

Corpora, Concordancing and Collocations

Kwanjira Chatpunnarangsee
Thammasat University

Abstract

This research used a case study and Data Driven Learning approach to address an existing gap in literature on corpus-based English language learning in the Thai context. Specifically, the researcher investigated students' problem-solving processes when they had direct access to corpus data and their views on using corpus data as a linguistic reference. Using data collected through think-aloud, interview, and survey techniques with two Thai students, the findings of this study suggest how corpus consultation may help students to acquire proficiency with English collocations to become more autonomous learners. However, some difficulties resulting from the students' use of the tool were also found, suggesting that students should be provided with step-by-step instructions in how to deal with corpus data.

Keywords: Corpora; Concordancing; Collocations

บทคัดย่อ

งานวิจัยชิ้นนี้ใช้วิธีการศึกษาเฉพาะกรณี และแนวทางการเรียนโดยใช้คลังข้อมูลภาษา (Data Driven Learning) เป็นกรอบในการศึกษาถึงกระบวนการแก้ปัญหาเมื่อผู้เรียนมีโอกาสใช้ข้อมูลในคลังข้อมูลโดยตรงโดยไม่ผ่านผู้สอน และทัศนคติของผู้เรียนต่อการใช้คลังข้อมูลเพื่อเป็นแหล่งอ้างอิงทางภาษา ผ่านการเก็บข้อมูลแบบกิจกรรมการบอกกระบวนการความคิด (Think-aloud) การสัมภาษณ์ และการสำรวจความคิดเห็นของผู้เรียน พบว่าการเรียนแบบใช้คลังข้อมูลอาจช่วยผู้เรียนเรียนรู้คำปรากฏร่วมภาษาอังกฤษ และสามารถเรียนแบบพึ่งพาตนเองได้มากขึ้น อย่างไรก็ตามผลการวิจัยยังพบว่าผู้เรียนมีอุปสรรคในการใช้เครื่องมืออยู่บ้าง จึงควรมีการสอนการใช้คลังข้อมูลอย่างละเอียดและเป็นขั้นตอน

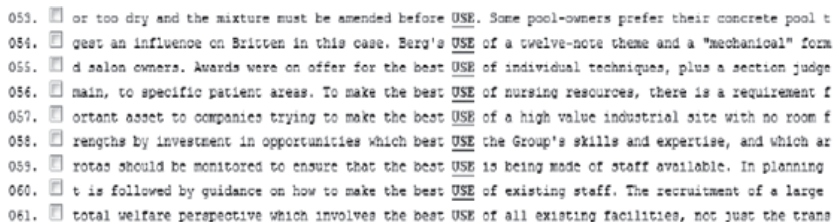
คำสำคัญ: คลังข้อมูล การเรียนโดยใช้คลังข้อมูลภาษา คำปรากฏร่วม

Introduction

Proficiency in English writing can play an important role in ESL/EFL students' access to good jobs or admission to higher education. However, while learners are often concerned about developing native-like oral English performance, the problem of writing English with a *foreign accent* may be of equal concern, especially in preparation for settings in which production of documents that are clear and idiosyncratically correct is important. However, EFL learners in particular, because they are not immersed in an environment in which English is regularly used, must depend on lexical resources and skills to aid their English writing. Traditionally, EFL writers have resorted mainly to monolingual and bilingual dictionaries, but today a more sophisticated resource that goes beyond translating or defining to contextual accuracy is available in the form of online corpus technology, which exposes learners to a wide variety of examples of language phrases and expressions as they are commonly used by native speakers (Chang & Sun, 2009; Sinclair, 1997; Sun, 2003; Varley, 2008; Yoon, 2008). When users consult a corpus by entering a keyword on a search input page, they receive a range of phrases or collocations illustrating various contexts in which the word is used by native speakers. This activity, known as "concordancing," serves both to support writing and develop language proficiency while encouraging learners' autonomy and control over their own processes of English acquisition. The purposes of the present study, then, are 1) to explore how Thai learners of English for academic purposes use language corpora to facilitate their learning of English collocations, and 2) to elicit their views on using corpus data as a linguistic reference. For these purposes, a case study approach with two learners at different levels of proficiency was used in order to capture both the benefits and difficulties of concordancing as a language acquisition strategy.

Corpus consultation as a tool for Language Learning

The use of corpora and concordancing in language teaching has grown along with the Internet from the early 1990s (Woolard, 2000). Corpus consultation provides examples of target words in various contexts, reflecting different usages and suggesting phrases that may be useful to language learners. The following excerpt from the British National Corpus (BNC) for the target noun “use” illustrates how a concordancer works:



053. ☐ or too dry and the mixture must be amended before USE. Some pool-owners prefer their concrete pool t
 054. ☐ gest an influence on Britten in this case. Berg's USE of a twelve-note theme and a "mechanical" form
 055. ☐ d salon owners. Awards were on offer for the best USE of individual techniques, plus a section judge
 056. ☐ main, to specific patient areas. To make the best USE of nursing resources, there is a requirement f
 057. ☐ ortant asset to companies trying to make the best USE of a high value industrial site with no room f
 058. ☐ ranches by investment in opportunities which best USE the Group's skills and expertise, and which ar
 059. ☐ rotes should be monitored to ensure that the best USE is being made of staff available. In planning
 060. ☐ t is followed by guidance on how to make the best USE of existing staff. The recruitment of a large
 061. ☐ total welfare perspective which involves the best USE of all existing facilities, not just the trans

Figure 1

This format gives learners the opportunity to identify a word's collocations, analyze language patterns, and identify the contextual factors that might influence variations among patterns (Hadley, 2002, as cited in Chang & Sun, 2009, p. 284). In the example above, for instance, learners can conclude from its frequent occurrence that “best use of” is a common English collocation. Using concordance-based materials calls for the application of both inductive and deductive reasoning. For example, if students are asked to look up collocations for “use” as a noun, they inductively learn possible collocated words of “use” in its common contexts, but if they are given a phrase “best use of” to look up in a concordance, they will deductively determine

that it is an acceptable collocation in English. Hence, concordance-based language teaching is flexible and can accommodate a variety of learning purposes.

