# Lexical Collocations in a Sample Corpus of Tourism Research Articles (SCTRA)

# คำปรากฏร่วมในคลังข้อมูลภาษาตัวอย่างในบทความวิจัยด้านการท่องเที่ยว

**กันตพัฒน์ ตรินันท์** **วชิราภรณ์ กิจพูนผล**

**Kantapat Trinant** **Wachiraporn Kijpoonphol**

คณะศิลปศาสตร์ มหาวิทยาลัยอุบลราชธานี

kantapat.t@ubu.ac.th and wachiraporn.k@ubu.ac.th

**Received:** March 6, 2021

**Revised:** June 17, 2021

**Accepted:** July 6, 2021

**Abstract**

This corpus-based study was conducted in an attempt to facilitate the teaching and learning of English for Specific Purposes (ESP) and benefit Thais who work in the field of tourism by exploring lexical collocations in Tourism Research Articles. A Sample Corpus of Tourism Research Articles (SCTRA), with a corpus size of over 1.5 million running words, was compiled from 240 research articles from six journals in the field of tourism and hospitality management. Nine hundred and thirty-two keywords were obtained under the set criteria using the AntConc version 3.5.8. The majority of the keywords were nouns (62.72%), followed by adjectives (18.37%), verbs (16.37%), and adverbs (2.53%) respectively. The keywords were then used as "nodes" to find their "collocates" which generated 2,989 pairs of lexical collocations with 16 combination types, where six combination types were in accordance with the set framework adapted from Benson, Benson and Ilson [1]. The majority of them were Noun + Noun (44.76%), Adjective + Noun (25.89%), and Noun + Verb (9.33%) respectively. The lists of the keywords and the collocations produced were also provided.

**Keywords**: **Corpus, corpus-based study, lexical collocations, tourism, tourism research articles**

**บทคัดย่อ**

การศึกษาคลังข้อมูลภาษานี้จัดทำขึ้นโดยความพยายามที่จะอำนวยความสะดวกให้กับการเรียนและการสอนภาษาอังกฤษ เพื่อวัตถุประสงค์เฉพาะและเป็นประโยชน์ต่อคนไทยที่ทำงานด้านการท่องเที่ยว โดยการสำรวจคำปรากฏร่วมในบทความวิจัยด้านการท่องเที่ยว คลังข้อมูลภาษาตัวอย่างในบทความวิจัยด้านการท่องเที่ยว (SCTRA) ซึ่งมีขนาดของคลังข้อมูลภาษามากกว่า 1.5 ล้านคำ ที่ได้รับการรวบรวมมาจากบทความ จำนวน 240 บทความ จากวารสารด้านการท่องเที่ยวและการบริหารกิจการการให้บริการ จำนวน 6 วารสาร มีการใช้โปรแกรมจัดการคลังข้อมูลภาษาชื่อ AntConc เวอร์ชั่น 3.5.8 ทำให้พบคำสำคัญ จำนวน 932 คำ คำสำคัญที่พบส่วนใหญ่เป็นคำนาม (62.72%) รองลงมาได้แก่ คำคุณศัพท์ (18.37%) คำกริยา (16.37%) และคำกริยาวิเศษณ์ (2.53%) จากนั้นได้นำคำสำคัญไปใช้เป็นคำหลักเพื่อหาคำปรากฏร่วม ซึ่งพบคำปรากฏร่วม จำนวน 2,989 คู่ จากคำปรากฏร่วมทั้งหมด 16 แบบ ซึ่งรูปแบบคำปรากฏร่วมจำนวน 6 รูปแบบที่พบนั้นสอดคล้องกับกรอบงานวิจัยที่ปรับปรุงมาจากของ Benson, Benson and Ilson [1] โดยคำปรากฏร่วมที่พบมากที่สุดคือคำปรากฏร่วมรูปแบบ คำนาม + คำนาม

(44.76%) คำคุณศัพท์ + คำนาม (25.89%) และคำนาม + คำกริยา (9.33%) ตามลำดับ ซึ่งได้มีการแสดงรายการคำสำคัญและคำปรากฏร่วมที่พบไว้ในงานวิจัยนี้

**คำสำคัญ: คลังข้อมูลภาษา การศึกษาคลังข้อมูลภาษา คำปรากฏร่วม การท่องเที่ยว บทความวิจัยด้านการท่องเที่ยว**

**1. Introduction**

For many years, tourism has been one of the most significant sectors since it has been considered by the Thai government as one of the most important sectors for income generation [2]. The income is mainly from English speaking countries or the countries that Thais use English to communicate with. Chulaphan and Barahona [2] mentioned that the tourist arrivals from Southeast Asia, i.e., India, were observed to lead to economic growth in Thailand. Since tourism sector has been a major source of income for Thailand, it is important for students in the field of tourism and hospitality to foster their competence in the use of English language before entering into their professional service arena. One important aspect to prepare them for reaching such goal, apart from other skills in English and beyond the knowledge of grammar and vocabulary, is the knowledge of collocations.

