

Using A Japanese-English Parallel Corpus for Teaching English Vocabulary to Beginning-Level Students

Kiyomi Chujo, Masao Utiyama, and Shinji Miura

Abstract

Although recognized by educators as a potentially useful tool, until now corpus application has been limited because the English concordance examples retrieved have been difficult for beginning-level learners to understand. The recent development of Japanese-English parallel corpus programs have exciting potential for not only eliminating this barrier, but also in facilitating inductive corpus-based language learning in ways that have not been possible in the past. This paper will present evidence, from an on-going research project, of the potential for using corpus techniques with beginning-level students in an English as a foreign language learning context. The research areas that the project investigated are the responses of the students to concordance-based teaching activities, the learning outcomes of concordance-based teaching activities in various learning contexts, and the optimum features of a computer interface between the corpus and the users.

1. Introduction

Over the last decade, there has been a growing interest in corpus applications in the classroom. While corpus use has been considered beneficial for language learning, few attempts have been made to use corpus-based activities directly in the classroom by foreign language teachers and learners in Japan other than students of linguistics (Tono 2003; Umesaki 2005) because of the difficulty students have in understanding the English concordance examples retrieved (Wible et al. 2000; Aston 2001; Thomas 2002; Tono 2003; Tian and Liu 2004). An English as a Foreign Language (EFL) learner looking at Figure 1 might easily be overwhelmed by not only the long list of examples, but by vocabulary too advanced to be useful. The recent development of Japanese-English parallel corpora has exciting potential for overcoming this problem and in turn, for facilitating inductive corpus-based language learning, i.e., data-driven learning (DDL), in ways that have not been possible in the past. The translation in the first language enables a learner to understand the target language concordance lines as shown in Figure 2, and provides a richer context, giving the learner “a better ‘feel’ for how the target language is used” (Hunston, 2002:184), and then enabling the learner to “tak[e] on the role of linguistic researchers to formulate and test their own hypotheses about language behaviour” (Gavioli, 2001:108). In

Figure 2, learners are provided with concentrated exposure to particular patterns of repetition not only in English but with a Japanese translation so that they are more easily able to extract linguistic knowledge and a cultural understanding of the use of the language.

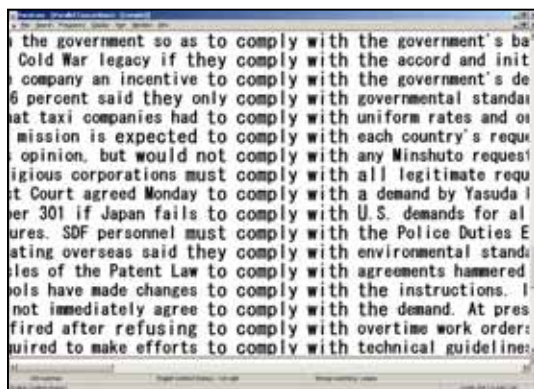


Figure 1. *Comply* in a monolingual corpus



Figure 2. *Comply* in a parallel corpus

This paper will demonstrate that combining parallel corpora usage with CALL-based activities can be very effective with beginning-level students in an EFL learning context. This research project investigates the learning outcomes and student responses to concordance-based teaching activities in various learning contexts, and explores the optimum interface features between the corpus and the user.

2. Procedure

Our team of associated researchers developed and combined a Japanese-English Yomiuri Shimbun parallel corpus (Utiyama and Isahara, 2003)¹ with “TOEIC Vocabulary 1, 2, 3” CALL teaching material (Chujo, 2003 and Chujo et al., 2004) to produce a set of corpus-based activities. These are vocabulary-based rather than grammar-based because as Gaskell and Cobb (2004:302) noted “adapting concordances for lower level learners’ grammar development is less straightforward than for lexical development.” The procedure of the study is outlined below.

2.1 Corpus

First, we created an appropriate corpus. The Japanese-English News Article Alignment Data, developed by Utiyama and Isahara (2003), is comprised of 150,000 translation pairs from the Japanese language *Yomiuri Shimbun* (having 6.1 million Japanese morphemes) and the English language *Daily Yomiuri* (having 4.9 million English words), and these pairs are automatically aligned.

This corpus is particularly useful for improving students’ TOEIC scores. For example, suppose that the target meaning of the word *file* that the teacher wants to teach is not the noun as

in ‘a paper file’, or the adverbial phrase ‘single file’ but the verb ‘to sue.’ The advantage of using a newspaper corpus as we have done is that when looking at 100 samples provided at random for *file*, 92 percent are given in the targeted or similar context in the newspaper corpus (Figure 3), compared to only a few percent from the more general-purpose language British National Corpus (Figure 4).



