Teaching Practice

Improving Student Writing with Paper-Based and Computer-Based Text Analysis

Kathryn Oghigian
Waseda University, Japan

Kiyomi Chujo
Nihon University, Japan

Abstract

In this study, mixed proficiency level classes of science and engineering university students performed both paper-based and computer-based text analysis using corpora to improve specific language forms such as noun, verb, and prepositional phrases, statements of intent, and results reporting in the production of a 2000-word academic research paper. Pre- and post-language samples for noun phrases and student feedback were analyzed to determine the effectiveness of this approach. Results were grouped into advanced and intermediate level students. Despite 5% and 8% overall gains respectively, more individual gains were seen with the advanced level students. All student responses to questionnaires regarding the usefulness of corpora and the various corpus tasks were positive. While advanced level students seemed to benefit the most from the corpus tasks, 90% of students reported they would use corpus analysis in the future, and 83% felt their writing had improved; thus, intermediate level students also benefited.

Corpus-Based Exercises in the L2 Classroom

Corpus-based text analysis has been shown to benefit L2 students in many ways. The learner controls the learning process (Braun, 2005; Huang, 2008), inductive thinking is encouraged (Johns, 1991), and there is a virtually limitless supply of data (Conrad, 2000). Since corpus-based text analysis is text-oriented and uses lexical patterns, it naturally lends itself to reading and writing (Flowerdew, 2002). Because grammar and vocabulary are interrelated (Sinclair, 1991), it is possible to clearly see common patterns and frequency of language use (Biber & Conrad, 2001). Yoon and Hirvela (2004) report that corpus analysis is increasing for English for Specific Purposes (ESP) courses in particular, since authentic texts provide specialized word patterns. Learners are able to see technical words in context, commonly occurring phrases and language chunks, and, as Yoon and Hirvela (2004, citing Odlin, 2001) noted, "where to put words into sentences."

In addition, various studies using corpus analysis in the L2 classroom report on particular language objectives. A sampling includes academic English vocabulary (Thurston & Candlin,

1998), the overuse of logical connectors (Milton & Tsang, 1993), basic grammatical structures such as noun and verb phrases (Chujo & Oghigian, 2008) and ESL university-level writing (Yoon & Hirvela, 2004). With the exception of Chujo and Oghigian, these studies have been conducted with intermediate or advanced level learners. In fact, there are very few studies at the beginner level (Boulton, 2008) or studies that incorporate data driven learning (DDL) in a class comprised of a range of levels. (For an excellent literature review focused on writing and student attitudes toward corpus use, see Yoon & Hirvela, 2004.)

The purpose of this study is to investigate the use of text analysis as an aid to EFL technical science and engineering research paper writing in a class of mixed proficiency level university students. The approach is task-based and employs paper-based and computer-based concordancing as well as text analyses of sample journal articles. To determine if the corpus activities had an impact on learning, an analysis of noun phrases was done on pre- and post-course writing samples. Feedback from students on end-of-term questionnaires was also collected and analyzed.

Case Study

Technical Writing

Technical Writing 1 (TW1) and Technical Writing 2 (TW2) are one-semester elective English classes in a science and engineering university faculty. The goal of TW1 is for students to produce a 2,000-word research paper on a topic related to their fields, which is IEEE (Institute of Electrical and Electronics Engineers) -cited and referenced. The goal of TW2 is to write a research paper based on primary-sourced data on a topic in their fields that is formatted to a relevant journal identified by each student. These are the first writing-focused courses offered in the English program, although students in the first year write lecture summaries, and second year students do a collaborative written research project. TW1 is not a prerequisite for TW2, but it is highly recommended.

Participants

Twenty-four students enrolled in TW2 in the fall of 2011 participated in this study. In addition to varied test scores (self-reported scores of 375-975 on the Test of English for International Communication [TOEIC]), students' listening, speaking, reading, and writing abilities also greatly varied, based on teacher observation and homework assessments. All were third- or fourth-year undergraduate students. Student majors included applied mathematics, civil and environmental engineering, medical bioscience, electrical engineering, chemistry, and applied physics. Weekly classes met for 90 minutes for 15 weeks, which comprised one semester. Ten classes were held in a regular classroom and five classes were held in a computer classroom, as dictated by each weekly objective.

Corpora and Corpus Tools

In a computer room in Week 2, students were shown how to access and use three online corpora: the Corpus of Contemporary American English (COCA) (http://corpus.byu.edu/coca/), Springer Exemplar (http://www.springerexemplar.com/), and the Professional English Research Consortium (PERC) corpus (http://scn.jkn21.com/~percinfo/). All three are both corpora and corpus tools. Although COCA is slightly more complex to use, sample concordances are easier to understand for lower proficiency level students. It is also possible to choose only academic sources for concordance lines. The user interfaces for Exemplar and PERC are very simple; however, the corpora used are taken from journals and professional books and are therefore at an advanced level.

In addition, students were shown how to download and use a corpus tool, Antconc (http://www.antlab.sci.waseda.ac.jp/software.html), and how to create their own corpora by converting sections from relevant journal articles into text files and uploading these into Antconc. They were also given a sample corpus in which the files were separated into title, abstract, introduction, method, results and discussion, conclusion, acknowledgements, and biography sections so that searches could be more specifically targeted. Creating their own corpora with a goal of 20 to 50 text files per section was encouraged and students were given class time and guidance to do so. More advanced students were able to create their own corpora so that searches would be more relevant to their fields; less advanced students had the option of using the provided corpus or the online corpora described earlier. By giving students a hands-on introduction to four different tools and opportunities for guided, supervised trial tasks, they had the option to employ the tool they were most comfortable using. This first computer class was supplemented by additional "gradual and guided" (Kennedy & Miceli, 2001) lessons, so students received demonstrations and supervision throughout the course.