Although collocation has been recognized as an important key to fluency, it has been a huge barrier for EFL/ESL learners to achieve a native-like fluency [3]. A number of studies on collocations have revealed that even high-level learners seem to be struggling with using and developing second language (L2) collocational knowledge [4]. Collocations have also been found to be the most frequent mistakes made by L2 learners. Additionally, collocational mismatches generally occur among L2 learners [5]. As a result of inadequate exposure to the language, the learners hardly encounter a word or combination of words to enable them to determine its range or narrow the item down to its more fixed partnerships. This difficulty mostly occurs with those semantic opaque combinations when the combination of words leads to the change of the original meaning or with specific field of the discourse. Moreover, it is evident that the study of collocations in the field of tourism, especially in research articles, is limited.

**2. Collocation**

There are various definitions of collocation given by a number of scholars [3], [6], [7], [8]. The similarities among them are mainly the combinations of two or more words. The varieties among them are those additional details given such as the co-occurrence happens repeatedly, naturally, predictably, structurally or syntagmatically, and selectively. However, it may be concluded that collocation is the co-occurrence of words which is predictable as they have tendency to occur together and are naturally co-selected by native speakers. Lexical collocation then refers to the co-occurrence of two lexical words which tend to occur naturally with statistical significance in terms of their association.

According to Schmitt [8] and Bahns [9], there are two basic types of collocations: syntactic/ grammatical collocations and semantic/ lexical collocations. The former are those collocations where a dominant word co-occurs with a grammatical word. The examples are *abide by*, *access to*, and *acquainted with*. The latter, however, usually

are those collocations which emerged from two equal words such as Noun + Verb, (*e.g., ball bounces*), Verb + Noun (e.g., *spend money*), and Adjective + Noun (*e.g., cheerful expression*). In the BBI Combinatory Dictionary of English, Benson, Benson and Ilson [1] view collocations as 'fixed, identifiable, non-idiomatic phrases and constructions' (p. xix). Similar to Schmitt [8] and Bahns [9], they categorize collocations into two major groups: grammatical and lexical collocations.

Lexical collocation framework for this research was adapted from that of Benson, Benson and Ilson [1] as it covered important pairs of lexical collocations. The six lexical collocation pairs adapted were: (1) Verb + Noun; (2) Adjective + Noun; (3) Noun + Verb; (4) Noun + Noun; (5) Adverb + Adjective; and (6) Verb + Adverb.

In terms of lexical collocations, there have been some studies conducted to investigate lexical collocations in various sources. Ackermann and Chen [10] built a corpus of 28 academic disciplines from journal articles and textbook chapters in order to develop the Academic Collocation List (ACL). The corpus comprises 25.6 million words produced 2,468 entries of lexical collocations with the majority combination being Adjective + Noun, Verb + Noun, Adverb + Adjective respectively. Molavi, Koosha, and Hosseini [11] examined lexical collocations used in three EFL textbooks: Interchange, American Headway, and American File. The results revealed that the majority of the lexical collocation found being Verb + Noun and Adjective + Noun from the total number of 362. Borucinsky and Kegalj [12] investigated collocations in marine engineering English which found that Verb + Noun and Adjective + Noun were the dominant combinations of collocations. Similarly, Demir [13] explored the use of lexical collocations in English language teaching articles among Anglophobic and Turkish writers. It revealed that articles written by native speakers of English used greater number of lexical collocations. Among the total number of 1,548 lexical collocations used, the majority of them being Adjective + Noun, Verb + Noun, Adverb + Adjective, Noun + Noun respectively.

In Thailand, Thongvitit and Thumawongsa [14] investigated the use of collocations in the abstracts of research articles in the field of liberal arts and humanities written by Thai EFL writers. They found that the majority of lexical collocations used were Adjective + Noun, Verb + Noun, and Noun + Verb respectively. Trinant and Yodkamlue [15] studied lexical collocations in a Sample Corpus of Nursing Research Articles (SCNRA). The study revealed that the majority of them were Noun + Noun, followed by Adjective + Noun, and Noun + Verb respectively.

This research article then attempted to facilitate the learning and teaching of lexical collocations in the field of tourism and hospitality and to be benefits to Thai people who work as tourist guides or other professionals in the field of tourism by building a Sample Corpus of Tourism Research Articles (SCTRA) to identify and classify keywords in the SCTRA as well as to identify and classify lexical collocations in the corpus using keywords extracted as 'nodes' to find their 'collocates'. Since lexical collocations are the co-occurrence of two content words which closest to single word vocabulary, the focus of this study was on this type of collocation. The research questions, therefore, were:

(1) What were the keywords in the SCTRA based on the frequency of occurrence at $\geq$ 100 and the keyness value at $\geq$ 100? What was the proportion according to their parts of speech? and

(2) What were the lexical collocations of those keywords in the SCTRA? What was the proportion according to each type of combinations?

