Non-English Major EFL Learners' Lexical Collocation Errors in a Chinese Context

Quping Hou and Issra Pramoolsook*

School of Foreign Languages, Suranaree University of Technology, THAILAND

Abstract

One of the difficulties a second language learner frequently experiences in writing is the choice of words to achieve native-like proficiency. This investigation reports a study of Chinese EFL learners' use of lexical collocations. The objective was to examine the problems of the students' use of lexical collocations, the results of which aim to help the researchers to decide how to utilize Corpus of lexical Contemporary American English (COCA) to raise their collocation awareness in a subsequent study. Two writing tasks were administered to 50 Non-English Majors (NEMs) in Kaili University (KU), resulting in a corpus of 100 essays. Lexical collocation errors in the texts were identified by using COCA as a reference corpus. The findings showed that the most frequent collocation errors were collocations with verbs as nodes and the second most frequent ones were collocations with adjectives as nodes. Misuses of quantifiers were also found in the corpus. Some suggestions were made by the researchers aiming at improving writing instruction through collocation awareness raising.

Keywords: Chinese learners, Non-English majors, lexical collocation, collocation errors, COCA, error analysis

บทคัดย่อ

เรื่องยากประการหนึ่งที่ผู้เรียนภาษาต่างประเทศประสบได้บ่อยในการเขียนคือการเลือกคำมาใช้ให้เหมาะสม เหมือนเจ้าของภาษา งานวิจัยนี้รายงานการศึกษาการใช้การปรากฏร่วมจำเพาะของคำศัพท์ของผู้เรียนภาษาอังกฤษใน ฐานะภาษาต่างประเทศชาวจีน วัตถุประสงค์ของการวิจัยนี้คือเพื่อศึกษาปัญหาของผู้เรียนดังกล่าวในการใช้การปรากฏ ร่วมจำเพาะของคำศัพท์ ซึ่งผลที่ได้รับจะช่วยนักวิจัยในการตัดสินใจว่าจะใช้ Corpus of Contemporary American English (COCA) อย่างไร เพื่อเพิ่มพูนความตระหนักในเรื่องการปรากฏร่วมจำเพาะของคำศัพท์ในงานวิจัยต่อยอดชิ้นถัดไป

ในการวิจัยนี้นักวิจัยได้ขอให้นักศึกษาที่ไม่ได้เอกวิชาภาษาอังกฤษ จำนวน 50 คน ที่มหาวิทยาลัยไค่ลี่ ประเทศ สาธารณรัฐประชาชนจีน เขียนชิ้นงานคนละ 2 ชิ้น ซึ่งทำให้ได้ชุดข้อมูลที่ประกอบไปด้วยความเรียง 100 ชิ้น จากนั้น นักวิจัยจึงทำการหาและระบุข้อผิดพลาดการปรากฏร่วมจำเพาะของคำศัพท์โดยใช้ COCA เป็นชุดข้อมูลอ้างอิง

ผลการวิจัยพบว่าข้อผิดพลาดการปรากฏร่วมจำเพาะที่พบบ่อยที่สุดคือ การปรากฏร่วมจำเพาะที่มีคำว่ากริยา เป็นแกนคำ และที่รองลงมาคือ การปรากฏร่วมจำเพาะที่มีคำคุณศัพท์เป็นแกนคำ นอกจากนี้ ยังพบว่านักศึกษายังใช้คำ เพื่อบอกปริมาณผิดพลาดอีกด้วย

ในตอนท้ายนักวิจัยได้ให้ข้อเสนอแนะที่มุ่งหวังที่จะพัฒนาการเรียนการสอนการเขียนภาษาอังกฤษโดยอาศัย การเพิ่มพูนความตระหนักในเรื่องการปรากฏร่วมจำเพาะของคำศัพท์

---

* *Corresponding author*
  E-mail address: issra@sut.ac.th

Introduction

English study in China is inspired by not only the desire to study abroad, but also a need to improve skills and find good jobs. Many universities in China offer English majors (EMs) and non-English majors (NEMs). Compared with EMs, NEMs' English competence is at a relatively lower level. College English is an important and compulsory course for students of NEMs in China. The objective of College English is to develop the students' ability to use English in a well-rounded way so that in their future studies and careers as well as social interactions, they will be able to communicate effectively in both oral and written forms.

