Teaching Specialized Vocabulary by Integrating a Corpus-Based Approach: Implications for ESP Course Design at the University Level

Hsiao-I Hou[1]

[1] Deaprtment of Applied English, National Kaohsiung University of Hospitality and Tourism, Taiwan

Correnpondence: Hsiao-I Hou, Department of Applied English, National Kaohsiung University of Hospitality and Tourism, Kaohsiung 812, Taiwan. Tel: 886-7-806-0505 ext. 1920. E-mail: houx0020@gmail.com

Abstract

The purpose of this study is to demonstrate how to integrate two in-house specialized corpora into a university-level English for Specific Purposes (ESP) course for nonnative speakers of English. The ESP course was an introductory level of wine tasting for Applied English Department students at a university specializing in hospitality in Taiwan. Two corpora of wine tasting notes selected from the official website of the Liquor Control Board of Ontario (LCBO) in Canada, one for red wine and one for white wine, were compiled. Lexical density and vocabulary compositions were analyzed. The results show that the lexical density and the percentages of specialized vocabulary of the wine corpora were higher than in other disciplines. In addition, wine reviewers used different vocabulary to describe the characteristics of white wine and red wine. From the keyword analysis, terms related to cooking methods and food names appeared in high frequencies. Based on the corpora analysis results, vocabulary lists, the LCBO website, and the in-house corpora were introduced to the students as supplementary materials. The pre- and posttest results for vocabulary indicate that the students enrolled in this program gained significant progress in both content and language knowledge. Based on the study results, recommendations for ESP teaching and materials development are discussed.

Keywords: specialized vocabulary, corpus linguistics, ESP teaching

1. Introduction

In the teaching and learning of a second language (L2) or foreign language (FL), vocabulary is an essential component of all comprehension (Folse, 2010). Whether in real-world tasks such as asking for travel information or in academic tasks such as textbook reading, vocabulary is essential for the success of language learners. Teachers in content-area classes (such as introduction to psychology and introduction to chemistry) frequently teach content-specific vocabulary but are not aware of the challenges that L2 and FL learners encounter because they do not know the specialized vocabulary or the basic words surrounding the content-specific vocabulary. Therefore, to help L2 and FL learners study more effectively in content-and-language integrated learning (CLIL), facilitating their vocabulary acquisition is a major focus in English for Specific Purposes (ESP) teaching and research.

In recent vocabulary teaching and learning research, three categories of words that FL learners and teachers should pay attention to based on the calculation of their frequencies have been identified (Hou, 2012). These words are (a) the General Service List (GSL) compiled by West (1953) which is commonly referred to as the 2,000 most frequent words in English; (b) the 570-word Academic Word List (AWL) compiled by Coxhead (2000), containing words that are found more frequently across a wide range of academic texts than in nonacademic texts; and (c) specialized or technical vocabulary. Word lists such as the GSL and the AWL were created over long periods of time to assist students and teachers, but these lists present a number of problems (Wang, Liang, & Ge, 2008). The GSL corpus is a collection of general texts from the 1950s, and the AWL was generated from texts in the late 1990s. Few efforts have been made to update either list for timeliness or with regard to the specific content domains encountered by college and university students in high academic or occupational-purpose classes such as tourism management or restaurant operation. Neither list accounts for the high level of lexical characteristics of many common English words and specialized vocabulary words. If

teachers and material writers have more knowledge of which common words or technical words, and specifically which meanings of those words, are more likely to appear in students' academic or occupational settings, then they will be better positioned to help students acquire the vocabulary needed for successful comprehension of the target language.

The purpose of this study is to demonstrate how to integrate specialized corpora from a particular hospitality field into an introductory wine tasting course and to identify core and specialized vocabulary word lists for practical ESP teaching materials for nonnative speakers of English at the university level. To prove the effectiveness of this teaching approach, the progress of learners will also be reported. In the following sections, essential literature and background information for this study are presented. The study then introduces the research questions and study results and proposes ESP vocabulary teaching implications.

2. Literature Review

To satisfy the research purposes, two major topics will be discussed in this section. First, issues in corpus linguistics, especially language teaching applications, will be presented. Second, specialized vocabulary and its teaching approaches will be discussed.

2.1 Corpora in Language Teaching

2.1.1 Corpus Definition

Corpus linguistics is a broad field that studies a representative body of language as it exists in authentic speech or text rather than language that is invented for the purpose of illustrating a certain linguistic construct or theory (Meyer, 2002; Murphy, 1996). Meyer (2002, p. xiii) referred to corpus linguistics as not only a linguistic field but also a methodology that presupposes that it is "important to base one's analysis of language on real data—actual instances of speech or writing—rather than on data that are contrived or 'made-up'". Other characteristics of corpus studies include the use of computers for quantitative analysis and the use of qualitative techniques by researchers. Corpus studies can be used to study a wide range of language phenomena, including historical and stylistic conventions, discourse patterns, language acquisition, grammatical constructions, and vocabulary analyses (Teubert & Cremakova, 2007).