Figure 3. File in the newspaper corpus



Figure 4. File in the British National Corpus

2.2 Concordance program

The second step in our procedure was to find a multilingual concordancer. We chose ParaConc (Barlow, 2002) to investigate the equivalences and contrasts between two languages. ParaConc is particularly useful since it permits a wide range of investigations of aligned texts between languages. The program includes routines for highlighting potential translations, including an automatic component called ‘hot words’ which uses frequency to provide information about possible translations of the search word (see Figure 5). In the example, the candidate translations for *regulations* such as *kisei*, *kisoku* and *kanwa* are displayed in a small window. Choosing the right translations, then highlighting and sorting by use of a component called ‘KWIC/highlight’ provides sorted translations with highlights as shown in Figure 6.



Figure 5. Candidate translation of regulations



Figure 6. Highlighted translation of regulations

2.3 Target vocabulary for DDL

Since one of the goals of our CALL courses is to improve learners' TOEIC scores, our next step was to design sets of corpus-based activities to teach learners how to use a corpus to identify recurring features of various words frequently appearing in practical communication such as TOEIC, and to show them how to induce generalizations from the samples. Using 640 words from our pre-existing CALL vocabulary teaching material 'TOEIC Vocabulary 1, 2, and 3' (Chujo, 2003; Chujo, et al., 2004), we created three sets of DDL lessons totaling 32 lessons in all (10 + 10 + 12 lessons). Thirteen of these lessons (4 + 5 + 4 lessons) were implemented as a DDL classroom case study as a part of this study.² Each lesson contains activities involving 'seven or eight target words,' providing a suitable amount of DDL activities for a typical class length of 90 minutes.

In analyzing the concordances to look for linguistic features, we needed to obtain a certain quantity of instances in which such features recur more or less consistently. After examining the concordances for the 640 words, the above-mentioned seven or eight target words were chosen for each lesson because they provide adequate samples from which learners may explore the recurring between-language lexical relationship patterns, such as the equivalent of *confirm* with *kakuninsuru*, and collocation patterns such as *prior to* frequently followed by a gerund.

2.4 Tasks

In a previous study, Chujo, Utiyama, and Nishigaki (in press) showed that the level of the newspaper texts is considered "Level III (difficult)" in both English and Japanese and is suitable for advanced level learners. To date there have been no studies showing to what extent the bilingual parallel feature reduces the difficulty level in actual use. Pre-editing the concordances for learners, given the huge size of the parallel corpus, would be a daunting if not impossible task. Therefore we've tried to reduce the difficulty of the texts by controlling the difficulty of tasks, and we have taken Aston's advice (2001:43-44) for choosing and grading appropriate tasks. These are to choose: (1) easy tasks which require a relatively superficial interpretation of the data; (2) tasks which lead to precise and not misleading answers; (3) simple tasks which require one search and limited manipulation of the output in order to categorize and sort citations; (4) tasks which yield an appropriate and manageable quantity of data to be analyzed; (5) tasks which yield few or no irrelevant citations; and (6) tasks which allow learners to help and support each other.

The next step in our procedure was to therefore develop a way to introduce the learners to hands-on concordancing in the least intimidating manner, using the guidelines above. The tasks are designed to move gradually from simple to more complex exploring. Since the targeted learners' English proficiency level was at the beginning level, the majority of the tasks were

fairly easy. These tasks were classified into six types listed from the easiest (1) to the most difficult (6) and are described as:

- (1) examining Japanese equivalents of an English word
- (2) examining English equivalents of a Japanese word
- (3) collecting frequently appearing citations of an English word
- (4) translating Japanese phrases into English
- (5) examining collocation patterns
- (6) observing recurrences and inducing generalizations

These are discussed in more detail below.

2.4.1 Examining Japanese equivalents of an English word

The easiest analysis for learners is to look at Japanese equivalents for an English word. This task is used (a) to provide an introduction to hands-on concordancing, and (b) to gradually explore the one-to-many relationships between the two languages. In searching for Japanese equivalents, for instance in the *decline* concordance, learners find this English word is translated into multiple Japanese equivalents such as *geraku*, *genshou*, *suitai*, *teika*, and *ochikomi*, and in the *promote* concordance, its translations are *sokushinsuru*, *suishinsuru*, *zoushinsuru*, and *shinkousuru*. In contrast, in the *client* concordance, the usual Japanese equivalent is *kokyaku*; in the *confirm* concordance, it's *kakuninsuru*; and in the *account* concordance, it's *kouza*. These words are used with specific meanings in this newspaper corpus. From these activities, learners become aware of the one-to-many relationships between the two languages as well as the one-to-one relationships. Although it is beyond the scope of this present study, it would be interesting to compare the results of this newspaper corpus with those of a general corpus such as the British National Corpus.