Kaili University (KU) is a local public university located in the southeast of Guizhou Province, China. It is the largest higher education institution in Qiandongnan Miao and Dong Autonomous Prefecture (QMDAP) with a student population of over 10,000. The main goal of KU is to cultivate teachers for schools in QMDAP. Most of the NEMs in KU cannot achieve the required level of college English education. Zheng (2000) stated that many college students with low reading speed cannot understand the contents of what they have read. There exist some common characteristics in English study, such as starting English learning late, L1 (national languages) influences, and lack of learning strategies, which impede the improvement of the ethnic undergraduates' listening and speaking (Wang, 2010). Among the four basic English skills, writing is the weakest for them. Many problems exist in students' compositions, and the learners usually do not know how to choose words to express themselves clearly (Wu, 2003).

English writing has always been an essential issue in English teaching. Several reports found that writing is the most difficult skill for Chinese learners to master. For example, a lot of college students in China lost scores in the writing part of the national English language test (CET Band-4), which negatively affects the pass rate of the test takers (Zhang, 2009). Also, college students in China pay more attention to grammar in writing but ignore how to express their meaning naturally (Sun, 2011). Actually, collocation is one of the factors responsible for Chinese EFL learners' inadequate writing competence (Meng & Li, 2005). For NEMs in KU, it is hard to achieve native-like written communication even though they have learnt thousands of words by heart. They often use unnatural English expressions that have the right word items but improper collocations (Wu, 2003). In China, though research

literature on the teaching and learning of vocabulary is very extensive, studies in collocation are still in need. The incorrect and inappropriate use of collocation in context is one of the main obstacles for learners to achieve native-like proficiency.

The College English Test (CET) is a national English language test for NEMs university students in China and is aimed at precisely measuring their comprehensive command of English. CET Band-4 and Band-6 are designed for the students who have completed the corresponding English courses which belong to basic and intermediate stages of college English learning (NCETC, n.d.). Almost every student in KU has the experience of taking CET but few of them can pass. The participants in the present study were intermediate level NEMs in KU who had scored high in a CET Band-4 but failed in CET Band-6. They had received grammatical training when they took CET Band-4. For these NEMs, one of the obstacles to learn English is the lack of competence in using collocations. Many sentences made by them, in most cases, are grammatically correct but do not make sense or sound natural. This fact led the present study to focus only on lexical collocations i.e. the predictable ways in which a noun, verb, adjective or adverb is combined with a word from another word class. Through identifying the errors, the researchers focused on a selection of NEMs' incorrect collocations in their production of language and attempted to find a way to help NEMs in KU to improve their English writing especially with regard to the use of English lexical collocations. Therefore, this study was conducted to examine the problems of students' use of collocations, the results of which will serve as a guideline for the researchers to utilize Corpus of Contemporary American English (COCA) to raise students' lexical collocation awareness in a subsequent study, and also will help the researchers to decide which types of collocations should be paid more attention to. The findings are thus of great pedagogical significance.

Methodology

Data and Data Collection

Fifty NEMs from KU participated in the study. They were intermediate students from eight different non-English majors. which was conducted in May, 2011.The participants were required to write two essays on the following topics: 1) Reduce

Waste on Campus and 2) How I Finance My College Education. The reason why these two topics were chosen was that the topics are related to the students' concerns and their college life, and they were expected to be able to complete the writing tasks with no trouble generating content and producing information. In order to get enough written data for the analysis, the task-takers were asked to write at least 150 words for each essay, which is similar to the length of the composition they are required to write in CET Band-6, each within one class period (50 minutes). To test their real performance, the students were required to write without help of dictionaries and other reference books when doing the tasks. The reason to set two compositions instead of one for this study was to reinforce the reliability of the data and to find out students' steady and actual performance in using lexical collocations.

