



Please cite this paper as follows:

Heidari Tabrizi, H., & Jamalzadeh, M. (2020). Academic Vocabulary in Tourism Research Articles: A Corpus-Based Study. *Journal of Language and Discourse Practice*, 1 (2), 25-44.

Research Paper

Academic Vocabulary in Tourism Research Articles: A Corpus-Based Study

Hossein Heidari Tabrizi^{1*}, Mehri Jamalzadeh²

¹Associate Professor, Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

heidaritabrizi@gmail.com

²Ph.D. Candidate, Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

m82jamalzadeh@yahoo.com

Abstract

Developing academic vocabulary word lists plays a significant role in academic writing. The present study aimed to establish a Tourism Academic Word List (TAWL) of the most frequently-used tourism academic vocabulary across different sub-disciplines in tourism by examining a written corpus of academic research articles in this field. This study also sought to determine whether and to what extent the words identified as high frequency in the tourism corpus have been identified as high frequency in West's General Service List (GSL) and Coxhead's Academic Word List (AWL). By analyzing a 3.7-million-word corpus, it was found that AWL words account for 12.34% of the Tourism Research Articles Corpus (TRAC). Most of the AWL word forms fit into the word families included in Coxhead's first and second sub-lists. The high word frequency and the wide text coverage of TAWL throughout Tourism RAs proved that TAWL plays an important role in tourism RAs. High frequent AWL items exhibited a different frequency order from those in Coxhead's AWL. Furthermore, there were many non-AWL content word families that occurred with high frequency in the corpus. The developed wordlist can increase students' exposure to academic vocabulary and facilitate its learning.

Keywords: *Academic Word List (AWL), Corpus Study, General Service (GSL), Tourism Academic Word List (TAWL)*

Introduction

The investigation of the specialized vocabulary used in academic discourse has attracted great attention lately because of the growing demand for instruction for non-native English speakers (NNES) in different backgrounds and contexts. Given the huge size of the vocabulary of a language, and to maximize the effectiveness of its learning, word lists of the recurrent vocabulary of academic texts have been created. These lists have been believed to provide the vocabulary required to function appropriately in academic settings where scientific English discourse is dominant.

Scientific English as described by Halliday (2004) refers to a generalized functional variety, or register of the modern English language. The variation can be summarized in terms of field, tenor and mode: By field, indicating extending, transmitting or exploring knowledge in the physical, biological or social sciences; by tenor, whether it is addressed to specialists, to learners or to laymen, from within the same group or across groups (e.g., lecturer to students); and by mode, he refers to phonic or graphic channel, most congruent or less so and rhetorical function variety—expository, hortatory, polemic, imaginative and so on.

Researchers communicate within specific academic discourse communities through the research article genre. Research articles are written by scientists and are addressed to different scientists within the same discipline. They are the pre-eminent genre of the academy restructuring the processes of thought and reported research to establish a discourse for scientific fact-creation (Hyland, 2010). As Hyland and Paltridge (2011) puts it, three major developments over the past 20 years have aroused interest in academic discourse, and particularly academic writing in English: changes in higher education leading to greater interest to the importance of writing; position of English as the international language of research and scholarship; and the emergence of theoretical perspectives with their emphasis on the centrality of academic discourses knowledge construction.

With the increasing non-native and native academics publishing in English, study of academic discourse and academic vocabulary in particular is inevitable. Academic vocabulary are words reasonably frequent in a wide range of academic genres but relatively uncommon in other kinds of texts (Coxhead & Nation, 2001). Academic vocabulary is one of four levels of vocabulary division by Nation (2001). They are high frequency words, academic vocabulary, technical vocabulary, and low frequency words. High frequency general words are core words used very frequently in most language use (Nation & Hwang, 1995). West's (1953) General Service List (GSL) reporting the 2000 most frequent word families are of this kind.

However, Engels (1968) and Richards (1974) criticized West's (1953) list for its size and age. The size criticisms question the necessity of the second 1000 words of the GSL because they usually cover only 4-5% of the running words in non-fiction texts compared to the 70% plus coverage of the first 1000. The report on which the GSL is based was prepared in the 1930s and since language changes overtime, the GSL is too old, it contains many words that are not essential and does not contain high frequent current words (Nation & Hwang, 1995).

Nation and Waring (1997) suggest that EAP students need to first learn the 2000 or so most 'general' words of English, and continue by a set of "academic" words common to all academic disciplines. General words refer to the 2000-word family GSL mentioned above and academic words to the 570-word family Academic Word List or AWL (Coxhead, 2000). About 90 percent of the running words in an academic text is covered by Academic words and general words together. (Coxhead, 2000; Coxhead & Nation 2001; Nation, 2001)

AWL is based upon the analysis of 3.5-million-word corpus of written academic English in 28 sub-disciplines in four main disciplines of Art, Commerce, Law, and Science. AWL is formed by 570-word families selected according to three criteria: a) frequency of occurrence (occurrence of at least 100 times in the corpus), b) range (occur not less than 10 times in each of the 4 disciplines and in 15 or more subject areas) and c) specialized occurrence (be outside the first 2000 GSL words). Coxhead categorized the AWL items into 10 sublists according to their frequency. All sublists contain 60-word families, with the exception of the tenth sublist, which contains 30-word families. The most frequent AWL items in the first sublist, with 3.6% coverage of the corpus; the last sublist comprises the least frequent ones with a coverage of only 0.1% of the corpus. The coverage, however, was not the same for all the 4 subject areas: the list provides the highest coverage over commerce (12.0%) and the lowest over science (9.1%).

Hyland and Tse (2007) pointed out that individual lexical items on the list often occur and behave in different ways depending on the disciplines in terms of range, frequency, collocation, and meaning. As they pointed out “the different practices and discourses of disciplinary communities reduces the usefulness of such lists” and suggested “that teachers help students develop a more restricted, discipline-based lexical repertoire” (Hyland & Tse, 2007, p. 235).