2.1.2 Corpora in Language Education

In addition to informing researchers, corpus linguistics has valuable classroom applications for language pedagogy. From a teaching perspective, based on the characteristics of a corpus, it can be categorized as a general or specialized corpus. A general corpus is most useful for developing an idea of the global behavior of vocabulary items or phrases and can be used for publishing dictionaries and textbooks. Unlike the corpora for dictionaries, a specialized corpus consists of language that is selected for specific purposes. A specialized corpus is particularly useful for understanding how language is used in a particular register of the language (Tribble, 1990). For instance, a corpus that consists of creative writing texts allows for an analysis of images and metaphors that were applied by specific writers and facilitates comparative studies of writers. Similarly, a corpus that consists of business letters can provide insight into the grammar and types of wording that are appropriate for a variety of politeness strategies (Cornnor, Davis, & DeRycker, 1995).

Today, both dictionaries and computerized corpora are key resources for L2 learners. With the development of user-friendly concordance programs and English corpora, attention has shifted to an approach using authentic language samples in the classroom (Johns, 2002). Wichmann (1995), for example, used corpora and a concordance program to teach German because she believed that dictionaries do not provide a sufficiently authentic meaning. A significant contribution that a corpus brings to L2 acquisition is that the materials that it provides are authentic examples. A corpus-based approach gives authentic examples while providing students with opportunities to be explorers (Dodd, 1997). With authentic examples, learners observe the true use of language. Flowerdew (1993) cautioned against using manipulated examples because of the possibility that learners may not observe the actual use of the target language. Therefore, many scholars promote the use of corpora in L2 and FL education as beneficial in developing reading and writing skills and in understanding and producing particular texts and types of texts (Aston, 2001; Flowerdew, 2001; Connor & Upton, 2004; Römer, 2011). From the perspective of language learners, corpus-based L2 and FL instruction can offer a means by which to increase their motivation and render them more autonomous (Woolard, 2000) while allowing them to mine language descriptions through the corpus (Aston, 2001). Several studies have reported the effectiveness of using corpora in teaching applications, including adopting concordancing software during the acquisition of L2 or FL, such as through collocation learning (Howarth, 1998; Sun & Wang, 2003), lexical acquisition (Cobb, 1999), writing (Sun, 2007; Tribble, 2001), and grammar (Sun, 2003).

In summary, researchers and practitioners have recognized the usefulness of corpora and various software applications in language classrooms and have developed course designs, materials, and classroom applications based on corpus-based language instruction. Corpus-based language learning can meet the needs of learners by stimulating their motivation to learn with authentic examples (Sun, 2007) through inductive thinking strategies, developing ESL learners' comprehension and transforming learned linguistic knowledge into reading and writing skills.

2.2 Specialized Vocabulary and Teaching Approaches

2.2.1 Definition of Specialized Vocabulary

Nation (2008) noted that specialized or technical vocabulary can come from high-frequency (HF) or low-frequency (LF) words in the GSL and in academia. He defined specialized vocabulary as words that are "recognizably specific to a particular topic, field or discipline" (2001, p. 198). Acquiring specialized vocabulary is crucial to learners to attain academic literacy and to become part of their chosen academic discourse communities. In recent years, numerous studies have focused on and contributed to the knowledge of specialized vocabulary (Chung & Nation, 2004; Fraser, 2004; Kaewphanngam, 2002; Rizzo, 2010; Sutarsyah et al., 1994). These studies have helped to define exactly what determines whether a word is specialized and how much specialized vocabulary is found in textbooks. Further studies have begun to address the issues of how well L2 students are acquiring this important vocabulary.

2.2.2 Specialized Vocabulary in ESP Research and Teaching

According to Waring and Nation (2004), a reader needs to know at least 95% of the surrounding vocabulary to successfully guess the meaning of a new word through context. If readers are not capable of doing so, then they will need to interrupt their reading to find the meaning of the word in a dictionary. If specialized vocabulary constitutes as much as 30% of a text, then it will be difficult for students to acquire these words incidentally when they first meet them, and it will be extremely slow and frustrating for them to be constantly using a dictionary. Chung and Nation (2003, 2004) found that the specialized vocabulary of an anatomy textbook consisted of 4,270 words, which is a large number of words for students to use a dictionary to find the meanings of words or to acquire them incidentally. The researchers (Chung & Nation, 2003) found that LF vocabulary constituted 11.8% of the words in a specialized anatomy text. If the level of LF vocabulary is relatively high in graduate texts, then the task of acquiring specialized vocabulary becomes even more difficult for nonnative speakers of English. As Chung and Nation (2004) observed, there is a need for further research focused on specialized vocabulary words because the number of these words and their characteristics are not well understood.