## 3. Research methodology

### 3.1 Research instruments

#### 3.1.1 The corpus: tourism journals and tourism research articles

Two hundred and fourty research articles were compiled from six academic journals in the field of tourism and hospitality accessible via Ubon Ratchathani University's Main Library. The six journals were purposefully selected based on their availability and accessibility at the time of study. Fourty research articles with the IMRD format and the length of at least 3,000 words were selected from the latest issues of each selected journal. These journals were: (1) Tourism Management; (2) Annas of Tourism Research; (3) Journal of Hospitality, Leisure, Sport & Tourism Education; (4) Journal of Hospitality and Tourism Management; (5) Journal of Outdoor Recreation and Tourism; and (6) Tourism Management Perspectives. Research articles collected were mostly published in 2019.

#### 3.1.2 A corpus analysis tool

The AntConc version 3.5.8 [16] developed by Laurence Anthony was used in this research. The AntConc is a computer-based freeware corpus analysis toolkit for concordancing and text analysis. The AntConc is also equipped with features for extracting keywords and collocations that appropriately serve the purpose of the study.

### 3.2 Procedures

#### 3.2.1 Identifying keywords

The keyness value and the frequency of occurrence were used as criteria for determining the keywords in the SCTRA. According to Baker [17], the higher the keyness score, the stronger the keyness of that word. The frequency of occurrence indicates how common the word is in the corpus. For the present study, the frequency of occurrence at $\geq$ 100 and the keyness value at $\geq$ 100 were applied.

In terms of keyness value, the British National Corpus (BNC) was used as the reference corpus as it is one of the largest corpora of general English and commonly used as a reference corpus in corpus-based studies as well as the word list is readily available online. With the BNC as the reference corpus, when the corpus tool processes for keywords in the SCTRA, all the running words in the SCTRA are compared with those words listed in the reference corpus, the BNC. The comparison of words from the two corpora results in the keyness value of the keywords found in the SCTRA. The higher the keyness value a word has, the higher uniqueness it belongs to the SCTRA.

**3.2.2 Identifying collocations**

The keywords earlier extracted were then used as 'nodes' to further identify their collocation pairs. The criteria for identifying the collocations were as follow:

1) The range of the collocates was within the 3-word span on the right of the node. This was to explore collocates that occurred after the nodes within this range of word span since this range was not too close and not too far for each pair to co-occur.

2) Only the lexical collocations were selected.

3) The association strength of each pair was measured on the basis of Mutual Information (MI) with the MI score set for the present study was $\geq 5$.

4) The frequency of occurrence of the pair was $\geq 10$.

5) In case, under the above set criteria, if there was no applicable collocates with the MI value $\geq 5$, the collocate with the highest MI value would be chosen.

6) In case, if there was no applicable collocates even with the MI value $\leq 5$, the reduction of the frequency of occurrence would be applied.

## 4. Results and Discussion

### 4.1 Keywords in the SCTRA

Under the set criteria, 932 keywords were extracted from the SCTRA. Most of the keywords generated have very high keyness value. The keyword with the highest keyness value is '*tourism*' with the keyness value at 61,173.73, while the lowest keyness value was '*frequently*' with the keyness value at 100.02, which was the lowest keyness value set. Each of the 932 keywords was then used as the node to identify their collocates. After the process of identifying collocates of the keywords had been completed, it was possible to categorize the keywords according to their parts of speech as shown in Table 1.

**Table 1** Keywords Generated from the SCTRA According to their Parts of Speech

| No. | Parts of speech | No. of keywords | Percentage |
|---|---|---|---|
| 1 | Noun | 594 | 62.72 |
| 2 | Verb | 155 | 16.37 |
| 3 | Adjective | 174 | 18.37 |
| 4 | Adverb | 24 | 2.53 |
| **Total** | | **947** | **100** |

It was noticeable from Table 1 that the total number of the keywords increased from 932 to 947. This was because along the process of identifying collocation pairs, some of the keywords functioned more than one part of speech. Keywords such as '*increase*' and '*fit*' were found functioning as nouns and verbs as well.

Among these keywords, the majority of them were the nouns ($N$ = 594), accounted for 62.72 percent. The adjectives ($N$ = 174) came second accounted for 18.37 percent. The verbs ($N$ = 155) are 16.37 percent. The smallest number among them was the adverbs ($N$ = 24), only 2.53 percent.

The prevalent number of content words such as nouns, adjectives, and verbs in the keywords seemed to be common in all corpora. This finding could be found in the studies conducted by other scholars in the field such as Coxhead's [18] Academic Word List; Ward's [19] Basic Engineering List generated from his Engineering Corpus; Yang's [20] Nursing Academic Word List; and Trinant and Yodkamlue's [15] SCNRA.

**4.2 Lexical collocations in the SCTRA**

The process of identifying the lexical collocations commenced once the keywords from the SCTRA had been identified. The summary of the findings was shown in Table 2.