According to Hyland and Tse (2007), the best way to prepare students for their academic studies is to provide them with an understanding of the features of the discourses they will encounter in their particular courses. As Coxhead, (2017) puts it, “Studying at university can mean exposure to several million running words a year through reading textbooks, source books, content and learning-based websites and other academic sources of information.” p. 90. Word list research has also been driven by the needs of particular groups of language learners and to help set learning goals (Nation, 2016).

Given the fact that the field of tourism has been neglected in vocabulary studies, the current study aims to fill this gap by investigating the distribution and frequency of the AWL (and non-GSL/AWL) items in tourism research articles. In view of the above remarks, for teaching and learning general academic courses, tourism vocabulary would be essential to any academic tourism program. Therefore, this study sought to establish a TAWL and determine whether and to what extent the words identified as high frequency in the tourism corpus have been identified as high frequency in West’s General Service List (1953) and Coxhead’s Academic Word List (2000).

Literature Review

Academic vocabulary points to the lexical items that are rather frequent across a wide range of academic texts but are infrequent in other genres (Coxhead & Nation, 2001). Some studies have developed academic word lists across disciplines. Champion and Elley (1971) and Praninskas (1972) developed the earliest lists of frequently used words students would encounter during their academic studies in a range of different university disciplines. Lynn (1973) and Ghadessy (1979) developed lists of difficult words for students who needed to read academic texts. The University Word List (UWL) contains 836 high frequencies non-GSL words across a wide range of disciplines was produced by Xue and Nation (1984) for university level students. They compiled the list from Champion and Elley (1971), Praninskas (1972), Lynn (1973), and Ghadessy (1979).

Coxhead (2000) developed the aforementioned academic word list whose frequency accounted for approximately 10% of tokens in academic texts. Coxhead (2000) set up a corpus of

3.5 million running words chosen from various academic journals and college course books in four primary subjects: the arts, commerce, law and natural science. The AWL list that contains 3,112 individual items, does not include words in the most frequent 2,000-word families in the English language and West's (1953) GSL.

Field-specific academic word lists or discipline-based lexical repertoires have been produced for different disciplines. Kwary and Artha (2017) created and tested a word list called the Academic Article Word List for Social Sciences. Lei Lei and Dilin Liu (2016) developed a new medical academic vocabulary list based on the results of a series of comparative analyses. Hsu (2013), Wang, Liang and Ge (2008), Chen and Ge, (2007), developed medical academic word lists too. Wang et. al presented a word list including 623 non-GSL word families occurring frequently across medical research articles. A coverage of almost one-tenth over their corpus was provided by the list. Only 342 of the 623 high frequency word families they identified concurred with those listed in AWL. Chen and Ge (2007) found that only 292 of 570 AWL's word families were frequent in medical research articles and 179 AWL items either did not occur or occurred infrequently in the corpus.

Now an increasingly sophisticated body of research and knowledge is available on various aspects of the discipline in Iranian context including Azadi and Chalak (2017); Elekaei, Faramarzi and Heidari Tabrizi (2015); Faramarzi, Elekaei and Heidari Tabrizi (2015); Heidari Tabrizi (2017); Jamalzadeh and Chalak (2019); Noorizadeh-Honami and Chalak (2018); Shirani and Chalak (2016, 2018); and Vaezi and Heidari Tabrizi (2015) to name but a few. For example, Jamalzadeh and Chalak (2019) identified 1450 high frequent academic word families in physiotherapy research articles and compared it with the distribution of high frequency words in Coxhead's (2000) Academic Word List and West's (1953) General Service. It was found that of the 570-word families in AWL, 562 occurred frequently in physiotherapy research articles and this provided a coverage of 11.51 of the tokens in the corpus.

Valipouri and Nassaji, (2013) developed an academic word list for chemistry discipline. They examined a 4-million-word corpus of chemistry research articles. They identified 1400-word families used with a reasonable frequently in the corpus. Muñoz (2015) and Martinez et. al.'s (2009) academic word lists included 1941-word families occurring frequently in agriculture corpus. Li and Qian (2010) profiled the presence of the AWL in a financial services corpus. It was found that only about one third of AWL items were frequently used in their corpus. Hyland and Tse (2007) compared the use of academic words in textbook chapters, academic book reviews, master's theses, and doctoral dissertations. Collectively, these studies outline a critical role for domain-specific word lists. They call for more research that examines field-specific corpora in order to develop such word lists for students studying in different disciplines. In the present study, a large written corpus of academic research articles in the field of tourism were examined to develop an academic word list that could be useful for tourism students. An academic word list exclusively for tourism students can be taught and directly studied in the same way as the words from the GSL. A tourism academic word list can also play an important role in helping EFL tourism students learn academic English more effectively.

Therefore, this study aimed to establish a Tourism Academic Word List (TAWL) of the most frequently-used tourism academic vocabulary across different sub-disciplines in tourism by examining a written corpus of academic research articles in this field. This study also sought to determine whether and to what extent the words identified as high frequency in the tourism corpus have also been identified as high frequency in West's (1953) General Service List (GSL) and Coxhead's (2000) Academic Word List (AWL) and vice versa. In other words, the second

objective was to check the overlap and the difference between our list and the current well-known general and academic word lists, and hence highlight its usefulness for tourism students. Thus, the following research questions were posed:

1. What are the most frequently used academic words in a corpus of tourism research articles?
2. What AWL word forms occur with high frequency in the Tourism Research Articles Corpus?
3. What non-AWL content word forms occur with high frequency in the corpus of Tourism Research Articles?

Methodology

Design and Context of the Study

A 3.7-million-word corpus of tourism research articles as explained in details in the following part was collected. This study as a quantitative research has a descriptive research design and has been conducted through gathering, analyzing, and presenting the collected data.