To identify specialized vocabulary words, Chung and Nation (2004) suggested four different approaches, which they then tested on an anatomy textbook to verify which approach was the most accurate. In their testing, Chung and Nation classified as specialized any words that occurred at least 50 times more often in the specialized text than in the general corpus. Overall, this approach was determined to be as reliable as using a dictionary, but the authors found that it did not identify specialized terms that were also commonly used outside of the field. Fraser (2005) replicated the research of Chung and Nation (2004) using a similar rating scale on a different applied linguistics text and a pharmacology text with similar results. Sutarsyah et al. (1994) used a corpus analysis on an undergraduate economics textbook and then checked their findings with a dictionary of economics and found that the number of vocabulary words was similar to that found in other undergraduate texts.

Nation (2008) estimated that the specialized vocabulary of a particular field consists of 1,000 to 5,000 words. Corpus linguistics analyses of university textbooks have shown that the GSL's 2,000 most common words are found in approximately 80% of academic texts and that academic vocabulary appears in 8.5% to 10% of texts (Coxhead & Nation, 2001; Sutarsyah et al., 1994; Ward, 1999). Nation (2001, p. 12) thus estimated that specialized vocabulary appears in approximately 5% of academic texts, with LF words constituting the final 5%. However, these analyses were conducted on introductory-level undergraduate textbooks.

Because specialized vocabulary was estimated to constitute only approximately 5% of these introductory-level texts, it has perhaps not drawn sufficient attention. However, some researchers have advocated that more attention should be given to teaching specialized vocabulary. Sutarsyah et al. (1994) advocated giving attention to HF, specialized vocabulary. The authors found that using texts from a variety of disciplines to introduce vocabulary led to learners acquiring a great deal of vocabulary that would be of no use to them in their own disciplines. Therefore, these researchers emphasized the value of ESP classes, in which students concentrate on the vocabulary that they need for their disciplines. In addition, more recently, there have even been suggestions that academic vocabulary should be taught in a more specialized manner. Hyland and Tse (2007) suggested that

academic vocabulary should be viewed as discipline-specific words ranging from those that are particular to one discipline to those that share some meaning and use with words in other fields, and they concluded that these should be the words that teachers introduce and spend time practicing in the ESP classroom. These researchers also suggested that corpus-informed lists should be established from the texts that students will need to read and the genres that they will need to write in for their various disciplines.

Specialized vocabulary definition and acquisition have been seriously lacking in research. In addition, current research outcomes are not sufficient to instruct on how to teach specialized vocabulary. One controversial issue has been whether to teach vocabulary implicitly or explicitly for better comprehension. Schmitt (2000) and Nation (2001) suggested that a combination of both approaches be utilized because certain vocabulary is best acquired incidentally through repeated exposure or specific learning activities, whereas other features, such as word forms or grammar features, should be explicitly explained.

The present study will demonstrate a case study to show how to integrate specialized corpora from the Internet and how to teach wine vocabulary to nonnative speakers of English who are hospitality majors at the university level. Implications for teaching specialized vocabulary and ESP material development in other disciplines will be recommended based on the study results. The following research questions will guide this study:

1) In the wine discourse, what are the characteristics of word composition?

2) After the implementation of this corpus-based approach in teaching, do hospitality students' English competence and subject knowledge of wine improve?

3) What implications for ESP teaching and material development can be made based on the study results?

3. Methodology

In this section, detailed descriptions of the course design and the specialized corpora used in this study will be presented. Subsequently, the linguistic features analyzed in this study will be discussed. Finally, the analysis tools and evaluation procedure will be described.

3.1 The Course Context and the Specialized Corpora

3.1.1 The Course Context

To enhance students' English learning and promote an internationalized curriculum, a content-and-language integrated learning (CLIL) program was implemented at a university specializing in hospitality in Taiwan where the researcher works. The CLIL approach involves combining content and language teaching, offering equal attention to each (Dalton-Puffer, 2011), and the basic concept is similar to the ESP teaching approach. After careful evaluation and selection, several introductory-level courses, including *Introduction to Hospitality, Food Culture*, *Introduction to Wine Tasting*, and *Service Management,* were chosen as model courses for demonstrating CLIL teaching approaches. Since 2011, these courses have been taught in the Department of Applied English and in the International Bachelor Program of Tourism Management at this university.