Before the writing samples were collected, the demarcations between lexical collocations and non-lexical collocations were made. According to Davies (2008), "typically, [MI] scores of about 3.0 or above shows [sic] a 'semantic bonding' between the two words". MI score is a measure of strength of collocation. Church & Patrick (1990) noted that pairs with scores above 3.0 can probably be considered collocations and below that, free combinations. They give the detailed information of MI, as follows:

"MI compares the probability of observing the joint probability of the two words x and y together with the probabilities of observing x and y independently (chance).

$$I(x, y) = \log 2 \; \frac{P(x,y)}{P(x)\,P(y)}$$

If there is a genuine association between x and y, then the joint probability P(x, y) will be much larger than chance P(x) P(y), and consequently I(x, y) > 0." (Church & Patrick, 1990, p. 23).

The higher MI score the word combination has, the stronger association between the two words. For instance: *soft drinks* (MI=6.67) is stronger than *soft voice* (MI=3.97). In other words, *soft* to collocate with *drinks* is stronger than to collocate with *voice.*

It is meaningless for students to learn language items they would probably never use. For the sake of convenience, in the present study, only word combinations with MI scores higher than 3.0 and raw frequency (RF) higher than 3 in COCA are classified as collocations. In the process of data collection, collocation errors were identified, counted, and the percentage calculated in terms of Benson et al' s (1997) lexical collocation patterns (Table 1).

Table 1: Lexical Collocation Types

Types	Examples
L1 Verb + noun	*to cancel an appointment;* *to reject an appeal;*
L2 Adjective + noun	*strong tea;* *reckless abandon;*
L3 Noun + verb	*bombs explode;* *bees sting;*
L4 Quantifier + noun (group or units of thing)	*a swarm of bees;* *a pack of dogs;*
L5 Adverb + adjective	*sound asleep;* *closely acquainted;*
L6 Verb + adverb	*run rapidly;* *argue heatedly;*

After the students had completed the writing tasks, 100 essays (50 participants × 2 essays) were collected as a learner corpus which included 19,140 words and the average length of the writing was 191 words per essay. All lexical collocations found in the learners' corpus, the predictable ways in which a noun, verb, adjective or adverb is combined with a word from another word class, were identified and underlined based on Benson et al' s (1997) lexical collocation types (Table 1). These collocations were then compared with COCA for detecting and analyzing collocation errors. As mentioned earlier, in the present study the collocation is limited to only lexical collocations, so grammatical collocations, i.e. word combinations consisting of the main word (a noun, an adjective, or a verb) and a preposition, or 'to+ infinitive', or '*that*-clause', e.g. *She was* <u>afraid that</u> *she would fail the examination,*

idioms e.g. *a wolf in sheep's clothing*, and free combinations, i.e. word combinations with MI scores lower than 3.0, e.g. *find work* (MI = 2.39), that appeared in the writing samples were excluded in this research.

Data Analysis

COCA is the largest and most recent corpus of English available for free, and the only large and balanced corpus of American English. Collocation errors in the current study referred to all the collocations which deviate from the norms of the target language. The present study only focused on lexical collocations. Collocations that cannot be found in COCA when MI is at 3.0 with RF at least 3 were considered collocation errors. Corder (1981) proposed the procedure for error analysis which includes three stages: 1. Data collection: Recognition of idiosyncracy; 2. Description: Accounting for idiosyncratic dialect; and 3. Explanation (the ultimate object of error analysis). Based on Corder's (1981) procedure, Chow (2006) observed Chinese students' grammatical errors in their writing and employed five steps to analyze these errors in her study: (1) identify errors; (2) categorize errors; (3) analyze errors; (4) evaluate errors; and (5) correct errors.