The Corpus

A specialized corpus was built to represent research articles genre, science register, and discipline of tourism. It was designed following the criteria proposed by Sinclair (1991, 2005) and Barnbrook (1996), considering representativeness, specificity of corpus, use of whole documents, and availability in electronic form. The relevant research articles were obtained from the electronic ISI journals of *Annals of Tourism Research*, *Journal of Hospitality and Tourism Management*, *Tourism Management* and *Tourism Management Perspectives*, all downloaded from Elsevier with full text. The articles included in the corpus were published between 2017 and 2018. Swales's (1990) model was followed and all the articles included in the corpus had identifiable Abstract, Introduction, Method, Result and Discussion sections (IMRD). Thus, if an article did not follow IMRD format, they were not selected.

A three-step selection was followed to choose the sample tourism research articles for the corpus. Firstly, only research articles focusing on empirical studies, written in the identifiable Introduction, Method, Result, and Discussion sections, were included in the TRAC.

Secondly, the research articles chosen had to have been published between 2017 and 2018. Thirdly, the length of the chosen articles must be longer than 2,000 running words and shorter than 10,000 running words. After this three-step selection a total of 400 articles were chosen for the TRAC. The articles were collected in their electronic version with their reference lists, appendices, captions, footnotes, and acknowledgments removed (Swales, 1990). The corpus represents a genre, experimental research articles, and a field, that is, tourism. The results show that the TRAC contains 3,711,779 running words.

Table 1

Description of the Corpus

Corpus	
Genre	Research Articles
Discipline	Tourism
Domain	Four ISI journals published online
Sources of the articles	Elsevier
Period (publication)	2017-2018
size	3,711,779 words – 400 articles
Writers	Mostly faculty

Readers	University students, specialists, and non-specialists with some
Purposes	expertise in Tourism To inform, instruct, explain

Data Processing

The research articles were in a PDF format and had to be converted into a .txt for further processing. After obtaining the corpus, the software called AntWordProfiler available from Laurence Anthony's Website was used for analysing the vocabulary load of texts. This tool generates vocabulary statistic and frequency information about a corpus of texts loaded into the program. It compares the files against a set of vocabulary level lists that can be plain frequency lists or 'family lists' based on the research of Paul Nation. (Anthony, 2014). Most of the previous studies on developing academic vocabulary have used Range software. AntWordProfiler 1.4.0w Vocabulary Analysis programs, is a much more modern version of the program with numerous extra features (Nation, 2014). Previously, Jamalzadeh, 2017 used another corpus analysis toolkit of the same series called AntConc for conducting a corpus-based study on cohesive conjunctions on medical research articles and it was found as an efficient and really user-friendly program. Thus using the program, the number of occurrences of each word, its range (i.e. in how many different texts each word has occurred) as well as the words shared with the AWL and GSL word lists were determined.

For a word family to be included in the list, Coxhead's (2000) procedure was followed. Coxhead's corpus for the AWL consisted of 3.5 million words and the non-GSL word families that occurred 100 times in the entire corpus and at least 10 times in each of the 4 disciplines were included in her list. Since she words with a frequency of 100 times in the whole corpus were selected, it becomes almost 28.5 times in a million words. There were 3,711,779 words, so it was decided that the cut-off frequency of members of a word family should be equal to or higher than 106 times in the whole corpus to be included in the list. All the content words that met the set criteria were identified. Function words such as pronouns, auxiliaries, articles, and numbers were excluded from the analysis as these were considered too general.

Results

Establishment of a Tourism Academic Word List

Identifying words that were frequently used by academic writers in the field of tourism concerned one of the main questions of this study. To answer this question, the two criteria of range and frequency used by Coxhead (2000) were applied. As noted earlier, all words with a frequency of 106 or more were considered as frequent.

After the elimination of the GSL word families 1002 content word families were left which constituted Tourism Academic Word List (TAWL). It includes 469 AWL word families plus 533 non-GSL/non-AWL word families. The full list of these word families is presented in the Appendix. Words shared with AWL are bold and the non-AWL and non-GSL words are in regular font.

AWL and TAWL Word Forms Used in the TRAC

The coverage of GSL, AWL, and non-GSL/non-AWL words (word families and word tokens) in TRAC corpus is depicted in Table 2. As Table 2 shows, the 2000 most frequent word families of GSL accounted for 2,480,109 tokens (66.81%) of the corpus. As for AWL words, of

the 570 AWL word families, 569 occurred in tourism corpus, out of which 469 met the cut-off frequency criterion. The frequency counts of tokens of these word families were 457,982 accounting for a coverage of 12.34% of TRAC. The AWL coverage in TRAC is higher than the coverage in both Coxhead's science corpus (i.e. 9.1%), which consisted of different scientific disciplines, and in Coxhead's multi-disciplinary corpus (10%). Out of the 1002-word families in TAWL, 533 word (53.47) were not among GSL and AWL.

Table 2*The Coverage of Different Base Word Lists Over TRAC*

Word lists	Tokens	% of TRAC
1st GSL	2,254,693	60.74
2nd GSL	225,416	6.07
AWL	457,982	12.34
Non-GSL/AWL	773,688	20.84
Total	3,711,779	100

Table 3*The First 10 Most Frequent Words in TAWL*

Word	Frequency	Word	Frequency
tourism	43974	study	14899
have	21127	journal	11525
research	18479	use	10922
manage	15718	experience	9408
tour	15489	relation	9357

Table 3 displays the 10 most frequent word families and their frequency in TAWL. The most frequent word in TAWL is the *physical* word family, which occurred more than 9,000 times in TRAC.

Discussion

The present study was an attempt to identify frequently used words in TRAC and develop a word list for tourism students. Furthermore, the list was compared with the academic word list (AWL) to explore its coverage. GSL or general academic words were excluded. Thus, a word list based on frequency and range of AWL word families and non- AWL/ GSL was constructed (presented in the appendix).

