JACET Journal 51 (2010) 1–13

Toward Creating a Specialized Vocabulary List for Tourism Majors: Analysis of its Profile and Receptive Knowledge Among University Students

FUJITA, Reiko
TSUSHIMA, Teruaki
Tokyo Keizai University

Abstract

This article reports preliminary results of a study aimed at creating a specialized English wordlist for tourism majors at universities, derived from a corpus that encompassed four different subfields (i.e., air travel, accommodation, travel firms and academic). First, the extracted EFT (English for Tourism) wordlist was analyzed in terms of its profile, specifically in relation to standard frequency lists (i.e., General Service Lists 1 & 2, and Academic Word List) as well as JACET 8000. It was found that about two-thirds of the vocabulary items in the EFT list were also considered important in general vocabulary learning, while the rest were highly technical items. Second, EFT vocabulary knowledge was investigated among university students with different levels of general vocabulary knowledge (N=131). Using vocabulary level tests, receptive knowledge of EFT and general vocabulary was measured in different frequency bands of JACET 8000. It was found that, although participants in different level-based groups showed varying levels of performance, their performance did not significantly differ on EFT and general vocabulary within the same frequency band. Then, the extracted items were classified based on their difficulty levels across the participant groups. Finally, implications for teaching EFT vocabulary are discussed.

Key Words: ESP, tourism, specialized vocabulary, corpus, vocabulary test

1. Introduction

Since the beginning of the 21st century, more and more Japanese universities have started tourism-related programs or faculties reflecting today's societal and educational needs. Because "tourism is an intercultural activity, constructed within and through language" (Jack & Phipps, 2005, p. 6), foreign language ability, particularly good command of English, is undoubtedly essential for students interested in finding a career in this industry. For these students, an English for Specific Purposes (ESP) approach may be the most effective, because an ESP approach places the highest priority on improving English abilities so as to meet the professional requirements of the field (Basturkmen, 2006).

An indispensable element in an ESP program is specialized vocabulary, as this "provides a sound basis for planning teaching and learning" (Nation, 2001, p. 205). In an English for Tourism (EFT) program, it is desirable to provide vocabulary appropriate to a wide variety of

tourism sectors (e.g., hotels, air travel, travel firms) so as to support students in a wide range of tourism-related work after graduation. Some tourism-related dictionaries or textbook glossaries available today, however, are mainly for vocational training and narrowly address a specialized occupation. For instance, a wordlist for the hotel and catering industry (Scott & Revell, 2004), and a glossary for in-flight cabin attendants (Beech, 1990) each offers a range of highly technical terms.

More recent research on computer-based analyses of large-scale language corpora has enabled researchers to create ESP vocabulary lists that each suit the needs of a relatively broad field, e.g., business (Chujo, Oghigian, Nishigaki, Utiyama, & Nakamura, 2007), or law (Ishikawa, 2004) to mention just two. Chujo et al., (2007), for instance, extracted a business vocabulary wordlist from spoken and written business components of the British National Corpus, on the basis of various statistical measures (e.g., log-likelihood ratio). They evaluated the validity of the extracted wordlist by comparing it with a business vocabulary dictionary and the U.S. grade level vocabulary list. They found that the different statistical measures used in the study effectively generated different levels of specialized vocabulary.

In the EFT field, Chujo, Utiyama & Oghigian (2006) created a corpus-based tourism wordlist for university students, generated from the Kyoto-Guide Corpus for promoting inbound tourism. The wordlist was classified by various statistical measures into three difficulty levels and was verified by comparing it to a grade level vocabulary list and a high school textbook vocabulary list. Although this study was valuable, as it successfully generated a wordlist using large-scale corpora and statistical measures, the list was narrowly targeted at a subfield within the travel industry.

More recently, Fujita (2009) made the first attempt to create comprehensive EFT wordlists for university tourism majors which encompassed four subfields (i.e., air travel, accommodations, travel-related firms and tourism-related academic journals), using a self-made corpus (N=40,000) of written English texts. These subfields were determined through an earlier needs analysis performed with the cooperation of students in a tourism program (Fujita, 2004). The study generated keywords for each subfield, and created four keyword lists and a combined list with the 150 most characteristic words. It was discovered that keywords in each list consisted of largely distinct sets of vocabulary.

