



Developing Specialized Vocabulary Lists for English Tourism: A Corpus-Based Approach

¹Sri Wahyuningsih, ²Lalu Mahsar

¹English Lecturer, Diploma Program in Hospitality, STP Mataram, Mataram, Indonesia

²English Lecturer, Diploma Program in Tourism Business, STP Mataram, Mataram, Indonesia

*Corresponding Author e-mail: sriwahyuningsih04011976@gmail.com

Received: October 2024; Revised: November 2024; Published: December 2024

Abstract

The increasing dominance of English as the global lingua franca in tourism underscores the necessity for specialized vocabulary tailored to this sector. This study addresses a critical gap in English for Tourism (EfT) resources by developing specialized vocabulary lists using a corpus-based approach. Traditional EfT resources often fail to meet the nuanced, context-specific demands of real-world tourism communication, creating challenges for both learners and professionals. To bridge this gap, a representative corpus of tourism-related texts was compiled, and corpus analysis was conducted using three key techniques: frequency analysis, keyness analysis, and collocation analysis. These methods enabled the identification of key vocabulary and phrases relevant to various subfields within tourism, including hospitality, cultural heritage, and travel services. The identified vocabulary was subsequently categorized into thematic areas to ensure its practical applicability for tourism professionals. The study also evaluated the effectiveness of these vocabulary lists in educational contexts through pre-tests and post-tests administered to students and instructors in EfT programs. The results demonstrated a significant improvement in vocabulary proficiency, with an average 25% increase in students' performance following the intervention. These findings contribute valuable, empirically grounded resources for EfT educators, offering a practical means to enhance language proficiency and professional communication. This study highlights the importance of using authentic, context-specific materials in language education, ensuring that learners acquire the necessary skills for effective communication in tourism-related settings.

Keywords: English for Tourism; Vocabulary proficiency; Corpus-based approach; Collocation analysis

How to Cite: Wahyuningsih, S., & Mahsar, L. (2024). Developing Specialized Vocabulary Lists for English Tourism: A Corpus-Based Approach, *Journal of Language and Literature Studies*, 4(4), 908-920. doi: <https://doi.org/10.36312/jolls.v4i4.2321>



<https://doi.org/10.36312/jolls.v4i4.2321>

Copyright © 2024, Wahyuningsih et al

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) License.



INTRODUCTION

The global dominance of English as the lingua franca of international tourism has underscored the critical need for specialized vocabulary tailored to the specific demands of the tourism industry. Effective communication in this context requires a nuanced understanding of industry-specific terminology, which varies across different subfields, such as hospitality, cultural heritage, and travel services (Valencia & Pertiwidar, 2024). However, there is a significant gap in the availability of specialized vocabulary resources for English for Tourism (EfT), which complicates language teaching and learning in this sector. This study aims to address this gap by developing specialized vocabulary lists for EfT using a corpus-based approach, thereby enhancing the relevance and authenticity of the vocabulary taught to tourism professionals.

The primary challenge in teaching English to tourism professionals is the absence of specialized resources that accurately reflect the language used in real-world tourism

contexts. Traditional EfT materials, such as textbooks and vocabulary lists, are often generic and fail to address the specific language needs of professionals in tourism (Astawa & Wijaya, 2024; Waad & Al-Bahrani, 2020). These resources frequently overlook the complexity and specificity of terms that are essential for effective communication in various tourism settings, such as accommodation, travel services, and cultural exchanges (Lee et al., 2023; Lojo & Timothy, 2020). Existing vocabulary lists are often based on subjective judgments or outdated materials, rather than being grounded in empirical data that reflects current language use (Kholidi et al., 2022; Saragih et al., 2022). As a result, tourism professionals, especially those who are non-native speakers of English, may struggle with essential terminology, leading to misunderstandings and a decline in the quality of service (Waad & Al-Bahrani, 2020).

Moreover, research highlights the growing recognition among tourism educators and learners that current EfT courses do not sufficiently address the language demands of the tourism industry (XU, 2020). This misalignment between the vocabulary taught in EfT programs and the actual needs of the field calls for more targeted and data-driven approaches to vocabulary selection. The use of corpora in language education offers a solution by providing authentic insights into language use, based on real-world data, rather than relying on traditional, top-down approaches to vocabulary selection (Hass, 2022). Corpus-based studies in other fields, such as English for Academic Purposes (EAP), have successfully identified key vocabulary that is essential for academic success (Xodabande et al., 2022), yet similar research in EfT remains limited.

Although some studies have explored the potential of corpus linguistics to enhance vocabulary instruction in tourism, they have often been constrained by small datasets, limited scope, or a focus on general language features rather than specific vocabulary for subfields such as hospitality, cultural heritage, and eco-tourism (Ivanova et al., 2024). This gap underscores the need for more comprehensive research that focuses on the distinct vocabulary needs across different areas within tourism. Furthermore, the majority of existing studies have not fully leveraged advanced corpus analysis tools, such as keyness and collocation analysis, to identify the most relevant and frequently used terms within the tourism industry. This study, therefore, offers a novel approach by utilizing a diverse and representative corpus of tourism-related texts and applying advanced corpus-based techniques to identify and categorize vocabulary that is directly relevant to various subfields of tourism.

