

A Corpus-Based Study of the Near-Synonyms: *Tourist*, *Traveler*, and *Visitor* for ESP Classrooms

Piyapong Laosrirattanachai¹, Phongsathorn Thongsawat¹, Montira Somsri¹,
Atittiya Obchoey¹, Piyanuch Laosrirattanachai^{1*}

*Corresponding author's email: piyanuch.la@ku.th

¹English for Service Industry Program, Faculty of Hospitality Industry
Kasetsart University, Kamphaeng Saen Campus, Thailand

Received: April 10, 2026

Revised: May 27, 2026

Accepted: May 27, 2026

Abstract

This study investigates the near-synonyms *tourist*, *traveler*, and *visitor* by examining their contextual usage and collocational behavior. Using the Corpus of Contemporary American English (COCA), the analysis focuses on genre distribution, formality, collocations, semantic preference, and semantic prosody. The findings show that *tourist* is commonly used in structured contexts such as academic writing and journalism, while *traveler* appears more frequently in narrative and lifestyle contexts. For *visitor*, it is the most frequent and versatile term, occurring across both formal and informal settings. Collocational patterns further reveal distinct semantic associations for each word, reflecting their specialized meanings. These results have important implications for English language teaching, particularly in English for Specific Purposes (ESP), by supporting more precise and context-sensitive vocabulary instruction. The study highlights the value of extending near-synonym research to specialized domains, especially for learners in the tourism industry.

Keywords: Corpus-based study, COCA, near-synonyms, tourism business, vocabulary

Introduction

Vocabulary plays a pivotal role in learning English and effective communication. It serves as the foundation for listening, speaking, reading, and writing in the language. However, mere knowledge of vocabulary meanings is insufficient for learners to communicate effectively. Mastery of appropriate and natural use of vocabulary is equally essential, particularly when dealing with synonyms. Synonyms, words with similar or identical meanings, play a critical role in understanding language variation and usage. Scholars often categorize synonyms into two primary subtypes: absolute synonyms and near-synonyms (Murphy, 2010). Absolute synonyms refer to words that are interchangeable in all contexts without any variation in meaning, connotation, or stylistic preference (Edmonds & Hirst, 2002). Absolute synonyms are rare in natural language due to the evolutionary efficiency of language, which avoids redundancy (Cruse, 1986). Near-synonyms, in contrast, are words with overlapping meanings that differ in degree, genres, or connotation (Dimarco et al., 1993). Near-synonyms enrich linguistic expression, allowing speakers to convey subtle distinctions. In communication, if a word choice is chosen and used appropriately, it can convey the meaning perfectly and communicate effectively (Liu & Espino, 2012). On the other hand, without a nuanced understanding of these differences, learners may misuse terms, leading to communication that sounds unnatural or inappropriate (Szudarski, 2018). One significant problem learners face is distinguishing between near-synonyms and applying them correctly in different contexts. These words often overlap in meaning, yet their differences can vary based on genre, formality, collocation patterns, semantic preferences, and prosody (Yang, 2016). For instance, choosing an unsuitable synonym may result in expressions that are either too formal or informal for the intended setting, potentially causing confusion or misinterpretation. Such difficulties highlight the need for focused and systematic studies to help learners understand the subtleties of near-synonyms, particularly in specialized fields where precise language is critical.

Previous research has thoroughly examined near-synonyms in general and academic English, providing valuable insights into their usage (Imsa-ard & Phoocharoensil, 2022; Phoocharoensil & Kanokpermpoon, 2021). However, the exploration of near-synonyms within specialized domains, such as English for Specific Purposes (ESP), remains sparse (Laosrirattanachai & Phoocharoensil, 2025). This gap is noteworthy given that the progression of English learning typically transitions from general vocabulary to academic and specialized terms (Nation, 2016). Similarly, studies on near-synonyms should extend beyond general and academic English to encompass ESP contexts. Addressing this gap is crucial for equipping learners with the linguistic competence required in specific fields.

Among various specialized domains, the tourism industry stands out as one of the most significant global industries. English fluency is indispensable for professionals working in tourism, as it serves as a lingua franca for international communication. The industry is replete with specialized terminology, requiring precise and contextually appropriate usage. Among these terms, the near-synonyms *tourist*, *traveler*, and *visitor* are among the most frequently used. They were selected for this

study for two primary reasons: their industry relevance and their inclusion in specialized domain word lists. First, these three words are the most common terms used to describe consumers in the tourism and hospitality sector. To provide a solid justification, a preliminary survey of 16 textbooks on English for Tourism—including both international commercial publications and textbooks written by university lecturers—was conducted. The survey revealed that all three terms consistently appeared in every textbook to refer to customers and service delivery types. This textbook-based evidence confirms their status as essential, high-frequency vocabulary for L2 learners. Second, the selection was informed by empirical evidence from specialized vocabulary research. All three target items appear in the Tourism Business Word List (TBWL) (Laosrirattanachai & Ruangjaroon, 2021). Within this established word list, they are key members of the core word families for tour, travel, and visit, proving their high pedagogical value for learners preparing to enter the tourism industry. Despite their importance, there is no definitive evidence to confirm that learners or professionals in the field can accurately explain or demonstrate their appropriate use. Misunderstanding or misusing these terms could lead to miscommunication, potentially affecting the quality of service and professionalism within the industry. However, before conducting the research to explore the differences among these three near-synonyms, it is essential to preliminarily verify whether differences actually exist. Therefore, the three near-synonyms were examined based on two preliminary aspects: their CEFR levels and definitions in dictionaries. Regarding CEFR levels and the definitions, the levels and meanings provided by dictionaries are presented in Table 1.

Table 1.

Definitions of Visitor, Tourist, and Traveler

Sources	<i>Tourist</i>	<i>Traveler</i>	<i>Visitor</i>
CEFR Level	A1	A2	A1
Longman Dictionary of Contemporary English	Someone who is <u>visiting</u> a place for pleasure on holiday	Someone who is on a journey or someone who <i>travels</i> often	Someone who comes to <u>visit</u> a place or a person
Oxford Advanced Learner's Dictionary	A person who is <i>traveling</i> or <u>visiting</u> a place for pleasure	A person who is <i>traveling</i> or who often <i>travels</i>	A person who <u>visits</u> a person or place
Cambridge Dictionary	Someone who <u>visits</u> a place for pleasure and	Someone who <i>travels</i>	Someone who <u>visits</u> a person or place

Sources	<i>Tourist</i>	<i>Traveler</i>	<i>Visitor</i>
	interest, usually while on holiday		
UN Tourism	A visitor is classified as a tourist if his/her trip includes an overnight stay, or as a same-day visitor otherwise.	A traveler is someone who moves between different geographic locations, for any purpose and any duration.	A visitor is taking a trip to a main destination outside his/her usual environment, for less than a year, for any main purpose.