In both the CLIL and ESP teaching principles, subject areas demonstrate specific language features or registers; therefore, learning specific subjects and registers is important for mastering specific content (Brinton et al., 2003). In addition, while learning a second or foreign language, learners should be exposed to real-life and authentic materials. Based on these concepts, two in-house corpora were compiled by the researcher and integrated as part of the teaching materials for the *Introduction to Wine Tasting* course. The teaching objectives for this wine course are as follows:

1) To understand the factors that influence the style and price of wine

2) To know the characteristics of the principal grape varieties

3) To provide information to customers about wines and spirits

4) To produce analytical tasting notes on wines using industry-recognized tasting terms

This wine course was offered in the Applied English Department for first-year university students during the winter 2012 term as an elective course. This course yielded two credits following successful completion and comprised 36 hours of class time that was divided into 2-hour sessions each week for 18 weeks. The 42 students enrolled in the class were aware that it was a CLIL program course, which meant that English would be the medium for instruction. Among the students, 5 of them (12%) were in CEF A2 level, 28 (67%) were in CEF B1, and 9 (21%) were in CEF B2. In addition to one required textbook, the in-house corpora were used as supplementary material for this course.

3.1.2 The Specialized Corpora

To compile the in-house corpora, the design methodology followed the suggestions of Coxhead (2000) and Biber et al. (1998) for specialized corpus design. As Coxhead (2000) stated, collecting various short texts increases the representativeness of the corpus and decreases the bias. In addition, the size should be approximately 100,000 words to reach acceptance requirements, and the texts should be collected from various sources. In addition, Biber et al. (1998) stated that the texts included in the corpus should be available for everyone to access. To satisfy the abovementioned principles and the course objectives, two specialized corpora were compiled, one for red wine (RC) and one for white wine (WC). All the wine tasting notes in these corpora were from the official website of the Liquor Control Board of Ontario (LCBO) in Canada. The LCBO is an Ontario government enterprise and is one of the world's largest buyers and retailers of alcoholic beverages. The LCBO website provides customers with complete information on wine products and services. The website is available in both English and French. The product search engine provides tasting notes, limited-time offers, daily store inventories, and store locations.

Unlike other research on wine criticism, this study did not select texts from magazines or newspapers as resources because those types of texts tend to be long and tend to cause comprehension difficulties for novice learners. The tasting notes on the LCBO website are short texts ranging from 30 to 130 words that are devoted to describing and evaluating wines. Through the search engine on the website, red wine and white wine tasting notes are easy to select. The following is a typical tasting note from the LCBO website:

(1) Earthy and fruity with plum, dark cherry and woodsmoke on the nose. The palate presents generous dark fruit accented by earth and smoke. Big fleshy tannins, but also a good backbone of acidity providing balance and structure. A boisterous and flavourful wine that will mellow impressively with another 2-4 years. Would make a nice addition to a meal featuring herbed rare steaks or game.

For the purpose of this study, to select appropriate texts for the corpora, keywords including product type, wine regions, and grape varieties were used while searching the LCBO websites. Only texts containing 50-90 words were selected. If a text is short—containing fewer than 50 words—then it likely cannot present sufficient information on a typical tasting note. If the text is longer than 91 words, then it is likely to be too difficult for novice readers. In addition, only texts that describe dry to off-dry wines were selected because sweet wine is in another wine category and its descriptors differ. From the texts, reviews that evaluated wines made from both old world (wine-producing countries in Europe) and new world wine regions (wine producing countries outside of Europe) were randomly selected. In addition, another wine instructor who has taught wine and spirits courses in the Foods and Beverages Management Department for twelve years at this university was invited as a consultant to participate in designing the corpora and in screening the analysis results.

Table 1 shows the distributions of the selected tasting notes according to where the wines were made. To compare the linguistic differences, the WC and RC were also combined as a Total Corpus (TC), which was used as a reference. As Table 1 shows, the TC contained 1,866 wine tasting notes (WC=933 and RC=933) and 124 665 words. The RC contained 65 421 words, and the WC contained 59 244 words. The average length of the red wine tasting notes was 71 words, whereas the average length for the white wine notes was 64 words.

Table 1. Number of tasting notes and reviewed wine production countries in the corpora

Regions	Countries	# of texts in RC	# of texts in WC
New World	USA	63	63
	Australia	63	63
	New Zealand	63	63
	Chile	63	63
	South Africa	63	63
	Argentina	63	63
	Canada	63	63
Old World	France	125	125
	Italy	125	125
	Spain	125	125

Germany	65	65	
Portugal	52	52	
# of Tasting Notes	933	933	
# of Words	65 421	59 244	

3.2 Selected Linguistic Features for Analyses

To answer the research questions, vocabulary compositions including the GSL, the AWL, specialized vocabulary, and keywords were analyzed. In addition, some linguistic features, including the personal pronouns, verb tenses, and verb forms presented in the corpora, were taught to the students explicitly or were assigned as homework for their own practice.