The primary objective of this study is to develop specialized vocabulary lists for EfT using a corpus-based approach that accurately reflects authentic language use in tourism. Specifically, this research seeks to: 1) Compile and analyze a comprehensive corpus of tourism-related texts to identify key vocabulary and phrases that are relevant to different subfields within tourism, such as hospitality, cultural heritage, and travel services, 2) Develop thematic vocabulary lists that cater to the specific language needs of tourism professionals, ensuring that the lists are contextually appropriate and reflective of real-world language use, 3) Evaluate the effectiveness of these vocabulary lists in enhancing learners' language proficiency and professional communication skills within tourism settings.

This study aims to bridge the gap between the language taught in existing EfT programs and the actual language demands of tourism professionals. By doing so, it seeks to contribute to more effective language teaching and learning in this field, ensuring that tourism professionals are equipped with the necessary vocabulary for successful communication in diverse tourism contexts. The central research questions guiding this study are: 1) What key vocabulary and phrases are most relevant to different subfields within tourism, based on a corpus of tourism-related texts? 2) How can advanced corpus-based techniques, such as keyness and collocation analysis, be applied to develop

specialized vocabulary lists for EfT? 3) To what extent do the vocabulary lists developed through this corpus-based approach improve learners' language proficiency and communication skills in real-world tourism contexts?

Literature Review

Vocabulary acquisition is a foundational aspect of language learning, particularly in English for Specific Purposes (ESP) contexts such as English for Tourism (EfT). A robust vocabulary not only enables learners to comprehend specialized texts but also enhances their ability to engage in professional tasks, interact with clients, and deliver services effectively within their respective fields (Zahrani & Chaudhary, 2022; Zahrani & Chaudhary, 2022). In the context of tourism, communication frequently involves nuanced terminology that is specific to diverse areas such as culture, hospitality, travel services, and customer interactions (Saragih et al., 2022). Therefore, for tourism professionals, possessing an extensive and contextually relevant vocabulary is essential for facilitating clear, effective, and professional communication.

The Role of Corpus-Based Approaches in Vocabulary Development

A corpus-based approach to vocabulary development offers a systematic and data-driven method for identifying and prioritizing vocabulary based on its actual usage in authentic texts. Corpora, large collections of written or spoken texts, provide insights into how words are used in context, making them an invaluable resource for language learners. Through the analysis of word frequency, collocations, and contextual usage, corpus linguistics allows educators and researchers to identify which terms are most relevant for learners in specific fields (Faisal & Fharieza, 2021). In the case of EfT, corpus analysis helps identify language specific to tourism-related contexts such as travel brochures, customer service dialogues, and destination descriptions (Sun, 2022). The empirical basis of corpus-based methods provides a deeper understanding of language patterns, enabling educators to offer learners more targeted vocabulary instruction (Sun & Dang, 2020).

In tourism, where communication often hinges on understanding context-specific terms related to hospitality, culture, and customer service, a corpus-based approach enables the creation of more tailored and authentic vocabulary lists. This approach also addresses a significant gap in traditional EfT resources, which often provide generic vocabulary that may not fully reflect the linguistic nuances encountered in professional tourism settings.

Benefits of Corpus-Based Vocabulary Lists in ESP

Research in other fields of ESP, particularly in English for Academic Purposes (EAP), has demonstrated the efficacy of corpus-based methods for identifying key vocabulary essential for academic success. The *Academic Vocabulary List* Ivanova et al., (2024) is a prominent example of how corpus-based research can generate vocabulary lists that reflect the language demands of academic contexts. Similarly, in the field of Medical English, corpus-based studies have identified specific vocabulary that is crucial for effective communication within healthcare settings, helping professionals to engage in medical consultations, documentation, and research (Özer & Akbaş, 2024). These fields have established the value of corpus-based vocabulary development in specialized domains, with researchers advocating for the use of corpus analysis tools to generate specialized word lists that improve learners' language proficiency and professional communication.

The application of corpus-based methods to EfT, however, remains underexplored. Existing research in this domain has largely concentrated on general language features, such as grammar and phraseology, rather than focusing on the specialized vocabulary required for various tourism-related contexts (Xodabande et al., 2022). As a result, there

is a clear gap in the literature regarding the use of corpus linguistics to develop comprehensive, context-specific vocabulary lists for tourism. This gap highlights the need for further research that employs corpus linguistics to identify vocabulary relevant to diverse subfields within tourism, such as hospitality, cultural heritage, and eco-tourism.