As can be seen in Table 1, when examining the differences in CEFR levels, it is found that *tourist* and *visitor* are allocated to the A1 level, while *traveler* is placed at the A2 level. This signifies the differences among them, as A1 vocabulary refers to a limited set of basic words, primarily consisting of everyday words and phrases, which typically cover simple and concrete topics, while A2 vocabulary refers to a broader range of familiar and everyday topics. Then, the definitions of the three near-synonyms were compared. As a result, the definitions of *tourist*, *traveler*, and *visitor* from the Longman Dictionary of Contemporary English, the Oxford Advanced Learner's Dictionary, and the Cambridge Dictionary highlight their circular and overlapping nature. A *visitor* is generally defined as someone who visits a person or place, relying on the verb *visit* without specifying purpose, making the term generic. A *tourist* is described as someone visiting for pleasure, often on holiday, but this overlaps with the general idea of visiting, blurring distinctions from *visitor*. A *traveler* is defined as someone traveling or frequently traveling, with vague overlap with *tourist* and *visitor*, especially in leisure contexts. These definitions fail to draw clear boundaries, relying on repetitive terms like *visit* and *travel*. This circularity obscures the nuanced meanings of the terms and limits their ability to distinguish the near-synonyms effectively. Therefore, such definitions may not effectively assist L2 learners in discerning potential variations in meaning. While the UN Tourism provides more specific definitions that somewhat highlight the different uses among the three synonyms, they remain insufficient for fostering a comprehensive understanding among L2 learners of their appropriate use in different contexts. Based on this information, it can be concluded that *tourist*, *traveler*, and *visitor* have distinct uses that require further investigation. The significance of this study lies in its aim to bridge this gap by employing a corpus-based analysis to investigate the differences among *tourist*, *traveler*, and *visitor*. Corpus-based methods have proven effective in revealing subtle linguistic patterns and usage tendencies that are difficult to capture through traditional approaches. According to the literature review, corpus-based methods frequently employed in analyzing the distinctions among near-synonyms consist of the following approaches.

The first method is genre distribution which refers to the categorization of vocabulary across various genres, such as fiction, news, academic, and spoken discourse. This distribution is a critical factor in corpus-based linguistic studies, as it highlights the contexts in which particular words are frequently used. In the analysis of near-synonyms, genre distribution provides insights into the stylistic preferences and functional tendencies of each word. For instance, in the Corpus of Contemporary American English (COCA), researchers can explore how specific near-synonyms appear across different genres. This approach allows for a nuanced understanding of lexical variation, showing how contextual factors influence word choice. Additionally, the frequency analysis within genres helps identify dominant trends and contexts that distinguish one synonym from another (Laosrirattanachai & Laosrirattanachai, 2025). For example, among the three near-synonyms *main*, *major*, and *primary*, *major* is most frequently found in academic texts and newspapers (Phoocharoensil, 2022). Another example is that *empty* is often found in fiction, web, TV/movies, and magazine genres, whereas *vacant* predominantly appears in the newspaper genre (Lertcharoenwanich, 2023).

Secondly, degrees of formality refer to the varying levels of linguistic appropriateness, ranging from formal to informal language, or even slang. This concept is pivotal in understanding the stylistic and contextual constraints governing word usage. When studying near-synonyms, the COCA enables researchers to evaluate how formality impacts lexical choices by revealing word frequencies across eight genres, ranging from formal contexts (e.g., the academic genre) to informal texts (e.g., TV/movies subtitles, fiction, and spoken genres). However, the remaining four genres—blog, web, magazine, and newspaper—are determined to fall somewhere in between and thus cannot clearly represent formal or informal degrees (Davies, 2021). The degree of formality of a word can be evaluated through its words per million (WPM) value in different genres in the COCA. This allows researchers to identify the contexts where these words are used and discern patterns that reflect their levels of formality. Through this approach, researchers can better appreciate the pragmatic and stylistic nuances that differentiate near-synonyms, providing deeper insights into language variation and choice.

Thirdly, collocation refers to the habitual co-occurrence of words within a given context, usually in combinations of two or more words in a text, forming natural lexical pairings (Sinclair, 1991). This phenomenon is essential for understanding the semantic and syntactic relationships between words. In the context of near-synonyms, analyzing collocations reveals subtle differences in usage and meaning. Using COCA, researchers can examine the collocational patterns of words, which demonstrate distinct tendencies that reflect contextual appropriateness. By studying word pairings, researchers can identify fixed expressions, idiomatic uses, and domain-specific applications, providing a comprehensive understanding of lexical behavior. Collocation analysis also aids in distinguishing near-synonyms by uncovering preferred semantic associations (semantic preference). For example, *disadvantages* and *downsides* commonly collocate with *considerable*, *perceived*, and *serious*, while *inherent* often occurs with *disadvantage* and *drawback* (Sumonsriworakun, 2022).

Fourthly, semantic preference describes a word's tendency to co-occur with specific semantic categories or concepts (Cheng, 2012). This aspect of linguistic analysis sheds light on the underlying meanings and associations of words. Semantic preference is related to collocations, which typically involves analyzing collocations that appear alongside near-synonym words to determine if these pairs can be grouped by meaning (Flowerdew, 2012). When analyzing near-synonyms, semantic preference helps distinguish subtle differences by identifying their preferred semantic contexts. By studying these patterns, researchers gain insights into the conceptual frameworks that govern word usage. Semantic preference analysis also contributes to understanding the cognitive and cultural factors shaping lexical choices. The vast corpus of COCA allows for systematic exploration of these tendencies, providing a robust basis for identifying and explaining distinctions between near-synonyms. Through this method, researchers can elucidate the nuanced relationships between words and their associated concepts. For example, Supanfai (2022) analyzed two near-synonyms, *people* and *persons*, demonstrating that *people* shows semantic preferences for negative actions, numbers, and ethnicity words, while *persons* frequently collocates with words from the semantic set of legislation.

Lastly, Semantic prosody refers to the evaluative or attitudinal associations that words develop through repeated contextual use (Flowerdew, 2013). Similar to semantic preference, the collocating words that appear alongside near-synonyms help determine the attitudinal aura of those near-synonyms. These associations can be positive, negative, or neutral, influencing a word's connotations and implications. However, identifying and interpreting whether a word carries a positive, negative, or neutral prosody is not always straightforward (Lindquist, 2009). The determination often depends on the researcher's chosen criteria, such as rational judgment, input from others, or linguistic corpus tools. In the study of near-synonyms, semantic prosody helps uncover differences in emotional or attitudinal overtones. This analysis enables a deeper understanding of the pragmatic implications of word choice, highlighting how semantic prosody shapes the perceived meaning and impact of near-synonyms. COCA's extensive contextual data allows for the precise identification of these patterns, offering a nuanced perspective on the interplay between meaning, context, and usage. Through semantic prosody analysis, researchers can better understand how lexical choices influence communication and interpretation. For instance, *persist* frequently co-occurs with words that convey negative meanings, often related to unpleasant situations that continue over time. In contrast, the words surrounding *persevere* tend to have positive meanings, emphasizing strong determination and significant effort in overcoming challenges (Phoocharoensil, 2021).

By examining these linguistic aspects, this study aims to provide valuable insights into the specialized use of the near-synonyms *tourist*, *traveler*, and *visitor*. Its findings are expected to enhance English language teaching and learning in ESP contexts, particularly for L2 learners in the tourism business. Furthermore, the study underscores the broader importance of extending near-synonym research to specialized domains, highlighting the need for precise and context-sensitive approaches to vocabulary instruction. To achieve this goal, the following research questions must be addressed.