Nevertheless, whether concordancing has advantages over using a dictionary that provides sample contexts has been a matter of debate. Some scholars (Wu, 2009) have pointed out the fragmentary nature of collocations, which in effect de-contextualizes the language in a corpus. The issue of corpus selection is also problematic. On the one hand, if students consult a very large corpus, too many examples might be confusing or lead to errors, and on the other hand, a small or specialized corpus may provide too few or even no examples for a particular word. Nevertheless, with proper training, EFL learners may find concordancing a useful way to access authentic contexts for English words and phrases.

Research on corpus use in Thai classrooms includes Sripicharn's (2002) evaluation of the use of teacher-designed Data Driven Learning (DDL) materials and Todd's (2001) examination of the effects of concordancing on students' induction ability. However, there is no research on how Thai students go about using corpus data to solve collocation problems when they have direct access to it or how useful they find concordancing for their English learning. The present study addresses this gap.

Conceptual Framework

The conceptual framework for this study is Data-Driven Learning (DDL), which Johns (1991b), describes as attempting to "cut out the middleman as far as possible and give direct access to data so that the learner can take part in building up his or her *own* profiles of meanings and uses"(pp. 30–31). Thus language learners can use data in electronic corpora autonomously as linguistic resources in ways that may also

improve their analytical and reasoning abilities. Johns outlines three effects of the DDL approach: (1) DDL can affect the language learning process by helping the learner develop the ability to see patterns in the target language, and by extension, to form generalizations to account for that patterning; (2) DDL changes the teacher's role to that of coordinator of student research; and (3) DDL supports *grammatical consciousness-raising*¹ by situating the learner's own discovery of grammar at the center of language learning (Johns, 1991a, as cited in Partington, 1998, p. 6). In DDL, the learners' role is very active as students become "language researchers," the assumption being that effective language learning is itself a form of linguistic research.

Methodology

The following questions guided this study:

1. What problem-solving processes and strategies do EFL learners use in web-based concordancing?
2. How do these view using corpus data as a linguistic reference?

Description of the Context and Participants

The study was conducted in a non-credit EFL writing workshop taught by the researcher at a university, which comprised four two-hour sessions on Sunday afternoons. *Paragraph Writing: A Process Approach* (Chalermmpatarakul, 2006) was used as the text. Of the six chapters, this research centered on two, "People Around Us" and "A Special Place," both focusing on description. Each chapter was organized to

¹ See Rutherford (1987) for insightful details of *grammatical consciousness-raising*.

follow eight steps in the writing process: *generating ideas, organizing ideas, developing cohesion, writing the first draft, revising your writing, editing your writing, writing the second draft, and developing your skills*. The students were asked to write a paragraph of approximately 200 words for each chapter following all steps of the writing process. Corpus-based lessons were introduced during the stage of *editing your writing*. Two Thai students, Nada and Karn, (pseudonyms) participated in the study. Nada, 31, possessed an undergraduate degree in Engineering from a Thai university. She had studied in a language school in the United States for 15 months. Nada reported having a TOEFL score of 550. Karn, 26, graduated from a Thai university majoring in Science. He had studied in an Intensive English Program in the United States for seven months. Karn reported having a TOEFL score of 540. It should be noted that even though their reported TOEFL scores were relatively close, Karn's language performance, as observed by the instructor, was much lower than Nada's.

Data Collection

Procedure of data collection

During the first class, the instructor taught Chapter 2, "People Around Us," and distributed a pre-project questionnaire to obtain students' demographic data, introduced a web-concordancer along with its searching techniques, and discussed the nature of collocations. A collocation pre-test was also administered in this session. The students were assigned to write a paragraph and turn it in by Thursday night so that the instructor could return the draft with feedback on Sunday. During the second class, along with content, organization and language feedback, the instructor highlighted mis-collocated words in students' written assignments and had them search for correct collocations on the Compleat Lexical Tutor website, <http://www.lex tutor.ca/conc/eng/>.

This online concordancer was used in the present study because it was practical and easy-to-use. After the students did corpus searches, they submitted a corpus search log and a written reflection on what they like, what they dislike, and how the web-concordancer help them to improve their English collocation learning, with each corrected draft. To ensure students' exposure to collocations before the post-test, they were given lesson-related concordancing worksheets for practice finding correct collocations in the web-concordancer. The students were assigned to submit the second draft on Thursday night. In the third class, the instructor taught Chapter 3, "A Special Place," and students worked on concordancing worksheets for chapter 3. In the last corpus-based lesson, a post-project questionnaire was distributed to obtain students' attitudes toward using a web-concordancer as a linguistic reference and a collocation post-test to see if there is a gain in scores from the collocation pre-test. The collocation pre- and post-tests were identical except that some items on the post-test were re-ordered. A focus group interview was conducted after the last session in Thai, the participants' first language, and then translated into English. The interview transcription was sent to the participants for member checking.

Although the pre- and post-tests were not directly related to the research questions, they are discussed in the findings part as they might be useful for data triangulation. Besides demographic information, the pre-project questionnaire elicited participants' background knowledge about web-concordancers and corpora. The post-project questionnaire, which contained 22 five-point Likert scale questions adapted from Yoon's (2004) instrument, elicited students' perceptions of the use of corpus consultation to improve their English collocations. Items 1 to 7 focused on benefits of corpus use while items 8 to 14 focused on difficulties with corpus use. Items 15 to 22 focused on overall evaluation of corpus use. The internal reliability of the post-project questionnaire

using Cronbach's Alpha is $r = 0.86$ indicating a relatively high level of internal consistency.