2.4.2 Examining English equivalents of a Japanese word

Another easy and useful analysis for exploring the one-to-many relationships between the two languages is to examine the English equivalents of Japanese words. If we take the Japanese word *joshi*, for example, students find multiple English equivalents such as *boss*, *supervisor*, *superior*, and *subordinate* in their corpus. In the Japanese loan word *panfuretto* concordance, the English equivalents are *pamphlet* and *brochure*. In the Japanese *kuruma* concordance, majority of the English equivalents are *car* and *vehicle*. In the Japanese *kaisha* concordance, most of the English equivalents are *company* and *firm*. Since Japanese learners often tend to have a determined predilection for precise single word-for-word translations, these activities are effective for developing learner awareness of the one-to-many relationships.

2.4.3 Collecting frequently appearing citations of an English word

Identifying a variety of recurring multi-word expressions is useful for learners to understand that there are repetitive expressions in language use in two languages. This task guides them through the typical patterns and meanings of an English word with its Japanese translations in order to be able to understand and use the targeted word. An example of this would be *measure* as “an action that is intended to achieve or deal with something.” In the English concordance of *measure*, there are expressions such as *assistance measures (shien-taisaku)*, *emergency measures (oukyuu-taisaku)*, *economic measures (keizai-seisaku)*, and *economic stimulus measures (keiki-seisaku)*. Learners understand the expressions consisting of two or more words have the common Japanese equivalents *taisaku* and *seisaku*. They also observe that *measure* is used in the plural form in these examples and understand that the *taisaku* or *seisaku* are usually plural actions. A student-favorite corpus-based activity is to find examples of various vehicles and Japanese equivalents such as *armored vehicle (soukousha)*, *commercial-use vehicle (shouyousha)*, *delivery vehicle (yusousha)*, *diesel vehicle (diizerusha)*, *electric vehicle (denki-jidousha)*, and *four-wheel drive vehicle (yonrin-kudousha)*. *Vehicle* is a rather difficult word rarely used in Japanese educational settings except that it frequently appears in university exams. So this is an excellent opportunity for learners to become familiar with this essential daily life word.

2.4.4 Translating Japanese phrases into English

The next activity requires learners to translate Japanese phrases into English by identifying similar corresponding citations of the Japanese and by learning from the examples. Since “there are inherent problems in translating from one language to another because of different grammatical resources of the two languages” (Loo Siang Yen, 2002), learners may not be able to translate the meaning accurately from one language to another even if they are provided with dictionaries and grammar books. Concordances can provide more information and clearer distinctions by supplying ample citations. For beginning-level students, for example, translating *hyakuman-yen iri-no fuutou* into English is challenging. They often use translation tools on the web from which they get an incorrect translation such as ‘1,000,000 yen enter the envelope,’ or an ambiguous translation such as ‘envelope of 1,000,000 yen.’ In the Japanese *fuutou* concordance, learners are shown the English equivalent *envelope* and its various recurring multi-word expressions, including “an envelope containing ... yen.” Trying to translate a phrase and compare their hypothesis with the concordance results shows learners that language is far too complex for simple, one-to-one translations from simple computer translation programs. Students are able to see for themselves the limits of these kinds of translation programs and the value of concordancing programs.

2.4.5 Examining collocation patterns

In looking at the concordance lines and observing the recurrences of similar forms and patterns, we may find recurring collocation patterns. For example, in the *accordance* concordance shown in Figure 7, of all the 239 lines, *accordance* appeared in a collocation pattern *in accordance with* and the Japanese equivalents *ni-shitagai*, *ni-shitagatte* and *ni-motozuite*. Similar patterns are observed, e.g. *due* is mostly followed by the preposition *to* and *prior to* is frequently followed by a gerund or noun phrase.



Figure 7. *Accordance* in Japanese-English parallel corpus

2.4.6 Observing recurrences and inducing generalizations

Observing recurrences and inducing generalizations are the most advanced level task provided in these corpus-based activities. Learners are asked to find recurring semantic categories such as the occurrence of ‘a rule or system’ similar to the pattern they found earlier with *in accordance with*. One example is *approve*, which recurs in the newspaper corpus as ‘*approve a plan*’, ‘*approve a bill*’, ‘*approve a request*’, and ‘*approve the proposal*.’ Here students see that *approve* is mainly used in discussions of Japanese government matters. Another good example of this type of activity is the semantic category of *downtown*. Japanese learners tend to translate the meaning of *downtown* as *shitamachi*, which are the traditional shopping, entertainment and residential districts of Tokyo such as Ueno and Asakusa, from the misinterpretation of the direct translation of each word section, i.e., *down* translated to *shita* and *town* to *machi*. In the *downtown* concordance shown in Figure 8, learners find that the majority of Japanese equivalents mean ‘*the center of a town or city*.’ This is an effective discovery learning technique for correcting learners’ erroneous beliefs.

