

**Hands-on Japanese-English Parallel Corpus Experience
in the EFL CALL Classroom**

Kiyomi Chujo
Nihon University
chujo@cit.nihon-u.ac.jp

Kathryn Oghigian
Waseda University
oghigian@gol.com

Chikako Nishigaki
Chiba University
gaki@faculty.chiba-u.jp

Abstract

Although corpus application has been considered beneficial for language learning by educators, corpus-based or data-driven language learning (DDL) has been limited because the English concordance lines retrieved have been difficult and overwhelming for beginner-level learners to understand (Tian & Liu, 2004). With the development of parallel corpus concordancing programs, the Japanese translation allows learners to understand the target language concordance lines and provides a richer context in both languages, enabling an inductive approach to understanding patterning in both languages in authentic data and to forming generalizations about language form and use. This study will present an effective way for Japanese beginner-level EFL university students to learn targeted grammar using guided DDL tasks designed for a bilingual newspaper corpus (Utiyama & Isahara, 2003) and bilingual concordancer (Barlow, 2004). Our findings will include: (1) the learning effect on specific grammatical features and structures; (2) the learning effect on TOEIC-formatted grammar questions; and, (3) learner evaluations of the hands-on corpus experience. Almost all of these evaluations showed this program was useful and effective for learning grammar basics and was favorably received by students.

Keywords: DDL, Japanese-English parallel corpus, grammar, beginner-level, EFL

Introduction

There have been interesting changes in Japanese educational reform over the last few decades, one being *yutori kyouiku* in 1978, which incrementally reduced subject content hours in an effort to reduce what was then perceived as too much academic pressure on students. Beginning at about the same time, there was a shift in English education from a traditional, grammar-based pedagogy to a more communicative one. Critics of *yutori kyouiku* complain that scholastic ability has drastically declined (Kariya, 2003), and there has been a clear decrease in students' knowledge and understanding of basic English grammar (Green, 2006). Chujo (2006) found, for example, that many beginner level college students do not recognize phrase structures and cannot conjugate verbs correctly. Only 25% could identify *a* and *the* as articles and only 20% knew that words ending in *-ly* are typically adverbs. Another study found that Japanese high school textbooks do not adequately prepare students for the grammar found on TOEIC (Test of English for International Communication) tests (Uchibori, et al, 2006). Add to this that the approach to grammar teaching seems to be shifting from traditional textbook grammar sequences, thought by some to be intuitive-based (Biber et al., 2004), to data-based grammar found in corpora. One only has to thumb through the hefty

Longman Grammar of Spoken and Written English (Biber et al., 1999) to understand that there is a clear difference between the sentence structures we think we commonly use and the sentence structures found in spoken and written corpora.

The reality for students in Japan is that at some point in their lives, they will be faced with proficiency tests such as the TOEIC, which is enjoying a surge in popularity among not only universities often using TOEIC scores as placement and annual achievement measures, but in businesses which use scores for hiring and promotion. In fact, even the Japanese Ministry of Education has announced the target scores for English teachers as TOEIC 630, TOEFL 550, and Eiken Pre-1 or higher (Tanabe, 2004). Our job as educators, then, is to ensure students can understand English grammar, and to teach them in a way that engages them in active learning and prepares them for challenges they will face on proficiency tests.

The purpose of this study has been to find an effective way to teach specific grammar, including verb forms, derivations, uncountable nouns, and most importantly, noun phrases and verb phrases, to beginner level Japanese university students.

Research questions

To date the temptation to use corpus application in the beginner-level language classroom has been tempered by the difficulty learners have in making sense of monolingual concordancing lines. With the development of parallel corpora, Japanese translation in the first language allows learners to understand the target language concordance lines and provides a richer context in both languages, enabling learners to be able “to formulate and test their own hypotheses about language behaviour” (Gavioli, 2001:108). Corpus-based, or data-driven language learning (DDL), involves the inductive learning of grammatical rules or regularities through the discovery of recurring patterns. This paper will present a case study using a DDL course for improving grammar basics, and our research questions focus on the effectiveness of the DDL lessons with beginner-level learners, and learners’ evaluation of the course.

1. Did learners improve their knowledge of grammar basics?
2. Were learners able to improve scores on TOEIC-formatted grammar questions?
3. How did learners rate the hands-on corpus experience?

