism. Common lexical items such as hotel, airport, room, guest and customer display that the materials contain a lot of input related to participants, facilities and service activities in tourism contexts. This pattern is further supported by auxiliary verbs (4.85%) and pronouns (4.09%) which are used in building up the interactional and transactional language needed in tourism service encounters, such as inquiries, reservations and customer communication.

However, the relatively small percentage of adjectives (1.28% of the total tokens), especially descriptive and evaluative adjectives (0.23% of the total corpus) recommend that the promotional and persuasive dimensions of tourism discourse are relatively underrepresented. Adjectives of operation such as available, single, double and total are mainly used for practical information about

the service. The adjectives of experience, quality, destination appeal and emotional evaluation are less frequent in the corpus. From an ESP perspective, the results revealed that the selected EFT coursebooks provide considerable linguistic support for transactional and operational communication, but limited chances to develop skills in persuasive and evaluative discourse, which are often required in tourism promotion, destination marketing, and experience-oriented communication. EFT materials should, hence, cater for a wider range of communicative functions, so that the linguistic diversity of the tourism industry is better reflected in the materials.

5.2 Limitations and Recommendations for Future Research

This study is an empirical investigation of the lexical features of English for Tourism (EFT) textbooks used in Thai tertiary education, but there are some limitations to be noted. First, the corpus was collected from 15 commercial EFT textbooks used in the academic year 2020–2021. The data includes materials used in a range of higher education contexts, but not every EFT textbook or other instructional material on tourism that is available. Secondly, the present study mainly dealt with lexical frequency, lexical distribution and part-of-speech patterns. Further linguistic features such as phraseological structures, multi-word expressions, collocations and contextual language use could be explored in future research. Further studies may also compare EFT textbook language with authentic workplace corpora, such as hospitality interactions, travel blogs and tourism promotional materials, to investigate the extent to which instructional resources reflect real-world tourism communication. Moreover, from the learner's point of view, future research may be directed to considering the effects of increased exposure to descriptive, evaluative and promotional language on students' communicative competence and their ability to use English effectively in a professional tourism context.

6. Conclusion

This study established a corpus-informed lexical profile of English for Tourism (EFT) coursebooks used in Thai higher education. A specialized corpus consisting of 181,303 tokens from 15 commercial coursebooks was analyzed to identify patterns of lexical diversity, part-of-speech distribution, and the functional orientation of tourism-related language input.

The results exhibit three main patterns, including a relatively low Type-Token Ratio (TTR) of 5.30%, which favors a high degree of lexical repetition. This hints that the chosen coursebooks concentrate on the repetition of the critical tourism vocabulary in order to strengthen it. However, differences in coursebook size and lexical volume indicate variation in the amount of language input provided by commercially available EFT materials. Secondly, the analysis revealed a massive dominance of transactional and operational language. The high percentage of nouns (41.00%) and the high frequency of lexical items relating to tourism and referring to entities, roles, facilities and service activities indicate that the selected materials are heavily oriented towards practical communication in the tourism context. Third, descriptive and evaluative language resources were found to occur less frequently. Although adjectives accounted for 1.28% of total corpus tokens, descriptive and evaluative adjectives represented only 0.23% of the corpus. This pattern suggests that the analyzed EFT coursebooks provide stronger representation of operational communication than persuasive and promotional discourse commonly associated with tourism marketing and destination description.

Overall, this study points to the importance of balancing transactional language and wider communicative functions in developing EFT materials. The coursebooks analyzed provide considerable lexical input for basic tourism communication, but the use of more descriptive,

evaluative and promotional language would add to the linguistic resources of the learners, thus better preparing them for different professional contexts in the tourism industry.

References

- Baker, P. (2018). *Language, linguistics and the real world: Discourse and corpus analysis*. Routledge.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education.
- Biber, D., Reppen, R., & Friginal, E. (2016). Research in English for Specific Purposes. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed., pp. 166–180). Routledge.
- Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press. <https://doi.org/10.1075/rs.20019.lyn>
- Flowerdew, J. (2016). English for Specific Academic Purposes (ESAP) and corpus studies. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed., pp. 445–460). Routledge.
- Flowerdew, L. (2015). Corpus-based research and pedagogy in English for academic purposes: From theory to practice. *Journal of English for Academic Purposes*, 20, 158–168. <https://doi.org/10.1016/j.jeap.2015.09.001>
- Friginal, E., Mathews, E., & Roberts, J. (2021). *Corpus-based approaches to teaching and learning English for specific purposes*. Routledge.
- Hyland, K. (2022). English for Specific Purposes: What is it and where is it taking us? *ESP Today*, 10(2), 202–220. <https://doi.org/10.18485/esptoday.2022.10.2.1>
- Khojasteh, L., & Mukundan, J. (2025). A systematic review of corpus-based methodologies in textbook analyses and evaluation. *Language Teaching Research Quarterly*, 47, 175–195. <https://doi.org/10.32038/ltrq.2025.47.10>
- Lacková, M. (2025). Teaching ESP Vocabulary with the Application of CorpusBased Materials at Universities: A Rapid Review of Literature. *Arab World English Journal*, 16(4), 273–288. <https://dx.doi.org/10.24093/awej/vol16no4.16>
- Laosirattanachai, P. & Ruangjaroon, S. (2021). Corpus-based creation of tourism, hotel, and airline business word lists. *LEARN Journal: Language Education and Acquisition Research Network*, 14(1), 50–86.
- McGrath, I. (2016). *Materials evaluation and design for language teaching* (2nd ed.). Edinburgh University Press
- Misnawati, M., Nur, S., & Tahir, S. Z. (2024). Corpus Linguistics Today: A Qualitative Approach. *Research and Innovation in Applied Linguistics (RIAL)*, 21, 45–62. <https://doi.org/10.31963/rial.v2i1.4486>
- Misnawati, M., Iskandar, I., & Tahir, S. Z. bin. (2025). Corpus Linguistics in English for Specific Purposes (ESP): A Systematic Literature Review. *GLENS: Global English Insights Journal*, 3(1), 47–67. <https://doi.org/10.66314/glens.v3i1.317>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
- Nguyen, T. T., & Boonsuk, Y. (2023). English as a lingua franca in Thai higher education: Pedagogy, materials, and communicative competence in tourism contexts. *Asian EFL Journal*, 25(3), 45–68.
- Römer, U. (2011). Corpus research in language teaching and textbook evaluation: State of the art and future perspectives. *Corpus Linguistics and Linguistic Theory*, 7(2), 205–225. <https://doi.org/10.1515/cllt.2011.010>

- Römer, U. (2019). Corpora and language teaching. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 477–499). Cambridge University Press. <https://doi.org/10.1017/9781108333603.021>
- Saragih, H. S. (2024). Genuine small talk in hospitality services encounter. *Consumer Behavior in Tourism and Hospitality*, 19(2), 237–252. <https://doi.org/10.1108/CBTH-09-2023-0164>
- Tono, Y. (2020). Corpus-based approaches to language teaching and learning. *Language Teaching Research*, 24(4), 475–492.
- Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching*, 45(2), 143–179. <https://doi.org/10.1017/S0261444811000528>
- Wahyuningsih, S., & Mahsar, L. (2024). Developing specialized vocabulary lists for English tourism: A corpus-based approach. *Journal of Language and Linguistic Studies*, 20(3), 115–130. DOI:10.36312/jolls.v4i4.2321
- Wong, C. C. (2021). Evaluating ESP coursebooks for tourism and hospitality studies: A register analysis perspective. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 28, Article 100301. <https://doi.org/10.1016/j.jhlste.2021.100301>