1. How do the near-synonyms *tourist*, *traveler*, and *visitor* vary in usage across different genres, and are they more commonly used in formal or informal contexts?
2. What collocations frequently appear with the near-synonyms *tourist*, *traveler*, and *visitor*?
3. How do these three near-synonyms co-occur with specific semantic concepts and attitudinal associations based on their patterns of co-occurrence?

Research Instrument and Methods

Data Collection

This study used the Corpus of Contemporary American English (COCA), which contains more than one billion words, to collect data. The data in COCA is continuously updated, ensuring its relevance and freshness (Davies, 2021). With its up-to-date content, COCA is widely used, particularly by ESL and EFL learners. COCA consists of eight genres: academic texts, newspapers, magazines, fiction, spoken language, TV/movie subtitles, webpages, and blogs. It provides extensive options for conducting corpus-based studies and offers a well-balanced distribution across genres. The data is categorized by text type, making it possible to analyze and distinguish specific target words of interest.

Data Analysis

To investigate the differences among the near-synonyms *tourist*, *traveler*, and *visitor* using a corpus-based approach, this study followed a structured analytical approach and utilized two corpus tools. The analysis consisted of five key aspects, as outlined below.

The first aspect involved analyzing genre distribution to determine in which genres these three near-synonyms were most commonly used. This analysis relied on data from COCA, where information is categorized into eight genres: academic, TV/movie subtitles, fiction, spoken, magazines, news, blogs, and web. When searching for any word in COCA, the system presents two types of frequency data: the frequency of the target word across different genres and the frequency per million (WPM) words within those genres. Since the total number of words varies across genres, WPM must be considered to accurately determine which words are more likely to be used in specific genres.

The next aspect examined whether these near-synonyms were more commonly used in formal or informal contexts, which can help L2 learners make informed decisions when selecting appropriate words for different levels of formality. Among the eight genres in COCA, the academic genre is classified as a formal context, while TV/movie subtitles, fiction, and spoken genres are categorized as informal contexts. The blog, web, magazine, and newspaper genres contain a mix of both formal and informal contexts (Davies, 2021), making it difficult to classify them definitively. Therefore, this study used these classifications to distinguish between formal and informal genres and evaluated the WPM values of the three near-synonyms in formal contexts (academic genre) and informal contexts (TV/movie subtitles, fiction, and spoken genres).

The next step involved analyzing collocations. Since *tourist*, *traveler*, and *visitor* are nouns, the analysis focused on three main parts of speech for their collocates: nouns, verbs, and adjectives. To extract these data, the “Collocates” function in COCA was utilized. To ensure the replicability of this study, specific statistical filters and extraction criteria were applied. First, the predefined collocation span was set to [4L, 4R]. According to Watson Todd (2018), a word span stretching up to four lexical items on either side of the node word is effective for capturing meaningful collocations. This standard allows the analysis to capture both immediate grammatical modifiers and broader contextual associations. Second, collocations were identified using Mutual Information (MI) scores. An MI score measures the strength of association between two words by comparing how often they appear together against how often they appear separately. A high MI score indicates a strong relationship, while a low score suggests a random association. Following Hunston (2002), MI scores of 3 or higher were established as the baseline threshold to filter out statistically insignificant word combinations. However, relying only on MI scores can sometimes highlight rare word pairs that happen to appear together but are not structurally important (Cheng, 2012). To prevent this, no minimum frequency cutoff was used; instead, an alternative sorting strategy was adopted. Collocates meeting the MI threshold were sorted by frequency, and the top 20 most frequently occurring collocates were selected within each part of speech category (nouns, verbs, and adjectives). Combining a minimum MI score with top-frequency sorting ensures that the selected collocates are both statistically strong and highly salient in real-world communication, making them pedagogically valuable for ESP learners.

To ensure structural accuracy and directly satisfy the need for pedagogical usability, a manual screening of the concordance lines was performed to filter the collocations. In corpus linguistics, collocations can be analyzed through two main frameworks: the frequency-based approach (words co-occurring within a loose statistical span) and the phraseological approach (words sharing a direct grammatical and adjacent relationship). In this study, once the initial 180 statistically significant collocates were identified using the frequency-based parameters (Top 20, $MI \geq 3$), the researchers carefully reviewed the concordance lines for each item. The final selection was then narrowed down to focus exclusively on phraseological collocations that displayed direct syntactic links to the node words (e.g., direct adjective-noun modifiers or active verb-noun object phrases). Any ambiguous, distant, or misclassified co-occurrences were removed. This rigorous refinement process ensured that only structurally stable and educationally relevant collocations were retained, allowing readers, teachers, and learners to gain the greatest benefit from understanding collocations identified through the phraseological approach.

The next step involved the analysis of semantic preference and semantic prosody. Because the primary aim of this analysis was to understand the semantic tendencies of words occurring around the nodes, collocations identified through the frequency-based approach were examined. This allowed for a broader overview of the semantic preference and semantic prosody patterns of the co-occurring words. Semantic preference refers to the grouping of words that belong to the same semantic domain, while

semantic prosody examines the attitudinal connotation of the collocates appearing with the three near-synonyms—whether the accompanying words convey a positive, negative, or neutral tone. However, previous studies on semantic preference and semantic prosody have often relied on manual analysis by researchers, which introduces subjectivity and may lead to biased judgments when determining whether a word has a positive, negative, or neutral connotation. To minimize such subjectivity, this study employed the UCREL Semantic Analysis System (USAS) (Rayson et al., 2004). The USAS is a tool for automatic semantic annotation that assigns words to 21 predefined semantic categories using a lexicon and contextual rules (see <https://ucrel.lancs.ac.uk/usas/>). It facilitates semantic categorization and sentiment analysis by identifying semantic meanings in textual data. Utilizing this tool enhances the consistency and reliability of the analysis.

To ensure coding reliability and manage cases where the automated system produced ambiguous or inconclusive results, a strict manual validation procedure was implemented. Two researchers from the five-member research team were assigned to independently cross-check the USAS outputs. Both researchers are specialists with over five years of experience teaching English for Tourism and Hospitality. In cases where there was a disagreement between the two coders regarding any collocate classification, two additional independent experts were invited to resolve the issue. These external experts also possess more than five years of experience in the English for Tourism and Hospitality context and were not part of the core research team. All four experts engaged in collaborative discussions until a 100% consensus was reached. This rigorous consensus-building process eliminates individual researcher bias and ensures high intercoder consistency, strengthening the validity of the final semantic and prosodic interpretations.

Results and Discussion

To determine the differences in the usage of the near-synonyms *tourist*, *traveler*, and *visitor*, it is essential to address three research questions. Therefore, the research findings and discussion are divided into five main sections: (1) Genre Distribution and Degrees of Formality, (2) Common Collocations, (3) Semantic Preference and Semantic Prosody, (4) Pedagogical Implications, and (5) Limitations of the Study and Recommendations for Future Research.