An analysis of the TAWL

This study found that the coverage of AWL word forms in the Tourism Research Articles Corpus was 12.34. This is higher than 9.3% coverage of AWL found in Hyland and Tse's (2007) science sub-corpus, 9.1% of Coxhead's (2000) science sub-corpus, 9.06% in Martinez et al.'s corpus of agricultural research articles and the 11.7% coverage over applied linguistics corpus in

Vongpumivitch, Huang, and Chang's (2009) study. These findings endorse the idea that AWL items are not equally useful for students in specific fields.

Regarding the first research question altogether, 1002-word families with a reasonable frequency were identified in the corpus. These words are called tourism Academic Word List (TAWL). A full list of these words is presented in the Appendix. Word families of *tourism*, *have* and *research* appear most often (43974, 21127 and 18479 times, respectively). In this study, the 10-word families with the highest frequencies account for 4.6% of running words in the corpus. Conversely, *architecture*, *bay booking*, *deem*, *hub*, *impairment*, *lovelock* and *reform* word families appear least often (106 times). In this study, the 10-word families with the lowest frequencies account for only 0.02% of the running words in the corpus.

Comparing the TAWL with the AWL

The second question in this research was about comparing the most frequently Tourism research articles academic words against Coxhead AWL and West GSL word lists. The TAWL list was compared with academic word list. (AWL) to explore its coverage. TRAC shares 569-word families with the AWL. Frequency analysis shows that there are 469 AWL word forms that occur more than 106 times in TRAC. While Following Coxhead (2000), the most frequent 60-word families in the corpus were calculated and also compared them with those in Coxhead's sublists. The headwords of these word families are shown in Table 4. The numbers in front of the words show the sublists of Coxhead's AWL they belong to. Coxhead divided the AWL into ten rank-ordered sublists, according to decreasing word family frequency. except sublist 10, each sublist contains 60 items. It can be seen that most of the frequently-occurring words in the TRAC come from Coxhead's first two sublists (sublists 1 and 2). Within the 60-word families, bold type indicates those that coincided with items in sublist 1 of Coxhead's AWL.

Among the first 60-word families, 29 coincided with Coxhead's sublist 1, 3 items more than that in Martinez et al.'s (2009) corpus of agricultural research articles, and 6 items fewer than that in Hyland and Tse's multi-disciplinary corpus. As explained by Coxhead (2000, p. 228), words in her first sublist "account for more than one-third of the total coverage of the list," which means that the words in her first sublist are the most frequently-occurring words in her AWL. Coxhead's next sublist, sublist 2, covers 1.8% of her academic corpus, while her last sublist, sublist 10, covers only 0.1% of her corpus. This decreasing coverage of AWL sublists in Coxhead's corpus is illustrated in the TRAC corpus as well. 18-word families come from Coxhead's sublist 2. Only a few of the top 60 AWL word forms found in the TRAC come from Coxhead's third, fourth, fifth and sixth seventh, sublists and no word form in the top 60 list comes from Coxhead's seventh, eighth, ninth and tenth sublists. Given the fact that Coxhead's corpus covers four disciplines (Arts, Commerce, Law, and Science) while our ALC corpus only covers research articles in applied linguistics, this similarity of AWL distribution is interesting. It emphasizes the importance of learning the frequently-occurring AWL words, especially those that come from Coxhead's first two sublists, regardless of the learners' field, and also validates the use of a corpus-based approach to create an academic word list, especially for specific fields of study.

To sum up, compared to the AWL, the TAWL includes more word families and can better reflect lexical features of tourism research articles.

Table 4*The First Most Frequent 60 AWL Word Families in the TAWL Compared with AWL Sublists.*

1. research 1	2. journal 2	3. culture 2	4. analyse 1
5. economy 1	6. perceive 2	7. vary 1	8. identify 1
9. data 1	10. theory 1	11. respond 1	12. environment 1
13. factor 1	14. significant 1	15. impact 2	16. participate 2
17. consume 2	18. image 5	19. process 1	20. role 1
21. positive 2	22. community 2	23. create 1	24. perspective 5
25. approac h 1	26. method 1	27. individual 1	28. focus 2
29. strategy 2	30. reside 2	31. area 1	32. motive 6
33. sustain 5	34. context 1	35. indicate 1	36. concept 1
37. structure 1	38. region 2	39. specific 1	40. affect 2
41. negate 3	42. construct 2	43. site 2	44. network 5
45. benefit 1	46. interact 3	47. resource 2	48. issue 1
49. attitude 4	50. policy 1	51. psychology 5	52. contribute 3
53. involve 1	54. dimension 4	55. item 2	56. source 1
57. sector 1	58. institute 2	59. author 6	60. attribute 4

Coverage of Unlisted Words in the Tourism Corpus

The third research question in this research was about non-AWL content word forms with high frequency in the corpus of Tourism Research Articles. Unlisted words are the words that appear in neither the AWL nor the GSL. Table 2 shows that unlisted words cover 20.84% of the running words in TRAC.

This study aims to highlight field-specific academic words, so unlisted words are only compared with those in the AWL here. Two factors potentially explain why unlisted words have relatively high coverage in TRAC. First, the AWL does not include some academic words that are commonly used in tourism research articles and some AWL word families seldom appear in the TRAC. Table 5 shows some examples of unlisted vocabulary that appear in TRAC, and their word frequency is quite high. For example, *tourism* occurs 43974 times in the TRAC. In contrast, a number of AWL vocabulary items have extremely low frequency. The word *clause* is a case in point: Its frequency of occurring is 8. Second, in contrast to Coxhead's (2000) Academic Corpus, the TRAC is a specialized one and contains a relatively higher number of field-specific terms that occur often and widely in the discipline of tourism.