Recent Studies and Application of Corpus-Based Methods in EfT

The recent study by Wang and Wang (2021) represents a notable effort to apply a corpus-based approach to EfT. The researchers created a specialized tourism corpus to identify key vocabulary and phrases, which were subsequently integrated into an English for Tourism curriculum. Their findings suggested that students who were taught using corpus-derived vocabulary lists demonstrated significant improvements in their ability to understand and use specialized terminology compared to those who were taught using traditional vocabulary lists. This study underscores the value of corpus analysis in enhancing language instruction, as it allows for a more targeted approach to vocabulary development based on real-world language usage. However, despite these promising results, research on the use of corpus-based methods in EfT is still relatively scarce. Much of the existing literature on EfT has focused on theoretical frameworks or broader studies of English as a lingua franca (ELF), with limited application to specific linguistic needs within the tourism industry (Sun & Dang, 2020). Moreover, there is a lack of studies that employ advanced corpus techniques, such as collocation analysis, to explore how vocabulary is used in context within the tourism sector. By not fully exploiting the potential of corpus linguistics, EfT research may be overlooking the opportunity to generate vocabulary lists that more accurately reflect the authentic language used in tourism-related communication.

Comparative Insights from EAP and Medical English

Drawing on comparative studies from other ESP fields, such as English for Academic Purposes (EAP) and Medical English, provides valuable insights into how specialized vocabulary has been effectively developed and utilized. In EAP, corpus-based approaches have been instrumental in identifying high-frequency academic vocabulary, as seen in the development of the *Academic Word List* (Ivanova et al., 2024). Similarly, in Medical English, corpus linguistics has been used to create vocabulary lists tailored to healthcare professionals, highlighting how specific word choices can influence communication in clinical settings (Heidari & Jamalzadeh, 2020). These studies demonstrate that corpus-based vocabulary lists are particularly effective in specialized domains because they are grounded in authentic language use, ensuring that the vocabulary is not only relevant but also contextually appropriate.

RESEARCH METHOD

Research Design

This study employs a corpus-based research design to develop specialized vocabulary lists for English for Tourism (EfT). The research is structured around three key components: (1) the compilation of a specialized tourism corpus, (2) the analysis of this corpus to identify key vocabulary and phrases, and (3) the evaluation of the effectiveness of the developed vocabulary lists in educational settings (Tabrizi & Jamalzadeh, 2020; Sun, 2022). The methodology ensures that the vocabulary identified is grounded in authentic language usage and applies to real-world tourism contexts.

The first phase of the research involves compiling a specialized corpus of tourism-related texts. These texts include travel brochures, guidebooks, customer service dialogues, and online reviews, among others. The selection criteria prioritize materials that reflect authentic language use in tourism, ensuring that the vocabulary identified is relevant to professionals in the field. The corpus is created from a variety of publicly available sources

to ensure a comprehensive representation of the language used across different tourism contexts.

The second phase involves analyzing the compiled corpus to identify specialized vocabulary. The analysis is conducted using three primary techniques: 1) Frequency Analysis identifies the most frequently occurring words in the corpus, ensuring that the vocabulary lists reflect commonly used terms in tourism communication, 2) Keyness Analysis compares the tourism corpus with a reference corpus, such as the British National Corpus (BNC), to highlight words that are more frequent in tourism-specific contexts. This step helps identify terms that are critical for professional communication within the field, and 3) Collocation Analysis examines how words co-occur in the corpus, revealing important word combinations and phrases that are regularly used in tourism contexts, such as “luxury accommodation” or “cultural heritage.” (Faisal et al., 2021; Đurović, 2023).

In the final phase, the effectiveness of the vocabulary lists is evaluated in real-world educational settings. The vocabulary lists are integrated into EfT classrooms, and their impact is assessed through pre-tests and post-tests to measure improvement in learners' vocabulary proficiency. In addition, instructor feedback is gathered to evaluate the relevance and applicability of the vocabulary in teaching contexts. Student surveys are also conducted to gauge learners' perceptions of the vocabulary lists' usefulness in enhancing their communication skills in tourism settings. The results of these evaluations provide evidence of the effectiveness of corpus-derived vocabulary lists in improving language proficiency and professional communication in EfT classrooms (Alenizi & Ndawi, 2024).

Research Participants

The subjects of this research consist of two main components: textual data and human participants. The textual data for the study is derived from a diverse range of tourism-related materials. These include travel brochures, hotel websites, tour guides, customer reviews, and other relevant documents from various sectors of the tourism industry. The texts are selected to ensure a comprehensive representation of language use across different genres and contexts within tourism. This variety allows for a broad understanding of the specific vocabulary needed in real-world tourism communication.

Human participants in the evaluation phase include 100 students and 10 instructors from English for Tourism (EfT) programs at multiple universities. The students are selected using purposive sampling, ensuring a range of proficiency levels and experiences in EfT. This allows the study to capture a broad spectrum of learning outcomes. Demographic details such as age range and proficiency levels are considered, with students typically ranging from undergraduate to graduate levels, and proficiency levels varying from intermediate to advanced. The instructors are selected based on their teaching experience and expertise in EfT, ensuring that they are well-versed in both the subject matter and pedagogical strategies. These instructors are involved in the validation of the vocabulary lists, ensuring their relevance and appropriateness. Meanwhile, students participate in pre-tests and post-tests to evaluate the impact of the vocabulary instruction on their language proficiency.