Genre Distribution and Degrees of Formality

To assess which genres are most appropriate for each of the three near-synonyms, this study examines their Per Million (PM) values, which indicate the number of times a word appears per one million words. Words with significantly high PM values are considered more suitable for use in a particular genre. The results are presented in Figure 1.

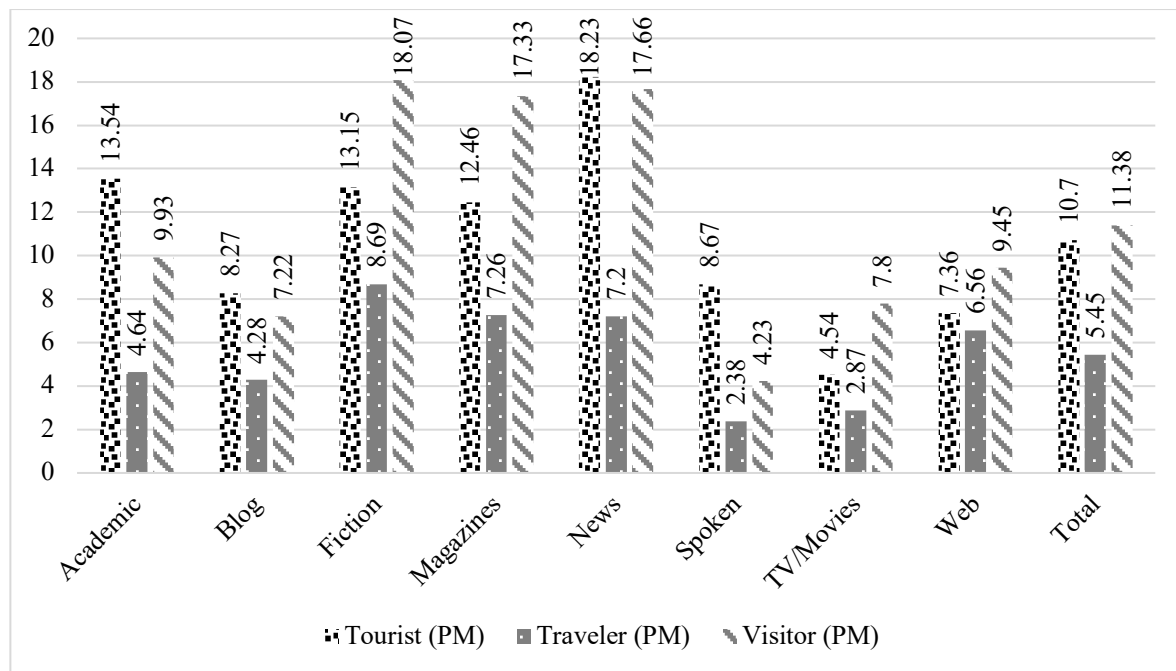


Figure 1. Genre Distribution of the Three Near-synonyms

According to Figure 1, *tourist* appears most often in news (18.23), academic texts (13.54), fiction (13.15), and spoken (8.67). This suggests that *tourist* is commonly used in structured discourse, particularly in academic writing and journalism, likely due to its association with discussions on tourism-related policies, statistics, and impacts. Additionally, the findings indicate that when choosing a word in spoken language, *tourist* is used the most among the three near-synonyms. *Traveler* occurs less frequently overall (5.45) and is least common in spoken language (2.38) and TV/movies (2.87). Its highest occurrences are in fiction (8.69) and magazines (7.26), indicating that *traveler* is often used in storytelling and lifestyle-oriented contexts, where it might evoke a sense of adventure and exploration rather than structured tourism. However, when considering all genres, it is evident that there is no genre in which *traveler* is the most frequently used. This may suggest that *traveler* is subject to other usage constraints, such as collocation or semantic preference, which require further investigation. *Visitor* has the highest overall frequency (11.38), surpassing *tourist* and *traveler*. It is most frequently used in fiction (18.07), magazines (17.33), TV/movies (7.8), and web (9.45). This suggests that *visitor* is commonly used in both narrative and journalistic writing, likely because it conveys a broader meaning, encompassing not only leisure travelers but also business visitors, temporary guests, and individuals visiting specific locations for various purposes. The findings thus indicate that even though they are nouns with the same core meaning, the occurrence of words in each genre cannot be the same (Jackson & Amvela, 2007).

After analyzing the genre distribution, the next step is to examine the degree of formality. This analysis considers the PM values of genres representing formal language (academic genre) and informal language (TV/movie subtitles, fiction, and spoken genres) (Davies, 2021). To calculate the PM values

for informal genres, the token sizes of TV/movie subtitles, fiction, and spoken genres are combined. This allows for a comparison based on the total frequency of each near-synonym across all three informal genres. The results are presented in Table 2.

Table 2.

Degrees of Formality of the Three Near-Synonyms

Degree	Genre	Size	Tourist		Traveler		Visitor	
		(Mil)	Freq.	PM.	Freq.	PM.	Freq.	PM.
Formal	Academic	119.8	1,622	13.54	556	4.64	1,189	9.92
	Total	119.8	1,622	13.54	556	4.64	1,189	9.92
Informal	Fiction	118.3	1556	13.15	1,028	8.69	2,138	18.07
	TV/Movie	128.1	528	4.12	368	2.87	999	7.8
	Spoken	126.1	1094	8.68	300	2.38	533	4.23
	Total	372.5	3,178	8.53	1,696	4.55	3,670	9.85

As can be seen in Table 2, the word *tourist* appears more frequently in formal contexts (13.54) than in informal ones (8.53). This suggests that *tourist* is commonly used in academic fields, likely due to its association with the tourism industry, research articles, economic reports, and policy-related topics. The lower frequency in informal contexts implies that *tourist* may be perceived as a more official or technical term rather than a casual one.

The term *traveler* has the lowest PM among the three words in both formal (4.64) and informal (4.55) contexts. The minimal difference between the two suggests that *traveler* maintains a relatively stable usage regardless of formality. This might be due to its more personal and experience-based connotation, which is commonly used in storytelling, lifestyle writing, and discussions about independent or long-term journeys rather than structured tourism.

The word *visitor* shows nearly equal PM in formal (9.92) and informal (9.85) contexts, indicating that it is highly versatile and widely applicable in different registers. This balanced distribution suggests that *visitor* is a more neutral and general term, suitable for both professional and casual discussions. It may be used in contexts involving guests, temporary stays, or non-tourism-related visits, making it broader in scope compared to *tourist* and *traveler*.

Common Collocations

To examine the similarities and differences among the three near-synonyms in terms of collocation, the noun, verb, and adjective collocates are first identified. Collocates with an MI value of ≥ 3 and ranked among the top 20 most frequent occurrences are considered in this study, as shown in Table 3.