Case study

Participants

This case study was conducted in a beginner-level (TOEIC 200-300) Japanese university English CALL class consisting of 21 students (15 male and 6 female), which met for one 90-minute class per week for twenty-two weeks. All were freshmen majoring in engineering and

received a total of 30 hours (20 weeks) of instruction over the course of two semesters. The first and last classes were used for administering the pre- and post- tests and questionnaires. Prior to this course, 91% of students reported that they did not feel confident in their English proficiency although 48% of the students had a favorable attitude about learning English. The students' course objective was to improve their communicative proficiency such as measured by TOEIC. In addition to this course, they took a concurrent Oral English course.

DDL target words and grammatical items

Since one of the goals of many CALL courses is to improve learners' TOEIC scores, we designed sets of corpus-based activities to teach learners how to use a corpus to identify recurring features of various words frequently appearing in TOEIC, and to show them how to induce generalizations from the samples. Targeted DDL search terms were chosen from our pre-existing web-based vocabulary teaching material 'TOEIC Vocabulary 1, and 2' (Chujo, 2003) and were incorporated into 20 DDL grammar lessons. The DDL activities were completed using the Japanese-English News Article Alignment Data (Utiyama & Isahara, 2003), which is based on a Daily Yomiuri newspaper corpus, and ParaConc (Barlow, 2004), a bilingual concordancer. From our previous studies (Chujo, Utiyama, & Miura, 2006), we learned that this Yomiuri newspaper corpus is particularly useful for providing students with TOEIC-related lexical and grammatical information.

The grammar basics targeted by the course were the basic patterns of noun phrases and verb phrases, and features and structures such as word classes, verb forms, derivations, and countable and uncountable nouns.

Classroom procedures

The classroom procedures are shown in **Table 1**. During a 90-minute class, the grammar-based DDL activities are integrated with the vocabulary-based CALL activities. For Step 1, a fifteen-minute review test for both DDL grammar and CALL vocabulary from the previous lesson is given. In the next ten-minute segment, the teacher checks the answers to the follow-up activities from the previous class and confirms their understanding of the language use.

In Step 2, learners work individually with a self-explanatory DDL worksheet consisting of several tasks for exploring the grammatical usage of five to seven targeted query words. In Step 3, students report their findings to the class and the teacher provides any necessary explanations on grammatical usage.

In Step 4, students use the CALL web-based vocabulary to study the 20 target words, which included the DDL targeted query words. At the end of the lesson, a web quiz is given to confirm their understanding of the vocabulary learned that day. At the end of the class, students are given a handout of follow-up homework activities.

Table 1 Experimental Class Procedure

Step 1: Review vocabulary test and feedback to previous follow-up activities	25 min.	Class work
Step 2: DDL activities • based on a self-explanatory DDL worksheet • students explore the usage of 7 target words • students perform the concordancing and write down the results	30 min.	Pair or Individual work
Step 3: Explicit explanations to the hypothesis formed by DDL • students report their findings to the class • teacher explains the patterns and rules they found	5 min.	Class work
Step 4: Web vocabulary & quiz • students review all 20 words (including the 7 DDL words) using the web-based program • a web quiz is given to confirm the learning • students are given follow-up activities as their homework assignment	30 min.	Individual work

DDL tasks

Since the targeted learners' English proficiency was at the beginner level, the majority of the assigned tasks were purposely made to be fairly simple. Examples of some DDL tasks are listed in **Table 2**.

Table 2 Examples of DDL tasks

DDL tasks	Examples
examining different verb forms and derivations	Search <i>develop</i> * and list both different verb forms and derivations such as <i>develop</i> , <i>develops</i> , <i>developed</i> , <i>developing</i> , and <i>development</i> .
examining countable and uncountable nouns	Search <i>furniture</i> * and <i>passenger</i> *. Uncountable nouns have only one form, e.g. <i>furniture</i> NOT <i>furnitures</i> .
examining word classes	Search a certain collocation such as <i>a * organization</i> and list the results such as <i>a new organization</i> to find what class of word comes between <i>a</i> and <i>organization</i> .
examining basic patterns of noun phrases	Search a word such as <i>quality</i> and list what words frequently come before and after.
examining the sentence position of adverbs	Search <i>canceled</i> and find <i>-ly</i> words or adverbs, and then find their sentence position. The <i>-ly</i> words are placed immediately before the main verb <i>canceled</i> .
examining basic patterns of verb phrases (1)	Search <i>enjoyed</i> and find which verb form frequently follows it: <i>to</i> -infinitives or gerunds.
examining basic patterns of verb phrases (2)	Search <i>discuss</i> . Find and write down <i>wh</i> -clauses.
examining basic patterns of verb phrases (3)	Search <i>agree</i> and count how many <i>agree that</i> phrases you can find.

