Beginner Level EFL DDL Using a Parallel Web-Based Concordancer

Kiyomi Chujo

Nihon University, Japan

chuujou.kiyomi@nihon-u.ac.jp

Kathryn Oghigian

Waseda University, Japan

k_oghigian@aoni.waseda.jp

Chikako Nishigaki

Chiba University, Japan

gaki@faculty.chiba-u.jp

Abstract

The use of corpus-based DDL (data-driven learning) in second language classrooms is expanding, but how effective is this approach for beginner level EFL? Using a new, free, web-based Japanese-English parallel concordance tool, a Japanese-English newspaper corpus and a guided, pair-based, student-centered four-step process, this paper reports beginner level EFL student gains and collects student feedback from a semester course in grammar basics. This tool, WebParaNews, is available to the public.

Keywords: DDL, data-driven learning, corpus-based, concordancer, WebParaNews

Introduction

Corpus linguistics has been used for the last few decades as a tool for lexicographers, translators and researchers but now corpus-based DDL, or data-driven learning, is being used more and more in the L2 classroom, usually at the intermediate and advanced proficiency levels. Few studies have used DDL at the beginner level, perhaps as Boulton (2008) suggests because "current research encourages the belief that DDL is only useful for advanced learners in a computer laboratory" (p. 38). There are various issues that must be addressed when using DDL with lower level proficiency students, and it has been found that using a parallel rather than monolingual corpus, having students work in collaboratively in pairs, and providing carefully constructed guidelines for searches creates an effective environment (Chujo & Oghigian, 2012).

At Nihon University in Japan, we have been using DDL with beginner level engineering students for six years. In 2005, the first four-week pilot study was aimed at teaching vocabulary to beginner level students using a CALL program. The specific vocabulary taught was identified in a previous study as appearing in TOEIC tests but not taught in Japanese junior or senior high schools (Chujo, 2003). Students responded enthusiastically, so in the next year, the program was expanded to one semester and included grammar (primarily noun phrases (NPs) and verb phrases (VPs)), which were identified in another study as appearing in TOEIC but not in Japanese school textbooks (Uchibori, Chujo & Hasegawa, 2006). This vocabulary and grammar is relevant because Japanese students and employees typically take TOEIC tests to measure English proficiency. By beginner

level, we mean students who score 70 out of 100 on the TOEIC Bridge Test. This is similar to the TOEIC 300-350 range. The goals were vocabulary and grammar for communication; and the students used a parallel corpus tool, a bilingual newspaper corpus and carefully constructed worksheets.

In each year since 2006, the program has been run for two semesters. Minor pedagogical modifications have been made each year, and we continued to receive positive feedback from students, who made gains each year in identifying and producing NPs and VPs. In 2010, we introduced paper-based DDL to compare gains with computer-based DDL, and in 2011, we measured gains using a combination of paper and computer-based DDL. In this study, we have introduced a web-based concordancing tool and are investigating how this tool may impact learning.

Previously, the only available parallel corpus tool was ParaConc (Barlow, 2004), which is a CD-ROM-based commercial software. A free, readily available and easy to use tool for students (and teachers) to use both in and out of the classroom called WebParaNews was developed (Anthony, Chujo & Oghigian, 2011). It was built on the AntWebConc-Bilingual server framework and runs on any standard browser such as Internet Explorer, Firefox, and Opera. It works for Windows, Mac and Linux and has been designed to be used with any two languages. The corpus used was a parallel Japanese-English newspaper corpus (Utiyama & Isahara, 2003).

2012: WebParaNews Case Study

The 2012 case study was a continuation of the previous studies, but the variables were the new concordancing tool and web-based tasks using this tool. The participants were the same type of participant, but a new group of 15 freshmen engineering students. They were beginner level, with an average TOEIC score of 350. All other variables remained the same, that is: the same goal, the same syllabus, the same environment, and the same procedure. We measured student gains with pre- and post- tests and asked for student feedback.

The goal of the study was to improve basic communication, as measured through student gains in identifying and producing NPs. The syllabus is shown in Table 1. Students began with lexical-based concepts, such as identifying word classes and derivations, then learned various NPs. Students used the vocabulary taught in each previous vocabulary lesson as the DDL search terms in the subsequent lesson, thus, vocabulary was spiraled through the curriculum. Grammatical structures were grouped by category (type of NP), and the vocabulary was grouped by topic, such as business, personnel, meetings, marketing, and transportation.

Table 1 DDL Syllabus

Week	Spring Semester	
	Grammar	Vocabulary
1	Pretest	Vocabulary [1]: Business
2	Word Classes	Vocabulary [2]: Personnel 1
3	Derivations and Inflections	Vocabulary [3]: Personnel 2

4	Non-Count Nouns	Vocabulary [4]: Meetings
5	NP: Art + Adj + N	Vocabulary [5]: Marketing
…	…	…
9	NP followed by *to*-infinitives	Vocabulary [9]: Daily Life
10	NP followed by *who*, *which*, *that*	Vocabulary [10]: Transportation
11	Various NPs	
12	Posttest	

Each 90 minute class was held in a CALL classroom. Students began by studying 20 new vocabulary words with CALL program, lasting for 30 minutes. This CALL program was developed in 2007 based on the TOEIC vocabulary study mentioned earlier (Chujo, 2003). The DDL portion of the lesson was for 60 minutes and was divided into four steps. In Step 1, students explored six of the 20 vocabulary words in a specific grammar context with DDL using WebParaNews. A sample of a guided exercise is shown in Figure 1. Students, working in pairs, typed in *lawyer *ing* with sampled hits set at "5." They observed in the search results that the noun *lawyer* was followed by present participles such as *acting*, *representing*, and *working*. Students wrote down the NPs on a worksheet (see Figure 2). The students were given five additional similar tasks, completing a total of six tasks per class. In Step 2, the teacher explained the grammar so students could confirm or correct the hypotheses they made through the inductive DDL tasks. In Step 3, they did practice and consolidation as homework. Finally, in Step 4, they did production practice. Because class time is limited and there is only so much homework we can assign, we have been looking at ways to increase production during class time. One of the reasons we developed the web-based concordancer was to be able to shift some of the concordancing work to homework so more production could be done in class.