The last meeting was for the students to perform a think-aloud task while trying to find correct collocations in the concordancer. After a training session, each student was presented with 10 items, 8 of which contained one mis-collocated phrase, with no requirement to finish all of them within their 45-minute think-aloud sessions. The task was based on the following 10 collocations which appeared in writing samples collected from students who took Paragraph Writing course: (1) in his late teens, (2) of medium height, (3) ashamed of, (4) on the phone, (5) between 10 p.m. and 2 a.m, (6) get so annoyed, (7) turn grey, (8) immediate reaction, (9) came running, and (10) one of the events, eight of which were replaced by the following incorrect collocations:

1. I met someone who was *about his late teens.
2. He is *a medium height.
3. He is now *ashamed from his conduct.
4. He talks *on a phone.
5. He talks *between 10 p.m. to 2 a.m.
7. Some days I think my hair could *get grey because of my roommate's behavior.
8. My roommate always has *immediately reaction to the expressions.
10. It was *one of events that showed me how thoughtful he is.

The participants were encouraged to use the British National Corpus and the Brown Corpus because they were considered to be general corpora.

Findings

Students' problem solving processes and strategies in a web-based concordancing setting

Case 1. Nada

Nada was a fast learner and a competent computer user. She could immediately navigate the input page (See Figure 2 for a screenshot) and she was able to finish answering all ten sentences correctly in 33 minutes. Before she started concordancing, she read the sentences carefully and made sure that she understood the meaning of the collocations. For items 3, 4, 5, 6,

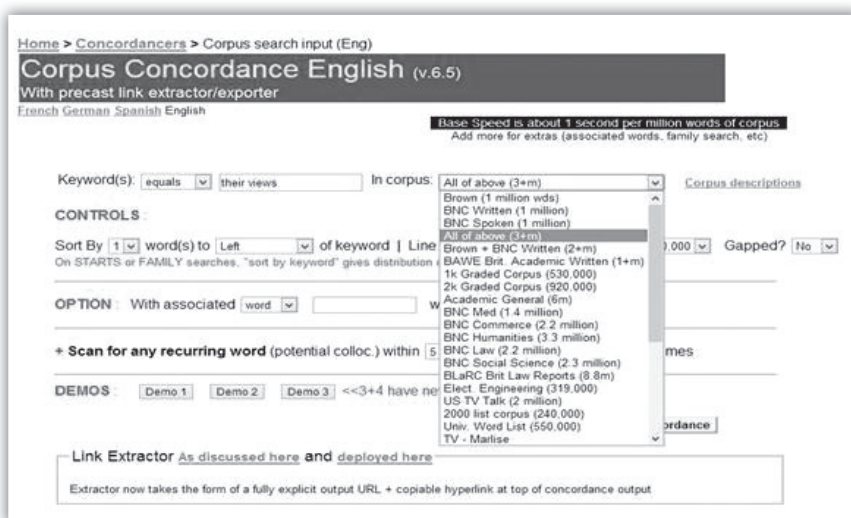


Figure 2

8, and 10, she gave the answer first and then searched the corpus data to confirm it. For Item 3, *He is now *ashamed from his conduct*, for example, after she entered all required information on the input page, the software listed 28 concordance lines on the output page, 16 of which contained the phrase, *ashamed of* and none of which showed the phrase, *ashamed from*. When she saw this pattern, it confirmed her guess, so she corrected item 3 to *He is now ashamed of his conduct*.

Sometimes her guess was wrong. For instance, for item 5, *He talks *between 10 p.m. to 2 a.m.*, Nada first thought the phrase was correct, and expressed surprise when she found three examples of *between (time) and (time)* on the output page (Figure 3).

Figure 3

Nada exclaimed,

I have always thought *between (time) to (time)* was correct because this phrase can be written with a dash and a dash to me means *to*. I am surprised. Now I am looking at other examples. I can see that *between ...and...* can be used with many different things such as

age (between 2 and 5 years old), distance (between 200 and 400 meters), percent (between 11 and 17 percent), etc.

Nada's comment showed that she could draw a conclusion from seeing patterns in corpus data, and she was pleased with this discovery. Item 4 *He talks *on a phone*, is another example that demonstrated how Nada discovered her misunderstanding of a phrase by reading examples on the concordance lines.

Nada also used her English grammatical knowledge to help her derive a correct answer. When she worked on item 6 *I get so annoyed with his talking and laughing*, she compared the bold phrase in her assigned sentence with her search results. A part of the search results is displayed in Figure 4.

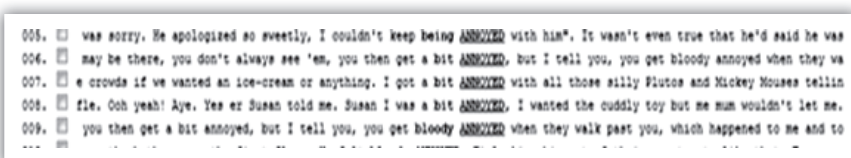


Figure 4

From such examples such as *get a bit annoyed* and *get bloody annoyed*, she concluded that, “It looks like *get annoyed* can have an adverb in the middle and adverb modify adjective *annoyed*. I see *get a bit annoyed* and *get bloody annoyed* so *get so annoyed* should be correct.” Based on corpus consultation in combination with her grammar knowledge, Nada also spotted the error in Item 8 *My roommate always has *immediately reaction to the expressions*.