Evaluation of the DDL activities

After completing all twenty DDL/CALL lessons, students were asked to fill out an online

questionnaire using a five-point rating scale from “strongly agree” (rating 5) to “strongly disagree” (rating 1). Students were also asked to take two types of post-tests, which were identical to the pre-tests. They took the pre-test during the first class in April 2007, and the post-test in the final class in December 2007. The first was a grammatical features test consisting of six groups of questions which specifically measure the learning effect on (1) word classes, (2) countable and uncountable nouns, (3) verb forms, (4) derivations, (5) noun phrases, and (6) verb phrases. The second was a 16-item TOEIC-formatted grammar test. These are discussed in detail below.

Results and Discussion

Did learners improve their knowledge of grammar basics?

In **Table 3**, the six types of questions, the pre-test scores, post-test scores, and the gain between the pre-test and post-test are shown. Since the number of questions for each item varied, each test score is expressed as a percentage. The result of a two-sample *t*-test indicates that the increase between the pre-test and post-test for five of the six grammatical features was statistically significant at the 1% level. This is shown in the *Gain* column in **Table 3** with the symbol “**”. No significant gains were made for “countable and uncountable nouns.” The details of the results are discussed below.

Table 3 Comparisons of the pre- and post-test scores on grammatical features

Grammatical features	Pre-test (%)	Post-test (%)	Gain
Word classes	69	81	12**
Countable & uncountable nouns	66	74	8
Verb forms	60	75	15**
Derivations	33	53	20**
Noun phrases	33	68	35**
Verb phrases	60	80	20**

Word classes. Students were given sentences and asked to write the name of the part of speech to which each word belongs. As a result of the DDL tasks corresponding to this grammatical feature, students were better able to recognize articles, nouns, and auxiliary verbs. However, they still had difficulty recognizing adverbs such as *very*. This finding provides us with another issue to address for our next study.

Countable and uncountable nouns. Students were given sixteen nouns and asked to decide if they were countable or uncountable. Students’ recognition of uncountable nouns such as *information*, *health*, and *darkness* improved considerably. However, the rate of correct answers for *clothing*, *cash*, and *furniture* was under 80%. These nouns were taught only in the spring semester and these scores emphasize the need to review vocabulary and key concepts

in both terms.

Verb forms. Although students were supposed to have learned verb conjugation in high school, we found that many cannot conjugate verbs correctly. We encouraged learners to observe the various verb forms in the contexts of the concordance lines to understand the different verb forms. In the pre-test, the percentage of correct answers was 60% on average, and 75% in the post-test, so we can see the improvement here. However, the percentage of correct answers for the different forms of *occur* remained at 10%. Students had difficulty spelling *occurred* and *occurring* correctly, and this is another issue to be addressed in the next study.

Derivations. Students were asked to fill in 17 derivative forms in a table of nouns, verbs, adjectives and adverbs having the same stems. Since we did a lot of tasks emphasizing derivations, students were able to write the correct adjective forms for typical adjective endings such as *financial* and *original*. However, the percentage of correct answers for adverbs such as *conceivably* remained at 33%. This is likely due to its irregular spelling (removing *-e* from *conceivable* and then adding *-ly*). We will address the derivation of adverbs in the future.

Noun phrases. According to Uchibori et al.'s 2006 study, the grammatical structures that frequently appear in TOEIC grammar sections are, in order of frequency, noun phrases (24.4%), verb phrases (13.0%), prepositional phrases (10.7%), subject-predicate relations (10.5%), and adverbs (6.5%). Thus, understanding these structures, and in particular noun phrases, is essential for improving English proficiency as measured by TOEIC.

Students were asked to underline the 32 noun phrases found in 16 sentences. The post-test scores showed an average increase of 35 points and the gain was statistically significant at the 1% level. We can say the enhanced noun phrase DDL instruction conducted in this study was effective.