**Table 2** Number of Lexical Collocations in SCTRA According to Types of Combination

| No. | Combination Types | No. of collocation pairs | (%) |
|---|---|---|---|
| 1 | Noun + Noun | 1,338 | 44.76 |
| 2 | Adjective + Noun | 774 | 25.89 |
| 3 | Noun + Verb | 279 | 9.33 |
| 4 | Verb + Noun | 211 | 7.06 |
| 5 | Noun + Adjective | 140 | 4.68 |
| 6 | Adjective + Adjective | 72 | 2.41 |
| 7 | Verb + Adjective | 50 | 1.67 |
| 8 | Adverb + Adjective | 26 | 0.87 |
| 9 | Adverb + Verb | 25 | 0.84 |
| 10 | Verb + Verb | 20 | 0.67 |
| 11 | Adverb + Noun | 17 | 0.57 |
| 12 | Adjective + Verb | 16 | 0.54 |
| 13 | Verb + Adverb | 11 | 0.37 |
| 14 | Noun + Adverb | 6 | 0.20 |
| 15 | Adverb + Adverb | 3 | 0.10 |
| 16 | Adjective + Adverb | 1 | 0.03 |
| | **Total** | **2,989** | **100** |

Table 2 showed that there were 2,989 collocation pairs generated from the keywords earlier extracted. The majority of the collocation pairs was the 'Noun + Noun' combinations with 1,338 pairs (44.76%). The 'Adjective + Noun' combination came second with 774 pairs (25.89%). The 'Noun + Verb' combination came third with 279 collocation pairs (9.33%). The 'Adjective + Adverb' generated the least number of combinations at only one pair (0.03%). Moreover, examples of collocation pairs of each combination type were shown in Table 3.

**Table 3** Examples of Lexical Collocations Extracted from the SCTRA

| Nodes | Collocates | Examples |
|---|---|---|
| Noun | Verb | results showed, findings suggest, participants (*were*) asked, tourist visit, employees need, willingness (*to*) engage, tourists seek |
| | Adjective | parks (*and*) natural, development (of) rural, employee psychological, customers prior, weather (*and*) climatic |
| | Adverb | scale (*was*) strongly, leadership (*was*) positively |
| Adjective | Noun | outdoor recreation, social media, sustainable tourism, local community, negative impacts, social responsibility |
| | Adjective | positive experiential, social (*and*) environmental, sustainable (*and*) responsible, important (*and*) memorable |
| | Verb | important (*to*) note, in-depth (*interview was*) conducted, likely (*to*) visit, similar (*cases were*) reported |
| Verb | Noun | perceived value, provide information, create value, improve quality, attract tourists, affect satisfaction |
| | Adjective | considered important, implement green, provide valuable, identify potential, engage (*in*) digital, develop sustainable |
| | Verb | conducted (*to*) examine, used (*to*) measure, visit (*and*) revisit |
| | Adverb | focuses only, noted earlier, located near, communicate effectively |
| Adverb | Verb | positively affects, negatively impact, frequently visited |
| | Adjective | positively predictive, statistically higher, highly competitive |
| | Noun | especially (*th*ose) countries, monthly income |

Altogether, there were 16 combination types of lexical collocations found in the SCTRA under this investigation. This means that there were more combination types than the set framework of the study. Therefore, the findings shown in Table 2 above could be categorized into two groups: lexical collocations with combination types according to the set framework (combination types numbers: 1. Noun + Noun; 2. Adjective + Noun; 3. Noun + Verb; 4. Verb + Noun; 8. Adverb + Adjective; and 13. Verb + Adverb) and lexical collocations with combination types not according to the set framework (combination types numbers: 5. Noun + Adjective; 6. Adjective + Adjective; 7. Verb + Adjective; 9. Adverb + Verb; 10. Verb + Verb; 11. Adverb + Noun; 12. Adjective + Verb; 14. Noun + Adverb; 15. Adverb + Adverb; and 16. Adjective + Adverb). Number of each collocation pairs and its percentage were shown in Table 2 and examples of Lexical Collocations extracted from the SCTRA were shown in Table 3.

The summary of the two groups was shown in Table 4.

**Table 4** Lexical Collocations According to Framework and Not According to Framework

| **Combination Types** | **No. of collocation pairs** | **(%)** |
|---|---|---|
| According to Framework | 2639 | 88.29 |
| Not According to Framework | 350 | 11.71 |
| **Total** | **2989** | **100** |

Table 4 revealed that over 88 percent of the lexical collocations extracted from the SCTRA were in accordance with the set framework, while the lexical collocations with combination types that were not in accordance with the set framework were just less than 12 percent. This showed that although there were more combination types that were not in accordance with the set framework, the numbers of collocation pairs from this group of collocations were far less than that of the set framework combination types.