In the present study, the researchers intended to analyze lexical collocation errors in Chinese students' writing. Therefore, based on Corder's (1981) procedure and taking Chow's (2006) steps into consideration, for using COCA to analyze learners' collocation errors, the present study took the following steps.

Step1. Collecting a sample of learner's language and identifying errors

This stage is a process of collecting a sample of learner's language and then identifying errors. As mentioned in the previous section, 100 essays were collected as a learner corpus. All the collocations were detected and classified by two raters, respectively. One rater was an experienced English teacher from the Foreign Language Institute in KU who had taught NEMs English for more than ten years. The other was an English language teacher and is currently studying for an MA in English Language Studies. Both were trained to use COCA to detect collocation errors and label them in students' compositions with error tags. At the beginning, the two raters read the essays twice and tried to understand the messages the subjects wish to express. Then, they underlined all of the word combinations which could be found in

the participants' compositions following the working definition of lexical collocation in the present study and Benson et al's (1997) lexical collocation types (Table 1). In this way, grammatical collocations were excluded in the present study.

Then, error identification began. In the process of detecting participants' collocation errors, COCA was used as a reference to analyze errors and provide suggestions for correction. Through utilizing COCA, it was easier for the raters to extract examples of common authentic usages from the corpus. One important thing that the raters bore in mind was that *free combinations*, such as *close window (MI=-2.48)*, *clean face (MI=0.87)*, and *dump food (MI =-0.14)*, should be filtered out. Once errors were identified, the raters labeled them with error tags (Table 2) based on the lexical collocation types:

Table 2: Lexical Collocation Error Domains and Categories

Error Domains		Error Categories		Examples
L1	Verb+noun	<L1>	Verb+noun collocation errors	fit the demand make advantage
L2	Adjective+noun	<L2>	Adjective+noun collocation errors	an attracting looking smooth English
L3	Noun+verb	<L3>	Noun+verb collocation errors	lions chime clock roar
L4	Quantifier+noun	<L4>	Quantifier+noun collocation errors	a line of hope a cloud of wind
L5	Adverb+adjective	<L5>	Adverb+adjective collocation errors	definitely value strongly polluted
L6	Verb+adverb	<L6>	Verb+adverb collocation errors	imminently required manage reasonably

The following illustrates how the raters detected the participants' collocation errors in their writing samples and provided suggestions for correction by utilizing COCA.

1. The raters found a suspicious L1 (Verb+noun) collocation error *"They think as a student, the most important duty is to **learn more knowledge** from the books."*,

in which the verb *learn* collocates with the noun *knowledge* in an unusual way.

2. The raters then searched COCA (MI=3.0) with the query command *learn knowledge*, but no solution was yielded. Since the components of free combinations are substitutable and their MIs are below 3.0 in COCA, to make sure whether it is a free combination, the raters queried COCA (disregarding the limitation of MI) with the query command again, but still no solution was displayed. So, they could decide it is a collocation error.

3. Finally, the raters searched in COCA with the word query command *knowledge* that collocates with verb *[V*]* to find appropriate verbs which co-occur with the word *knowledge*. The appropriate words are *acquire* and *gain*. Examples are shown in Table 3.

Table 3: Examples extracted from COCA

---

*1. By this approach, student teachers are encouraged to **acquire** professional **knowledge** and skills by making personal efforts.*

*2. He can not **acquire** the full **knowledge** which would make mastery of the events possible.*

*3. She stated that constant exposure to the media was a way to **gain knowledge** and a sense of control about the war.*

*4. Do children with visual impairments **gain** scientifically accurate **knowledge** using inquiry-based approaches?*

Step 2. Comparing and classifying errors

As soon as errors were identified, the next step was to describe them properly. The description of learners' errors involved two aspects: comparison and classification. As discussed in Step 1, collocation errors can be found through utilizing COCA. In this Step, the raters checked the errors by comparing the learners' errors with the reconstruction of the sentences in the target language (COCA) and then tagged the errors. Take this sentence as an example: *"They think as a student, the most important duty is to **learn** more **knowledge** from the books." "Learn knowledge"* here was labeled as *"<L1>"* because it is a Verb + noun lexical collocation error. Finally, errors can be classified in terms of Benson's (1997) lexical collocation patterns, counted and calculated in terms of percentage.