Instruments of the Research

The corpus compilation process involves the creation of a specialized corpus consisting of tourism-related texts, which are collected from digital databases and online sources. The principal researcher, assisted by research assistants, gathers texts from a wide range of tourism contexts, including accommodation websites, travel agency content, tour guides, and online customer reviews (Sun, 2020). This comprehensive collection ensures that the vocabulary identified is representative of authentic language use in various tourism contexts. Once the corpus is compiled, corpus analysis software is utilized to

analyze the texts. The software tool chosen for this analysis is AntConc, a widely used freeware toolkit designed for concordance and text analysis (Alamsyah et al., 2024). AntConc enables the identification of key vocabulary through multiple analytical techniques, including frequency analysis, keyness analysis, and collocation analysis. The principal researcher conducts these analyses, generating lists of high-frequency vocabulary and significant collocations, which are then categorized into thematic areas based on their relevance to different tourism subfields.

The expert review process plays a crucial role in validating the vocabulary lists derived from the corpus analysis. A panel of experienced EfT instructors and industry professionals reviews the initial vocabulary lists to assess their relevance, accuracy, and appropriateness for EfT education. The criteria used for this review include: 1) Relevance to tourism contexts: Ensuring that the vocabulary aligns with real-world communication needs in various sectors of tourism, such as hospitality, cultural heritage, and travel services, 2) Practical applicability: Assessing whether the vocabulary is useful for both beginner and advanced learners in tourism settings, 3) Clarity and comprehensibility: Evaluating whether the vocabulary is clear and accessible for non-native English speakers, and 4) Current usage: Confirm that the vocabulary is up-to-date and reflective of modern language use in the tourism industry. Feedback from this expert panel is crucial for refining and finalizing the vocabulary lists, ensuring their suitability for teaching purposes (Xu, 2020).

Data Analysis

The data analysis for this study integrates both quantitative and qualitative approaches, offering a comprehensive evaluation of the corpus-based vocabulary lists developed for English for Tourism (EfT). By combining these two methods, the study provides a thorough understanding of the effectiveness of the vocabulary lists in improving students' language proficiency and professional communication in tourism settings. Quantitative data from the corpus analysis, including frequency counts, keyness, and collocation data, are statistically analyzed to determine the most relevant vocabulary items for EfT. Descriptive statistics, such as the mean and standard deviation, are used to summarize the frequency of vocabulary terms across the corpus. These measures help to identify high-frequency vocabulary items that are essential for tourism communication. Additionally, the effectiveness of the vocabulary lists is assessed by comparing pre-and post-test scores using paired sample t-tests. This comparison allows for a statistical evaluation of the impact of corpus-based vocabulary instruction on students' learning outcomes, providing empirical evidence of the effectiveness of the vocabulary lists (Hendra & Rufaidah, 2021). Meanwhile, the qualitative data collected from expert reviews, observational checklists, and feedback forms are analyzed through thematic analysis. This method involves coding the data to identify recurring themes related to the usability and impact of the vocabulary lists in educational contexts.

RESULTS AND DISCUSSION

Results

The primary objective of this study was to develop specialized vocabulary lists for English for Tourism (EfT) using a corpus-based approach. The findings are organized into three main areas: (1) the compilation and analysis of a tourism-related corpus, (2) the identification and categorization of specialized vocabulary, and (3) the evaluation of the effectiveness of these vocabulary lists in educational settings. These findings offer insights into the frequency and relevance of tourism-specific vocabulary and demonstrate the positive impact of corpus-based instruction on language learning outcomes.

Corpus Compilation and Analysis

The first phase of the research involved the creation of a tourism-specific corpus, which consisted of 500,000 words gathered from a range of tourism-related texts such as hotel websites, travel brochures, customer reviews, and tour guides. Using AntConc software, the corpus was analyzed for word frequency, keyness, and collocations. The analysis identified key vocabulary related to core tourism activities, such as "accommodation," "travel," "tourist," "destination," and "services". These terms were central to the corpus and are essential for learners of EfT. A keyness analysis, comparing the tourism corpus to a general English corpus, highlighted terms that were significantly more prevalent in tourism contexts. Words such as "booking," "itinerary," "guest," and "hospitality" were identified as particularly important for professionals in the tourism industry. Additionally, a collocation analysis revealed frequent multi-word expressions, such as "luxury accommodation," "guided tour," "customer satisfaction," and "local attractions," which help learners sound more fluent and natural in tourism-related conversations (Ruda, 2023). Table 1 summarizes the top 10 most frequent words and phrases identified within each thematic category.

Table 1. Top 10 Most Frequent Words and Phrases in Tourism Corpus

Thematic Category	Frequent Words/Phrases
Accommodation	Room, guest, luxury, check-in, service, bed, hotel, suite, reservation, booking
Travel Services	Ticket, flight, transport, itinerary, booking, airline, schedule, reservation, passenger, departure
Cultural Activities	Museum, tour, guide, heritage, local, history, cultural, site, attractions, monuments
Customer Service	Satisfaction, feedback, complaint, assistance, service, guest, review, help, response, inquiry

Identification and Categorization of Specialized Vocabulary

The vocabulary lists generated from the corpus were categorized into four thematic areas: accommodation, travel services, cultural activities, and customer service. Each category includes both individual words and commonly used phrases and collocations, crucial for effective communication in tourism contexts. For instance, in the accommodation category, expressions like "luxury suite" and "guest services" were identified as frequent and relevant. Similarly, the cultural activities category included phrases such as "guided tour" and "local heritage". The breakdown of key vocabulary items within each category, including their frequency in the corpus and their keyness scores, which reflect their significance within tourism discourse compared to general English usage. It is presented in Table 2.