When considering the combination types according to the framework of the study, it was found that the majority of the collocations were under the set framework. This high in number and percentage of the combination types which were in accordance with the set framework confirmed that the combination types found from the study were common combination types of collocations. Thus, they should also be useful for students in tourism and in related fields of study to know and be aware of. For the collocations which were not according to the set framework, there were 10 combination types of 350 collocation pairs. These uncommon combination types of collocations generated might be the result of the set criteria for identifying collocations where the collocates were identified within the range of 3-word span on the right side (3R) of each node. Such wider word span led to more combination types to occur. In other words, with the set word span in the study, it allowed more combination types to occur. Among them, there were uncommon combination types which were not commonly found and not recommended by scholars and researchers from previous studies in the field. This might be the reason they were excluded in the combination types suggested by Benson, Benson and Ilson [1] and Hill [3], except the Adverb + Verb combination that was found suggested by Hill [3]. This could be concluded that those uncommon combination types could be ignored since they did not commonly occur. However, the awareness of their existence was still valued.

It was also visible that the noun nodes were more likely to co-occur with noun collocates with some possibility to co-occur with verb collocates. The adjective nodes were more common to take noun collocates with a high chance to co-occur with other adjectives and some chance to co-occur with verb collocates. The verb nodes had higher possibility of co-occurrence with noun and adjective collocates and might co-occur with other verbs. The adverb nodes, although found in small number of collocation pairs, had slightly equal possibility to co-occur with noun, adjective, and verb collocates.

When ranking the lexical collocations according to the frequency of occurrence, it was evident that the majority of the 200 most frequent collocations are the Noun + Noun and Adjective + Noun combinations. Therefore, it should be useful to compile the list of most frequent collocations of these two combinations. The list of 100 most frequent collocations of Noun + Noun was shown in Table 5.

**Table 5** 100 Most Frequent Noun + Noun Collocations from the SCTRA

| No. | Nodes | Collocates | No. | Nodes | Collocates |
|---|---|---|---|---|---|
| 1 | climate | (-,xx) change | 51 | revisit | (-,xx) intention(s) |
| 2 | tourism | Industry | 52 | hotel | reviews |
| 3 | tourism | (x) hospitality | 53 | quality | (x) life |
| 4 | service | failure(s) | 54 | sport | event(s) |
| 5 | service | (-,xx) recovery | 55 | tourist | experience |
| 6 | value | co-creation | 56 | recovery | performance |
| 7 | tourism | Development | 57 | product | (x,xx) service |
| 8 | customer | Satisfaction | 58 | sport | (x,xx) program(s) |
| 9 | hospitality | (-,xx) industry(ies) | 59 | decision-making | process(es) |
| 10 | sport | (-,x,xx) management | 60 | sales | (-,xx) course(s) |
| 11 | service | (-,xx) quality | 61 | satisfaction | (x) loyalty |
| 12 | destination | Image | 62 | knowledge | domain(s) |
| 13 | data | Collection | 63 | consumer | (-,xx) behavio(u)r |
| 14 | research | question(s) | 64 | brand | experience |
| 15 | service | provider(s) | 65 | employees | (x) customers |
| 16 | tourism | Destinations | 66 | knowledge | (x,xx) skill(s) |
| 17 | brand | Equity | 67 | sport | sales |
| 18 | tourist | destination(s) | 68 | customer | (-,x) knowledge |
| 19 | tourism | Sector | 69 | cruise | line(s) |
| 20 | customer | (-,x) behavior(s) | 70 | tour | operator(s) |
| 21 | data | Analysis | 71 | students | (-,x,xx) learning |
| 22 | tourism | Demand | 72 | recreation | area(s) |
| 23 | management | agency(ies) | 73 | leisure | literacy |
| 24 | focus | group(s) | 74 | home-sharing | lodging(s) |
| 25 | literature | Review | 75 | management | programs |
| 26 | hotel | (-,xx) industry | 76 | relationship | intentions |
| 27 | tourist | arrival(s) | 77 | risk | (-,xx) management |
| 28 | factor | analyses(is) | 78 | simulation | game |
| 29 | products | (x,xx) services | 79 | study | abroad |
| 30 | learning | outcome(s) | 80 | destination | governance |

| No. | Nodes | Collocates | No. | Nodes | Collocates |
|---|---|---|---|---|---|
| 31 | customer | (-,x,xx) value | 81 | satisfaction | (xx) intention(s) |
| 32 | service | (-,xx) performance | 82 | customer | (-,xx) citizenship |
| 33 | leisure | activity(ies) | 83 | fit | index(ices) |
| 34 | hospitality | (-,x,xx) students | 84 | brand | awareness |
| 35 | convention | Centre | 85 | accommodation | (-,xx) provider(s) |
| 36 | recreation | site(s) | 86 | consumer | trust |
| 37 | customer | (-,x,xx) loyalty | 87 | marketing | (-,xx) strategy(ies) |
| 38 | attitudes | (x) behavio(u)r(s) | 88 | csr | activities |
| 39 | business | model(s) | 89 | factor | loading(s) |
| 40 | hospitality | (-,x,xx) management | 90 | gender | age |
| 41 | sample | size(s) | 91 | community | (-,xx) members |
| 42 | measurement | model(s) | 92 | travel | agency(ies) |
| 43 | tourist | Trust | 93 | university | extension |
| 44 | citizenship | behavior(s) | 94 | employee | brand |
| 45 | residents | attitude(s) | 95 | career | (-,x) intentions |
| 46 | heritage | site(s) | 96 | visit | (-,xx) intention(s) |
| 47 | customer | (-,x,xx) service | 97 | brand | image |
| 48 | tourist | attraction(s) | 98 | model | fit |
| 49 | residents | (-,x,xx) perceptions | 99 | tour | guide(s) |
| 50 | customer | need(s) | 100 | content | analysis |