Step 3. Explaining and evaluating errors

This stage dealt with explaining errors and evaluating errors, which was very important because some errors could reflect learners' attempts to perform the task. As illustrated in the example above, in Step 2 the raters tagged all the collocation errors in the learner corpus. In this Step, these errors were reported according to their occurrences in students' writing and calculated in terms of percentage. The number and percentage of collocation errors can be used to discuss which types of collocations should be focused on when teaching at the subjects' intermediate level.

Results & Discussion

As discussed above, the present study using COCA to analyze learners' collocation errors followed three steps. In Step 1, the two raters underlined all the word combinations in the learner corpus separately. Then, both queried all the underlined combinations in COCA and preliminarily judged which word combinations were collocation errors. In Step 2, the researchers compared collocation errors from the learner corpus with the reconstructed sentences in COCA to check whether the two raters' judgment and classification of errors were correct and whether they tagged errors in learner corpus without making mistakes. At the end, they found 214 lexical collocations in the subjects' writing, which included 106 error-free collocations and 108 incorrect collocations. The findings are shown in Table 4.

Table 4: Error-free collocations and incorrect collocations in subjects' writing

Type	Total number	Error-free collocation Example:	Number (%)	Incorrect collocation Example:	Number (%)
L1—Verb+noun	107	*solve problem; attend college;*	52 (48.6%)	*learn knowledge; fit the demand;*	55 (51.4%)
L2—Adjective +noun	64	*negative effects; serious problem;*	35 (55%)	*smooth English; social sign;*	29 (45%)
L3—Noun+ verb	0	…		…	0

Table 4:   Error-free collocations and incorrect collocations in subjects' writing (cont.)

L4—Quantifier + noun	5	*A pair of jeans*	1 (20%)	*a line of hope; a cloud of wind;*	4 (80%)
L5—Adverb +adjective	12	*extremely high; highly educated;*	8 (66%)	*definitely value; intensely fresh;*	4 (33%)
L6—Verb +adverb	26	*affect deeply; reading widely;*	10 (38%)	*manage reasonably; arrange properly;*	16 (62%)
In Total	214 (100%)		106 (49.5%)		108 (50.5%)

     Some collocations repeatedly appeared in the learner corpus, such as: *learn knowledge, daily life, and add parents' burden.* In this case, they were only counted once. Some collocations such as *practice their thrifty conscious* (raise their thrifty consciousness) were counted twice, and *practice conscious* and *thrifty conscious* were considered two errors.

     As shown in Table 4, there was no L3 Noun+ verb lexical collocation that the subjects produced in this learner corpus. This is probably because L3 lexical collocations, such as *lion roar, clock chime,* and *bell ring,* were rare for them to use in writing. Errors amounted to a large percentage in L4 Quantifier + noun lexical collocations, but it does not mean that this type of collocations is the most difficult one. Actually, the compositions indicated that participants seldom used this L4 pattern of collocations. The participants employed only five L4 lexical collocations in their writings. The most frequent errors were related to the use of L1 Verb+noun collocations. The participants made 55 errors in this type of lexical collocations. Figure 1 below displays the distribution of lexical collocation error domains found in the present learner corpus.
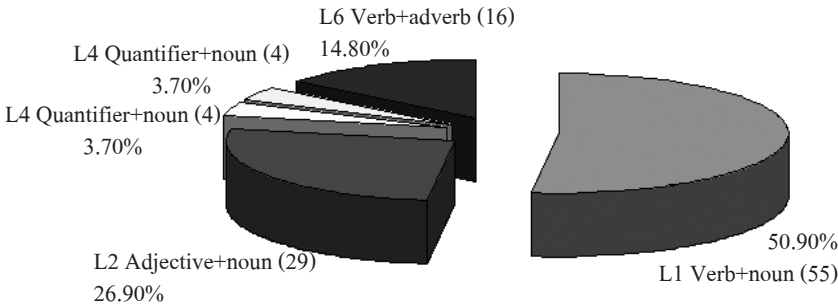
L6 Verb+adverb (16) 14.80%