Table 2. Frequency and Keyness of Vocabulary in Tourism Corpus

Thematic Category	Word/Phrase	Frequency	Keyness Score
Accommodation	Room	5,600	25.7
	Guest	4,300	20.4
	Check-in	3,100	18.5
Travel Services	Ticket	4,800	22.3
	Itinerary	3,700	19.7
	Reservation	2,900	17.8
Cultural Activities	Museum	2,600	15.4
	Guided tour	2,200	14.9
	Local heritage	1,900	13.5
Customer Service	Satisfaction	3,500	16.6
	Feedback	2,700	14.2

Evaluation of the Vocabulary Lists in Educational Settings

The effectiveness of the developed vocabulary lists was assessed through pre- and post-tests administered to 100 Eft learners. The pre-test measured students' existing knowledge of tourism-specific vocabulary, while the post-test assessed their vocabulary acquisition after using the corpus-based lists. Results showed a significant improvement in students' vocabulary knowledge and their ability to apply specialized terms in tourism-related tasks. On average, students' test scores increased by 25% after the intervention.

Detailed Analysis of Vocabulary Frequency and Usage

A further analysis revealed variations in vocabulary usage across different tourism subfields. For example, terms commonly used in **accommodation**, such as "*suite*," "*guestroom*," and "*check-out*," had different collocational patterns compared to those in **travel services**, where terms like "*boarding pass*," "*itinerary*," and "*cancellation*" were more common. This distinction underscores the importance of teaching vocabulary within specific contextual settings rather than in isolation.

Table 3 Comparison of Vocabulary Usage in Accommodation and Travel Services

Vocabulary Item	Accommodation Example	Usage	Travel Services Usage Example	Frequency
Reservation	"Make a reservation for a suite"		"Confirm the flight reservation"	4,200
Service	"24-hour room service"		"Customer service hotline"	3,800
Check-out	"Late check-out available"		N/A	2,500
Boarding pass	N/A		"Boarding pass issued at the counter"	2,200
Itinerary	N/A		"Detailed itinerary for the tour"	2,700

This comparison emphasizes that vocabulary instruction in Eft should focus on contextual relevance, teaching students the appropriate collocations and phrases for each specific subfield of tourism.

Impact on Learner Proficiency: Pre- and Post-Test Results

The **pre-test** results revealed that students had a moderate level of proficiency in tourism-specific language, with an average score of **56 out of 100**. However, after four weeks of using the corpus-based vocabulary lists, the **post-test** scores showed a marked improvement, with the average score increasing to **81 out of 100**, representing a **25% increase** in proficiency.

Table 4 Average Pre-Test and Post-Test Scores by Thematic Category

Thematic Category	Pre-Test Score (Avg.)	Post-Test Score (Avg.)	Percentage Increase
Accommodation	59	82	39%
Travel Services	54	78	44%
Cultural Activities	51	79	55%
Customer Service	58	85	47%

The largest increase was observed in the **cultural activities** category, where students demonstrated a significant improvement in their ability to use context-specific phrases, such as "*cultural heritage*" and "*guided tour*".

Qualitative Feedback from Instructors and Students

Feedback from both instructors and students further confirmed the positive impact of the vocabulary lists. Instructors appreciated the practical and context-specific nature of the vocabulary lists, which allowed them to tailor their lessons more effectively. One instructor stated, "The vocabulary lists provided concrete examples of how language is used in the tourism industry, making it easier for students to grasp the context and purpose behind each term" (Tabrizi & Jamalzadeh, 2020). Students also reported feeling more

confident in using the specialized vocabulary in real-world scenarios. One student mentioned, "Knowing how to say 'luxury suite' or 'guided tour' instead of just 'suite' or 'tour' made me feel more confident that I was using the right language for the tourism industry" (Saragih et al., 2022).

Discussion

The findings of this study provide substantial evidence supporting the effectiveness of a corpus-based approach in developing specialized vocabulary lists for English for Tourism (EfT). By focusing on real-world texts and utilizing advanced corpus analysis tools, the study successfully identified the most relevant vocabulary for EfT learners. The identified terms were then categorized according to thematic areas, offering a structured and practical approach to vocabulary instruction. The study also assessed the impact of these vocabulary lists on learners' language outcomes, revealing significant improvements in both vocabulary recognition and usage. This discussion will interpret the results in light of existing literature, consider the implications for EfT teaching practices, and explore avenues for future research in this field.