Moreover, the list of 100 most frequent collocations of Adjective + Noun combinations was illustrated in Table 6.

**Table 6** 100 Most Frequent Adjective + Noun Collocations from the SCTRA

| No. | Nodes | Collocates | No. | Nodes | Collocates |
|---|---|---|---|---|---|
| 1 | outdoor | Recreation | 51 | international | (-,x,xx) student(s) |
| 2 | social | (-,xx) media | 52 | memorable | (-,x,xx) experience(s) |
| 3 | previous | (-,x,xx) studies | 53 | mock | trial |
| 4 | behavio(u)ral | (-,xx) intention(s) | 54 | local | (-,x,xx) residents |
| 5 | future | (-,x,xx) research | 55 | subjective | (-,xx) norm(s) |
| 6 | positive | (-,x,xx) effect(s) | 56 | inbound | (-,x) tourists |
| 7 | sustainable | (x,xx) tourism | 57 | federal | land |
| 8 | online | (-,x) review(s) | 58 | federal | (x) management |
| 9 | positive | (-,x,xx) impact(s) | 59 | structural | (-,x) model(s) |
| 10 | previous | (-,x,xx) research | 60 | cultural | (-,x,xx) heritage |

| No. | Nodes | Collocates | No. | Nodes | Collocates |
|---|---|---|---|---|---|
| 11 | current | (-,x,xx) study | 61 | social | exchange |
| 12 | present | (-,x) study | 62 | rural | (-,x,xx) area(s) |
| 13 | empowering | Leadership | 63 | domestic | (-,xx) tourism |
| 14 | high | (-,x,xx) level(s) | 64 | negative | (-,xx) effect(s) |
| 15 | local | (-,x,xx) community(ies) | 65 | online | (-,xx) course(s) |
| 16 | local | (-,xx) government(s) | 66 | environmental | (-,xx) impacts |
| 17 | rural | (-,xx) tourism | 67 | experiential | relationship |
| 18 | negative | (-,x,xx) impact(s) | 68 | in-depth | (-,x,xx) interview(s) |
| 19 | significant | difference(s) | 69 | smart | (-,xx) destinations |
| 20 | higher | Education | 70 | positive | (-,xx) emotions |
| 21 | significant | (-,x,xx) effect | 71 | experiential | satisfaction |
| 22 | social | (-,xx)responsibility(ies) | 72 | different | (-,x,xx) levels |
| 23 | social | (-,xx) capital | 73 | natural | (-,xx) environment(s) |
| 24 | digital | Leisure | 74 | responsible | (-,x,xx) tourism |
| 25 | social | (-,xx) interaction(s) | 75 | online | learning |
| 26 | overall | (x,xx) satisfaction | 76 | online | hotel |
| 27 | authentic | Leadership | 77 | positive | (-,x,xx) attitude(s) |
| 28 | future | (-,x) studies | 78 | local | (-,x,xx) people |
| 29 | inbound | Tourism | 79 | qualitative | (-,x,xx) research |
| 30 | indirect | (-,xx) effect(s) | 80 | developing | (-,xx) countries |
| 31 | economic | (-,x,xx) development | 81 | competitive | advantage(s) |
| 32 | sustainable | (-,x,xx) development | 82 | behavio(u)ral | control |
| 33 | protected | area(s) | 83 | economic | growth |
| 34 | social | (-,xx) norm(s) | 84 | natural | (-,xx) areas |
| 35 | social | (-,xx) network(s) | 85 | qualitative | data |
| 36 | dependent | variable(s) | 86 | online | (-,x,xx) programs |
| 37 | social | (-,x) theory | 87 | outdoor | activity(ies) |
| 38 | different | (-,x) types | 88 | significant | (-,xx) impact |
| 39 | international | (-,xx) tourist(s) | 89 | economic | (-,xx) impacts |
| 40 | natural | (-,x,xx) resource(s) | 90 | moderating | effect(s) |
| 41 | direct | (-,x,xx) effect(s) | 91 | personal | norm |
| 42 | positive | (-,x,xx) relationship | 92 | physical | (-,xx) environment |
| 43 | smart | (-,xx) tourism | 93 | direct | (-,x,xx) impact(s) |
| 44 | important | (-,xx) role(s) | 94 | behavio(u)ral | loyalty |
| 45 | multiple | (-,x) level(s) | 95 | positive | (-,xx) engagement |

| No. | Nodes | Collocates | No. | Nodes | Collocates |
|---|---|---|---|---|---|
| 46 | internal | (-,xx) branding | 96 | experiential | (-,xx) learning |
| 47 | social | (-,xx) science(s) | 97 | prior | (-,x) research |
| 48 | indigenous | People | 98 | urban | regeneration |
| 49 | federal | (-,xx) agencies | 99 | pro-environmental | (-,xx) behavio(u)r(s) |
| 50 | systematic | (-,x,xx) review(s) | 100 | dynamic | capabilities |