As Hunston noted, “With very advanced students, or students who have worked with concordance lines before, this may be a successful strategy. With other students, more detailed instructions may have to be given” (2002:174). Therefore with beginning-level learners, the task might be divided into several stages such as (1) to list the Japanese equivalents of *downtown*;

(2) to list the English citations including *downtown* which appear more than twice, and attach their Japanese equivalents; and (3) to find recurring semantic categories common to *downtown*.



Figure 8. *Downtown* in Japanese-English parallel corpus

3. Combining DDL with CALL

At first glance, CALL and DDL are opposites. The CALL vocabulary teaching material that we developed focuses on the explicit learning of word lists. In contrast, the DDL, or corpus-based, vocabulary teaching which we described above involves the inductive learning of words through the discovery of recurring patterns. CALL can teach more words than DDL, though admittedly at a more superficial level. DDL, however, is too slow for the usual class time available. Another difference is that CALL is individualized learning, and the DDL conducted in this study is pair-based so that students can help and support each other.

We believe both methods are complementary and in this study we combined the strengths of both methods. Specifically, we explored a third of the targeted words (of twenty) through DDL and then reviewed all twenty words with the CALL material. It was assumed that this system was able to provide students with both an understanding of the vocabulary and the systematic patterns that lie behind both English and Japanese words.

4. Research questions

Our research questions focus on the effectiveness of concordancing tools with beginning-level learners, and with what affect and attitude they used these in a CALL environment:

1. Did learners become accustomed to using the parallel concordance tool?
2. Was the concordancing activity useful for learners?
3. Did learners discover the one-to-many relationships between the two languages?
4. Did learners explore the lexicogrammatical and collocation patterns in English?

5. Was the concordancing tool easy to use and does it show information in a way that learners can manage easily with the minimum of instruction?

5. Experiment

5.1 Participants

Our case study was conducted in three Japanese university beginning-level English CALL courses which met for one 90-minute class per week for four or five weeks. There were 24 freshmen participants in Class 1, and 28 in Class 2. Class 3 had 20 graduate students. Each class had one female student and the rest were male. All were engineering students. They did not feel confident in their English proficiency, but they had sufficient computer skills to use the concordancing programs. The students' course objective was to use the CALL listening material and vocabulary learning material to improve their communicative proficiency.

5.2 Classroom procedure

The 90-minute course followed the procedure shown in Table 1 and the DDL activities were integrated with the CALL activities. A review test in both listening and vocabulary from the previous lesson was given in the first 15 minutes, followed by a CD-ROM listening activity for 30 minutes.

Next, the easy-to-gradually-more-difficult DDL tasks described in 2.4 were done for 25 minutes. After a 10-minute DDL introduction in Week 1, learners worked in pairs with a self-explanatory DDL handout consisting of about ten tasks for exploring the usage of seven target words. Some target words were provided with several tasks. The tasks were arranged from easy to difficult. Since progress differed from class to class, 83 tasks in total from the pre-set tasks developed for the DDL course were used during this experiment. In pairs, one student performed the concordancing and the other wrote down the results on the handout. They consulted with each other about how to perform the concordance and how to interpret the patterns and tendencies in the data they found. Each pair was required to submit the handout after filling out the answers to each task question. They also were asked to jot down their daily reaction to performing the DDL activity in order to assist us in evaluating the transitions in learner interest and ability. These evaluations are referred to as the "Daily Evaluations".

As soon as they finished the DDL activities, they had 15 minutes to use the CALL vocabulary CD-ROM to study the 20 target words which included the seven DDL targeted words. At the end of the lesson, a CALL CD-ROM quiz was given to confirm their understanding of the vocabulary learned that day.

Table 1. Incorporating DDL activities with CALL

<i>Classroom Procedure</i>	<i>Time (min.)</i>	<i>Activity</i>
Review test	15	Individual
CD-ROM Listening	30	Individual
DDL activities	25	Pair Work
CD-ROM Vocabulary	15	Individual
Quiz	5	Individual

5.3 Evaluation of the DDL activities

After completing all thirteen DDL lessons (four for Classes 2 and 3, and five DDL lessons for Class 1), students were asked to fill out a questionnaire (“Final Questionnaire”) using a five-point rating scale. They were asked if they were able to get used to using the concordancing tool and if they had gained anything from the concordancing activities. Responses were ranked from “strongly agree” (rating 5) to “strongly disagree” (rating 1). In addition to this scale, open-ended comment questions were used. The Daily Evaluations, Final Questionnaire and students’ remarks were originally written in Japanese and were later translated into English. In all the evaluations and in this article, the term “concordance tool” is used to include both the concordance program and the corpus data.