Table 5*Unlisted Words in the Tourism Corpus*

tourism	43974	destination	6942	hospitality	6699	heritage	2345
online	2285	fig	2019	authenticity	1901	brand	1758
rural	1716	leisure	1641	organizational	1485	engagement	1129

urban	1267	emotional	1184	emotions	1154	interviews	1009
cruise	1114	retrieved	1033	Climate	1027	travelers	849
cognitive	927	experiential	925	governance	858	crisis	770
collaboration	828	competitive	810	personality	791	regression	690
mobile	743	Festival	737	resort	722	emotion	658
stakeholder	674	airport	668	Spatial	661	socio	632
career	656	outsourcing	651	indigenous	639	van	601
conservation	619	multi-617		pilgrimage	617	spirituality	567
scholars	593	questionnaire	583	interview	568	barriers	532
carbon	556	managerial	550	professor	543	recreation	488
engage	531	internet	517	mobility	495	Entrepreneurship	486

Conclusion

In this study, a 3.7-million-word corpus of research articles in tourism was examined. The aim was to identify frequently used words in tourism research articles and develop a word list for tourism students. The list was also compared with the academic word lists (AWL) to explore its coverage. In total, 1002-word families with a reasonable frequency in the corpus were identified. They were called words Tourism Academic Word List (TAWL). A full list of these words is presented in the Appendix. A comparison of TAWL with AWL showed that many of the AWL items were not used frequently in the subject area examined. High frequent AWL items had a different frequency order from those in Coxhead's AWL, indicating that academic words are not used similarly across disciplines. In addition, there were many non-AWL content word families that occurred with high frequency in our corpus. This supports the idea of developing field-specific vocabulary lists which derive from the target genres and texts that students need to read and write in their own academic discipline (Hyland & Tse, 2007; Wang et al., 2008; Martinez et al., 2009).

Hyland and Tse (2007) recommend that "teachers help students develop a more restricted, discipline-based lexical repertoire" (p. 235). Based on the findings of this study, students of tourism are recommended to direct their attention to the first two subsets of AWL as well as the list of words provided through the analysis of this study. Field-specific word lists for students in different disciplines is recommended. Field-specific lists can help students learn the necessary words specifically important for their field of study. For teaching and learning general academic courses, these word lists are considered as one of the best efficient and practical methods. It would be of special significance for tourism students/instructors and professionals in learning or using tourism academic vocabulary in reading and writing.

This research is only a preliminary study on the tourism academic vocabulary used in tourism RAs. If possible, the TAWL needs to be rechecked in larger corpora or in other genres of tourism, such as tourism textbooks or spoken tourism academic English. The occurrence of the words in different contexts was not examined and TAWL consisted of isolated words. Knowing words in isolation does not guarantee having knowledge of how to use or understand them in context. In addition, students need to know not only the meanings of words but also how they co-occur (collocations) with other words in specific domains. It is possible that each word may be associated with different words and that the frequency of these associations may vary in different contexts and disciplines. Thus, further research is needed to examine how words collocate with other words and how their specific meanings and functions vary in different disciplines. It is

hoped that the availability of exercises and tests based on the TAWL will promote effective and efficient teaching and learning of tourism academic vocabulary.

References

- Azadi, G., & Chalak, E. (2017). The Frequency of Macro/Micro Discourse Markers in Iranian EFL Learners' Compositions. *International Journal of English and Education*, 6(1), 20-40.
- Campion, M. E., & Elley, W. B. (1971). *An academic vocabulary list*. New Zealand Council for Educational Research.
- Chen, Q., & Ge, G. C. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, 26(4), 502-514.
- Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a foreign language*, 15(2), 103.
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), 213-238.
- Paltridge, B., Starfield, S., & Coxhead, A. (2017). *Vocabulary and English for specific purposes research: Quantitative and qualitative perspectives*. Routledge.
- Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. *Research perspectives on English for academic purposes*, 252-267.
- Elekaei, A., Faramarzi, S., & Tabrizi, H. H. (2015). Genre-based Discourse Analysis of Death Announcements. *Editorial team*, 23, 37.
- Faramarzi, S., Elekaei, A., & Tabrizi, H. H. (2015). Genre-based discourse analysis of wedding invitation cards in Iran. *Journal of Language Teaching and Research*, 6(3), 662-668.
- Farrell, P. (1990). Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary. CLCS Occasional Paper No. 25.
- Ghadessy, P. (1979). Frequency counts, word lists, and materials preparation: A new approach. In *English Teaching Forum* (Vol. 17, No. 1, pp. 24-27).
- Halliday, M. A. K. (2002). *The collected works of MAK Halliday*. Continuum.
- Heidari Tabrizi, H. (2017). A Comparative Study of Interactive and Interactional Metadiscourse Markers in Sales Contract Written by English Natives vs. Iranian Non-natives. *Journal of Research in Applied Linguistics*, 8(Proceedings of the Fourth International Conference on Language, Discourse and Pragmatics), 75-83.
- Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, 17(4), 454-484.
- Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL quarterly*, 41(2), 235-253.
- Hyland, K. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes*, 9(2), 116-127.
- Jamalzadeh, M. (2017). A Corpus-based Study of Cohesive Conjunctions in Medical Research Articles Written by Iranian and Non-Iranian Authors. *Journal of Teaching English for Specific and Academic Purposes*, 669-686.
- Jamalzadeh, M., & Chalak, A. (2019). A corpus-based study of academic vocabulary in Physiotherapy research articles. *Language Teaching Research Quarterly*, 9, 69-82.
- Kwary, D. A., & Artha, A. F. (2017). The academic article word list for social sciences. *Mextesol Journal*, 41(4), 1-11.

- Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42-53.
- Li, Y., & Qian, D. D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System*, 38(3), 402-411.
- Lynn, R. W. (1973). Preparing word-lists: a suggested method. *RELC journal*, 4(1), 25-28.
- Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for specific purposes*, 28(3), 183-198.
- Muñoz, V. L. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes*, 39, 26-44.
- Nation, I. S. (2001). *Learning vocabulary in another language*. Ernst Klett Sprachen.
- Nation, P. (2014). How Much Input Do You Need to Learn the Most Frequent 9,000 Words? *Reading in a Foreign Language*, 26(2), 1-16.
- Nation, P., & Kyongho, H. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23(1), 35-41.
- Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, 14, 6-19.
- Praninskas, J. (1972). *American university word list*. Longman.
- Noorizadeh-Honami, L. & Chalak, A. (2018). Comparative analysis of architecture research article abstracts written by native and non-native authors: A crosslinguistic, cross-cultural study. *Theory and Practice in Language Studies*, 8(3), 325-330.
- Omoniyi, T., Hyland, K., & Paltridge, B. (2011). Discourse and identity. K. Hyland & K. Paltridge, *The continuum companion to discourse analysis*, 260-276
- Pauwels, P. (2017). Making and Using Word Lists for Language Learning and Testing. Nation, ISP (2016). *ITL-International Journal of Applied Linguistics*, 168(1), 134-139.
- Richards, J. C. (1974). Word lists: problems and prospects. *RELC journal*, 5(2), 69-84.
- Shirani, S., & Chalak, A. (2016). A genre analysis study of Iranian EFL learners' master theses with a focus on the introduction section. *Theory and Practice in Language Studies*, 6(10), 1982-1987.
- Shirani, S., & Chalak, A. (2018). Genre Analysis of Iranian TEFL Students' Master Theses. *Advances in Language and Literary Studies*, 9(2), 31-37.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Vaezi, R., & Tabrizi, H. H. (2016). ESP writing for tourism: A genre-based approach in teaching writing to Iranian university students. *Mediterranean Journal of Social Sciences*, 7(2 S1), 470.
- Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, 12(4), 248-263.
- Vongpumivitch, V., Huang, J. Y., & Chang, Y. C. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33-41.
- Wang, J., Liang, S. L., & Ge, G. C. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27(4), 442-458.
- Manuel, H. T. (1954). A General Service List of English Words with Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language learning and communication*, 3(2), 215-229.

Appendix:
Tourism Academic Word List (TAWL)

Note: AWL words are bold, non-GSL/non-AWL words are in regular font.

1. tourism	44. construct	87. academy
2. research	45. site	88. framework
3. journal	46. network	89. finance
4. culture	47. benefit	90. previous
5. analyse	48. heritage	91. media
6. online	49. destinations	92. organizational
7. hospitality	50. interact	93. major
8. economy	51. resource	94. estimate
9. perceive	52. issue	95. ethic
10. vary	53. attitude	96. medical
11. identify	54. policy	97. evident
12. data	55. psychology	98. stakeholders
13. theory	56. contribute	99. facilitate
14. respond	57. involve	100. empirical
15. environment	58. dimension	101. project
16. factor	59. fig	102. link
17. significant	60. item	103. overall
18. impact	61. authenticity	104. consist
19. participate	62. source	105. require
20. consume	63. sector	106. investigate
21. image	64. institute	107. transport
22. role	65. author	108. tradition
23. positive	66. attribute	109. conduct
24. community	67. locate	110. technology
25. create	68. statistic	111. challenge
26. perspective	69. evaluate	112. outcome
27. approach	70. potential	113. select
28. method	71. brand	114. orient
29. individual	72. assess	115. promote
30. focus	73. globe	116. urban
31. strategy	74. job	117. function
32. reside	75. similar	118. theme
33. area	76. define	119. aspect
34. motive	77. rural	120. category
35. sustain	78. available	121. period
36. context	79. innovate	122. seek
37. indicate	80. survey	123. interpret
38. concept	81. leisure	124. emerge
39. structure	82. communicate	125. emotional
40. region	83. hypothesis	126. establish
41. specific	84. access	127. implicate
42. affect	85. design	128. final
43. negate	86. gender	129. voluntary

130. emotions
131. engagement
132. **invest**
133. rely
134. **qualitative**
135. **relevant**
136. **valid**
137. **conclude**
138. **range**
139. **physical**
140. **accommodate**
141. **enhance**
142. **conflict**
143. retrieved
144. climate
145. **publish**
146. **reveal**
147. **mediate**
148. interviews
149. **generate**
150. **income**
151. **diverse**
152. **dynamic**
153. **achieve**
154. **highlight**
155. **element**
156. **distribute**
157. **complex**
158. **integrate**
159. cognitive
160. experiential
161. **obtain**
162. **attach**
163. **emphasis**
164. **section**
165. **goal**
166. **demonstrate**
167. **primary**
168. **index**
169. **implement**
170. governance
171. **component**
172. **alternative**
173. **corporate**
174. travelers
175. **constrain**
176. **adapt**
177. **proceed**
178. collaboration
179. **consequent**
180. **code**
181. **occur**
182. competitive
183. **professional**
184. **unique**
185. personality
186. **regulate**
187. **status**
188. **predict**
189. **labour**
190. **partner**
191. **contemporary**
192. **criteria**
193. crisis
194. **domestic**
195. **insight**
196. mobile
197. **assume**
198. festival
199. resort
200. **confer**
201. **aware**
202. **utilise**
203. **evolve**
204. **commit**
205. **cooperate**
206. regression
207. **encounter**
208. **confirm**
209. **feature**
210. stakeholder
211. airport
212. **trend**
213. spatial
214. **purchase**
215. emotion
216. career
217. **revenue**
218. **dominate**
219. outsourcing
220. **correspond**
221. **enable**
222. **furthermore**
223. indigenous
224. **generation**
225. **instance**
226. **bias**
227. socio
228. **equate**
229. **distinct**
230. **despite**
231. conservation
232. multi
233. pilgrimage
234. **comment**
235. **phenomenon**
236. **fund**
237. **maintain**
238. **intense**
239. **stress**
240. **internal**
241. van
242. **ensure**
243. **administrate**
244. **transform**
245. **external**
246. scholars
247. **migrate**
248. **appropriate**
249. **hence**
250. **prior**
251. **target**
252. questionnaire
253. **publication**
254. **contact**
255. **norm**
256. **percent**
257. interview
258. **initial**
259. spirituality
260. **normal**
261. sage
262. carbon
263. **capacity**
264. **illustrate**
265. **maximise**
266. managerial