3.3 Analytical Tools

Lexical Tutor and *AntConc* were the tools used for the analyses in the corpus study. These two application programs are available free on the Internet. *Lexical Tutor* can be helpful in analyzing the word compositions in a corpus by categorizing words into GSL, AWL, and off-list words. In addition, this program can determine the defining lexis as keywords in a specialized corpus by comparing with a reference corpus. *AntConc* was used for the word frequency and concordance analyses.

To evaluate the program's effectiveness and to monitor students' learning progress, pre- and posttests on vocabulary and reading comprehension of wine discourse were implemented. The tests were designed by the researcher and the subject expert. The pretest was implemented during the second week of the classes, and the same test was delivered in the 17th week of the classes as a posttest. The results were checked and scored by the researcher. *SPSS 14.0* was used for the ANOVA analysis for the pre- and posttests.

4. Results

4.1 Vocabulary Profiles

Based on the *Lexical Tutor* analysis, the vocabulary profiles of the corpora are presented in Table 2. Overall, in the corpora, the K1 and K2 (GSL) words constituted 70% of the texts. Approximately 4% of the words appeared on the AWL, and more than 25% of the vocabulary consisted of off-list words. The red wine corpus appeared to contain more academic and off-list words than did the white wine corpus. Compared with other studies (Coxhead & Nation, 2001; Sutarsyah et al., 1994; Ward, 1999) of university academic textbooks, the specialized corpora contained lower percentages of GSL and AWL words but higher percentages of off-list words. In other words, off-list words appeared relatively frequently in these texts, which could cause difficulties for the learning processes of nonnative speakers of English learning processes (Chung & Nation, 2003). The lexical density was higher in the RC than in the WC. According to Stubbs (2001), the typical level of lexical density in written texts is approximately 40%, ranging from 36% to 57%. The lexical density values for the RC (67%) and the WC (65%) were much higher than 57%, indicating that these wine tasting notes contained greater percentages of content words in the texts and thus that the texts could be understood only by members of the wine community. This type of text was likely challenging for novice first-year university students.

Table 2. Vocabulary profiles of the RC, WC, and TC

	RC		WC		TC	
	Tokens	%	Tokens	%	Tokens	%
K1 Words (1–1000)	38 389	58.68	35 114	59.27	73 502	58.96
Function Words	(21 406)	(32.72)	(20 854)	(35.20)	(42 261)	(33.90)
Content Words	(16 983)	(25.96)	(14 260)	(24.07)	(31 241)	(25.06)
K2 Words (1001–2000)	6882	10.52	6872	11.60	13 751	11.03
K1 + K2 Words		(69.20)		(70.87)		(69.99)
AWL Words (Academic)	2820	4.31	2079	3.51	4899	3.93
Off-List Words	17 330	26.49	15 179	25.62	32 513	26.08
Total	65 421	100	59 244	100	124 665	100
Lexical Density		67		65		66

Although there were low percentages of AWL words in both of the corpora, these types of words are important in nonnative speakers' academic learning. As Kaewphanngam (2002) defined, word families in a specialized corpus that belong to an AWL category can be identified as subspecialized vocabulary. Based on the *Lexical Tutor* data, Table 3 presents the 20 most frequent AWL words in the RC and the WC. The table shows that words in this category tended to appear across different disciplines and did not reflect the nature of the specific discourse. Among the listed words, 10 words appeared in both the RC and the WC: *medium, complex, core, layer, intensity, feature, style, concentration, display,* and *classical*. The table also shows that reviewers describing red wine and white wine used different words from the AWL to write their tasting notes.

Table 3. The 20 most frequent AWL words in the RC and the WC

Rank	RC	WC	Rank	RC	WC
1	medium	medium	11	display	persistent
2	structure	intensity	12	plus	considerable
3	complex	style	13	evolve	create
4	core	classical	14	focus	display
5	layer	complex	15	approach	emphasis
6	intensity	feature	16	classical	layer
7	feature	concentration	17	estate	precise
8	potential	core	18	mature	restraint
9	style	notion	19	reveal	tension
10	concentration	energy	20	benefit	anticipate

In a previous study, Kaewphanngam (2002) identified specialized vocabulary as word families that do not belong to function words, the GSL, or the AWL. From the *AntConc* analysis, Table 4 lists the 20 most frequent specialized vocabulary words from the RC and the WC. After the GSL, AWL, and function words were deleted from the corpora, the specialized vocabulary constituted approximately 24.6%. The subject expert was consulted, and he agreed that the lists could have reflected the nature of the language used in wine discourse. The most frequent specialized vocabulary in these two corpora differed. Most of the words were related to the nouns that are used to describe the aromas and flavors of the wines. On both lists, among the top 5 words, *palate* and *aroma* could be identified as HF specialized words in the wine community. In describing the sensation of wine in the mouth, the level of *tannin* is an important modifier for red wine, and *acidity* is an important criterion for evaluating the sense on the palate for white wine. A study of the 20 most frequent specialized vocabulary words found that words related to berry fruits and to burnt and wooden notes were frequently used in the red wine notes. In the white wine tasting notes, citrus and stone fruits were widely referenced.