Fujita (2010) further examined the degree of overlap between tourism industry vocabulary and tourism academic vocabulary, using an expanded corpus. The former was based on written texts (N=90,000) found in websites related to airlines, hotels and travel firms, while the latter on texts from tourism-related academic journals (N=50,000). It was found that the overlap of vocabulary in these two subfields was only 14%, indicating that tourism industry vocabulary and tourism academic vocabulary are largely distinct from each other. The results implied that teachers must select from or combine the two vocabulary lists, depending on the particular student needs.

These studies (Fujita, 2009, 2010), however, were limited in that the corpora were relatively small, and no indication was given as to the difficulty level of each word in the wordlists. The current study attempted to remedy these gaps by refining the specialized EFT wordlist on the basis of expanded corpora. It examined its profile, specifically in relation to the frequency bands in the JACET 8000 (JACET, 2003), the General Service List 1 & 2 (West,

1953) and the Academic Vocabulary List (Coxhead, 2000). It then examined receptive knowledge of the specialized EFT vocabulary among university students. It was believed that the present data would have important implications in understanding the traits of the specialized EFT vocabulary required in order to design an ESP curriculum.

2. Creating a corpus-based EFT wordlist

2-1. Method

2-1-1. Targets, representativeness and size of the corpus

First of all, it should be noted that the present specialized vocabulary list was targeted for students at colleges or universities. In order to secure representativeness of the overall field, a wide variety of frequent and important text categories were proportionally sampled (McEnery, Xiao, & Tono, 2006). Finally, approximately 120,000 words related to tourism were harvested in a corpus with adequate attention paid to its diversity within the specific fields and to its design[1]. A comparison corpus harvested nearly 240,000 words from a variety of websites not related to tourism. Table 1 shows the balance and diversity of the sources.

Table 1. Text Sources of the Tourism Corpus and General Corpus

Subfield	Word Count	Text Sources
Air travel	30,000	Two Japanese airlines, 4 non-Japanese airlines, 5 airports (in Japan, America, Europe, Asia), international aviation associations, customs (in America, Asia): 2,000 words were taken from each site.
Accommodations	30,000	Five hotel chains in Japan and abroad, 1 lodge, 1 resort hotel, 1 motel, 2 local hotels, 3 hotel associations (America, Middle East, Europe), 3 hotel operation related firms, 2 hotel reservation related firms, hotel and restaurant organizations: 1,500–2,000 words were taken from each site.
Travel-related firms & associations	30,000	Two travel agencies, 3 tourist offices, 3 travel reservations and information sites, international travel associations, 3 travel jobs and operations, world heritage/tour/theme parks: Approximately 2,000 words were taken from each site.
Tourism academic	30,000	Articles and abstracts in tourism or hospitality related academic journals (20,000 words), university tourism related course syllabus and descriptions (10,000 words).
General	240,000	1) Online news: 40,000 words. Taken from 20 topics on the menu bar, 2,000 words for each topic. 2) Online encyclopedia, 30,000 words: 2,000 words each from different categories. 3) Variety of websites opened randomly, 170,000 words: 2,000 words from each site.

2-1-2. Vocabulary list

Four sub-corpora were combined to constitute a tourism corpus. In order to generate a wordlist from the corpus, concordance software, AntConc 3.2.1., (Anthony, 2007) was used. In preparation, Lemma List File (Someya, 1998) was used so that the inflectional variants of words were reduced to their respective lemmas. For instance, the lemmas of fly, flew, flying

flown and flies were coded as a single lemma, FLY, and counted as the same lexis.

The corpus data were analyzed by using the "keyword list" function of the concordance software. This function compared the two corpora, tourism and non-tourism, and identified the words that were characteristic to the tourism corpus by applying log-likelihood ratio statistics[2]. The words identified as having high "keyness" were listed as products. Initially, the total number of word tokens was 122,041 and the total number of word types was 7,883. Then, 927 words were identified as keywords. This wordlist included proper nouns (e.g., London, Hilton) and non-words (e.g., km, hr), which were manually eliminated. Concurrently, basic level words which were categorized as JACET Level 1 (first 1,000 words for junior and high school students) were deleted, since these words were considered too basic to be included in the specialized vocabulary list. After the elimination, the final list contained 553 keywords.