L4 Quantifier+noun (4) 3.70%

L4 Quantifier+noun (4) 3.70%

50.90%
L1 Verb+noun (55)

L2 Adjective+noun (29) 26.90%

Figure 1   Distribution of Domains of Lexical Collocation Errors

Of all the collocation error types, the most frequent one was the collocation errors in L1 Verb+noun i.e. 55 (50.90%). On one hand, this is because L1 Verb + noun collocations are used most frequently in this learner corpus. On the other hand, a verb in a collocation has a restricted sense, which makes its correct use more difficult when learners cannot fully distinguish subtle differences among verb candidates, for instance, *tell jokes, tell a story* but not *say jokes, say a story.* Therefore, Chinese EFL learners have more trouble in choosing a proper verb in collocations (Chen, 2002). The second most frequent error type is L2 Adjective+noun collocations i.e. 29 (26.90%). This is probably because it was hard for students to distinguish and select appropriate adjectives to express their meanings.

Collocation errors were detected and classified by two raters, respectively. Cronbach alpha reliability test indicated that the inter-rater reliability (a=0.854) was quite significant. In order to know how frequently a certain pattern of collocation errors appeared in the learner corpus, the researchers listed the findings of lexical collocation error occurrences and percentage in Table 5.

Table 5: Breakdown of Lexical Collocation Error Domains

Tag	Number of Errors	Number of Occurrences (Token)	Percent (%)
L1	55	88	57.9
L2	29	37	24.3
L4	4	5	3.3
L5	4	5	3.3
L6	16	17	11.2
Total	108	152	100

Note: Error tags are shown in Table 2.

From Table 5, a total number of 88 tokens found representing 55 <L1> errors, which accounted for 57.9% of the total tokens. That is, 11 errors occurred more than once. There were 29 <L2> errors realized by 37 tokens in <L2> type, which accounted for 24.3%. That is, 7 errors were repeated. When comparing <L5> Adverb+ adjective with <L6> Verb+ adverb, participants made more errors in <L6> pattern. That is, 16 <L6> errors were realized by 17 tokens, which accounted for 11.2% and ranked the third most

frequent one in the learner corpus. <L5> and <L4> patterns of errors were the least frequent errors in the learner corpus with 5 tokens representing 4 errors each, which accounted for 3.3%, respectively.

All error types that occurred in the writings are discussed below in details:

Some students made errors because they translate Chinese into English literally according to the Chinese linguistic conventions. These expressions are understandable in Chinese but are not grammatically acceptable in English. Here are a few of such errors.

e.g. 符合要求 (fu he yao qiu) *fu he* means *fit*

*We can't **fit the demand** <L1> (to meet the demand) of finding a job.*

增加负担 (zeng jia fu dan) *zeng jia* means *add*

*I don't want to **add** my parents' **burden** <L1> (impose a burden) and make them work hard for me.*

一线希望 (yi xian xi wang) *yi xian* means *a line*

*I still have **a line of hope** <L4> (a glimmer of hope) I can earn money to pay the tuition fee by my own, but parents didn't think so.*

开灯 (kai deng) *kai* means *open*

*Many students **open the lights** <L1> (turn on the lights) even though they go out to play.*

学习知识 (xue xi zhi shi) *xue xi* means *study or learn*

*We should try our best to make full use of study time, instead of wasting is reading more widely and **studying more knowledge** <L1> (acquire/gain knowledge).*

*The most important is that I can **learn** all kinds of **knowledge** <L1> (acquire/ gain knowledge).*

Students who had learnt that some verbs with *-ing* and *-ed* endings function as adjectives and known some collocations e.g. *a missing boy, a knowing smile,* probably considered that verb with *-ing* equaled adjective, hence errors occur, which are illustrated by these examples.