Table 4. The 20 most frequent specialized vocabulary in the RC and the WC

Rank	RC	WC	Rank	RC	WC
1	cherry	aromas	11	raspberry	floral
2	aromas	acidity	12	ruby	lemon
3	palate	palate	13	blackberry	lime
4	tannins	citrus	14	vanilla	Chardonnay
5	vintage	crisp	15	cellar	melon
6	bodied	peach	16	blueberry	spice
7	plum	body	17	Pinot	grapefruit
8	acidity	score	18	cedar	texture
9	spice	oak	19	smoky	creamy
10	oak	pear	20	elegant	apricot

In the study of specialized corpora, another important feature for analysis is the keyword. According to Scott (1997, p. 237), a keyword is "a word which occurs with unusual frequency in a given text". To calculate the unusual frequency of words in a specialized corpus, a large general corpus is typically used as a reference for contrast. In the present study, the Spoken British National Corpus, a general corpus, was used to perform the analysis. From the *Lexical Tutor* analysis, Table 5 shows the 20 most frequent keywords from the RC and the WC. The list shows that most of the words were related to the descriptions of the smell and taste of the wines. The wine reviewers primarily used fruits and vegetables to describe the perceptual experience, but the types of fruits and vegetables to describe the white and red wines were quite different. In addition, providing suggestions for food pairing is an important aspect of wine tasting notes. Therefore, certain words were related to cooking methods and food terms. For example, on the RC keyword list, *braise* and *tenderloin* were ranked 8th and 12th, whereas *sear, sushi,* and *seafood* were ranked 8th, 10th, and 16th, respectively.

Table 5. The 20 most frequent keywords from the RC and the WC

Rank	RC	WC	Rank	RC	WC
1	tannins	citrus	11	hickory	herbaceous
2	palate	palate	12	tenderloin	guava
3	blueberry	aromas	13	mocha	nectarine
4	aromas	racy	14	garnet	floral
5	medallion	pith	15	allure	gooseberry
6	cassis	asparagus	16	blackberry	seafood
7	espresso	elegant	17	weight	vibrant
8	braise	sear	18	peppercorn	fragrant
9	anise	melon	19	clove	ferment
10	elegant	sushi	20	opaque	zest

4.2 Students' Learning Progress

In the first lecture of the class, the LCBO website and the *AntConc* software were introduced to the students, who were taught how to search for wine information and tasting notes on the LCBO website. The students were also trained on how to use the functions of the software, including the concordance and word lists, using the in-house wine corpora. Figure 1 illustrates an example of the concordance lines of *palate*.

Figure 1. Screen shot of a concordance line in *AntConc*

First, the students were told to pay attention to the verb tenses used in the concordance lines. For example, in Figure 1, all the verbs after *palate* are in present tense. Second, the students were advised to be aware of the prepositions used in the concordance lines. As Figure 1 shows, *on* is the preposition used before *the palate*. These analytical skills were explained thoroughly and repeatedly to the students. A variety of activities and tasks integrated with using the LCBO website, the two in-house corpora, and *AntConc* were assigned as homework for the following 14 weeks. These activities included checking the definitions of the vocabulary words and identifying the typical use and meaning patterns of the vocabulary words, phrases, and grammatical points found in the RC and WC. Some practices were discussed and taught explicitly in the class, and some were designed as previewing or reviewing activities.

In the second lecture, lists of the 50 most frequent AWL and 100 most frequent specialized vocabulary and keywords generated by the researcher from the two corpora were given to the students. In addition, vocabulary and reading comprehension tests were administered. On the tests, 30 multiple-choice questions were related to specialized vocabulary and keywords, and 10 multiple-choice questions were related to the AWL. Two tasting notes were selected from the LCBO website and used as reading comprehension passages. The students had to answer 10 multiple-choice reading comprehension questions based on these two passages. The total score for the test was 100 points. The test questions had been reviewed by the subject expert, and the tests were scored by the researcher. During the 17th week, the same vocabulary and reading comprehension tests were administered again to investigate the students' progress.

The pre- and posttest ANOVA results clearly show that the students' performance on the vocabulary and reading comprehension tests progressed. The average score on the pretest was 59.12, and the average score on the posttest was 81.91. The students scored significantly higher for all three test items on the posttest compared with the pretest. This result indicates that the students gained both vocabulary and content knowledge of the wine community during the 18 weeks of course instruction.