Table 3*Noun, Verb, and Adjective Collocates of Visitor, Tourist, and Traveler in COCA*

Tourist			Traveler			Visitor		
Noun	Verb	Adjective	Noun	Verb	Adjective	Noun	Verb	Adjective
<i>attraction</i>	visit	popular	business	<i>cater</i>	fellow	center	attract	unique
<i>destination</i>	attract	<i>foreign</i>	air	strand	<i>frequent</i>	park	draw	<i>foreign</i>
town	<i>flock</i>	accidental	check	attract	international	site	greet	<i>frequent</i>
industry	<i>cater</i>	curious	airport	encounter	weary	bureau	welcome	regular
local	lure	wealthy	airline	advise	seasoned	million	invite	first-time
<i>hotel</i>	crowd	nearby	leisure	lure	stranded	convention	entertain	monthly
<i>visa</i>	wander	visiting	travel	urge	experienced	museum	guide	unexpected
trap	<i>stroll</i>	<i>occasional</i>	<i>tourist</i>	time	geographic	<i>resident</i>	<i>flock</i>	welcome
shop	<i>frequent</i>	unsuspecting	<i>hotel</i>	book	intrepid	website	accommodate	casual
<i>resident</i>	gawk	crowded	<i>destination</i>	cash	registered	tour	lure	alien
trade	venture	<i>adventurous</i>	holiday	board	savvy	exhibit	<i>stroll</i>	interactive
bus	descend	attracting	vacation	venture	trusted	gallery	sample	distinguished
spot	flood	middle-aged	adventure	screen	<i>adventurous</i>	<i>tourist</i>	<i>beckon</i>	<i>occasional</i>
beach	swarm	thriving	solo	<i>entice</i>	avid	local	crowd	unwanted
visitor	overrun	bustling	diarrhea	<i>beckon</i>	sophisticated	exhibition	venture	out-of-town
<i>guide</i>	mill	lucrative	highway	depart	interstate	<i>visa</i>	escort	unwelcome
resort	jam	quaint	delay	crowd	veteran	<i>guide</i>	tour	overnight
boat	mingle	well-heeled	fare	detain	lone	log	immerse	welcoming
crowd	evacuate	overseas	local	beware	disabled	<i>attraction</i>	browse	extraterrestrial
traffic	cruise	burgeoning	passport	<i>frequent</i>	mysterious	stream	<i>entice</i>	overseas

While Table 3 presents the initial list of collocations extracted through the frequency-based approach, these automated statistical results include words that merely co-occur within the wide [4L, 4R] window frame. To ensure high pedagogical value, a manual qualitative screening of the concordance lines was performed. This step filtered the data using a phraseological approach, focusing exclusively on collocates that share a direct, adjacent grammatical relationship with the node words (such as direct adjective-noun modifiers or active verb-object phrases). Through this careful manual verification, structurally ambiguous or distant co-occurrences were removed. Out of the initial 180 statistically significant pairs, a total of 139 collocations were confirmed as structurally stable and valid phraseological combinations (see Table 4).

Table 4.*Part-of-Speech Classification of Valid Collocates for Tourist, Traveler, and Visitor under Phraseological Criteria*

Tourist			Traveler			Visitor		
Noun	Verb	Adjective	Noun	Verb	Adjective	Noun	Verb	Adjective
attraction	attract	popular	business	<i>cater</i>	fellow	center	attract	unique
destination	<i>flock</i>	<i>foreign</i>	air	strand	<i>frequent</i>	park	draw	<i>foreign</i>

Tourist			Traveler			Visitor		
Noun	Verb	Adjective	Noun	Verb	Adjective	Noun	Verb	Adjective
town	<i>cater</i>	accidental	check	attract	international	site	greet	<i>frequent</i>
industry	lure	curious	leisure	encounter	weary	bureau	welcome	regular
hotel	crowd	wealthy	holiday	advise	seasoned	convention	invite	first-time
visa	wander	visiting	vacation	lure	stranded	museum	entertain	monthly
trap	<i>stroll</i>	<i>occasional</i>	adventure	urge	experienced	website	guide	unexpected
shop	<i>frequent</i>	unsuspecting	solo	board	intrepid	exhibit	<i>flock</i>	welcome
trade	gawk	<i>adventurous</i>	diarrhea	venture	registered	gallery	accommodate	casual
bus	venture	middle-aged		screen	savvy	exhibition	lure	alien
spot	descend	well-heeled		<i>entice</i>	trusted	log	<i>stroll</i>	distinguished
beach	flood	overseas		<i>beckon</i>	<i>adventurous</i>	stream	<i>beckon</i>	<i>occasional</i>
<i>guide</i>	swarm			depart	avid		crowd	unwanted
resort	mill			crowd	sophisticated		venture	out-of-town
boat	jam			detain	interstate		escort	unwelcome
crowd	mingle			<i>frequent</i>	veteran		tour	overnight
traffic	evacuate				lone		immerse	extraterrestrial
					disabled		browse	overseas
					mysterious		<i>entice</i>	

Table 4 provides a comprehensive list of noun, verb, and adjective collocates associated with the near-synonyms *tourist*, *traveler*, and *visitor*. Note that bold words indicate words that collocate with all three near-synonyms, while italicized words indicate words that collocate with two near-synonyms. Examining these collocations allows for a clearer differentiation among the three words and aids L2 learners' understanding, particularly in ESP contexts such as English for hospitality.

Some collocates appear across all three near-synonyms. These include verbs such as *attract*, *crowd*, *lure*, and *venture*. An interesting observation is that four verbs can collocate with all three near-synonyms, whereas no adjective is found to collocate with all three. When considering collocates shared between two near-synonyms, the following results emerge. *Tourist* and *traveler* share three collocates, including verbs (*cater* and *frequent*), and the adjective *adventurous*. Similarly, *traveler* and *visitor* share three collocates, including the verbs *beckon* and *entice*, and the adjective *frequent*. The overlap between *tourist* and *visitor* is reflected in four collocates, including verbs (*flock* and *stroll*), and adjectives (*foreign* and *occasional*).

These findings highlight the close connection between *tourist*, *traveler*, and *visitor*, as words with similar meanings often share common collocates (Sinclair, 2004; Szudarski, 2018). However, when viewed holistically, the proportion of collocates that appear with at least two of the near-synonyms is relatively small (16 out of 102 collocation types). This emphasizes that although the three words are near-synonyms, they are not always interchangeable, particularly in terms of collocational usage.

Certain collocates appear exclusively with one of the near-synonyms, demonstrating their distinct meanings. For instance, *trap* (as in *tourist trap*) is unique to *tourist*, indicating a place that attracts many tourists but is often overpriced or designed primarily to make money from visitors rather

than providing an authentic or valuable experience. Similarly, *flood* (*tourist flood*) refers to an overwhelming influx of tourists, causing overcrowding and environmental or social strain. *Traveler* collocates with *diarrhea* (*traveler's diarrhea*), highlighting health risks associated with long-term or international travel, and *air* (*air traveler*), which refers to a passenger who travels by airplane for business, leisure, or other purposes. *Visitor* uniquely collocates with *bureau* (*visitor bureau*), reflecting its use in official or administrative contexts, and *regular* (*regular visitor*), which describes a visitor who frequently returns to a place, such as a tourist site, business, or event. For L2 learners, especially in ESP classrooms focusing on hospitality, recognizing these collocational patterns helps them select the most appropriate term in different scenarios. Understanding these distinctions enables learners to use the terms more precisely in professional and academic settings.