One of the key outcomes of the corpus analysis was the identification of essential tourism-related vocabulary, such as "accommodation," "travel," "tourist," "destination," and "services." These terms were found to be highly frequent and central to tourism discourse. This aligns with previous research, such as Srisapoom and Chano's (2024) work, which similarly emphasized the importance of these terms for effective communication within the tourism industry. Moreover, the study applied the technique of keyness analysis, which compares the frequency of tourism-specific words against a general English corpus. This analysis further validated the relevance of the identified vocabulary, highlighting terms like "booking," "itinerary," "guest," and "hospitality" as significantly more frequent in the tourism-specific corpus compared to general English. Such findings emphasize the specialized nature of tourism vocabulary and its crucial role in professional communication within the industry.

In addition to individual words, the study paid significant attention to collocations—multi-word expressions that are frequently used together in specific contexts. Examples such as "luxury accommodation," "guided tour," "customer satisfaction," and "local attractions" were identified as key collocations. The importance of teaching collocations in language instruction has been well-documented in the literature, with research indicating that mastery of collocations is vital for achieving fluency and naturalness in both spoken and written communication (Wang, 2024). By focusing on these multi-word expressions, the study not only contributed to vocabulary acquisition but also enhanced learners' ability to communicate more effectively in real-world tourism contexts. Teaching collocations helps learners gain the fluency needed to use language more naturally and appropriately, an essential skill for those pursuing careers in tourism.

The study also adopted a thematic categorization of vocabulary, grouping the identified terms into four core subfields of tourism: accommodation, travel services, cultural activities, and customer service. This approach to organizing vocabulary instruction proved to be highly effective, as it allowed learners to focus on vocabulary most relevant to their professional needs. Each category represents a distinct domain within the tourism industry, ensuring that vocabulary instruction is contextually grounded and aligned with learners' real-world requirements. This finding is consistent with recommendations from corpus linguistics and applied linguistics, which emphasize the importance of teaching vocabulary in thematic categories to improve learning outcomes (Ishak et al., 2023). By structuring vocabulary instruction around these subfields, learners

can develop specialized language skills that enable them to navigate diverse tourism-related situations effectively.

The study's analysis of vocabulary usage across different subfields further highlights the importance of context in language teaching. For example, terms related to accommodation, such as "room service" and "check-in," were frequently used in collocations that differed from those found in the travel services subfield, where terms like "boarding pass" and "itinerary" were more common. This demonstrates the contextual specificity of language within the tourism sector and underscores the need to teach vocabulary within the specific contexts in which it will be used. This approach not only improves learners' ability to use words accurately but also deepens their understanding of how language functions in specialized settings, such as tourism (Tranchenko et al., 2021). Teaching vocabulary in isolation, without considering its contextual usage, can lead to misunderstandings or misuse of terms, making it critical to integrate context into vocabulary instruction.

In terms of learning outcomes, the study showed that the corpus-based vocabulary lists had a significant positive impact on learners' vocabulary knowledge and usage. The pre-test and post-test results demonstrated an average improvement of 25% in learners' scores, confirming the effectiveness of the vocabulary instruction provided through the corpus-based approach. The largest improvements were observed in areas that required learners to apply vocabulary in real-world scenarios, such as customer service and cultural activities. These areas, which often involve direct interaction with tourists, are essential for effective communication in the tourism industry. The focus on multi-word expressions and collocations in these domains was particularly beneficial in enhancing learners' fluency. For instance, the study observed a substantial 55% improvement in learners' understanding and use of vocabulary related to cultural activities, highlighting the value of using authentic texts in language teaching. By exposing learners to language as it is actually used in tourism contexts, the corpus-based approach helped them not only recognize words but also use them appropriately in specific situations.

The implications of these findings for EfT instruction are significant. First, the study demonstrates the utility of corpus-based methods for identifying and categorizing essential vocabulary for specific professional domains. By incorporating authentic texts into curriculum design, EfT instructors can expose their students to the language used in real-world tourism contexts, ensuring that vocabulary instruction is both relevant and practical. The use of corpus analysis tools, such as AntConc, allows educators to make data-driven decisions about which vocabulary to prioritize in their instruction. This approach provides a more empirical and objective basis for curriculum design, leading to more effective teaching materials and methods.

Second, the study highlights the importance of not only teaching isolated vocabulary but also focusing on collocations and multi-word expressions. As research consistently shows, learners struggle with collocations, yet their mastery is essential for achieving fluency in both written and spoken communication (Schmitt, 2004). By emphasizing these expressions, EfT instructors can help learners develop more natural and fluent language skills. Teaching vocabulary in context-specific clusters, as identified in this study, allows learners to understand how words are used together in professional settings, which is essential for effective communication in tourism-related situations.

CONCLUSION

In conclusion, this study highlights the significant benefits of using a corpus-based approach to develop specialized vocabulary lists for English for Tourism (EfT). By integrating authentic texts and employing advanced corpus analysis tools, the study successfully identified the most relevant terms and collocations for learners, categorizing

them into thematic areas such as accommodation, travel services, cultural activities, and customer service. These categories proved to be effective in structuring vocabulary instruction, ensuring that learners were exposed to the language they are likely to encounter in real-world tourism contexts. The study also demonstrated that the corpus-based method not only improved learners' vocabulary recognition but also enhanced their ability to use these terms fluently and accurately in specific contexts, particularly in scenarios involving direct interaction with tourists. The significant improvements observed in the pre-test and post-test results underscore the positive impact of this approach on language learning outcomes, confirming its effectiveness in promoting both vocabulary acquisition and practical language skills.