Table 6. The ANOVA results from the pre- and posttests

Test Items	Pretest		Posttest		F
	Mean	S. D.	Mean	S. D	
Specialized Vocabulary & Keywords	35.91	5.33	51.64	3.63	130.91***
AWL	12.27	2.17	14.45	1.67	8.531*
Reading Comprehension	11.82	2.89	14.82	2.02	15.96***
Total Score	59.12	8.90	81.91	5.72	93.82***

5. Conclusions

This study demonstrates how to integrate specialized corpora into a CLIL course for nonnative speakers of English at the university level. To satisfy the teaching objectives, two in-house corpora were compiled and used as supplementary materials for the students' practice and homework. The vocabulary analyses demonstrated that more specialized corpora tend to have a higher frequency of content words. This observation was also reflected in the high lexical density (greater than 66%) and the strong presence of specialized vocabulary (more than 24%). Although the passages that were chosen for the wine corpora were fairly short, roughly 50 to 90 words per passage, compared with Rizzo's (2010) study of the telecommunication engineering corpus (TEC), the lexical study in the present study was much higher than that of the TEC. In addition, as Chung and Nation (2003) observed, when specialized vocabulary words constitute as much as 30% of a text, it is difficult for students to acquire these words. In addition, GSL words are HF words that constitute up to 80% of running texts (Nation, 2001), but the wine corpora analyses showed that GSL words actually constituted approximately 70% of the wine tasting notes. Compared with normal general English texts, wine tasting notes appear to be more challenging for nonnative speakers.

As Sinclair (1991) noted, analysis of each specialized corpus demonstrates the individual characteristics because of the unique objectives and emphases of each corpus. Because providing only the core vocabulary lists as supplementary material to students is not sufficient, the reading skills of learners should be emphasized. Therefore, in addition to the word lists and the corpora, the present study also integrated the LCBO website as part of the supplementary materials. Through the exercises and homework, the students should have become familiar with the language used in the specific community. From the analyses of vocabulary compositions, the

wine tasting notes in the corpora could have been difficult for the students; however, based on the pre- and posttest analyses, student performance in vocabulary and reading comprehension improved significantly.

In this study, the teachers played two roles: one role as the facilitator expected to know about the availability of useful and authentic resources to fit the students' needs and the teaching objectives and another role as a guide to assist students in identifying the language patterns and behaviors of words and phrases on their own as language researchers. In addition, teachers should focus on teaching the specialized vocabulary and expressions that students will use or be expected to use in their subject areas. This can be done by helping learners gain the more general skills of recognizing specialized words, interpreting definitions, relating senses to a core meaning, and learning word parts (Chung & Nation, 2003). Teachers can provide learners with the tools for dealing with specialized vocabulary. This strategy can help teachers not to get involved in trying to teach in a specialized area, but can direct students' attention to vocabulary strategies.

For other academic disciplines, ESP teachers and material developers are encouraged to use the results from the specialized corpus analysis as inputs for course content design. Many interactive exercises can be developed based on these corpus analysis results, including creating gap-filling exercises for students to explore specific vocabulary, such as adjectives, verb tenses, and prepositions. Other exercises may encourage students to explore a particular discourse by identifying the grammatical features and usage patterns of the vocabulary and key phrases in a specific corpus. When designing these types of activities, both ESP teachers and learners should have proper training in accessing the corpora and using the analytical tools. If the ESP teacher is not an expert in the subject area, then a subject expert should be involved in the corpus design and material development processes.

This study emphasizes the importance of helping learners increase their awareness of how language features are used in their subject areas. The results of the present study aim to empower teachers and students to use authentic materials as tools to understand and explore their own discourse communities. For future research, it would be useful to investigate the same linguistic features in the same field but using different text resources, such as wine review articles in newspapers or magazines. By comparing the research results, researchers can obtain a broader understanding of wine language.

Acknowledgements

References

Aston, G. (2001). Learning with corpora: An overview. In G. Aston (Ed.), *Learning with corpora* (pp. 1-45). Houston, TX: Athelstan.

Biber, D., Conrad, S., & Repper, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511804489

Brinton, D., Snow, M., & Wesche, M. (2003). *Content-Based Second Language Instruction*. Ann Arbor: The University of Michigan Press.

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialized texts. *Reading in a Foreign Language*, *15*, 103-116. Retrieved from http://nflrc.hawaii.edu/rfl/October2003/chung/chung.html

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, *32*, 251-263. http://dx.doi.org/10.1016/j.system.2003.11.008

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, *12*(4), 345-360. http://dx.doi.org/10.1076/call.12.4.345.5699

Connor, U., Davis, K., & DeRycker, T. (1995). Correctness and clarity in applying for overseas jobs: A cross-cultural analysis of U.S. and Flemish applications. *Text*, *15*, 457-475. http://dx.doi.org/10.1515/text.1.1995.15.4.457

Connor, U., & Upton, T. A. (Eds.). (2004). *Discourse in the Professions. Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl.16

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213-238. http://dx.doi.org/10.2307/3587951

Coxhead, A., & Nation, P. (2001). The specialized vocabulary of English for Academic purposes. In J. Flowerdew, & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252-267).