Nada also made use of the full text or KWIC (key word in context) option provided in the concordancer, especially when she was

not certain of the answer. By clicking on the blue font in the middle of a concordance line, she accessed the text excerpt from which the concordance line came. For example, for item 2, *He is *a medium height*, she entered *height* in the keyword field and found 118 concordance lines. She looked at the ones that were relevant to her query, as shown in figure 5, and then clicked on HEIGHT in line 55 to see its full text, as shown in Figure 6. She decided that *of medium height* was the correct answer because it can be used to describe a person's appearance, and item 2 seemed to have the same purpose.

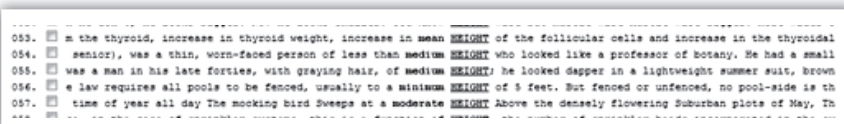


Figure 5

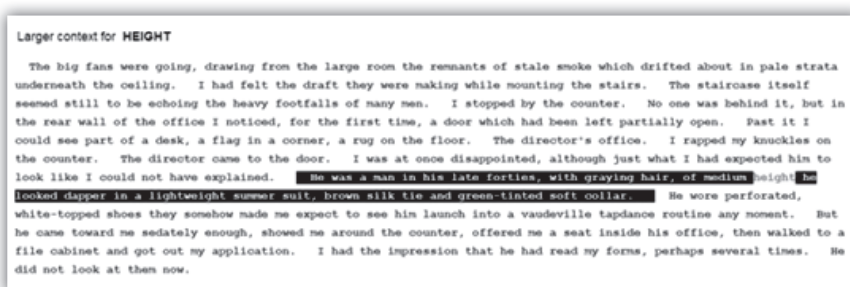
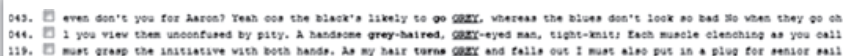


Figure 6

In general, overall, Nada's problem-solving using the concordancer went smoothly; however, sometimes there were either too many or too few results on the output page. She preferred too many examples to too few. To find the correct collocation(s), she needed to read vertically first, looking at the keyword and then looked at the

word(s) on either side. Too few results did not give her enough information, so she needed to go back to the input page and change settings in some fields. This effort getting to the right data caused some frustration for her.

The second problem was the issue of choosing a keyword that would produce the right set of data. This was evident when she worked on item 7, *Some days I think my hair could *get grey because of my roommate's behavior*. At first, she entered *get* as a keyword and received 4,976 concordance lines, none of which had *got* followed by *grey*. She then extended her query to *get* followed by the words for colors, but she did not find any examples. One reason for her difficulty seemed to be that she didn't understand the phrase, as Thais refer to hair becoming white, not grey, and she asked the researcher for the meaning. Then she entered *grey* and this time received 133 concordance lines, three of which were relevant (figure 7). After she spent another five minutes on the task, she answered with *turn grey* which was a correct answer. This item took her longer than the other items.



043. ☐ even don't you for Aaron? Yeah one the black's likely to go GREY, whereas the blues don't look so bad No when they go oh
 044. ☐ I you view them unconfused by pity. A handsome grey-haired, GREY-eyed man, tight-knit: Each muscle clenching as you call
 119. ☐ must grasp the initiative with both hands. As my hair turns GREY and falls out I must also put in a plug for senior sail

Figure 7

Nada's last problem was choosing a corpus from several offered by the website (see figure 2 for a list of corpora), including specialized corpora such as BNC Law or BNC Medicine:

I don't know which corpus I should use besides the three corpora that the teacher has suggested, the one that is labelled All of above (3+M). It is complicated and it takes time to find the answer. If I can't find an

answer in one corpus, I have to get out and try another corpus.

Nada mentioned the issue of choosing a corpus in the interview as well. On the positive side, she believed that the more time and effort she spent on the task, the more she tended to retain her new knowledge.

Along with the Lextutor concordancer, Nada used Google and a dictionary program called *Longman*. She wanted to get immediate answers to her questions. She carefully read and ensured that she understood the meaning of the sentences so as to know what collocations she was looking for. When she could not find the information she needed from other sources, she asked the researcher.

Case 2. Karn

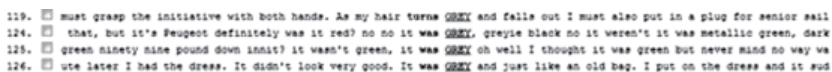
Karn took forty-five minutes to finish his think-aloud task and got six out of ten items correct. He usually chose to search in three corpora which were labelled as *All of above (3+M)*. Karn had used another website, *Corpus of Contemporary American English (COCA)* before and preferred some of its features, such as the fact that it gives information about the parts of speech associated with a word. Karn was quite nervous and needed a few minutes to work alone, so the researcher left the room at the beginning of his think-aloud session.

Some of Karn's problem-solving processes during concordancing were found to be similar to those of Nada's. First, he guessed the correct answer whenever he could and then used the concordancer to check his assumption. This was evident when he worked on item 6 *I get so annoyed with his talking and laughing*. On this item, he guessed that *get so annoyed* was wrong and should be *am annoyed* or *feel annoyed*. He seemed surprised but satisfied when he learned *get*

so *annoyed* was also correct. Karn went through the same process for items 3, 5, 8, 9, and 10, spending more time and effort on some items than others, depending on the results on the output page.