2-2. Results

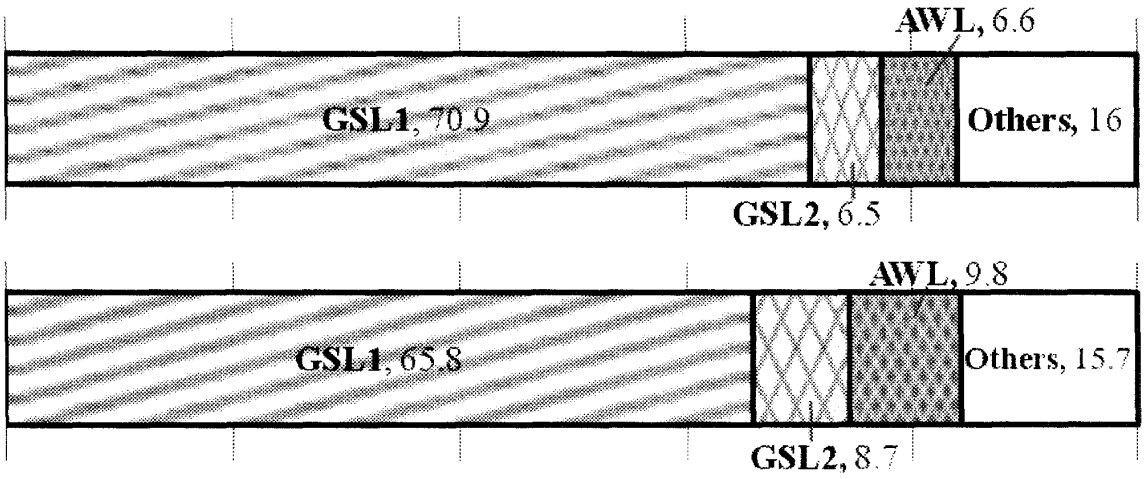

Figure 1. Profiles of the General Corpus (above) and the EFT Corpus (below).

The EFT corpus and the general corpus were compared to General Service List 1 (GSL 1: the most frequent 1,000 words), General Service List 2 (GSL 2: the 2nd most frequent 1,000 words) and Academic Word List (AWL) (see Figure 1). These are the largest and most commonly used word frequency lists (the "standard lists", henceforth), often used to measure vocabulary levels and sizes[3]. AWL contains 570 words that are not in the first 2,000 words but are frequent in university texts from a wide range of subjects. All of these lists include the base forms of words and derived forms. The first 1,000 words thus consisted of about 4,000 forms or types. The chi-square test[4] found that the frequency distribution of vocabulary items across the word frequency lists significantly differed between the two corpora (p=.000). In particular, the proportion of the basic vocabulary (i.e., GSL 1) was larger in the general than the tourism corpus, while the reverse was true with the intermediate-level vocabulary (i.e., GSL 2 and AWL).

Next, 553 individual keywords on the EFT wordlist were compared to GSL1, GSL2 and AWL. As is shown in Table 2, the EFT list included a rather high percentage of academic words (i.e., 32.6%), and the words covered by the standard lists accounted for about two-thirds of the EFT vocabulary (i.e., 64.0 %). The words outside the standard lists (i.e., 36.0%) were

highly technical terms of tourism (e.g., "affiliate", "airway", "amenity", "alliance", "beverage", "buffet", "camel"), which were highly specific to tourism contexts but much less common in general contexts.

Table 2. EFT vocabulary present in GLS 1 & 2, and AWL

List	GSL 1	GSL 2	AWL	Absent	Total
Count	84	90	180	199	553
%	15.2%	16.3%	32.5%	36.0%	100.0%

Furthermore, the 553 keywords were compared to the JACET 8,000 word list ("JACET 8000", henceforth), which is one of the largest and most readily available word frequency lists in Japan. The words in the JACET 8000 are listed in eight levels (from LV1 to LV 8), depending on the frequency and significance of each word. The words categorized as LV1 had already been eliminated from the EFT wordlist. Table 3 shows the percentage of words belonging to LV2 through LV9 (i.e., beyond LV8). It was found that the tourism vocabulary occurred most frequently at LV2 (31.8%), and that approximately two-thirds of the EFT vocabulary (i.e., 67 %) was within LV2, LV3 and LV4.