Cambridge: CUP. http://dx.doi.org/10.1017/CBO9781139524766.020

Dalton-Puffer, C. (2011). Content and language integrated learning: From practice to principles. *Annual Review of Applied Linguistics*, *31*, 182-204. http://dx.doi.org/10.1017/S0267190511000092

Dodd, B. (1997). Exploiting a corpus of written German for advanced language learning. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 131-145). New York, NY: Longman.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, *21*(2), 231-244. http://dx.doi.org/10.1016/0346-251X(93)90044-H

Flowerdew, J. (2001). Concordancing as a tool in course design. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. 71-93). Amsterdam: John Benjamins Publishing Company.

Folse, K. (2010). Is explicit vocabulary focus the reading teacher's job? *Reading in a Foreign Language*, *22*, 139-160. Retrieved from http://nflrc.hawaii.edu/rfl/April2010/articles/folse.pdf

Fraser, S. (2004). The lexical characteristics of technical texts. In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.), *JALT2004 conference proceedings* (pp. 318-327). Tokyo, Japan: Japan Association for Language Teaching.

Hou, H. (2012). A Corpus approach to discourse analysis of newspaper restaurant reviews: A preliminary analysis. *Studies in Literature and Language*, *5*(3), 95-100. http://dx.doi.org/10.3968/j.sll.1923156320120503.1422

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, *19*(1), 24-44. http://dx.doi.org/10.1093/applin/19.1.24

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, *41*, 235-253.

Johns, T. (2002). *Data-driven learning: The perpetual challenge*. Paper presented at The Teaching and learning by doing corpus analysis, Graz, 19-24 July, 2000.

Kaewphanngam, C. (2002). *A corpus analysis of psychology texts as a basis for the development of teaching materials in English for academic purposes*. Bangkok: Mahidol University.

Meyer, C. (2002). *English corpus linguistics: An introduction*. Cambridge: CUP. http://dx.doi.org/10.1017/CBO9780511606311

Murphy, B. (1996). Computer, corpora and vocabulary study. *Language Learning Journal*, *14*, 53-57. http://dx.doi.org/10.1080/09571739685200391

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: CUP. http://dx.doi.org/10.1017/CBO9781139524759

Nation, P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, Heinle.

Rizzo, C. R. (2010). Getting on with corpus compilation: From theory to practice. *ESP World Issue*, *9*(1), 1-23.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, *31*, 205-225. http://dx.doi.org/10.1017/S0267190511000055

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: CUP.

Scott, M. (1997). PC analysis of key words—and key key words. *System*, *25*(2), 233-245. http://dx.doi.org/10.1016/S0346-251X(97)00011-0

Sinclair, J. (1991). *Corpus, concordance, collocation*. UK: Oxford University Press.

Stubbs, M. (2001). *Word and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Sun, Y. C. (2003). Extensive reading online: An overview and evaluation. *Journal of Computer Assisted Learning*, *19*(4), 438-446. http://dx.doi.org/10.1046/j.0266-4909.2003.00048.x

Sun, Y. C., & Wang, L. Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, *16*(1), 83-94. http://dx.doi.org/10.1076/call.16.1.83.15528

Sun, Y. C. (2007). Learner perceptions of a concordancing tool for academic writing. *Computer Assisted Language Learning*, *20*(4), 323-343. http://dx.doi.org/10.1080/09588220701745791

Sutarsyah, C, Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based study. *RELC Journal*, *25*(2), 34-50. http://dx.doi.org/10.1177/003368829402500203

Teubert, W., & Cermakova, A. (2007). *Corpus linguistics: A short introduction*. NY: Continuum.

Tribble, C. (1990). Concordancing and an EAP writing program. *CAELL Journal*, *1*(2), 10-15.

Tribble, C. (2001). Small corpora and teaching writing. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 381-408). Amsterdam: John Benjamins.

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, *27*, 442-458. http://dx.doi.org/10.1016/j.esp.2008.05.003

Ward, J. (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language*, *2*, 309-323.

Waring, R., & Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English Speaking World*, *4*, 97-110.

West, M. (1953). *A general service list of English words*. London: Longman.

Wichmann, A. (1995). Using concordances for the teaching of modern languages in higher education. *Language Learning Journal*, *11*, 61-63. http://dx.doi.org/10.1080/09571739585200201

Woolard, G. (2000). Collocation-encouraging learner independence. In M. Lewis (Ed.), *Teaching collocation: Further development in the lexical approach* (pp. 28-46). Hove, England: Language Teaching Publications.