Apart from classification based on types of combination, it was also noticeable that the lexical collocations found from the SCTRA could be categorized into two main groups: tourism specific collocations and general academic collocations. For the ease of use in terms of pedagogical implications, the list of 100 most frequent tourism specific lexical collocations was provided in Table 7.

**Table 7** 100 Most Frequent Tourism Specific Lexical Collocations from the SCTRA

| No. | Nodes | Collocates | No. | Nodes | Collocates |
|---|---|---|---|---|---|
| 1 | climate | (-,xx) change | 51 | recreation | site(s) |
| 2 | tourism | Industry | 52 | customer | (-,x,xx) loyalty |
| 3 | tourism | (x) hospitality | 53 | attitudes | (x) behavio(u)r(s) |
| 4 | outdoor | Recreation | 54 | business | model(s) |
| 5 | social | (-,xx) media | 55 | international | (-,xx) tourist(s) |
| 6 | service | failure(s) | 56 | natural | (-,x,xx) resource(s) |
| 7 | service | (-,xx) recovery | 57 | hospitality | (-,x,xx) management |
| 8 | behavio(u)ral | (-,xx) intention(s) | 58 | positive | (-,x,xx) relationship |
| 9 | value | co-creation | 59 | tourist | trust |
| 10 | tourism | Development | 60 | citizenship | behavior(s) |
| 11 | customer | Satisfaction | 61 | smart | (-,xx) tourism |
| 12 | hospitality | (-,xx) industry(ies) | 62 | important | (-,xx) role(s) |
| 13 | sustainable | (x,xx) tourism | 63 | internal | (-,xx) branding |
| 14 | sport | (-,x,xx) management | 64 | residents | attitude(s) |
| 15 | service | (-,xx) quality | 65 | heritage | site(s) |
| 16 | destination | Image | 66 | customer | (-,x,xx) service |
| 17 | service | provider(s) | 67 | indigenous | people |
| 18 | previous | (-,x,xx) research | 68 | tourist | attraction(s) |
| 19 | tourism | Destinations | 69 | residents | (-,x,xx) perceptions |
| 20 | empowering | Leadership | 70 | memorable | (-,x,xx) experience(s) |
| 21 | high | (-,x,xx) level(s) | 71 | local | (-,x,xx) residents |
| 22 | local | (-,x,xx) community(ies) | 72 | customer | need(s) |

| No. | Nodes | Collocates | No. | Nodes | Collocates |
|---|---|---|---|---|---|
| 23 | local | (-,xx) government(s) | 73 | inbound | (-,x) tourists |
| 24 | brand | Equity | 74 | revisit | (-,xx) intention(s) |
| 25 | tourist | destination(s) | 75 | perceived | (-,x,xx) quality |
| 26 | tourism | Sector | 76 | quality | (x) life |
| 27 | rural | (-,xx) tourism | 77 | federal | (x) management |
| 28 | customer | (-,x) behavior(s) | 78 | sport | event(s) |
| 29 | social | (-,xx)responsibility(ies) | 79 | tourist | experience |
| 30 | social | (-,xx) capital | 80 | recovery | performance |
| 31 | digital | Leisure | 81 | product | (x,xx) service |
| 32 | tourism | Demand | 82 | cultural | (-,x,xx) heritage |
| 33 | management | agency(ies) | 83 | economic | (x,xx) social |
| 34 | social | (-,xx) interaction(s) | 84 | perceived | authenticity |
| 35 | authentic | Leadership | 85 | social | exchange |
| 36 | inbound | tourism | 86 | perceived | (-,xx) behavio(u)ral |
| 37 | hotel | (-,xx) industry | 87 | rural | (-,x,xx) area(s) |
| 38 | tourist | arrival(s) | 88 | domestic | (-,xx) tourism |
| 39 | products | (x,xx) services | 89 | sport | (x,xx) program(s) |
| 40 | perceived | (-,x,xx) value | 90 | decision-making | process(es) |
| 41 | economic | (-,x,xx) development | 91 | sales | (-,xx) course(s) |
| 42 | customer | (-,x,xx) value | 92 | environmental | (-,xx) impacts |
| 43 | service | (-,xx) performance | 93 | satisfaction | (x) loyalty |
| 44 | leisure | activity(ies) | 94 | knowledge | domain(s) |
| 45 | sustainable | (-,x,xx) development | 95 | consumer | (-,xx) behavio(u)r |
| 46 | protected | area(s) | 96 | experiential | relationship |
| 47 | social | (-,xx) norm(s) | 97 | brand | experience |
| 48 | social | (-,xx) network(s) | 98 | smart | (-,xx) destinations |
| 49 | hospitality | (-,x,xx) students | 99 | employees | (x) customers |
| 50 | convention | centre | 100 | knowledge | (x,xx) skill(s) |