Table 3. EFT vocabulary classified by JACET 8000 level

Level	LV2	LV3	LV4	LV5	LV6	LV7	LV8	LV9	Total
Count	176	78	119	39	25	21	16	77	553
%	31.8%	14.0%	21.5%	7.1%	4.5%	3.8%	2.9%	14.4%	100.0%

3. Vocabulary test

3-1. Rationale

The analyses on the coverage of the tourism corpus and the EFT wordlist found that the corpus covered a relatively large proportion of the standard lists (i.e., GLS 1 & 2, AWL), and that the EFT wordlist contained a high proportion of high-frequency words (JACET LV2-LV4). In order to relate these findings more directly to a curriculum that would include teaching the EFT wordlist, the present study further investigated EFT vocabulary knowledge among university students in Japan. It specifically examined how the receptive knowledge of EFT vocabulary differed among students with different levels of general vocabulary knowledge and across different frequency bands of JACET 8000. It also examined whether the level of knowledge differed between general and EFT vocabulary in the same frequency band of JACET 8000, as there was a possibility that the EFT vocabulary would be relatively difficult, insofar as some lexical items may be associated with meanings exclusively used in specific tourism contexts.

3-2. Method

3-2-1. Participants

Participants in the vocabulary test were 131 students at a private university in Tokyo,

Japan. They were enrolled in different levels (e.g., advanced and intermediate) of the first-year general English course. One student was excluded from the analyses because his performance in the test was extremely low[5].

3-2-2. Test format

Two vocabulary levels tests were created, one for tourism items and the other for general items. The design of the test was based on "The Vocabulary Levels Test" (Nation, 2001), which "has been widely used and well researched" (Nation, 2008, p. 141) as well as Mochizuki's vocabulary size test for Japanese students (Mochizuki, 1998). The present tests used the JACET 8000 as a scale of frequency. They were divided into three levels; LV2 (2nd 1,000 words), LV3 (3rd 1,000 words), LV4 and above (4th to 8th, for a total of 5,000 words). Each level of the test contained 18 items. One section of the test had three questions in the form of Japanese translation and six possible choices in the form of English words to choose from.

1. ___ 四角　2. ___ 薬　3. ___ 気候

a. climate　b. trace　c. square　d. data　e. yard　f. medicine

The non-tourism vocabulary test was created by sampling 18 words randomly from the corresponding levels of the JACET 8000. The most commonly used meaning of each sampled word was translated into Japanese. The other items for multiple-choice were also randomly taken from the same levels. For the tourism vocabulary test, each word in the EFT wordlist was matched with the level of JACET LV2-LV8, then 18 samples were randomly taken from three test levels. Items for multiple-choice were also randomly taken from the same-level words in the EFT wordlist. Each word was translated into Japanese words which would be used in tourism situations. Each test contained three levels, 18 items to be tested for each level, totaling 54 words in one test. These tests were named as Type A (non-tourism) and Type B (tourism).

3-2-3. Test procedure

The test was given in eight separate English classes on the last day of the course in the academic year of 2009. A separate sheet was prepared for each test, Type A and Type B. To avoid confusion in administering the tests, all the students in one class took the two tests in a fixed order. The order of test was varied such that about half the participants took the test in one order (i.e., from Type A to Type B: N=66) and half in the other (i.e., from Type B to Type A: N=65).

3-3. Results

3-3-1. Reliability of the receptive vocabulary test

Reliability analyses were conducted to assess the reliability of the test format and test items used in the receptive vocabulary test. A reliability measure, Cronbach's alpha, was computed on all the 108 items in the two tests. The results showed that the reliability coefficient was .93, indicating that the receptive vocabulary test was adequately reliable.