*e.g. In order to have an **attracting looking** <L2> (an attractive looking) they spend much money.*

*Besides it can also create a **healthing environment** <L2> (healthy environment) for us.*

Some verbs such as *take, make, do* are de-lexicalized verbs. They have a very wide applicability but have little meaning on their own. Besides, they share some common meanings but they are used with different nouns e.g. *make the bed, do the dishes,* and *take a shower.* Due to the lack of collocation knowledge, some students assume that these verbs could replace one another. Also, some learners use synonyms to replace nodes or collocates. Therefore, the participants made errors, such as:

*People can **make advantage** <L1> (take advantage) of saving time.*

*We should **take a profit** <L1> (make a profit) on doing part-time job.*

*We **do plans** <L1> (make plans) for summer holidays.*

*We also need to **treasure every moment** <L1> (cherish every moment) in our daily life.*

*We should care for the environment and try to **less** unnecessary **waste** <L1> (reduce waste).*

*I think that it is **definitely value** <L5> (definitely worth) the risk to take the part-time job.*

*The social support **provides** many **chances** <L1> (provide opportunities) to student.*

*Few of them can speak **smooth English** <L2> (fluent English).*

*I believe that we can reduce the waste on campus with our **insistent efforts** <L2> (ceaseless efforts).*

Some students were used to learning words in isolation. They only understood the basic meanings of a word but did not know what words it would go with. They were not able to produce the right collocation or they just put words together randomly.

*The most important is to **plant the conscious** <L1> (raise consciousness) of reducing waste in daily life.*

*It's easy to find that most students try to **practice** their **thrifty conscious** <L1> (raise their thrifty consciousness).*

*It takes lifetime **to prove promise** <L1> (to fulfill promise).*

*If we have a good habit of save energy and foods, **keeping the environment balance** <L2> (keeping the ecological balance), we would have a lovely planet.*

*I stay in classroom when my classmates like **a cloud of wind** <L4> (a gust of wind) to go to supermarket to buy goods after class.*

*Waste resource is a main reason of causing **strongly polluted** <L5> (badly / seriously /heavily polluted) environment.*

*Time like arrow, we should grasp it and **manage reasonably** <L6> (manage effectively or properly).*

Conclusion & Implications for Future Studies

This study has shown that through utilizing COCA to conduct error analysis, one hundred and eight lexical collocation errors were found in 100 NEMs' writings in KU. They were classified, counted, and calculated in terms of percentage. These collocation errors were not distributed evenly across different lexical collocation types. Collocation errors in L1 Verb + noun collocations accounted for most of the errors because the usage of verbs is the most complex in English for them. Collocation errors in adjective accounted for a large percentage of collocation errors, too. So, students should pay attention to the choice of adjectives to express their ideas. Only five L4 Quantifier + noun collocations were found in the participants' writings, but four of them were errors, which may indicate that NEMs in KU did not pay much attention to what quantifiers go with a given noun.

For error analysis in this study, the inter-rater reliability (a=0.854) was quite high. Actually, the two raters arrived at an agreement when using COCA to detect collocation errors. For instance, two raters decided "span" for a query key (4 words on the left and 4 words on the right of a "node" word) when they detected what words go with a given word. However, some practical limitations occurred. The two raters were free users of COCA, so inevitably sometimes the corpus was unstable when query results could not be yielded. They needed to exit and then log in again. Also, the number of queries was limited for free users each day, so each rater had to register two accounts to avoid slowness in conducting searches. Since COCA can be used as an assistant tool to raise students' collocation awareness through a variety of exercises in a subsequent study, the researchers will take all the limitations into consideration.