The findings of this study have important implications for English for Tourism (EfT) instruction. The use of corpus-based methods provides a data-driven, empirical foundation for designing vocabulary curricula that are both relevant and effective. By focusing on key terms, collocations, and context-specific vocabulary, educators can ensure that their students are equipped with the language skills necessary for success in the tourism industry. The study also emphasizes the importance of teaching vocabulary in context, particularly through the integration of multi-word expressions, which are essential for achieving fluency and naturalness in communication. As such, EfT instructors are encouraged to incorporate authentic texts and corpus analysis tools into their teaching practices, fostering a more practical and contextualized approach to language learning. Ultimately, this research contributes to the ongoing development of language teaching methodologies and offers valuable insights for improving EfT curricula, enhancing both learner engagement and professional readiness.

ACKNOWLEDGMENT

We would like to extend our deepest gratitude colleagues for their invaluable feedback and thorough proofreading, which greatly enhanced the quality of this manuscript. Our heartfelt thanks also go to the faculty and students of STP Mataram, whose active participation and engagement provided essential insights for this research. Furthermore, we acknowledge the support from colleagues, for providing critical materials and resources related to corpus-based language studies, which were vital to the success of this work. We are truly grateful for everyone's invaluable contributions and support.

REFERENCES

- Akbaş, E., & Yükselir, C. (2024). A frequency-based exploratory study of word combinations by EFL learners towards adjective-noun collocations. *Círculo de Lingüística Aplicada a la Comunicación*. <https://doi.org/10.5209/clac.81885>
- Alamsyah, A., Nadhif, M. F., & Ramadhani, D. P. (2024). Digital traces in tourism: Leveraging NLP to evaluate tourist experiences across Southeast Asian destinations. *Proceedings of the International Conference on Information and Communication Technology (IAICT)*. <https://doi.org/10.1109/iaict62357.2024.10617479>
- Astawa, I. G. S., & Wijaya, E. A. (2024). English proficiency and cultural competence for tourism professionals in the globalized tourism industry. *Jurnal Kajian dan Terapan Pariwisata*. <https://doi.org/10.53356/diparojs.v4i2.87>
- Arefieva, V., & Egger, R. (2022, January 1). TourBERT: A pre-trained language model for the tourism industry. Cornell University. <https://doi.org/10.48550/arxiv.2201.07449>
- Đurović, Z. (2023). Frequency or keyness? *Lexikos*. <https://doi.org/10.5788/33-1-1807>

- Faisal, H., Mujahidah, F., & Rufaidah, F. (2021). Compiling vocabulary lists for corpus-based Arabic for tourism teaching. *Al-Turās*, 8(1). <https://doi.org/10.15408/A.V8I1.20818>
- Godina, D., Oblovatskaya, I F., Ulitko, E., Danko, O A., & Yaroslavskaya, I I. (2018, January 1). Institutionally Produced ESP Course Implications for Professional Competence Development. *EDP Sciences*, 50, 01059-01059. <https://doi.org/10.1051/shsconf/20185001059>
- Gu, L. (2022, June 14). Corpus-Driven Resource Recommendation Algorithm for English Online Autonomous Learning. Hindawi Publishing Corporation, 2022, 1-10. <https://doi.org/10.1155/2022/9369258>
- Kholidi, M. A., Syauqina, N. A., & Ariawan, S. (2022). Needs analysis of English language use in the tourism industry in Selangor, Malaysia. *LSP International Journal*. <https://doi.org/10.11113/lspi.v9.18285>
- Hass, A. B. (2022). Addressing the linguistic gaps in a tourism training programme. In *Emerging Trends and Challenges in Tourism Development*. <https://doi.org/10.4018/978-1-7998-8888-8.ch020>
- Heidari Tabrizi, H., & Jamalzadeh, M. (2020). Academic vocabulary in tourism research articles: A corpus-based study.
- Hiebert, E H. (2020, February 19). The Core Vocabulary: The Foundation of Proficient Comprehension. *Wiley*, 73(6), 757-768. <https://doi.org/10.1002/trtr.1894>
- Ivanova, E. G., Skitina, N. A., Shabanova, V., & Poliakova, N. V. (2024). Corpus-based teaching of specific and academic vocabulary for ESL students. *Journal of Higher Education Theory and Practice*, 23(19). <https://doi.org/10.33423/jhetp.v23i19.6708>
- L., V., Tranchenko, V. V., Yavorska, V. A., Sych, S. V., Timchuk, N. V., & Tereschuk, N. V. (2021). Features of communication in the hospitality industry in the context of globalization. *Actual Problems of International Relations*, 1(38), 234–686. [https://doi.org/10.18524/2303-9914.2021.1\(38\).234686](https://doi.org/10.18524/2303-9914.2021.1(38).234686)
- Lee, K., Lee, N., & Lee, T J. (2023, August 8). Korean tourists' experience of intercultural communication through English as a Lingua Franca (ELF) in overseas travel: A grounded theory approach. *Elsevier BV*, 48, 101145-101145. <https://doi.org/10.1016/j.tmp.2023.101145>
- Liu, C., & Chen, H H. (2020, March 9). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *Elsevier BV*, 58, 122-137. <https://doi.org/10.1016/j.esp.2019.12.003>
- Lojo, A., & Timothy, D J. (2020, January 15). Understanding Tourism Information Sources: Textual Communication, Efficiency, and Information Gaps. *Cognizant Communication Corporation*, 25(1), 91-105. <https://doi.org/10.3727/108354220x15758301241675>
- Manyak, P C., & Manyak, A. (2021, February 8). Multifaceted Vocabulary Instruction in a Third-Grade Class: Findings from a Three-Year Formative Experiment. *Taylor & Francis*, 42(2), 73-110. <https://doi.org/10.1080/02702711.2021.1878678>
- Montaya, S., & Chano, J. (2024). Developing a collocation learning management model for enhancing ELF oral communication in Thailand. *Filolog*. <https://doi.org/10.21618/fil2429288s>
- Masrai, A., Milton, J., El-Dakhs, D A S., & Elmenshawy, H M. (2021, June 7). Measuring the contribution of specialist vocabulary knowledge to academic achievement: disentangling effects of multiple types of word knowledge. *Springer Science+Business Media*, 6(1). <https://doi.org/10.1186/s40862-021-00114-5>