6. Results and discussion

6.1 Did learners become accustomed to using the parallel concordance tool?

Since the students were generally familiar with computers, they quickly gained confidence for doing concordance searches. To address the first research question, we had asked the learners to jot down their responses to using the concordancing tool at the end of each DDL session (Daily Evaluations). Typical feedback from Week 1 to Week 4 is shown in Table 2. We can see that the learners gradually did get used to using the tool as they learned its basic functions.

Table 2. Examples of typical remarks on concordance usage (Class 1)

<i>Session</i>	<i>Daily evaluation remarks</i>
Week 1	“It seems to take time to get accustomed to the tool.”
Week 2	“Sort and KWIC are convenient functions for looking for answers.”
	“We got used to concordancing after the first week.”
Week 3	“Changing the font size made it easier to read the results.”
	“Our searching speed improved. It’s fun.”
Week 4	“We gradually got used to the parallel concordancing tools.”
	“The tool is convenient, easy, and fun to use.”

In addition, the results of the Final Questionnaire were evaluated (see Table 3). In total, 63 students completed it. (From Class 3, we collected questionnaires from only eleven graduate students out of twenty because nine students were preparing for and participating in conferences.) Sixty-three percent (25%+38%) of students reported they became accustomed to using the tool and 40 percent (5%+35%) indicated it was easy to use. In addition, the percentage of students who had difficulty with teacher instruction, handout usage, and concordance tasks totaled 16% or less. Overall we can conclude that the learners were able to use the parallel concordance tool.

Of course there is always room for improvement, as indicated by the 24% (19%+5%) of students who reported that they were not able to fully use the tool, and 32% (22%+10%) of students who disagreed that the tool was easy to use. We'll address our modifications in sections 6.5 and 7.

Table 3. Final Questionnaire results

Statements	Evaluation				
	5 <i>strongly agree</i>	4 <i>agree</i>	3 <i>neutral</i>	2 <i>disagree</i>	1 <i>strongly disagree</i>
I got used to the tool.	16 25%	63% 24 38%	8 13%	12 19%	3 5%
The tool was easy to use.	3 5%	40% 22 35%	18 28%	14 22%	6 10%
Teacher's instruction was understandable.	24 38%	71% 21 33%	12 19%	6 10%	0 0%
Handout was easy to use.	11 17%	49% 20 32%	22 35%	9 14%	1 2%
The tasks were easy.	7 11%	46% 22 35%	26 41%	6 10%	2 3%

Note: the upper number shows the number of students who responded.

6.2 Was the concordancing activity useful for learners?

To determine if the concordancing activity was useful to learners, we looked at the evaluations from the Final Questionnaire shown in Table 4. Sixty-two percent of students (24%+38%) agreed that the DDL activity was useful for vocabulary learning, versus 17% (14%+3%) who did not. On the whole, learners responded with a positive attitude to the experience. This is reflected by the fact that only three students of 63 responded negatively in the students' remarks.³

In spite of the high percentage of students who believed the DDL activity was useful, only 36% (11%+25%) felt it was useful for vocabulary retention. Although some students discerned that the process of searching for words did help them retain words longer, overall 36% is rather low. We believe this is due to a lack of class time for learners to report their findings and share the knowledge they acquired from the concordance activities with peers. 'Students reporting on

results' (Tribble and Jones, 1997) is a necessary classroom feature for giving learners positive feedback and to allow them to consolidate their learning outcomes. This shortcoming will be addressed in future studies.

In evaluating if pair work was useful for concordancing, 51% of students (21%+30%) responded positively. More specifically, they indicated that (1) pair work was effective (21%); (2) working together allows them to help each other (21%); (3) working together was fun, and thus helped them retain vocabulary (17%); (4) they could teach each other what they didn't understand (14%); (5) they could communicate with each other (14%); (6) they made friends (7%); and, (7) they could correct and reduce mistakes (7%). Overall we felt that pair work was a useful aspect for the DDL activities. Sixteen percent (11%+5%) of students didn't value the pair work, and indicated they preferred to work individually. It should be noted that the concordancing program used requires paid licensing per user and another reason we used pairs was to reduce the cost and stay within our budget.

This research question was further addressed by the learners' responses to the open-ended Final Questionnaire question "How would you like to use this tool?" Eighty-eight percent of students answered that they'd like to use it like a dictionary to find ample examples, to clarify the meaning of words, and to translate Japanese sentences into English. Since they are already used to using online search tools such as Yahoo and Google as part of their daily routines, we believe that if the concordance tool were available in a similar way at no cost, then learners would use it.