Second, he also used his English grammatical knowledge to help him find a correct answer. However, he sometimes relied on grammar so much that he failed to pay attention to patterns on the output page. For example, when he worked on item 7 *Some days I think my hair could *get grey because of my roommate's behavior*, he found 133 hits after he submitted his query, but he did not see *get grey* included in any concordance lines. He knew that sometimes *to be* and *to get* could be used interchangeably, so he changed *get grey* to *be grey* after only a quick glance at the concordance lines, assuming that because some of the concordance lines showed *was grey*, his grammar-based assumption was wrong. It seemed that he made this swift decision because he saw *was grey* on some of the concordance lines, and that confirmed his initial grammar assumption about this item. In this case, reliance on grammar knowledge might have hindered him from paying attention to the context of the item and the examples on the output page,

Unlike Nada, Karn did not look closely at the contexts of the collocations, which varied in relevance to his search as shown in Figure 8.



119. ☒ must grasp the initiative with both hands. As my hair turns **grey** and falls out I must also put in a plug for senior rail
 124. ☐ that, but it's Peugeot definitely was it red? no no it was **grey**, greyie black no it weren't it was metallic green, dark
 125. ☐ green ninety nine pound down isn't? it wasn't green, it was **grey** oh well I thought it was green but never mind no way wa
 126. ☐ ute later I had the dress. It didn't look very good. It was **grey** and just like an old bag. I put on the dress and it sud

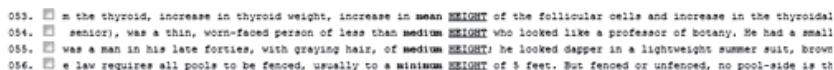
Figure 8

If he had paid attention to context, he would have noticed that line 119 referred to hair and was the only one applicable to his query.

Karn was also discouraged when he encountered a large amount of data in a format he was not accustomed to reading:

Sometimes the concordancer got back to me with a lot of results. It was difficult to read. The output page looked daunting with a lot of words on it. Even though I started reading at the keyword in the middle of the lines, it was still hard, and my eyes were blurred with a thousand words on the page. Plus those lines were difficult to understand. They have many vocabularies that I don't know.

This confusion might explain his mistake on item 2 *He is *a medium height*. After he entered height as a keyword on the input page, he received 118 hits, of which 53, 54, 55, and 56 (Figure 9) were directly relevant to his query, but he decided that *a medium height* was correct.



053. [] m the thyroid, increase in thyroid weight, increase in mean HEIGHT of the follicular cells and increase in the thyroidal
 054. [] senior), was a thin, worn-faced person of less than medium HEIGHT who looked like a professor of botany. He had a small
 055. [] was a man in his late forties, with graying hair, of medium HEIGHT; he looked dapper in a lightweight summer suit, brown
 056. [] e law requires all pools to be fenced, usually to a minimum HEIGHT of 5 feet. But fenced or unfenced, no pool-side is th

Figure 9

Other challenges he encountered were quite similar to Nada's, including getting too few or no results from a search, choosing an inaccurate keyword, and misspelling a keyword. In the first case he had to go back to the input page and fill in the required fields again. When he chose the wrong keyword, the data he got were irrelevant. Karn also found that even omitting an *s* at the end of a word would yield the wrong results, as when he entered *teen* instead of *teens*.

In his linguistic research process, Karn used Google and COCA to confirm the answers from Lextutor's concordancer as well as to obtain more information. He also used Longdo Dictionary online (<http://dict.longdo.com/>). He said that he needed to know the meaning of a phrase before answering, and reading concordance lines did not help him.

Students' views on using corpus data as a linguistic reference

The main themes that emerged from data analysis are discussed in the following section.

The perceived advantages of concordancing

Overall, both students expressed favorable views toward using corpus data for collocation learning. The advantages they perceived included the following: (1) they could take a more active role in learning; (2) they could be more autonomous learners; and (3) they remembered the new knowledge better. First, the students found that *learning by doing* engaged them more with the materials than listening to a teacher. Karn mentioned the following in the interview:

Concordancing helped me learn English collocations. This is because instead of being told by someone, I had to concordance to find the collocations myself. This is a new way of learning to me. I usually learned English by listening to the teacher. This time I actually did the work.

Nada agreed with Karn; she also added that:

For some items, I had to concordance many times because I could not find the answers. I had to

think of a good keyword which could take me to the answer. I think I was acting like a detective trying to find the answers, and because of that, once I found the correct answers, I remembered them automatically.

Secondly, they felt more autonomous because, once they understood how to read the output page; they could access corpus data independently to find an answer to any question that they might have. If they could access the Internet, they could access the data in a corpus. After they learned and they became independent, Karn added, “When I asked people, some people could give me answers but some could not.” Finally, both participants agreed that concordancing might help them remember the new collocations better. When talking about their post-test scores, which had improved by 20% for Nada and 46% for Karn, the students expressed the idea that, by going through the repeated process of concordancing, they were more likely to retain correct collocations. As Karn shared,

I think it helps me to memorize the collocations because I actually performed the search myself. I was first introduced to collocations when I was in high school. I had a book and the teacher told me to remember, for example, the phrase *commit a crime*. It took a few seconds but I forgot it in a few days. With concordancing, I have to search and read a lot and that helps me remember collocations better than having someone tell me.

Other advantages of concordancing they mentioned included the availability of having the full text and the layout of the output page, which enabled them to see the collocations more easily.

Perceived disadvantages of concordancing

The two main issues the students encountered were (1) the complex process involved in using the tool; and (2) its uninviting visual display. The first issue appeared to be their biggest concern. Both concurred that concordancing was complicated and difficult at first. It took time and effort to get to the data and to derive correct collocations, due to the difficulty level of the examples and the amount of data in the corpora. Karn recalled his frustration:

At first, I was frustrated [about] when was I going to find an answer. It took so long just to find only one word. I got used to learning new vocabulary and phrases or what we called *collocations* here by a teacher telling me. So at first I thought, c'mon', just tell me'.