3-3-2. Classification of participants

For each participant, mean percents correct were calculated across the two levels of vocabulary types (i.e., general and tourism vocabulary) and three levels of JACET frequency bands (i.e., LV2, LV3 and LV4 & above). Using the six mean scores as variables, a hierarchical cluster analysis procedure was conducted to classify all the participants (N=130)[6]. The analyses used z-standardized scores with the Ward's method as a clustering method and the squared Euclidian distance as a distance measure[7]. Examination of the results indicated that the participants could best be classified into four groups (see Table 4). Looking vertically down at the grand means averaged across the vocabulary types and frequency bands (in the rightmost column), the four cluster groups (termed Participant Cluster, "PC") were ordered from top to bottom according to the average performance.

Table 4. Percents Correct by Clusters of Participants, Type and Level8

Type		General			Tourism			
Level		LV2	LV3	LV4~	LV2	LV3	LV4~	
	N	*M*	*M*	*M*	*M*	*M*	*M*	*GM*
PC1	33	94.8	92.4	74.7	92.4	96.1	81.3	88.6
PC2	50	89.1	86.1	60.8	84.2	82.1	63.9	77.7
PC3	20	76.4	70.3	42.2	86.7	76.1	48.3	66.7
PC4	27	70.8	66.9	40.5	69.5	53.9	38.3	56.7
Total	130	84.8	81.3	57.3	83.6	78.9	60.6	74.6

Notes. PC= Participant Cluster, LV=Level, M=Mean, GM=Grand Mean

Overall, the participants performed best at LV2, followed by LV3 and LV4 & above, as expected. Their performance at LV4 & above was particularly poor because the lexical items at this level were drawn from all the frequency bands at LV4 and above LV4. With regard to comparisons between general and tourism vocabulary, PC3 showed consistently higher mean percents correct for the tourism vocabulary across the frequency bands, while PC4 showed the opposite pattern. However, the statistical analyses, using Wilcoxon signed ranks tests[9], found that the difference was significant only at LV2 in PC3 (p=.001) and at LV3 in PC4 (p=.000). The overall results indicated that, within the same frequency band, the receptive knowledge of participants did not significantly differ between tourism vocabulary and general vocabulary.

3-3-3. Classification of the tourism vocabulary

The following analyses classified the tourism vocabulary used in the test (N=54) according to the level of difficulty. For each lexical item, mean percents correct were calculated across the three levels of the frequency bands and the four levels of the PC groups. With these 12 data points, hierarchical cluster analyses were conducted, using the same analysis procedures as described in 3-3-2. The results indicated that the vocabulary items could best be divided into four cluster groups (termed Vocabulary Cluster: "VC"), as shown in Table 5. VC1 (N=14)

included the vocabulary items which the participants in all the PC groups found the least difficult, with mean percents correct being over 90% (e.g., flight, tour, customer). VC4 (N=10), on the other hand, was a group of the most difficult vocabulary items, with mean percents correct below 50% in three of the PC groups (e.g., aviation, monastery, accommodation).

Table 5. Percents Correct by Clusters of Vocabulary and Participants

		PC1		PC2		PC3		PC4		LV2	LV3	LV4~
VC	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*N*	*N*	*N*
VC1	14	98.9	2.6	96.9	3.8	99.3	1.8	91.5	7.5	11	3	0
VC2	17	95.2	6.2	85.3	8.7	81.8	7.9	57.3	9.6	3	9	5
VC3	13	94.2	3.4	71.8	11.4	54.2	10.6	33.3	9.4	1	6	6
VC4	10	63.0	10.6	40.4	14.1	31.5	16.0	22.2	7.6	3	0	7
Total	54	90.0	14.3	76.7	21.8	70.4	26.5	53.9	27.3	18	18	18

Notes. VC= Vocabulary Clusters; PC= Participant Clusters

Table 5 also shows the number of lexical items that belong to each JACET frequency band. In the clusters of easier vocabulary items, the greater proportion of lexical items belonged to the lower frequency bands, as expected. However, some notable exceptions were found: some items from higher frequency bands were included in the clusters of easier vocabulary items and vice versa. For example, five items that belonged to LV4 & above (i.e., seasonal, ferry, lounge, departure and receipt) were included in the VC2. These items obtained high percents correct because, for most of them, there were corresponding loanwords which have similar meanings in Japanese, and are written in Japanese *katakana*. On the other hand, three items in LV2 were found in the VC4 (i.e., arrange, charge and fair). These lexical items have one common meaning with which they are frequently used, as well as another much less common meaning that is exclusively used in some tourism contexts. Some other examples found in the wordlist are listed in Table 6.