267. professor
268. barriers
269. core
270. channel
271. authority
272. cite
273. task
274. contrast
275. technique
276. series
277. internet
278. phase
279. text
280. input
281. compute
282. acknowledge
283. undertake
284. error
285. volume
286. initiate
287. topic
288. abstract
289. mobility
290. occupy
291. mental
292. web
293. recreation
294. edit
295. entrepreneurship
296. keywords
297. assist
298. entrepreneurs
299. ethnic
300. via
301. document
302. determinants
303. secure
304. virtual
305. visual
306. revise
307. references
308. reference
309. cycle
310. constitute
311. team
312. subsequent
- 313. derive**
314. vacation
315. output
316. existential
317. tourman
318. subjective
319. expert
320. notion
321. geography
322. whereas
323. mechanism
324. ecotourism
325. correlation
326. exclude
327. annual
328. pilgrims
329. authentic
330. imply
331. revisit
332. expand
333. stable
334. symbol
335. lifestyle
336. hierarchy
337. shift
338. settings
339. quantitative
340. summary
341. muslim
342. digital
343. option
344. domain
345. homepage
346. demographic
347. discourse
348. landscape
349. yoga
350. mode
351. halal
352. objective
353. turnover
354. wellbeing
355. interviewees
356. capable
357. energy
358. route
- 359. convene**
360. react
361. pearce
362. incorporate
363. museum
364. principle
365. panel
366. dental
367. antecedents
368. branding
369. airports
370. cluster
371. pursue
372. forecasting
373. senior
374. workplace
375. display
376. enterprises
377. spa
378. accurate
379. bond
380. beach
381. sex
382. constant
383. decade
384. sociology
385. comprise
386. coefficient
387. festivals
388. practitioner
389. style
390. empowerment
391. profile
392. foundation
393. congruity
394. scores
395. yang
396. additionally
397. moderating
398. append
399. exhibit
400. shark
401. pacific
402. infrastructure
403. adjust
404. sufficient

405. **flexible**
 406. graduates
 407. narratives
 408. airline
 409. souvenir
 410. **differentiate**
 411. emissions
 412. graduate
 413. interpersonal
 414. score
 415. **transfer**
 416. mega
 417. **approximate**
 418. **transit**
 419. competence
 420. eco
 421. **parameter**
 422. **recover**
 423. handbook
 424. organic
 425. asset
 426. **appreciate**
 427. backpackers
 428. coefficients
 429. versus
 430. turkey
 431. geographical
 432. **random**
 433.
 434. competitiveness
 435. booking
 436. platforms
 437. **ratio**
 438. technologies
 439. **technical**
 440. turner
 441. **comprehensive**
 442. **retain**
 443. **adequate**
 444. **logic**
 445. **relax**
 446. disaster
 447. fuzzy
 448. **extract**
 449. **restrict**
 450. linear
451. resorts
 452. wildlife
 453. objectives
 454. **mutual**
 455. **priority**
 456. **underlie**
 457. **philosophy**
 458. narrative
 459. goods
 460. interdisciplinary
 461. **crucial**
 462. **intrinsic**
 463. efficacy
 464. coastal
 465. luxury
 466. **acquire**
 467. **proportion**
 468. ecological
 469. jones
 470. brands
 471. **induce**
 472. **version**
 473. **deviate**
 474. questionnaires
 475. **medium**
 476. **vision**
 477. **expose**
 478. **prospect**
 479. commodification
 480. county
 481. disability
 482. collaborative
 483. unesco
 484. **consult**
 485. **debate**
 486. **minor**
 487. household
 488. resilience
 489. outbound
 490. airlines
 491. google
 492. segment
 493. assets
 494. meditation
 495. temporal
 496. **nevertheless**
497. segmentation
 498. usage
 499. peak
 500. surf
 501. engaged
 502. **intervene**
 503. marine
 504. **advocate**
 505. **legal**
 506. **paradigm**
 507. **anticipate**
 508. favorable
 509. seasonality
 510. touristic
 511. budget
 512. peer
 513. sentiment
 514. **coordinate**
 515. causal
 516. platform
 517. trip advisor
 518. **modify**
 519. feedback
 520. likert
 521. **adult**
 522. **margin**
 523. faculty
 524. inbound
 525. supervisor
 526. **lecture**
 527. **reinforce**
 528. **ignorant**
 529. novelty
 530. mar
 531. **reject**
 532. entrepreneurial
 533. gaming
 534. ratings
 535. **thereby**
 536. **visible**
 537. taylor
 538. tween
 539. backpacker
 540. capture
 541. causality
 542. geographies

- 543. grant**
 544. mainland
 545. semi
 546. competencies
 547. cronbach
548. tense
 549. alpha
 550. mart
551. fundamental
552. label
553. ultimate
 554. dependence
 555. butler
 556. photo
557. considerable
558. convert
 559. correlations
 560. sharks
561. aggregate
562. contract
563. monitor
 564. agenda
 565. martin
 566. ski
567. decline
568. scenario
 569. empowering
 570. alienation
 571. gross
 572. predictors
573. entity
 574. expenditures
 575. matrix
 576. reviewers
577. ambiguous
578. assign
579. concentrate
580. formula
 581. loadings
 582. appraisal
 583. clusters
 584. equity
 585. software
 586. reciprocity
 587. relational
588. consent
- 589. complement**
590. explicit
591. minimum
592. scope
 593. novel
 594. citizenship
 595. wellness
596. couple
597. ministry
 598. coping
 599. density
 600. expenditure
 601. supervision
602. incentive
 603. engaging
 604. devel
 605. overnight
 606. whale
607. device
 608. cornell
609. resolve
 610. electronic
 611. hedonic
 612. humour
 613. postmodern
 614. standardized
 615. intangible
 616. disabilities
617. denote
618. sum
 619. anthropology
 620. cooper
 621. province
 622. culinary
 623. segments
 624. twitter
 625. forecast
 626. geographic
 627. islamic
628. supplement
629. aid
 630. promotional
 631. vietnam
 632. bramwell
 633. threshold
 634. timothy
- 635. colleague**
636. exploit
637. overseas
638. remove
 639. port
 640. weaver
 641. cuisine
 642. lynch
 643. norman
 644. rituals
 645. zone
646. assure
 647. crises
 648. informants
 649. quest
 650. sponsoring
 651. tangible
 652. aesthetic
 653. beverage
654. scheme
655. equip
656. rational
 657. museums
 658. null
659. sequence
660. brief
661. clarify
 662. meta
 663. squared
 664. tsunami
 665. latent
 666. simultaneously
667. allocate
668. alter
 669. pilgrim
670. justify
671. minimise
 672. sociological
 673. careers
 674. clustering
 675. embedded
676. neutral
677. ongoing
678. discrete
 679. mapping
 680. personnel