Nada added her comments

I agree with Karn. It was frustrating at first. I did not know how to use it [the concordance]. Until I knew I didn't have to read everything and that made my life easier. For example, to check if *fresh breeze* is a correct collocation, I just need to scan for the words starting with letter *f* before the keyword *breeze*. Those words are in alphabetical order. That's one thing. Other problems are some examples were too difficult to understand and sometimes I could not find some words even though I knew they exist.

The other main issue, the visual display, was problematic for both the input and output pages. The input pages were not attractive in appearance and required users to fill in many fields, and the output

pages were crowded with words in a format that was tedious to read. Nada acknowledged that the output page layout was designed for a purpose, but she still wished that it was more user friendly. This comment was repeated in her written reflections.

The students' future plans for concordancing

When asked whether they would continue to use the tool, both Nada and Karn replied that they would and gave the following reasons:

Nada: I think I will use it. Some phrases/collocations I didn't know before. I might come up with language that is not actually used.

Karn: I will use it when I want to check if the phrases I wrote are in fact used by native speakers. For example, I found *many styles of paintings* instead of *many kinds of paintings*. I think it sounds better.

However, both participants thought that corpus consultation might be appropriate only for students with high English proficiency. Beginning students might struggle with excessive amounts of data, much of which might be beyond their current levels.

Finally, descriptive statistics in SPSS indicated that the results of the post-project questionnaires corresponded to the interview responses in many ways. Tables 1, 2, and 3 below show the mean scores of three constructs, benefits, disadvantages, and overall assessment of corpus consultation:

	Mean	Std. Deviation
Q1 Easy search technique	4.00	0.00
Q2 Helpful than dictionary	3.50	2.12
Q3 Help learn vocabulary	3.00	0.00
Q4 Help learn collocations	4.50	0.70
Q5 Help learn grammar	3.50	2.12
Q6 Improve writing skill	4.00	1.41
Q7 Usually give info I look for	3.50	0.70

Table 1 Descriptive Statistics: Benefits of corpus use (1st construct)

	Mean	Std. Deviation
Q8 Problem due to no access to Internet	1.50	0.70
Q9 Require time and effort	3.00	1.41
Q10 Unfamiliar vocabulary	2.00	0.00
Q11 Cut-off sentences	2.50	0.70
Q12 Too many sentences	4.00	0.00
Q13 Too few sentences	3.50	0.70
Q14 Difficult search technique	2.00	0.00
Q15 Difficult texts	2.00	0.00

Table 2 Descriptive Statistics: Problems or difficulties of corpus use (2nd construct)

	Mean	Std. Deviation
Q16 Use corpus by own choice	4.50	0.70
Q17 Will use corpus for other courses	3.50	2.12
Q18 Learn more, like more	4.00	0.00
Q19 Will use corpus in future	4.50	0.70
Q20 Corpus increase writing confidence	4.50	0.70
Q21 Should use in other courses	4.50	0.70
Q22 Very useful resource	4.50	0.70

Table 3 Descriptive Statistics: Overall evaluation of corpus use (3rd construct)

These tables show that students valued the tool primarily as a helpful aid to language learning but might be frustrated by too much or too little data, and that overall they would use the resource in the future.

Discussion and Conclusion

The results of this study allow us to understand how the participants used and viewed corpus consultation as a linguistic reference. Overall, corpus consultation seemed to assist their learning of English collocations and supported them in *learning by doing* which, in turn, could help them retain new knowledge. In addition, the tool enabled them to find answers to their queries without having to depend on teachers, increasing their autonomy as learners. However, using corpus resources did not come easily at first, but required sufficient practice, time, and effort. Students needed to familiarize themselves with how to access and analyze corpus data, especially when they had to deal with massive amounts, often at a level of difficulty beyond their current knowledge. They also had to deal with lack of relevant data for some words.

This brings us back to DDL's claim regarding the benefit of having direct access to data so that the learner "can take part in building up his or her *own* profiles of meanings and uses" (Johns, 1991b, pp. 30–31). The present study shows that learners engage in this constructive process upon depending on their English proficiency. The student with higher English proficiency in the study demonstrated stronger capability in analyzing corpus data and inductively solving collocation problems than the lower proficiency student, although both found the resource valuable. However, I recognize that the small number of participants places a limitation on this study's generalizability. Future research could examine whether larger confirmatory studies would yield similar or different results.

This look at the concordancing behaviours and perspectives of the participants suggest ideas for further teaching and research. In terms of teaching, it is recommended that students with lower English proficiency be introduced to concordancing with step-by-step instructions of how to analyze corpus data and scaffolding activities, such as beginning with words that would yield a moderate amount of useful data and gradually progressing to more complicated searches. In terms of suggestions for future research, firstly, more types of collocations, both lexical collocations and grammatical collocations, should be included such as adjective+ noun, noun+verb, verb+ adjective, noun+ preposition. Also, items for the think-aloud task should include more items for which participants are likely to try more than one keyword, such as item 7 *get grey. These types of prompts could more clearly distinguish the difference between problem-solving processes of high and lower proficiency students. Lastly, it would be useful to keep track of the time for each concordance session, to determine whether searches become more efficient with practice.

Biodata

Kwanjira Chatpunnarangsee is an Assistant Professor at the Department of English, Faculty of Liberal Arts, Thammasat University. Her research interests center around corpus-based language learning and its application to ESL/EFL teaching, second language writing, bilingualism and bilingual education, and discourse analysis in second language classrooms.