Table 6. Examples of high frequency words with technical meanings

Word (J-Level)	General meaning	Meaning in tourism
tip (LV2)	the pointed end of something long and narrow	gratuity
gate (LV2)	a door in a fence or hedge	an exit from an airport building to an aircraft
minor (LV3)	lesser importance	child below the age of legal responsibility
transfer (LV3)	move or displace something	change from one vehicle to another
confirm (LV3)	establish the correctness or truth	make an arrangement definite
attraction (LV4)	appeal, feeling of liking someone	some place that people go to for enjoyment

4. Discussion and teaching implications

The present study provided some important findings in terms of the profiles of the tourism corpus and the extracted EFT wordlist, the difficulty levels of the words in the list, as well as implications for teaching EFT. First, the results of the analyses on the profiles of the tourism corpus showed that it contained a relatively large proportion of vocabulary from the standard lists (i.e., GSL 1 & 2 and AWL), as compared with certain corpora of the other specialized fields. For example, the text coverage of the standard lists in the tourism corpus was approximately 84%. By contrast, Nation's investigation on technical vocabulary of anatomy using the frequency and range-based approach found that the standard lists had as low as 64.6 % coverage (Nation, 2008, p. 136). Second, the examination of the extracted wordlist showed that a fair proportion of tourism vocabulary items was not highly technical but was within the range of what is commonly considered as general vocabulary (see Figure 1). For example, the vocabulary in the standard lists accounted for 64% of the EFT wordlist, with the amount of AWL being particularly large (32%). In addition, approximately 67 % of the vocabulary was within the JACET LV2, 3 and 4[10].

The overall findings suggested that the first essential step to teaching EFT vocabulary in any curriculum is to ensure that the students master the high frequency words up to LV4 in the JACET 8000 as well as the academic words in AWL. At this level, students should be encouraged to learn both tourism and general vocabulary. The teaching goal of the participants in PC3 and PC4, for example, is to acquire these levels of vocabulary (see Table 4 and 5). This general learning would involve learning the vocabulary in the EFT list probably up to VC3 (see Appendix A). The second step is to teach highly technical, mostly low-frequency words (e.g. alliance, buffet, casino, excursion etc.), which are exemplified by the items in VC4 (see Appendix A). At this level, target lexical items might be selected based on the results of needs analysis of students in a particular program. As described in the introduction, Fujita (2009, 2010) found that the keyword lists of different subfields of tourism comprised fairly distinct sets of vocabulary. For instance, if a majority of students in a program are interested in the hotel industry, words associated with the hotel subfield may be selected as material for focused vocabulary teaching.

Finally, the results on the vocabulary test showed that some lexical items (i.e., arrange, charge and fair) in a low frequency band in JACET 8000 (i.e., LV2) were found in the vocabulary cluster of a difficult level (i.e., VC #4). These lexical items had one common meaning with which they are generally used, as well as another much less common meaning that is frequently used in some tourism contexts (see Table 6). These items should be listed and treated differently in the process of classification. At a practical level, teachers are strongly encouraged to pay attention to this type of item. Nation emphasized the importance of helping learners to recognize the "connections and differences between the high-frequency meanings and the technical uses" (2001, p. 19) of such vocabulary. Such words are best learned as a part of input activities such as reading or listening (Basturkmen, 2006; Nation, 2001, 2008). With special attention by the teachers, students are able to understand how the word fits into a framework of specific knowledge.

Future research may be directed toward further expanding the size of corpora. Although

the major findings of the present study may well remain intact, further increase in the size of corpora would bring about a more accurate EFT wordlist, especially with respect to the lexical items with relatively low keyness. Another step in this line of research is to obtain further data on the difficulty levels of the EFT vocabulary. It would be very useful if one could determine a list of the top 100 lexical items with respect to "keyness" in each of the four subfields of tourism, and classify them by difficulty level.