Table 4. The evaluation for the concordancing activity

<i>Statements</i>	<i>Evaluation</i>				
	<i>5 strongly agree</i>	<i>4 agree</i>	<i>3 neutral</i>	<i>2 disagree</i>	<i>1 strongly disagree</i>
It's useful for vocabulary learning.	15 24%	62% 24 38%	13 21%	9 14%	2 3%
It's useful for retaining vocabulary.	7 11%	36% 16 25%	15 24%	18 29%	7 11%
Pair work was useful for concordancing.	13 21%	51% 19 30%	21 33%	7 11%	3 5%

Note: the upper number shows the number of students who responded.

6.3 Did learners discover the one-to-many relationships between the two languages?

We asked students the open-ended question "What did you find by using parallel corpus concordancing?" One of the most frequent findings given by students was that they noticed the one-to-many relationships between the two languages (see Table 5).

Table 5. What learners found through concordance activities

<i>What students found</i>	<i>Number of respondents (%)</i>
I found a word has many meanings (translations).	22 (37.9%)
The quantity of concordance results depends on words.	6 (10.3%)
The concordance provided ample authentic examples.	5 (8.6%)
I found this an interesting learning method.	4 (6.9%)
It's a convenient tool to find appropriate translations.	4 (6.9%)
I'd like to use this all the time.	3 (5.2%)
If it had sound, that'd be nice.	3 (5.2%)
Other	11 (19.0%)

In addition to the Final Questionnaire, we had also asked learners to jot down their responses to the concordancing activities after every DDL lesson (Daily Evaluations) so that we were able to observe the transition of students' responses. We can also see from these responses that learners attained this goal.

Table 6. Typical remarks on concordance findings (Class 1)

Week 1	"We were surprised to find a large quantity of various meanings by searching only one word."
Week 2	"We were able to find frequently used examples."
	"One Japanese word has many English translations."
Week 3	"Some words have a lot of examples and others do not."
	"Looking at both examples and their meanings is useful for learning words."
Week 4	"We can frequently find both a word's meaning and part of speech along with the word."
	"We realized there are meanings we knew and others we didn't know."

6.4 Did learners explore the lexicogrammatical and collocation patterns in English?

Students reported they had learned new words or specific information regarding the lexical and collocation relationships between words, as written in the students' reflective responses (Daily Evaluations) to each lesson. These included:

- "We were able to find frequently used lexical collocations."
- "We found that English word combinations produce various meanings."
- "We noticed that slightly different expressions are used to express similar meanings."

- “We thought it was a convenient tool for finding which parts of speech come after a word.”
- “We learned the word *due* is mostly followed by *to*.”
- “We learned that the meaning of words could be decided on in the context while carrying out the task.”

Overall, students were able to explore the collocation patterns as we had hoped, although this process taught us two important things. First, learners had difficulty distinguishing the chunks of examples and their boundaries in context. This task seemed “easy” at first but we realized that to extract examples requires knowledge of grammatical structures such as noun phrases (NP), verb phrases (VP), and prepositional phrases (PP), which are the larger chunks into which words are grouped. From another study we found that Japanese college students might not have sufficient knowledge to do this task (Uchibori, Chujo and Hasegawa, 2006). For example, in a task requiring a student to look for examples which include the word *merchandise*, the expected answers were such as *merchandise management* and *merchandise lines*. However, some students answered *yen of merchandise management* or *from them affiliates over merchandise lines*. We had unconsciously expected them to find the minimal meaningful phrases that contain the target word. We learned that we must add some examples to each task in the handout for clarification. In a subsequent lesson, for example, in the case of *currency*, we put examples such as *Asian currency* and *foreign currencies*. Another way to make this task more do-able is to use easier words for tasks such as *insurance*. Providing Japanese equivalents to *health insurance* (*kenkou-hoken*), and *medical insurance* (*iryuu-hoken*) at the beginning stage of DDL clearly shows the parameters of what is and what is not included in the examples and helps students to understand what to look for.

The second issue involved grammatical collocations, and was an interesting learning experience for us as teachers. In the first class we asked students which parts of speech or words come after *prior to*. We expected the answer to be a noun or gerund, or a noun phrase. Students’ answers were *a*, *an*, and *the*. We carefully looked at the concordance results and found that *prior to* is actually followed by articles like *the* and *an* such as *prior to the summit* and *prior to an earthquake*, and in frequency is followed by nouns such as *prior to maturity*, and then is followed by *gerunds* such as *prior to attending the summit* and *prior to making such a move*. In the second class, we revised the task question to “Which verb form comes after *prior to*?” and were able to obtain answers such as a *gerund* or *-ing*. We learned the importance of making tasks very specific so that learners don’t stray too far in their searches.

6.5 Was the concordancing tool easy to use and does it show information in a way that learners can manage easily with the minimum of instruction?