English collocation is important in receptive as well as productive language competence (Cowie, 1994). Collocational knowledge is essential for students to

distinguish grammatically well-formed sentences that are "natural" from those that are "unnatural" (Malligamas & Pongpairoj, 2005). Error analysis has demonstrated that adequate knowledge and effective use of collocations are significantly associated with the production of collocation and quality of written communication. The more English collocations students master, the higher writing quality students will produce. Taking the conclusions into consideration, some recommendations for NEMs and teachers in KU are made.

1. Vocabulary should be learnt by means of collocations, so that learners will notice "how words co-occur together";

2. Vocabulary should be learnt in contexts, which is helpful for learners to master collocations;

3. Teachers can expect students to produce more collocations by providing more collocation input in classes;

4. Corpora and collocation dictionaries not only are helpful tools for learners to learn collocations but also useful for them to correct their collocation errors and enhance the accuracy.

5. For the treatment of collocation errors the students make, attention should be paid to the most frequently made error types first, which is L1 type in this present study, before moving on to the less serious ones.

Research has shown that EFL learners face problems with collocations, resulting in producing erroneous ones. The findings of this study are not conclusive. Thus, the need for more research on collocations is urgent. The researchers offer some recommendations for further studies on collocations in EFL contexts.

First of all, large scale studies covering more participants are needed so as to make the findings more generalizable.

Second, future studies can be wider in scope to include all the possible lexical collocations, such as Noun + noun collocations, so that the data could be more representative of the participants' knowledge of collocations.

Third, future research should consider grammatical collocations and its effectiveness in improving writing fluency.

Finally, the focus of future research could be the effects of using COCA to raise student's collocation awareness on different levels of English proficiency.

References

Benson, M., Enson, E., & Ilson, R. (1997). *The BBI Combinatory Dictionary of English (Rev. ed.).* Amsterdam: John Benjamins Publishing Company.

Chen, W. X. (2002). Collocation errors in the writings of Chinese learners of English. *Journal of PLA University of Foreign Languages*, 25(1), 60-62.

Chow, P. K. (2006). *Tense and aspect in interlanguage: error analysis in the English of Cantonese-speaking secondary school students.* Master's Degree Thesis. University of Hong Kong.

Church, K. & Patrick, H. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22-29.

Corder, S. P. (1981). *Error Analysis and Interlanguage.* Oxford: OUP.

Cowie, A. P. (1994). Phraseology. In Asher, R.E. & Simpson, J. (Eds.), *The Encyclopedia of Language and Linguistics* (pp. 3168-71). Oxford: Pergamon Press.

Davies, M. (2008). Corpus of Contemporary American English (COCA). http://www.americancorpus.org (accessed October 23, 2010).

Malligamas, P. & Pongpairoj, N. (2005). Thai learners' knowledge of English collocations. *HKBU Papers in Applied Language Studies*, 9, 1-28.

Meng, J., & Li, L. (2005). A tentative analysis of collocational errors in Chinese EFL learners' composition. *Journal of Shangdong Electric Power Institute*, 20-22.

Sun, L. (2011). Reflections on the teaching of English writing. *China-Asean Exposition*, 5, 52-53.

The National College English Testing Committee (NCETC) (n.d.). *College English Test Band-4 and Band-6.* Retrieved from: http://www.en.cet.edu.cn/

Wang, H. Y. (2010). An analysis of obstacles of undergraduates' listening and speaking in ethnic communities. *Journal of Kaili University*, 28(2), 79-81.

Wu, Z. (2003). Developing students writing strategies: an investigation of language errors from students' compositions. *Journal of Southeast Guizhou National Teacher's College*, 21(1), 114-115.

Zhang, X. Y. (2009). On improving English writing teaching and raising student's writing ability in higher vocational colleges. *Science & Technology Information*, 7, 641-642.

Zheng, H. P. (2000). On today's English reading problems and improvement measures. *Journal of Southeast Guizhou National Teacher's College*, 18(l), 84-85.