Acknowledgements

This research was supported by a grant from Tokyo Keizai University. The authors are grateful to the anonymous reviewers for their insightful feedback and advice on improving the manuscript. The authors are also grateful to Aika Miura, Sayoko Maswana, Chikako Nakagawa and Alan Stoke for their contributions at various stages in the preparation of the manuscript.

References

Anthony, L. (2007). AntConc 3.2.1. Retrieved from http://www.antlab.sci.waseda.ac.jp/antconc_index.html

Basturkmen, H. (2006). *Ideas and options in English for specific purposes.* Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers.

Beech, J. G. (1990). Glossary. *Thank you for flying with us : English for in-flight cabin attendants* (pp. 124–128). New York: Prentice Hall.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics : investigating language structure and use.* Cambridge ; New York: Cambridge University Press.

Chujo, K., Oghigian, K., Nishigaki, C., Utiyama, M., & Nakamura, T. (2007). Creating E-learning Material with Statistically-Extracted Spoken & Written Business Vocabulary from the British National Corpus. *Journal of the College of Industrial Technology, Nihon University 40,* 1–12.

Chujo, K., & Utiyama, M. (2006). Selecting level-specific specialized vocabulary using statistical measures. *System, 34*(2), 255–269.

Chujo, K., Utiyama, M., & Oghigian, K. (2006). Selecting Level-Speciic Kyoto tourism Vocabulary using statistical measures. *New Aspects of English Language Teaching and Learning* (pp. 126–138). Taipei: Crane Publishing Company Ltd.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34*(2), 213–238.

Fujita, R. (2004). Hospitality English: Needs analysis and course design as an ESP course. *The journal of Rikkyo University Language Center 9,* 31–45.

Fujita, R. (2009). Using an Ad Hoc Corpus to Create an ESP Vocabulary List for Students Majoring in Tourism. Paper presented at the JACET 48th Annual Convention Program, Hokkai-Gakuen University.

Fujita, R. (2010). Generating English vocabulary lists for students majoring in tourism. *Journal of Tourism Research 51.* (in press)

Ishikawa, S. (2004). Development of an ESP Vocabulary List for College Law Students : Based on the Statistical Comparison of Word Frequencies in US Judicary English Corpus and Freiburg-Brown Corpus. *Journal of the School of Languages and Communication, Kobe*

University 1, 13–27.
JACET. (2003). *JACET List of 8000 Basic Words*. Tokyo: JACET.
Jack, G., & Phipps, A. M. (2005). *Tourism and intercultural exchange : why tourism matters*. Clevedon ; Buffalo: Channel View Publications.
McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies : an advanced resource book*. London ; New York: Routledge.
Mochizuki, M. (1998). Nihonjin gakushusha notameno goi tesuto [A Vocabulary size test for Japanese Students]. *The IRLT bulletin 12*, 27–53.
Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge ; New York: Cambridge University Press.
Nation, I. S. P. (2008). *Teaching vocabulary : strategies and techniques*. Boston, MA: Heinle.
Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
Scott, T., & Revell, R. (2004). Wordlist. *Highly recommended: English for the hotel and catering industry* (pp. 97–109). Oxford: Oxford University Press.
Someya, Y. (Producer). (1998). English lemma list (version 3). Retrieved from http://www.lexically.net/wordsmith/version3
West, M. (1953). *A general service list of English words* (Rev. and enl. ed.). London: Longmans.
Yamamori, K., Isoda, T., Hiromori, T., & Oxford, R. (2003). Using cluster analysis to uncover L2 learner differences in strategy use, will to learn, and achievement over time. *International Review of Applied Linguistics in Language Teaching, 41*, 381–409.

Notes

[1] It is generally believed that larger corpora can generate more accurate wordlists, and so are particularly desirable in lexicographical studies. Bier, Conrad & Reppen (1998, pp. 248–249), however, claimed that more important than size is the diversity and design. It is inevitable that every corpus has limitations due to time and financial restrictions, but a "well-designed corpus will still be useful for investigating a variety of linguistic issues (Biber, Conrad & Reppen: 1998: p. 250)."