As shown in Table 3 above, 40% of students agreed (5%+35%) and 32% disagreed

(22%+10%) with the statement “The tool was easy to use”. Generally, students were able to manage easily with ten minutes of instruction in Week 1. A few students forgot how to use the tool in Week 2 or Week 3 and we had to explain it to them individually again. As a reference for students, we developed a manual for using ParaConc with a Japanese-English parallel corpus and posted it on our webpage as a free download.⁴

The general complaints from students about the concordance tool were (1) the concordancing program was unable to produce certain kanji characters so these looked garbled (for example some kanji characters such as 機, 織, 法 were replaced with a character such as “@”;⁵ and (2) students had to do the initial setting of both languages (English and Japanese) and fonts every week before they could start concordancing, and this was a nuisance.⁶ To address these issues, we decided to develop a Japanese-English bilingual concordancer on our own. Details are described below.

7. The online bilingual concordancer

Although the Paraconc concordancer used in this study was hugely beneficial for illustrating the potential for concordancing in data-driven learning, its limitations for Japanese-English applications are (1) it's expensive for classroom use with a limited budget; and (2) it cannot handle certain kanji characters since it was not designed exclusively for the Japanese language. To address these issues, we developed our own Japanese-English on-line bilingual concordancer, called *Text-Searcher*©, using the newspaper parallel corpus. We also incorporated students' feedback on optimal user interface features between the corpus and the user.⁷ This is free and available at <http://www.kotonoba.net/~snj/cgi-bin/text-search/text-search.cgi>.⁸ Figure 9 shows a screenshot of *Text-Searcher*© concordancing the term *take care of*.



Figure 9. *Take care of* in the *Text-Searcher*©

We are currently implementing various features including (1) highlighting accurate word and phrase level translations, (2) displaying the citations in KWIC format (showing the query in the center of the citations), (3) sorting the citations according to its text difficulty levels based on our study (Chujo, Utiyama and Nishigaki, in press), and (4) other features useful for vocabulary learning and ELT. Future modifications will be made to later versions as our research continues to identify effective features.

8. Conclusion

Vocabulary building is an essential component to language learning, and new developments in technology provide new tools for learners. Using high-speed computers, it is now possible to use corpora in the classroom as a means to develop vocabulary and understand language structures in real language contexts. In this study, corpus-driven language learning using Japanese-English parallel corpus allowed learners to compare the vocabulary and language patterns between two languages to not only acquire new vocabulary, but understand patterns and usage in both languages such as the one-to-many relationships between-languages and collocation patterns. One of the strengths of concordancing use was the discovery aspect students are not “taught” these language patterns but find them themselves (with guidance from handouts), and this kind of discovery can be both powerful and highly motivating. Corpus-based activities have generally been very difficult for beginning-level students and as such, this kind of tool has had limited use. However, incorporating bilingual concordancing lessons with appropriate CALL vocabulary teaching material provides a successful learning tool for these EFL students.

The activities seemed to be well received by students and although the evaluation of the learning effect in this study was mainly based on learners’ impressions, we were able to discern from their feedback that this system was both meaningful and useful. Further research will explore parallel-corpus applications to grammar in a CALL environment, and the on-going development of *Text-Searcher*©, our online bilingual concordancer.

In a nutshell, we would like to point out the importance of focusing on reference skills in the learning process in a college context. In this IT age, students need reference skills to obtain necessary information about foreign languages even after finishing their coursework. Eighty-eight percent of the learners in this study reported that they’d like to use a bilingual concordancing tool as easily and as often as a dictionary, so our long-term objective is to create an effective, accessible tool for an autonomous learning environment.

Notes

* Part of this study is based on a presentation given at the 25th JAECS Conference, April 23rd, 2005, in Kyoto, Japan, and the Fifth Foreign Language Education and Technology (FLEAT) Conference, August 10th, 2005, in Provo, Utah, USA.

1. <http://www2.nict.go.jp/jt/a132/members/mutiyama/jea/index-ja.html>
2. The 640 words and ten of these DDL lessons discussed in this section are available at <http://www5d.biglobe.ne.jp/~chujo/>.
3. They complained that the activity was time-consuming, that they weren't able to understand its value, and that they preferred to learn words with the CALL method.
4. As Anthony (2004:7) pointed out, "Strangely, few of these [concordance] programs have been designed specifically for learners in a classroom context. Rather, they have tended to be aimed at researchers." There are several barriers such as overly complex user interfaces which need to be addressed for successful classroom use. A concise manual in Japanese for using ParaConc is available at <http://www5d.biglobe.ne.jp/~chujo/>
5. After consulting with Michael Barlow, we found a solution to this problem by simply deleting "@" from the initial setting in 'search options'. For more details, please refer to our manual listed in Note 4.
6. Most of these initial settings are retained in each computer, however the setting changes are reset to its default every day in usual school computer rooms. This is why students had to do initial settings each class time. Recently we found a solution to this problem.
7. The students' requests for improvement on interface were as follows: 1) to display citations with the correct kanji; 2) to display menus in Japanese, not in English; 3) to make its operations easier and simpler; 4) to retain its initial settings on languages, fonts, and others; 5) to work SORT function properly (sometimes SORT doesn't work); 6) to design the graphical user interface into a more familiar one; and, 7) to search terms faster.
8. *Text-searcher*© was released for public use in December 2004. A growing number of people use it to search various words and phrases. The numbers of distinct hosts that accessed *Text-searcher*© from 1 January 2005 to 22 September 2005 increased from 32 to 463 per month. The number of distinct words or phrases searched by users increased from 216 (January 2005) to 3,826 (September 2005). A steady increase of those numbers indicates that users use *Text-searcher*© for their daily activities. We might conclude that *Text-searcher*© is useful for searching bilingual corpora.