Figure 1. A screen shot for *lawyer *ing* using WebParaNews

Unit 6		Noun Phrase followed by -ing	

1）lawyer *ing を Search（Hits 5）しよう。**lawyer の前後の説明部分を書き出しましょう。**

	限定詞	名詞	後置修飾語句（説明部分）
1	a	lawyer	acting as judge
2		lawyer	
3		lawyer	
4		lawyer	
5		lawyer	

Figure 2. Sample of a Guided Student Worksheet for Identifying NPs

A pre- and post-test was given on the first and last day of the semester, respectively. The same test was used for the pre- and post-test but the order of the questions was different and the students were not given the answers at any time. The test had 45 questions and took about 30 minutes to complete. There were three types of NP questions: (1) identifying an NP using higher level (TOEIC) vocabulary; (2) understanding a complex (TOEIC-type) NP; and (3) producing a simple NP using easy (high frequency) vocabulary. Each question consisted of three NP patterns: (1) article + adjective + noun, (2) noun + prepositional phrase, and (3) noun + to / -ing / -ed.

We looked at the score difference between the three types of NP questions and applied the Wilcoxon test. The students made gains in identifying NPs and TOEIC-type NPs with a difference significant at the 1% level, but in producing NPs at the 5% level. The reason might be that we added concordance search tasks in the follow-up homework and decreased tasks for producing NP type tasks. It may be that the concordance *search* tasks at home are effective for identifying TOEIC-type noun phrases, although the decreased number of *writing* tasks for homework may have undermined student gains in production.

Table 2 Student Scores (%) for the Three Types of NP Questions

	Pre-test Mean（SD）	Post-test Mean（SD）	Gain
NP1: Identifying NP	46.2（20.2）	81.8（13.2）	35.6**
NP2: TOEIC-type NP	45.3（ 9.8）	61.3（ 8.8）	16.0**
NP3: Producing NP	57.6（21.9）	66.7（20.9）	9.1*

** $p < .01$ *$p < .05$

Table 3 shows the student feedback regarding WebParaNews. We can see that students rated the tool quite highly for accessibility, intuitive use and screen design. Speed got a reasonable mean and very few students experienced problems. These results are quite different from the feedback we received from previous students using Paraconc. Paraconc required new settings each time it was used, it sometimes garbled Japanese characters, and because it was designed for researchers, not language learners, it was difficult for students initially to learn to use. Also, for copyright reasons it could only be used in tightly controlled classrooms. Because WebParaNews is free and online, it is accessible anywhere, so we were able to assign DDL-based homework.

Table 3 Student Responses to WebParaNews

Feature: 1 (strongly disagree) - 5 (strongly agree)	Mean	SD
Easy to access	4.5	0.9
Intuitive	4.0	0.8
Screen design	4.0	0.9
Speed	3.6	1.1
Experience problems	1.9	1.1

Conclusion

DDL can be effective at the beginner level in teaching basic grammar such as noun phrases when used in conjunction with a parallel corpus, collaborative pair work and guided worksheets. Students in this study showed significant gains using WebParaNews and responded positively to this new parallel web-based concordancer and web-based tasks. We have made the parallel website available to the public with a bilingual newspaper corpus from http://www.antlab.waseda.ac.jp/webparanews/.

References

Anthony, L., K. Chujo, & K. Oghigian (2011). 'A novel, web-based, parallel concordancer for use in the ESL/EFL classroom' in J. Newman, H. Baayen & S. Rice (eds.) (2011), *Corpus-based Studies in Language Use, Language Learning, and Language Documentation.* Amsterdam/New York: Rodopi Press.

Barlow, M. (2004). ParaConc [Computer software]. Houston: Athelstan. http://www.athel.com/mono.html

Boulton, A. (2008). 'DDL: Reaching the parts other teaching can't reach?' in A. Frankenberg-Garcia, I. Rkibi, M. R. Cruz, R. Carvalho, C. Direito, & D. Santos-Rosa (eds.)(2008), Proceedings of TaLC 8 – Lisbon, 8th Teaching and Language Corpora Conference. Lisbon, Portugal: Associação de Estudos e de Investigação Cientifíca do ISLA-Lisbo Fichier.

Chujo, K. (2003). 'Eigo shokyuusha-muke TOEIC-goi 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., & K. Oghigian (2012). 'DDL for EFL beginners: A report on student gains and views on paper-based concordancing and the role of L1' in J. Thomas & A. Boulton (eds.) (2012), *Input, Process and Product: Developments in Teaching and Language Corpora.* Brno, Czech Republic: Masaryk University Press.

Uchibori A., K. Chujo, & S. Hasegawa (2006). 'Toward better grammar instruction: Bridging the gap between high school textbooks and TOEIC.' *The Asian EFL Journal, 8*(2), 228-253.

Utiyama, M., & H. Isahara (2003). 'Reliable measures for aligning Japanese-English news articles and sentences' in E. Hinrichs & D. Roth (eds.) (2003), *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*. Sapporo: The Sapporo Convention Center.