Semantic Preference and Semantic Prosody

After obtaining all 180 collocates through the frequency-based approach, these words were analyzed for semantic preference and semantic prosody using the USAS, as a collocation list helps identify the various associations of a target word and reveals its semantic preference. The program identifies the semantic categories of the collocates and determines whether their tone is positive, negative, or neutral. The analysis results are presented in Table 5.

Table 5.

Semantic Preferences and Prosody of Visitor, Tourist, and Traveler

	Tourist			Traveler			Visitor		
	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative
[ATTRACT]	attract, attracting, lure, popular, quaint			attract, beckon, entice, lure, sophisticated			attract, beckon, draw, entice, immerse, lure,		
[BUSINESS]	lucrative	industry, shop, trade			business				
[DESIRE]	thriving		accidental, unsuspecting	avid	book				unexpected, unwanted
[DOCUMENT]		visa			passport, registered			log, visa	
[ENTERTAINMENT]		adventurous, venture		leisure	adventure, adventurous, venture		entertain	venture	
[HEALTH]						diarrhea, disabled, weary			
[HELPING/HINDERING]	cater		trap	advise, cater, urge		detain	accommodate, escort, guide (x2)		
[IT & COMPUTING]								browse, website	
[KNOWLEDGE]				experienced, savvy, veteran		mysterious			
[MONEY]	wealthy, well-heeled				fare, cash				
[PEOPLE]		guide, visitor			board, fellow, tourist			tourist	

	Tourist			Traveler			Visitor		
	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative
[PLACE & LOCATION]	attraction	beach, destination, foreign, local, nearby, overseas, spot, town			air, airport, destination, geographic, international, interstate, local		attraction, center,	convention, exhibition, foreign, local, overseas, park, site	alien, extraterrestrial, out-of-town
[QUANTITY]		burgeoning, crowd (x2), flock, flood, mingle, swarm	bustling, crowded, descend, jam, overrun		crowd	lone, solo		crowd, flock, million, sample, stream	
[RESIDENCE & BUILDING]		hotel, resident, resort			hotel			bureau, gallery, museum, resident	
[SAFETY/DANGER]			evacuate	intrepid, trusted	check,	beware, strand, stranded			
[SIGHT]		curious, gawk, visit, visiting			encounter, screen			exhibit	
[SOCIALIZING]							greet, interactive, invite, welcome (x2), welcoming		unwelcome
[SPECIALNESS]							distinguished, unique	casual	
[TIME]	frequent		occasional	frequent (x2), holiday, vacation	seasoned, time	delay	frequent	first-time, monthly, overnight regular	occasional
[TRANSPORTATION]		boat, bus, cruise, traffic			airline				
[TRAVELING]		mill, stroll, wander			depart, highway, travel			stroll, tour (x2)	
[MISCELLANEOUS]		middle-aged							
Total (%)	12 (20.00%)	38 (63.33%)	10 (16.67%)	19 (31.67%)	30 (50.00%)	11 (18.33%)	22 (36.67%)	31 (51.67%)	7 (11.67%)

From Table 5, it is found that in terms of semantic preference among the three near-synonyms, there are both shared semantic domains and those that occur exclusively with certain near-synonyms. In the shared domain, all three words can be used with vocabulary related to [ATTRACT], [DESIRE], [ENTERTAINMENT], [PLACE & LOCATION], [QUANTITY], [TIME], and [TRAVELING], among others. This finding is unsurprising, as all three terms refer to individuals engaged in movement, whether for leisure, business, or other purposes. However, despite significant similarities in semantic preference, the analysis reveals distinct nuances (Hornby, 2014). For instance, only *traveler* collocates with vocabulary in the [HEALTH] domain, which may be attributed to the fact that *traveler* is often associated with extended or independent journeys, where concerns about health risks, vaccinations, and medical preparedness are more relevant. In contrast, *visitor* uniquely collocates with vocabulary in [IT & COMPUTING], [SOCIALIZING], and [SPECIALNESS]. This may reflect the fact that visitor is commonly used in digital and social settings, such as website visitors or guests at events, highlighting its broader contextual applicability beyond physical travel.

While the USAS framework provides a valuable and highly detailed automated semantic analysis, these abstract categories can sometimes be too technical for direct classroom use. To make these findings more practical for ESL/ESP teachers and learners, this study further filters the data to focus specifically on tourism-related semantic domain preferences. Table 5 reorganizes the key collocates that appear frequently within actual tourism contexts, grouping them into distinct communicative themes.

Table 6.

Tourism-Related Semantic Domain Preferences and Key Collocates

Tourism-Related Semantic Domain	Word	Key Collocates
[STRUCTURED INDUSTRY & COMMERCIAL LEISURE]	Tourist	<i>attraction, resort, destination, trap, visa, industry</i>
[PERSONAL MOBILITY & EXPERIENTIAL EXPLORATION]	Traveler	<i>adventure, seasoned, savvy, solo</i>
[INSTITUTIONAL ACCESS & TEMPORARY ENGAGEMENT]	Visitor	<i>center, gallery, museum, guide</i>

This thematic framing highlights a clear conceptual distinction that is highly useful for ESP pedagogy. Instead of looking at abstract linguistic codes, students can see how word choice reflects a speaker's communicative intent and interpersonal stance. As shown in Table 6, *tourist* has a strong semantic preference for structured tourism systems and commercialized leisure. Its collocates often relate to organized spaces (e.g., *resort, destination*) or industry structures (e.g., *industry*). On the other hand, *traveler* shows a clear preference for domains related to personal agency, deep exploration, and individual experience. This is evident from collocates like *adventure* and *savvy*, which imply a higher degree of intimacy with the journey. Finally, *visitor* displays a preference for domains connected to institutional access and temporary stays, often appearing with words related to controlled environments (e.g., *center, gallery, museum*). Introducing this framework in the ESP classroom allows learners to move beyond simple dictionary definitions. Teachers can guide students to select the right term based on whether they want to emphasize participation in a tourism system (*tourist*), a deeper personal exploration (*traveler*), or a temporary, purpose-neutral visit (*visitor*).

Moving on to semantic prosody, overall, the three synonyms—*tourist, traveler, and visitor*—are predominantly used in neutral contexts, as indicated by the proportions of occurrence with neutral words at 63.33%, 50.00%, and 51.67%, respectively. One possible reason is that the three near-synonyms are nouns; therefore, their collocating words are so widespread that no specific meaning can be determined (Jarunwaraphan & Mallikamas, 2020). Clearly, *tourist* is the most frequently used in neutral contexts. This could be due to the fact that tourism is a well-established and structured industry,

with tourist generally carrying a neutral or commercially oriented connotation. Additionally, the definition of neutral in this analysis implies that these words do not inherently convey positive or negative sentiment but can be used in both contexts depending on the situation. When focusing specifically on positive and negative associations, all three words are less frequently used in negative contexts. Among them, *visitor* has the lowest co-occurrence with negative words at 11.67%, contrasting with its co-occurrence with positive words at 36.67%. This suggests that *visitor* is more likely to be used in welcoming contexts, such as museum or event visitors, where the sentiment is generally positive. In contrast, *tourist* shows a smaller gap between positive (20.00%) and negative (16.67%) contexts, reinforcing its more neutral usage. The relatively balanced distribution of positive and negative associations for *tourist* may be attributed to the varying perceptions of tourism; while it brings economic benefits, it can also be linked to environmental concerns, overcrowding, and cultural disruption. This suggests that although nouns, including *tourist*, *traveler*, and *visitor* in this study, may differ in terms of distribution across genres, degree of formality, collocation, or semantic preference, they may share the same level of semantic prosody (Hou, 2023). To make these findings easier to understand, Table 6 provides a clear, integrated summary. It highlights the core shared meaning of these three near-synonyms, their key conceptual differences, and how they behave in the corpus.