[2] Oakes (1998, p. 174) found that, among some statistical measures tested, irrespective of the corpus size, the log-likelihood ratio worked the best in identifying technical terms. Chujo & Utiyama (2006) further showed that the log-likelihood ratio was an appropriate statistic by which to distinguish intermediate-level vocabulary.

[3] The Nations Range Program is downloadable from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx.

[4] All the statistic tests were conducted using SPSS software, Version 17.

[5] The mean percent correct in the two tests was 22.2.

[6] The cluster analyses were employed for the following reasons. First, there were no external measures which could be used to group the participants in terms of their vocabulary knowledge. Second, using the averaged scores to divide the participants may have failed to detect a group of participants with relatively high scores on tourism vocabulary and low scores on general vocabulary, or vice versa.

[7] The analysis procedure followed a recommendation put forward by Yamamori, Isoda,

Hiromori, & Oxford (2003).

[8] The values for standard deviation could not be shown due to insufficient space. They ranged from 4.9 to 19.0.

[9] As a Shapiro-Wilks test showed that the sample distribution of the percents correct was significantly deviant from normality ($p<.05$), a non-parametric, Wilcoxon signed ranks test was conducted. The Wilcoxon test was used whenever the assumption of normality was violated in the following analyses.

[10] An anonymous reviewer pointed out that the following two sets of data appeared to contradict each other. On the one hand, a fairly large proportion of the extracted EFT vocabulary was categorized as "general" vocabulary. On the other hand, each EFT subfield consisted of a fairly distinctive set of vocabulary. However, this is exactly what was found. For example, the word "adventure" is included in the standard lists and LV2 as it frequently occurs in general contexts such as stories. But, when the EFT vocabulary was extracted from the tourism corpus and separated into different subfields based on keyness, the item remained only in the travel industry list, and was absent from the other subfields. Other examples of such items include "breakfast", "daily" and "luxury" in the hotel industry; and "equipment", "delay" and "entry" in the airline industry (see Appendix A).

Appendix A. List of lexical items classified by group level

Group	JCT	Words	Subfield	Group	JCT	Words	Subfield
VC1	LV2	adventure	T	VC3	LV2	estimate	AC
BSC	LV2	airport	H/A/T	INT2	LV3	confirm	A
	LV2	breakfast	H		LV3	cruise	T
	LV2	conference	H/T		LV3	destination	H/A/T/AC
	LV2	cultural	T/AC		LV3	document	A
	LV2	customer	H/A/AC		LV3	heritage	T
	LV2	daily	H		LV3	proceed	A
	LV2	equipment	A		LV4~	equip	H
	LV2	flight	A/T		LV4~	exceed	A
	LV2	restaurant	H/T/AC		LV4~	refund	A/T
	LV2	tour	H/T		LV4~	regional	H/T/A/AC
	LV3	outdoor	H/T/A/AC		LV4~	resident	A/T/AC
	LV3	overseas	T		LV4~	terminal	A
	LV3	vacation	H/T/AC				
VC2	LV2	available	H/A/T	VC4	LV2	arrange	H/T/AC
INT1	LV2	award	H/A	ADV	LV2	charge	H/A
	LV2	purchase	H/A		LV2	fair	T
	LV3	delay	A		LV4~	accommodation	H/A/T/AC
	LV3	entry	A		LV4~	aviation	A
	LV3	luxury	H		LV4~	cuisine	H
	LV3	participate	H/T		LV4~	currency	A
	LV3	preference	AC		LV4~	domestic	A
	LV3	promotion	H		LV4~	hospitality	H/AC
	LV3	recommend	A		LV4~	monastery	T
	LV3	reservation	H/A/T				
	LV3	transfer	H/T/A/AC				
	LV4~	departure	A/T				
	LV4~	ferry	T				
	LV4~	lounge	H/A				
	LV4~	receipt	A				
	LV4~	seasonal	H/T/A/AC				

Notes. JCT=JACET, BSC=Basic, INT=Intermediate, ADV=Advanced, H=Hotel, A=Airline, T=Travel, AC=Academic