Some of the types of noun phrases which students had difficulty answering correctly were: (1) noun phrases that include prepositional phrases such as *a dinner with Mr. Suzuki*, (2) noun phrases that are included in prepositional phrases such as *to his new address*, and (3) complex noun phrases that include subordinate clauses such as *salespeople who travel every week*. We will incorporate these types in our next study.

Verb phrases. Students were asked to answer forty fill-in-the-blank questions related to verb phrases such as (1) transitive and intransitive verbs, (2) gerunds, (3) *to*-infinitives, (4) past tense verbs, (5) passive voice, (6) the sentence position of adverbs, (7) *that*-clauses, and (8) *wh*-clauses. The post-test scores showed an average increase of 20 points and the gain was statistically significant at the 1% level. We can say the enhanced verb phrase DDL instruction conducted in this study was effective.

Were learners able to improve scores on TOEIC-formatted grammar questions?

Students were asked to answer sixteen fill-in-the-blank questions, which were taken from a retired TOEIC Part VI test (The Chauncey Group International, 2000). An example question is as follows, and the answer must be chosen from four choices given for each question:

We will () have a trip to China next spring. {probable, prove, probably, probability}

The answer is a certain adverb. It is taught that an adverb may be placed before a main verb, and that many adverbs end in *-ly*. In order to choose the correct answer, it is important to have an understanding of these two concepts. Compared to the other six types of features shown in **Table 3**, which focus on a single grammar issue, this particular type of TOEIC-formatted test question is complex and requires students to bring together knowledge of more than one aspect of grammar. In the example, the respondent would have to understand not only that an adverb is required, but that an adverb generally ends in *-ly* and that the adverb could appear in the space indicated.

The pre-test and post-test scores of the TOEIC-formatted test were compared and are shown in **Table 4**. The post-test scores showed an average increase of 12 points ($t = 2.431^*$, $df = 20$, $p < 0.05$)—from 39 to 51 at the conclusion of this study. The result of the two-sample *t*-test indicates that the gain between the pre-test and post-test was statistically significant at the 5% level (symbolized by “*”).

Table 4 Comparisons of the pre- and post-test scores on TOEIC-formatted grammar questions

TOEIC-formatted grammar questions	Pre-test (%)	Post-test (%)	Gain
	39	51	12*

From this observation, we see the students were rather successful in applying explicit knowledge of English grammar acquired through the DDL lessons learned in this course to the questions that appear on TOEIC. In a similar case study done the previous year, the difference between the pre-test and post-test was not significant and the gain was smaller than we had expected. In this study, we tried to give students more in-depth follow-up activities to consolidate learning. As Moran and Diniz (2005) remind us, “follow-up gives them an opportunity to practice using the target form in a meaningful way, which helps students internalize the pattern or rule they discovered in the previous steps.” To better prepare students for these kinds of test questions, more complex grammar issues can be added to the DDL program, as well as additional follow-up activities.

How did learners rate the hands-on corpus experience?

First, we asked the learners to give us feedback regarding grammar and vocabulary learning. Looking at **Table 5** we see that 62% of students agreed that the DDL activity was useful for grammar learning, versus 10% who did not. Fifty-eight percent of students agreed

that the DDL activity was useful for vocabulary learning, versus 5% who did not. On the whole, learners responded with a positive attitude to the experience. This is underscored by the fact that only one student of 20 responded with a ‘strongly disagree’ in **Table 5**.

Table 5 The evaluation for the DDL activity

Statements	Average	21 students						
		5 strongly agree	4 agree	3 neutral	2 disagree	1 strongly disagree		
It's useful for grammar learning.	3.6	10%	62%	52%	29%	5%	10%	5%
It's useful for vocabulary learning.	3.6	10%	58%	48%	38%	5%	5%	0%

Next we asked students how they specifically viewed the DDL tasks, follow-up activities and consolidation explanations. As shown in **Table 6**, 75% of students agreed and 5% disagreed with the statement “DDL tasks were useful.” Eighty-one percent of students thought “follow-up activities were useful,” and 71% of students indicated the teacher’s consolidation of the follow-up activities was useful.

Overall, we can conclude that most learners considered this DDL course material and teaching to be useful and that it helped students at the beginner level to improve their understanding of grammar.