References

- Chalermpatarakul, C. (2006). *Paragraph writing: A process approach*. Bangkok: Thammasat University Press.
- Chang, W. & Sun, Y. (2009). Scaffolding and web concordancers as support for language learning. *Computer Assisted Language Learning*, 22(4), 283–302.
- Johns, T. (1991a). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Johns & P. King (Eds.), *Classroom Concordancing, ELR Journal*, 4 (pp.27–45). Birmingham: University of Birmingham.
- Johns, T. (1991b). Should you be persuaded: Two samples of data-driven learning. *English Language Research Journal*, 4, 1–46.
- Marriam, S. (2002). *Qualitative Research in Practice*. San Francisco, CA: Jossey-Bass.
- Partington, A. (1998). *Patterns and meaning: Using corpora for English language research and teaching*. Amsterdam: Johns Benjamins.
- Rutherford, W. (1987). *Second Language Grammar: Learning and Teaching*. London: Longman.
- Sinclair, J. S. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery & J. Knowles (Eds.), *Teaching and language corpora* (pp. 27–39). London: Longman.
- Sripicharn, P. (2002). Evaluating data-driven learning: the use of classroom concordancing by Thai learners of English. Unpublished PhD dissertation, The University of Birmingham.
- Sun, Y. (2003). Learning process, strategies, and web-based concordancers: A case study. *British Journal of Educational Technology*, 34(5), 601–613.
- Todd, R. (2001). Induction from self-selected concordances and self-correction. *System*, 29, 91–102.

- Varley, S. (2008). I'll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133–152.
- Woolard, G. (2000). Collocation – encouraging learner independence. In M. Luwis (Ed.), *Teaching collocation: Future Development in the Lexical Approach* (pp. 28–46). Hove: Language Teaching Publications.
- Wu, W. (2009). The marriage between corpus-based linguistics and lexico-grammar instruction:
Using *advise*, *recommend*, and *suggest* as an example. Retrieved on November 3, 2012 from <http://www.j-let.org/~wcf/proceedings/d-072.pdf>
- Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning and Technology*, 12(2), 31–48.
- Yoon, H. & Alan, H. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13, 257–283.

Appendices

Pre-project Questionnaire

1. General information

Name_____ Faculty and major_____

Year_____ Gender : ☐ Male ☐ Female

Age ☐ under ☐ 19 ☐ 19-20 ☐ 21-22 ☐ over 22

English subject score on the Entrance examination_____

Your EL 172 Grade_____ Exempted, check here_____

2. How long have you been learning English?

☐ 1-3 ☐ 4-6 ☐ 7-9 ☐ 10-12 ☐ more than 12 years

3. How many English courses have you taken so far at TU (Including this semester)?

☐ 1-3 ☐ 4-6 ☐ 7-9 ☐ 10-12 ☐ more than 15

4. How would you rate your overall English proficiency?

☐ Poor ☐ Fair ☐ Good ☐ Very good

5. How would you rate your knowledge of computers in general?

☐ Poor ☐ Fair ☐ Good ☐ Very good

6. How often do you use computer for personal purpose (e.g. e-mail)?

☐ Seldom ☐ About once a month ☐ About once a week

☐ About 2-5 times a week ☐ Several times a day

☐ Others (please specify)_____

7. How often do you use computer for school work (e.g. writing a paper)?

☐ Seldom ☐ About once a month

☐ About once a week

☐ About 2-5 five times a week ☐ Several times a day

☐ Others (please specify)_____

8. When you use computer, do you use English or Thai?

☐ English

☐ Thai

☐ Both

10. How much of your total computer time is in English?

☐ Almost none

☐ About 25%

☐ About 50%

☐ About 75%

☐ Others (please specify) _____

11. Do you have Internet access at home?

☐ Yes

☐ No

12. Do you use a dictionary for English writing?

☐ Yes

☐ No

If yes, what kind of dictionary do you use often? (Check all that apply)

☐ Bilingual (e.g. English–Thai)

☐ Monolingual (English–English)

☐ Paper dictionary

☐ Online dictionary

☐ Electronic program dictionary (e.g. talking dictionary or installed on computer)

13. Had you heard about corpora before you took this class?

☐ Yes

☐ No

If yes, which corpora did you use? _____

Post-project Questionnaire

Name _____

Please complete this questionnaire and return it to the teacher โปรดกรอกแบบสอบถามนี้ให้ตรงกับความคิดเห็นของคุณ และส่งให้อาจารย์เมื่อทำเสร็จ

	Statement	Strongly Agree เห็นด้วย อย่างมาก	Agree เห็นด้วย	No opinion ไม่มี ความเห็น	Disagree ไม่เห็นด้วย	Strongly Disagree ไม่เห็นด้วย อย่างมาก
1	The searching technique was easy to learn. เทคนิคการสืบค้นคำ สามารถเรียนรู้ได้อย่างง่าย ๆ					
2	The corpus is more helpful than a dictionary for my English writing. คลังข้อมูล (corpus) มีประโยชน์ในการเขียน ภาษาอังกฤษมากกว่า พจนานุกรม					
3	Using the corpus is helpful for learning vocabulary. การใช้คลังข้อมูลช่วยฉันเรียนรู้คำศัพท์					
4	Using the corpus is helpful for learning collocations. การใช้คลังข้อมูลช่วยฉันเรียนรู้คำ ปรากฏร่วม (collocations)					

	Statement	Strongly Agree เห็นด้วย อย่างมาก	Agree เห็นด้วย	No opinion ไม่มี ความเห็น	Disagree ไม่เห็นด้วย	Strongly Disagree ไม่เห็นด้วย อย่างมาก
5	Using the corpus is helpful for learning grammar. การใช้คลังข้อมูลช่วยฉันเรียนรู้หลักภาษา (grammar)					
6	Using the corpus improved my English academic writing ability. การใช้คลังข้อมูลช่วยฉัน พัฒนาการเขียนเชิงวิชาการ					
7	Whenever I search for information in the corpus, I usually get the information that I need. เมื่อไรก็ตามที่ฉันค้นข้อมูลในคลังข้อมูล ฉัน มักจะได้ข้อมูลที่ ฉันต้องการ					
8	I have some difficulty in using the corpus due to limited access to computer/Internet. ความยากลำบากของฉันในการใช้คลังข้อมูล มาจาก การที่ฉันไม่ค่อยมีโอกาสใช้คอมพิวเตอร์ หรือ อินเทอร์เน็ต					