Table 7

An Integrative Summary of Core Meanings and Distinct Features of Tourist, Traveler, and Visitor

Dimension	Tourist	Traveler	Visitor
Core Shared Meaning	Individual temporarily outside their usual environment	Individual temporarily outside their usual environment	Individual temporarily outside their usual environment
Conceptual Orientation	Institutionally framed & consumption-oriented: Closely linked to the commercial structures of the tourism industry	Experience-oriented & individualistic: Associated with personal mobility, adventure, and exploration	Functionally broad & purpose-neutral: Highly context-flexible, encompassing leisure, business, and digital realms
Dominant Genres (COCA)	News, Academic, Fiction, and Spoken	Fiction and Magazines	Fiction, Magazines, Web, and TV/Movies
Degree of Formality	Primarily formal contexts	Stable usage across both formal and informal contexts	Balanced and highly versatile in both formal and informal contexts

Dimension	Tourist	Traveler	Visitor
Exclusive Semantic Domains (USAS)	[BUSINESS], [TRANSPORTATION]	[HEALTH], [KNOWLEDGE]	[IT & COMPUTING], [SOCIALIZING], [SPECIALNESS]
Semantic Prosody Tendency	Predominantly neutral, with a balanced distribution of positive and negative associations	Predominantly neutral, reflecting a focus on stable, experience- based narratives	Significantly positive, frequently occurring in welcoming or appreciative contexts

As synthesized in Table 7, even though these three words share a basic meaning, they have very different usage patterns, collocations, and emotional tones. These corpus-based findings do more than just repeat what we already know; they explicitly extend and add nuance to the traditional dictionary definitions and CEFR classifications discussed earlier in this paper.

First, while traditional dictionaries often provide overlapping and circular definitions for these three words, the corpus data helps separate them by revealing specific functional, experiential, and institutional boundaries. For example, *tourist* is clearly extended beyond a simple definition to show a heavy institutional and commercial focus. *Traveler* is shown to move away from mere movement toward a highly experiential and individualistic meaning. Meanwhile, *visitor* is revealed as a highly flexible, purpose-neutral word that adapts to various contexts, including digital environments. These specific differences are rarely captured in traditional reference books.

Second, these findings offer a new way to look at the CEFR level classifications (A1 for *tourist* and *visitor*, and A2 for *traveler*). While CEFR ranks *traveler* as a higher-level word (A2), the corpus data challenges this simple view. The data shows that *traveler* actually has a stable and predictable usage pattern across genres. On the other hand, *visitor* is used in highly diverse semantic domains and complex digital contexts. This suggests that the complexity of actual language usage does not always align perfectly with conventional CEFR levels. By making these connections clear, this study demonstrates how corpus linguistics can validate and refine traditional lexical tools, providing much more precise guidance for ESL and ESP classrooms.

To see how these differences look in real-world English, you can review the authentic corpus examples (1–15) listed below.

1. Most countries will not authorize to work on **tourist visa** for short term. [DOCUMENT-Neutral]
2. The spectacle not only draws **curious tourists** but also hundreds of bald eagles. [SIGHT-Neutral]
3. They try to **lure tourists** to New York and then tourists have nice pocketbooks. [ATTRACT-Positive]

4. Balinese residents have developed strategies to profit from the **tourist industry** while maintaining cultural integrity. [BUSINESS-Neutral]
5. It's a frustrating reality for a team that had the look of **accidental tourists** venturing into Mercury Morris'. [DESIRE-Negative]
6. My **fellow traveler** had practically danced down the stairs. [PEOPLE-Neutral]
7. Noah offers temporary respite for the **weary traveler**. [HEALTH-Negative]
8. You are a **savvy traveler** and it sounds like this piece does not resonate with you in a big way. [KNOWLEDGE-Positive]
9. **International travelers** have their baggage examined by U.S. Customs. [PLACE & LOCATION-Neutral]
10. The Americans with Disabilities Act do for **disabled travelers** on jets. [HEALTH-Negative]
11. The boardwalk, masonry pathways and canopy trees **immerse** the **visitors** in the surrounding garden. [ATTRACT-Positive]
12. He's a **regular visitor** and has even written the historical book of what lies between his past and present. [TIME- Neutral]
13. The report said 83 percent of **first-time visitors** said they would return. [TIME- Neutral]
14. Recently they showed up as **unwelcome visitors** in Orange County. [DESIRE-Negative]
15. To most **foreign visitors**, Tokyo looks uncannily familiar. [PLACE & LOCATION-Neutral]

Pedagogical Implications

Teachers can use these research findings to teach students studying English for the service industry or English for the tourism business. The simplest method is for teachers to create fill-in-the-blank exercises, where students complete sentences by filling in the missing words. This exercise helps students understand and use the most appropriate near-synonyms in given contexts (Popham, 2011). It is also possible to integrate corpus-based learning into this teaching approach by selecting sentences from COCA and having students choose the most appropriate word among the three near-synonyms, ensuring consistency with the findings of this research. For example:

Look at the sentences and fill in the blanks with the correct word: *tourist*, *traveler*, or *visitor*.

1. American _____ **encounter** the mystery when traveling to Cuba, wondering whether it is legal to go there in the first place. (Answer: *traveler*)
2. Varna is one of the most popular **resorts**, with _____ arriving to enjoy the beach. (Answer: *tourist*)
3. A **lone** _____ was walking through the village and happened to peer inside a house. (Answer: *traveler*)
4. The Crested Butte Mountain Resort **welcomes** summertime _____ in conjunction with the annual Crested Butte Wildflower Festival. (Answer: *visitor*)

Conclusion

This study provides valuable insights into the specialized use of the near-synonyms *tourist*, *traveler*, and *visitor*, offering a corpus-based analysis that enhances our understanding of their contextual differences. By utilizing the COCA, the study systematically examines genre distribution, degrees of formality, collocational patterns, semantic preference, and semantic prosody to highlight the distinct usage of these terms. The findings reveal that while the three words share some overlapping meanings, their distribution across genres, formal and informal contexts, and collocational tendencies demonstrate significant differences that should not be overlooked.

The results indicate that *tourist* is predominantly used in structured and academic discourse, frequently appearing in news articles and research publications related to tourism. *Traveler*, on the other hand, is more commonly associated with storytelling and lifestyle contexts, often evoking a sense of adventure and long-term exploration. *Visitor* is the most versatile of the three, used in both professional and casual discussions, encompassing a broader range of scenarios beyond tourism, including business and digital interactions.