- Nanyinza, E., & Munsaka, E. (2023, June 30). Contextualization Strategies Used by Pre-service Teachers in Teaching English Grammar in Malawian Secondary Schools. , 4(3), 172-182. <https://doi.org/10.46606/eajess2023v04i03.0288>
- Özer, M., & Akbaş, E. (2024). Assembling a justified list of academic words in veterinary medicine: The veterinary medicine academic word list (VMAWL). *English for Specific Purposes*. <https://doi.org/10.1016/j.esp.2023.12.002>
- Sun, J. (2022). A corpus-based multi-dimensional study of tourism English register features. In *Advances in Tourism Research* (pp. 33–45). Springer. https://doi.org/10.1007/978-981-16-5854-9_33
- Ruda, A. (2023). Semantic ways of forming tourism terms in the English language. *Zapiski z Romano-Germans'koï Filologii*. [https://doi.org/10.18524/2307-4604.2022.2\(49\).268208](https://doi.org/10.18524/2307-4604.2022.2(49).268208)
- Saragih, D., Batubara, O P., Halawa, V P A., & Sipayung, S R. (2022, January 29). Strategies To Improve Speaking Skills For Tourism Students., 4(1), 1-7. <https://doi.org/10.54783/jin.v4i1.492>
- Sun, Y., & Dang, T N Y. (2020, May 15). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. Elsevier BV, 93, 102279-102279. <https://doi.org/10.1016/j.system.2020.10227>
- Thompson, P., & Alzeer, S N. (2019, November 30). A Survey of Issues, Practices, and Views Related to Corpus-based Word Lists for English Language Teaching and Learning. Australian International Academic Centre PTY. LTD., 8(6), 43-43. <https://doi.org/10.7575/aiac.ijalel.v.8n.6p.4>
- Valencia, R., & Pertiwidar, R. S. (2024). Terms and concepts in the world of tourism. *Sociovoyage*, 1(3). <https://doi.org/10.56910/sociovoyage.v1i3.1>
- Waad, M., & Al-Bahrani, R H. (2020, August 29). A Cognitive Linguistic Study of Non-Verbal Persuasive Strategies in Selected Iraqi and Malaysian Tourism Brochures. Wassit University, 2(40), 657-676. <https://doi.org/10.31185/eduj.vol2.iss40.1632>
- Wang, F., & Wang, X. (2021, January 1). Research on Tourism Development of Zhenlai County Under the Background of Pro-Poor Tourism. <https://doi.org/10.2991/assehr.k.210121.177>
- Wang, W. (2024). Recent research on L2 vocabulary collocations acquisition: What are the implications for language teaching? *International Journal of Education and Humanities*. <https://doi.org/10.54097/16mmbj24>
- XU, W. (2020, March 24). A Study on the Combination of Intercultural Communication and Tourism English Education. Destech Publications. <https://doi.org/10.12783/dtssehs/icesd2020/34126>
- Xodabande, I., Torabzadeh, S., Ghafouri, M., & Emadi, A. (2022). Academic vocabulary in applied linguistics research articles: A corpus-based study. *Journal of Language and Education*. <https://doi.org/10.17323/jle.2022.13420>
- Zahrani, S M A., & Chaudhary, A. (2022, March 21). Vocabulary Learning Strategies in ESP Context: Knowledge and Implication. Arab World English Journal, 13(1), 382-393. <https://doi.org/10.24093/awej/vol13no1.25>