	Statement	Strongly Agree เห็นด้วย อย่างมาก	Agree เห็นด้วย	No opinion ไม่มี ความเห็น	Disagree ไม่เห็นด้วย	Strongly Disagree ไม่เห็นด้วย อย่างมาก
9	I have some difficulty in using the corpus due to time and effort spent on analyzing the data. ความยากลำบากของฉันในการใช้คลังข้อมูลเนื่องมาจากในกาวิเคราะห์ข้อมูลนั้นใช้เวลาและความพยายามมากเกินไป					
10	I have some difficulty in using the corpus due to unfamiliar vocabulary on concordance output. ความยากลำบากของฉันในการใช้คลังข้อมูลเนื่องมาจากฉันไม่รู้คำศัพท์ในตัวอย่างประโยคที่ให้					
11	I have some difficulty in using the corpus due to cut-off sentences in concordance output. ความยากลำบากของฉันในการใช้คลังข้อมูลเนื่องมาจาก ตัวอย่างภาษาไม่ได้เป็นประโยคสมบูรณ์					
12	I have some difficulty in using the corpus due to too many sentences in concordance output. ความยากลำบากของฉันในการใช้คลังข้อมูลเนื่องมาจาก ตัวอย่างภาษามีมากเกินไป					

	Statement	Strongly Agree เห็นด้วย อย่างมาก	Agree เห็นด้วย	No opinion ไม่มี ความเห็น	Disagree ไม่เห็นด้วย	Strongly Disagree ไม่เห็นด้วย อย่างมาก
13	I have some difficulty in using the corpus due to the limited number of sentences in concordance output. ความยากลำบากของฉันในการใช้คลังข้อมูลเนื่องจาก ตัวอย่างภาษามีไม่เพียงพอ					
14	I have some difficulty in performing the search technique. ฉันมีความยากลำบากในการป้อนข้อมูลเพื่อการสืบค้น					
15	The full texts in the corpus are too difficult to understand. ประโยคเต็มในคลังข้อมูลนั้นยากเกินไปที่จะเข้าใจ					
16	Whenever I have problems in English writing, I search for help in the corpus. เมื่อไรก็ตามที่ฉันมีปัญหาเกี่ยวกับการเขียนภาษาอังกฤษ ฉันมักจะใช้คลังข้อมูลช่วย					
17	I use the corpus when writing papers for other courses too. ฉันใช้คลังข้อมูลเพื่อช่วยในการเขียนภาษาอังกฤษในวิชาอื่นๆ ด้วย					

	Statement	Strongly Agree เห็นด้วย อย่างมาก	Agree เห็นด้วย	No opinion ไม่มี ความเห็น	Disagree ไม่เห็นด้วย	Strongly Disagree ไม่เห็นด้วย อย่างมาก
18	As I have learned more about the corpus, I have come to like them more. ยิ่งรู้จักคลังข้อมูลมากเท่าไร ฉันยิ่งชอบใช้มันมากเท่านั้น					
19	I will use the corpus for my English writing in the future. ฉันจะใช้คลังข้อมูลต่อไปในอนาคต					
20	Learning about the corpus has increased my confidence about writing in English. คลังข้อมูลทำให้ฉันมีความมั่นใจในการเขียนภาษาอังกฤษมากขึ้น					
21	The corpus should be introduced in all English writing courses. ควรแนะนำการใช้คลังข้อมูลในวิชาการเขียนภาษาอังกฤษทุกวิชา ให้กับผู้เรียน					
22	Overall, the corpus is a very useful resource form English writing. โดยรวมแล้ว คลังข้อมูลเป็นแหล่งข้อมูลที่มีประโยชน์สำหรับ การเขียนภาษาอังกฤษของฉัน					

Additional comments ความเห็นอื่นๆ

Written Reflections

Name_____ Date_____

Estimate time of concordancing/ เวลาที่ใช้_____ minutes/ นาที

Please describe your experiences using the web-concordancer and corpus data to find correct collocations for your draft./ โปรดบรรยายประสบการณ์ของนักศึกษาในการใช้โปรแกรมสืบค้นและการใช้ข้อมูลจากคลังภาษา เพื่อแก้ไขคำผิด นักศึกษาสามารถบรรยายเป็นภาษาไทยได้

1. What you like สิ่งที่คุณชอบ

2. What you don't like (Any difficulties?) สิ่งที่คุณไม่ชอบ หรือปัญหาที่พบ

3. How the web-concordancer has helped you to improve your English collocation learning คุณคิดว่าโปรแกรมสืบค้นนี้ช่วยให้คุณเรียนรู้ คำภาษาอังกฤษที่มักจะใช้ร่วมกัน (collocations) อย่างไร

Interview Questions for Student Participants

1. Describe your overall experiences in using corpus consultation to learn English collocations.
2. Describe how you felt about using the web-concordancer.
3. Give an example of a time you thought it was useful/ not useful.
4. What strategies did you use when you were consulting the corpus/ corpora?
5. What, if any, are the advantages of corpus consultation?
6. What, if any, are disadvantages of corpus consultation?
7. Can you think of any other ways you are getting exposed to English collocations besides concordancing?
8. How did it compare with a dictionary other tools you use to learn collocations?
9. In what way, if any, do you think it helps you to retain collocation knowledge? (Do you think concordancing help you remember collocations better?)
10. Tell me if you think you will use it in the future? Why or why not?
11. Will you recommend it to other English learners to help them improve their collocation usage? Why/Why not?
12. If you had to tell someone what concordancer is, what would you tell them?
13. Any other comments you would like to make?