**5. Conclusion and recommendations**

As it is crucial for those who are in the field of tourism and hospitality to be efficient in English for providing efficient and impressive service, this corpus-based study attempted to facilitate the learning and teaching of EFL/ESL as well as ESP in the field of tourism and hospitality by filling the gap that prevent effective and natural use of English. The keywords and the lexical collocations extracted from the SCTRA are believed to be useful for tourism and hospitality students as well as EFL/ ESL and ESP teachers and learners as these keywords and collocations could well facilitate their academic and professional development. By extracting keywords and collocations and classifying them into the lessons such as by means of content-based or corpus-based instructions, this makes it easier for the learners to notice and be aware of the keywords and collocations.

**References**

[1] M. Benson, E. Benson, and R. Ilson, *The BBI Combinatory Dictionary of English: Your Guide to Collocations and Grammar* (3rd ed.), John Benjamins, Amsterdam, The Netherlands, 2010.

[2] W. Chulaphan, and J. F. Barahona, "Contribution of Disaggregated Tourism on Thailand's Economic Growth," *Kasetsart Journal of Social Sciences*, 39(3), pp. 401-406, 2018. Available: https://www.sciencedirect.com/science/article/pii/S2452315116301898

[3] J. Hill, "Revising Priorities: From Grammatical Failure to Collocational Success," in M. Lewis (Ed.), *Teaching Collocation: Further Developments in the Lexical Approach* (pp. 47-69), Language Teaching Publications, London, UK, 2000.

[4] P. Farrokh, "Raising Awareness of Collocation in ESL/EFL Classrooms," *Journal of Studies in Education, 2*(3), pp. 55-74, 2012.

[5] M. Martin, "Advanced Vocabulary Teaching: The Problem of Synonyms," *Modern Language Journal, 68*(2), pp. 130-137, 1984.

[6] W. Cheng, *Exploring Corpus Linguistics: Language in Action*, Routledge, London, UK, 2012.

[7] M. Lewis, *Implementing the Lexical Approach: Putting Theory into Practice*, Heinle, Thomson Corporation, London, UK, 2002.

[8] N. Schmitt, *Vocabulary in Language Teaching*. Cambridge University Press, Cambridge, UK, 2000.

[9] J. Bahns, "Lexical Collocations: A Contrastive View," *ELT Journal,* 47(1), pp. 56-63, 1993.

[10] K. Ackermann, and Y. H. Chen, "Developing the Academic Collocation List (ACL) – A Corpus-driven and Expert-judged Approach". *Journal of English for Academic Purposes, 12*(4), pp. 235-247, 2013.

[11] A. Molavi, M. Koosha, and H. Hosseini, "A Comparative Corpus–based Analysis of Lexical Collocations Used in EFL Textbooks," *Latin American Journal of Content and Language Integrated Learning, 7*(1), pp. 66-81, 2014. Available: doi:10.5294/laclil.2014.7.1.4 eISSN 2322-9721.

[12] M. Borucinsky, and J. Kegalj, "Collocations in Marine Engineering English," *Scripta Manent, 10*(2), pp. 36-51, 2015.

[13] C. Demir, "Lexical Collocations in English: A Comparative Study of Native and Non-native Scholars of English," *Journal of Language and Linguistic Studies, 13*(1), pp. 75-87, 2017.

[14] S. Thongvitit, and N. Thumawongsa, "A Corpus-Based Study of English Collocations Found in the Abstracts of Research Articles Written by Thai EFL Writers," *International Journal of Social Science and Humanity, 7*(12), pp. 751- 755, 2017.

[15] K. Trinant, and B. Yodkamlue, "Lexical Collocations in a Sample Corpus of Nursing Research Articles (SCNRA)," *Suranaree Journal of Social Science,* 13(1), pp. 45-72, 2019.

[16] L. Anthony, *AntConc (Version 3. 4. 4)* [Computer Software]. Tokyo, Japan: Waseda University, 2019. Available: http://www.laurenceanthony.net/.

[17] P. Baker, *Using corpora in discourse analysis*, Continuum, London, UK, 2006.

[18] A. Coxhead, "A New Academic Word List," *TESOL Quarterly, 34*(2), pp. 213-238, 2000.

[19] J. Ward, "A Basic Engineering English Word List for Less Proficient Foundation Engineering Undergraduates," *English for Specific Purposes, 28*, pp. 170-182, 2009.

[20] M. N. Yang, "A Nursing Academic Word List," *English for Specific Purposes,* 37, pp. 27-38, 2015.