681. landscapes
682. vendors
683. exceed
684. archaeological
685. optimal
686. patients
687. temporary
688. bureau
689. mindfulness
690. souvenirs
691. worldwide
692. gaze
693. mobilities
694. shaw
695. duration
696. principal
697.
698. embodied
699. lens
700. stressors
701. golf
702. retail
703. correlated
704. empathy
705. feminist
706. interviewed
707. solidarity
708. vital
709. intelligence
710. register
711. ecology
712. exogenous
713. industry
714. liberal
715. pose
716. enterprise
717. facebook
718. forum
719. indices
720. ness
721. secular
722. springer
723. video
724. budgeting
725. robust
726. traffic
727. proactive
728. entrepreneur
729. gendered
730. inherent
731. specify
732. apparent
733. commission
734. forecasts
735. gray
736. memorable
737. supervisors
738. municipality
739. oaks
740. seasonal
741. incidence
742. welfare
743. somewhat
744. overview
745. verbal
746. atmosphere
747. attendees
748. credibility
749. holistic
750. mission
751. retailing
752. centrality
753. cosmetic
754. proximity
755. sabotage
756. guideline
757. impose
758. obvious
759. bridging
760. pike
761. recreational
762. attitudinal
763. database
764. discriminant
765. stake
766. stimuli
767. territory
768. vulnerability
769. predominant
770. export
771. intercultural
772. traits
773. circumstance
774. interval
775. blogs
776. elasticities
777. scenic
778. multivariate
779. nostalgia
780. outdoor
781. profiles
782. release
783. survive
784. corruption
785. nationality
786. prentice
787. species
788. coopetition
789. frontline
790. predictive
791. persist
792. infer
793. antecedent
794. crouch
795. marketers
796. peripheral
797. layer
798. fee
799. likewise
800. cues
801. longitudinal
802. traveller
803. competitors
804. appeal
805. profitability
806. vice
807. trigger
808. vehicle
809. ideology
810. sole
811. thesis
812. foster
813. predictor
814. prominent
815. embeddedness
816. paranormal
817. pilot
818. routine

- 819. accompany**
820. eventual
821. passive
 822. cation
 823. dale
 824. generational
 825. typology
 826. undergraduate
 827. societal
828. isolate
829. precise
 830. discourses
 831. endogenous
 832. expatriate
 833. mediator
 834. moderated
 835. doctoral
836. whereby
 837. niche
 838. census
 839. christian
 840. municipalities
841. accumulate
 842. intermediaries
 843. rethinking
 844. convergent
 845. metrics
 846. static
847. quote
848. overlap
849. substitute
 850. cognition
 851. confirmatory
 852. intra
 853. photos
 854. proposition
 855. vocational
 856. determinant
 857. disasters
 858. menu
 859. transaction
860. classic
 861. econometric
 862. superior
 863. augmented
 864. ben
 865. cope
 866. homestay
 867. curriculum
 868. dialogue
 869. intellectual
 870. murphy
 871. terrorism
872. mature
 873. captured
 874. carnival
 875. lag
 876. median
 877. periphery
 878. dataset
 879. heterogeneity
 880. itineraries
 881. workforce
882. contrary
883. schedule
884. attain
885. conform
886. detect
887. file
 888. arguably
 889. fieldwork
 890. hos
 891. mindful
 892. pop
 893. amenities
 894. sectional
 895. spillover
 896. surgery
897. restore
898. compensate
 899. focal
 900. offline
 901. pine
902. regime
 903. email
 904. healthcare
 905. nodes
 906. paternalistic
 907. ports
 908. territorial
 909. transparency
 910. traveler
 911. trustworthiness
 912. bachelor
 913. biodiversity
 914. overload
 915. smart
 916. sponsorship
 917. tucker
 918. aboriginal
 919. binary
 920. dredge
 921. dual
 922. guidance
 923. households
 924. inspired
 925. scholarship
 926. scholarships
 927. transformative
 928. vulnerable
 929. zones
930. definite
931. contradict
 932. buddhist
 933. ethnographic
 934. negotiation
 935. tech

936. vector
 937. von
 938. executive
 939. footprint
 940. griffin
 941. islam
 942. muslims
 943. spectrum
 944. weber
945. immigrate
 946. intrapersonal
 947. catering
 948. disruptive
 949. haul
 950. integrative
 951. sensory
952. converse
953. founded
954. trace
 955. bulletin
 956. era
 957. interestingly
 958. moderator
 959. ward
 960. award
 961. capitalism
 962. employability
 963. motivational
964. implicit
965. manipulate
966. simulate
967. fluctuate
 968. normative
 969. propensity
 970. stimulate
 971. trail
 972. barrier
 973. climatic
 974. congruence
 975. demographics
 976. mood
 977. international
978. instruct
979. devote
980. integrity
981. legislate
982. refine
 983. anti
 984. campus
 985. extant
 986. recruitment
 987. vacations
 988. authentication
 989. critique
 990. extrinsic
 991. superhost
992. eliminate
993. licence
 994. architecture
 995. bays
 996. bookings
 997. deemed
 998. hub
 999. impairment
 1000. lovelock
 1001. nomads
 1002. reform