Acknowledgements

Part of this study is funded by a Grant-in-aid for Scientific Research (No.17520401) from the Japan Society for the Promotion of Science and Ministry of Education, Science, Sports and

Culture. It is also partly supported by the Council for Improvement of Education through Computers (CIEC) Project.

References

- Anthony, L. (2004) "AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit." *Proceedings of IWLeL (An Interactive Workshop on Language E-Learning) 2004*: 7–13.
- Aston, G. (2001) *Learning with Corpora*. Houston: Athelstan.
- Barlow, M. (2002) ParaConc (A Concordancer for Parallel Texts) (Computer Software).
- Chujo, K. (2003) "Eigo Shokyuusha-muke TOEIC-go 1 to 2 no Sentei to Sono Kouka (Selecting "TOEIC Vocabulary 1 and 2" for Beginning-level Students and Measuring Its Effect on a Sample TOEIC Test)." *Journal of the College of Industrial Technology, Nihon University* 36: 27–42.
- Chujo, K., T. Ushida, A. Yamazaki, M. Genung, A. Uchibori, and C. Nishigaki. (2004) "Visual Basic niyoru TOEIC-you Goiryoku Yousei Software no Shisaku (3) (The Development of English CD-ROM Material to Teach Vocabulary for the TOEIC Test (Utilizing Visual Basic Part 3)." *Journal of the College of Industrial Technology, Nihon University* 37: 29–43.
- Chujo, K., M. Utiyama, and C. Nishigaki. (in press) "Towards Building a Usable Corpus Collection for the ELT Classroom." In E. Hidalgo, L. Quereda, and J. Santana (eds.) *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi.
- Gavioli, L. (2001) "The Learner as Researcher: Introducing Corpus Concordancing in the Classroom." In G. Aston (ed.) *Learning with Corpora*. Houston: Athelstan, pp. 109–137.
- Gaskell, D. and T. Cobb (2004) "Can Learners Use Concordance Feedback for Writing Errors?" *System* 32, 3: 301–319.
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Loo Siang Yen, A. (2002) "English to Japanese—to What Extent Can Translation Be Accurate?" *Translation Journal* 6, 3 (<http://accurapid.com/journal/21japanese.htm>).
- Thomas, J. (2002) "Concordancing for all." Paper presented at the Fifth Teaching and Language Corpora (TALC) Conference, Bertinoro, Italy.
- Tian, S. and L. Liu. (2004) "Course-specific Corpora in the Classroom: a News Media English Class in Taiwan." *The Journal of ASIA TEFL* 1, 1: 267–290.
- Tono, Y. (2003) "Corpus wo Eigo Kyoiku ni Ikasu [What corpora can do for language teaching]." *English Corpus Studies* 10: 249–264.
- Tribble, C. and G. Jones. (1997) *Concordances in the Classroom*. Houston: Athelstan.

- Uchibori, A., K. Chujo, and S. Hasegawa. (2006) "Toward Better Grammar Instruction: Bridging the Gap between High School Textbooks and TOEIC." *The Asian EFL Journal* 8, 2. (<http://www.asian-efl-journal.com/index.php>).
- Umesaki, A. (2005) "Eigo no Jugyou ni okeru Corpus no Riyou [Using Corpora in the ELT Classroom]." *Ritsumeikan Studies in Language and Culture* 16, 4: 115–145.
- Utiyama, M. and H. Isahara. (2003) "Reliable Measures for Aligning Japanese-English News Articles and Sentences." *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL-2003)*: 72–79.
- Wible, D., F. Chien, C. Kuo, and C. Wang. (2000) "Adjusting Corpus Searches for Learners' Level: Filtering Results for Frequency." Paper Presented at the Fourth Teaching and Language Corpora (TALC) Conference, Graz, Austria.

(Nihon University: chujo@cit.nihon-u.ac.jp)

(National Institute of Information and Communications Technology: mutiyama@nict.go.jp)

(Kobe University: snj@users.sourceforge.jp)