Furthermore, the study underscores the importance of considering collocational and semantic aspects when teaching vocabulary, particularly in ESP contexts. By incorporating data-driven approaches such as corpus-based analysis, educators can provide learners with precise and context-sensitive instruction, enabling them to use these near-synonyms more accurately and effectively. The findings also highlight the need for further research on near-synonym distinctions across specialized domains, emphasizing the role of corpus linguistics in refining vocabulary instruction for L2 learners in the tourism industry and beyond.

Limitations of the Study and Recommendations for Future Research

The first limitation is that this study used COCA, which focuses on American English. As a result, the findings may not be applicable to other varieties of English. To clarify, this study examined *traveler*, which is used in American English, while in the inner-circle language, i.e., British English, *traveller* is used. Another limitation is that the researchers limited the word collocates to only the top 20 nouns, verbs, and adjective collocates with MI score values ≥ 3 . This may have excluded additional word collocates that could better reflect the results of the data for these three target words.

Since this study focused on data from COCA, which primarily represents American English, future research could incorporate other corpora, such as the British National Corpus (BNC). This would allow for a comparative analysis of *tourist*, *traveler*, and *visitor* across different English varieties, thereby providing a broader perspective on their usage patterns. Additionally, future studies could examine the British English equivalent of *traveler* to determine whether its collocational behavior differs significantly from its American English counterpart. Lastly, future research could extend the scope by considering a larger set of collocates, both in terms of quantity (>20) and word classes.

References

- Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
- Davies, M. (2021). *The new Corpus of Contemporary American English (COCA 2021)*.
<https://www.english-corpora.org/coca/>
- Dimarco, C., Hirst, G., & Stede, M. (1993). The semantic and stylistic differentiation of synonyms and near-synonyms. In *AAAI Spring Symposium on Building Lexicons for Machine Translation*, 114–121, Stanford, California.
- Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics*, 28(2), 105–144.
- Flowerdew, L. (2012). *Corpora and Language Education*. Palgrave Macmillan.
- Flowerdew, L. (2013). Corpus-based discourse analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 174–188). Routledge.
- Hornby, D. (2014). *Linguistics: A complete introduction*. Hachette.
- Hou, S. (2023). A Comparative Study of Synonyms Semantic Prosody of Effect, Impact and Influence Based on Corpus. *International Journal of Linguistics, Literature and Translation*, 6(11), 131–136.
- Hunston, S. (2002). *Corpora in applied linguistics* (1st ed.). Cambridge University Press.
- Imsa-ard, P., & Phoocharoensil, S. (2022). “A whole new world... Wait, is it a whole, entire, or total world?”: The extraction of collocations for the three English synonym discrimination. *3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies*, 28(2), 67–82.
- Jackson, H., & Amvela, E. Z. (2007). *Words, meaning and vocabulary: An introduction to modern English lexicology*. (2nd ed.). Bloomsbury Publishing.
- Jarunwaraphan, B., & Mallikamas, P. (2020). A corpus-based study of English synonyms: Chance and Opportunity. *rEFLECTIONS*, 27(2), 218–245.
- Laosrirattanachai, P., & Laosrirattanachai, P. (2025). Unveiling the distinction of near synonymy: A Corpus-based analysis on attempt, endeavor, strive, and try. *PASAA*, 70, 132–163.
- Laosrirattanachai, P., & Phoocharoensil, S. (2025). Differentiating the near-synonyms using a corpus-based approach: booking vs reservation. *TEFLIN Journal*, 36(2). 350–371.
<http://dx.doi.org/10.15639/teflinjournal.v36i2/350-371>
- Laosrirattanachai, 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.

- Lertcharoenwanich, P. (2023). Analysis of Collocations and Semantic Preference of the Near-Synonyms: “Blank”, “Empty”, and “Vacant”. *LEARN Journal: Language Education and Acquisition Research Network*, 16(1), 365–383.
- Lindquist, H. (2009). A corpus study of lexicalized formulaic sequences with preposition + hand. In R. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (Eds.), *Formulaic language. Volume 1.: Distribution and historical change* (pp. 239-256). John Benjamins.
- Liu, D., & Espino, M. (2012). Actually, genuinely, really, and truly: A corpus-based behavioral profile study of near-synonymous adverbs. *International Journal of Corpus Linguistics*, 17(2), 198–228.
- Murphy, M. L. (2010). *Lexical meaning*. Cambridge University Press.
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.
- Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. *Eurasian Journal of Applied Linguistics*, 7(1), 240–258.
- Phoocharoensil, S. (2022). Primary, main, and major: learning the synonyms through corpus data. *GEMA Online® Journal of Language Studies*, 22(4), 76–89.
- Phoocharoensil, S., & Kanokpermpoon, M. (2021). Distinguishing the near-synonyms ‘increase’ and ‘rise’: Genre and collocation investigation. *Kasetsart Journal of Social Sciences*, 42(4), 968–975.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator’s confession. *The Teacher Educator*, 46(4), 265–273.
- Rayson, P., Archer, D., Piao, S. L., McEnery, T. (2004). The UCREL semantic analysis system. In *proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 25th May 2004, Lisbon, Portugal, pp. 7–12.
- Sinclair, J. M. (Ed.) (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
- Sumonsriworakun, P. (2022). A Corpus-Based Investigation of English Synonyms: “Disadvantage”, “Downside”, and “Drawback”. *LEARN Journal: Language Education and Acquisition Research Network*, 15(2), 649–678.
- Supanfai, P. (2022). People or persons?: A corpus-based study. *rEFlections*, 29(3), 603–620.
- Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. Routledge.
- Watson Todd, R. (2018). Exploring the direction of collocations in eight languages. *Canadian Journal of Linguistics*, 64(1), 146–154.
- Yang, B. (2016). A corpus-based comparative study of learn and acquire. *English Language Teaching*, 9(1), 209–219.

Dr. Piyapong Laosrirattanachai is an Associate Professor of English at the Department of Service Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, Kamphaeng Saen, Thailand. His research studies focus on corpus-based studies, teaching and learning vocabulary, word list development, and move analysis.

ORCID: <https://orcid.org/0000-0002-9283-5202>

Phongsathorn Thongsawat is studying in the English for Service Industry program at the Department of Service Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, Kamphaeng Saen, Thailand. Their research interests comprise English for the tourism industry and vocabulary.

ORCID: <https://orcid.org/0009-0006-8445-5273>

Montira Somsri is studying in the English for Service Industry program at the Department of Service Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, Kamphaeng Saen, Thailand. Their research interests comprise English for the tourism industry and vocabulary.

ORCID: <https://orcid.org/0009-0001-5299-0814>

Atittiya Obchoey is studying in the English for Service Industry program at the Department of Service Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, Kamphaeng Saen, Thailand. Their research interests comprise English for the tourism industry and vocabulary.

ORCID: <https://orcid.org/0009-0002-2617-7528>

Piyanuch Laosrirattanachai currently serves as an Assistant Professor in the English for Service Industry program at the Department of Service Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, Kamphaeng Saen, Thailand. Her research interests include corpus-based studies, English for tourism and hospitality businesses, and collocations.

ORCID: <https://orcid.org/0000-0002-6261-0176>