Table 6 Students’ assessment for the DDL tasks, follow-up activities and consolidation explanations

Statements	Average	21 students						
		5 strongly agree	4 agree	3 neutral	2 disagree	1 strongly disagree		
DDL tasks were useful.	4.0	25%	75%	50%	20%	5%	5%	0%
Follow-up activities were useful.	4.1	29%	81%	52%	19%	0%	0%	0%
Consolidation of follow-up activities was useful.	4.3	57%	71%	14%	29%	0%	0%	0%

Conclusion

In this 2007 case study, learners explored grammatical structures and features in DDL concordancing lines and consolidated their understanding of both grammar and vocabulary through a CALL program and a series of follow-up activities. A comparison of pre- and post-test scores showed gains were made, although we had hoped the post-test scores would be higher. By examining areas in which gains were low or not made, we will continue to modify the course design. As Conrad (2005:404) tells us, there is a “need for more empirical studies

on the impact of using corpus materials and techniques in the classroom.” The information obtained from this case study adds to the body of knowledge for parallel concordancing use in the classroom.

Acknowledgement:

This study is funded by a Grant-in-aid for Scientific Research (No.19520515) from the Japan Society for the Promotion of Science and Ministry of Education, Science, Sports and Culture.

References

- Barlow, M. (2004). ParaConc [Computer software]. Houston: Athelstan.
<http://www.athel.com/para.html>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, S. (1999). *Longman Grammar of Spoken and Written English*. Edinburgh Gate: Pearson Education Limited.
- Biber, D., Conrad, S., Reppen, R., Byrd, H.P., Helt, M., Clark, V., Cortes, V. Csomay, E., & Urzua, A. (2004). Representing language use in the university: analysis of the TOEFL 2000 spoken and written academic language corpus. ETS TOEFL Monograph Series MS-25; RM-04-03. Princeton, New Jersey: Educational Testing Service.
<http://www.ets.org/research/researcher/RM-04-03.html> [Access date 2/10/08]
- Chujo, K. (2003). Eigo shokyuusha-muke TOEIC-go 1 to 2 no sentei to sono kouka [Selecting TOEIC vocabulary 1 & 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. (2006). Corpus ni motozuita syllabus design to sono jissen [Corpus-based syllabus design and ELT]. *International Institute of Language and Culture Studies, Ritsumeikan University*, 17 (4), 119-128.
- Chujo, K., Utiyama, M. & Miura, S. (2006). Using a Japanese-English parallel corpus for teaching English vocabulary to beginning-level students. *English Corpus Studies*, 13, 153-172.
- Conrad, S. (2005). Corpus linguistics and L2 teaching. In Eli Hinkel, Lawrence Erlbaum Associates, Inc. (Eds.) *Handbook of Research in Second Language Teaching and Learning* (393-409). Mahwah: New Jersey,
- Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with Corpora* (109-137). Houston: Athelstan.
- Green, B. (2006). A framework for teaching grammar to Japanese learners in an intensive English program. *The Language Teacher*, 30 (2), 3-11.
- Kariya, T. (2003). *Naze Kyouiku Ronsou wa Fumou-Nanoka [Why is Educational Argument Unproductive?]*. Tokyo: Chuoukouron Shinsha.

25th International Conference of English Teaching and Learning
2008 International Conference on English Instruction and Assessment

- Moran, K. & Diniz, L. (2005). Teaching with concordances. *On CALL Newsletter* 22 (2).
[Electronic version.] Retrieved March 10, 2005 from http://www.tesol.org/s_tesol/article.asp?vid=192&DID=3503&sid=1&cid=768&iid=3491&nid=3303
- Tanabe, Y. (2004). What the 2003 MEXT Action Plan proposes to teachers of English. *The Language Teacher*, 28(3), 3-8. Retrieved 9/15/2006 from <http://www.jalt-publications.org/tlt/articles/2004/03/tanabe>.
- The Chauncey Group International. (2000). *TOEIC Koushikig Guide & Mondaijshuu [The Official TOEIC Test-Preparation Guide]*. Tokyo: IIBC.
- Tian, S. & 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.
- Uchibori A., Chujo, K. & Hasegawa, S. (2006). Toward better grammar instruction: Bridging the gap between high school textbooks and TOEIC. *The Asian EFL Journal*, 8 (2).
[Electronic version.] Retrieved February 10, 2006 from <http://www.asian-efl-journal.com/index.php>
- Utiyama, M. & Isahara, H. (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.