Tasks

Corpus searches were done in class and as homework. As in-class group work in a computer classroom, students followed the teacher in doing specific types of searches on a large demonstration screen and the findings were discussed as a group. Subsequently, they did similar individual searches using the corpus of their choice. Corpus work included paper-based non-concordance type text analysis (an examination of texts printed on paper, with no direct computer interaction), computer-based concordancing (an examination of various concordance lines found by typing searches directly into a computer program), and a combination of paper- and computer-based concordancing (viewing paper-based concordance lines and typing additional searches directly into a computer program). In all cases, students did a text analysis, made assumptions about the lexical or grammatical target feature, and then produced practice sentences using their own topic-related technical vocabulary. They were encouraged to search various language forms and produce sentences that would be directly applicable to their final research papers.

Text Analysis of Journal Articles

Non-concordance type text analysis was done with a sample journal article from a journal in their fields provided by each student, and a sample article provided by the teacher. The former was used to understand formatting and organization. The provided article ("the Tesla paper") was examined in detail to analyze various writing aspects such as purpose and style, and specific aspects such as sentence structure (noun phrases [NPs], verb phrases [VPs], prepositional phrases [PPs] and clauses), logical connectors, referents, and hedging.

Computer-Based Concordancing

In one assignment, students were given a paper-based example of how to find common NPs for technical words using COCA and the "clusters" function in Antconc, and were then asked to search ten keywords related to their topics and write down common NPs. In a related task, they were asked to write practice sentences using these NPs with particular attention to correct articles. In another task, students were asked to search a corpus for *clearly [v*]* and note common verb phrases; they then wrote practice sentences using verbs from the Academic Word List (Coxhead, 2011). See Figure 1.

verb phrases

A **verb phrase** (VP) is the verb and all the words around it such as adverbs and function words, for example: *is, seems to be, would likely indicate, clearly has been shown, is highly unlikely, is critically important.* Notice that the verb phrase includes all the modifiers (adverbs) and function words (articles, pronouns, conjunctions and prepositions) connected to it. Sometimes a verb phrase (e.g., gerund or infinitive) can take the position of a subject or an object. In these example sentences below, the main verb phrase is in **bold** and verb phrases are in *italics*. The subjects are underlined. Even though these underlined phrases act as subjects, they are also verb phrases.

Replacing the water channel with a copper wire **yielded nearly** the same output. [1]

Maintaining a stable temperature **is critically important.**

To maintain a stable temperature **is critical.**

It **is important** *to write* in a formal, academic style.

An important point **is** *knowing how to paraphrase.*

PRACTICE 1. Search *clearly [v*]* in the COCA corpus or *clearly* in your own corpus with Antconc. Find four VPs using this adverb and write them below.

1. *clearly defined*   3. *clearly shows*

2. *clearly stated*   4. *clearly demonstrated*

PRACTICE 2. Think of three AWL verbs that you might use in your paper. Check a corpus to see what common VPs are used with these verbs. Write sentences using these VPs.

Example: *calculate/ were calculated using/ The results were calculated using Equation 1.*

1. *illustrate / dramatically illustrate / These results dramatically illustrate that u/P rate affects gene transfer*

2. *indicate / not necessarily indicate / This data does not necessarily indicate that gene transfer rate is zero percent.*

3. *predict / can accurately predict / From these results we can accurately predict that fibronectin has something to do with cell adhesion.*

Figure 1. A student example of sentences produced from searching verb phrases in a corpus.

Several computer-based searches were also done in class. Some searches were demonstrated on a large screen and discussed as a group; other searches were done individually, with students choosing the corpus. Examples (not shown) for discussing how to write the method section included prepositional phrases such as *at, by, during, from, into, on, through, to, under,* and *via,* and the terms *figure, Fig.,* and *Table.*

Combined Paper- and Computer-Based Concordancing

Two tasks used combined paper- and computer-based concordancing: statements of intent and titles. For the statement of intent, students were shown paper-based concordance lines on a large classroom screen and on a handout for *the purpose of this* and *In this paper,* which were taken as a screen shot from COCA academic sources. As a group, these were examined to understand how the purpose of a paper could be expressed. For homework, students were

asked to use Antconc and the introduction files from their own corpora or any corpora of their choosing to search *purpose, aim, objective, goal,* and *paper*. Next, they were asked to look specifically at results for *in this paper* and write down subjects or verbs commonly following this phrase such as *in this paper, we briefly describe*, or *in this paper, we address*. Finally, they were asked to write statements of intent for their papers using both *the purpose of this paper* and *in this paper*. A student example is shown in Figure 2.

Figure 2. An example of student work for defining a statement of intent.

For addressing titles, students were given a list of previous student research paper titles. In class, they were asked to identify hanging, NP + PP and VP styles and to find and highlight at least five NPs + PPs, and at least five VPs. As homework, using either their own corpora or the mini-corpora in Antconc, they were next asked to look at the titles in the "file view" function

and note any conclusions about the general style of the titles. In the next exercise, they were asked to search *tion and *ing with the "concordancing" function and note any common NPs or VPs. Finally, they were asked to write three titles for their papers, one using each style. A student example is shown in Figure 3.

titles

PRACTICE 1
Look at the attached list of titles for 2009 TW papers written by Waseda science and engineering students.
1. Identify three that use a hanging style. Circle or highlight them on your list of titles.
2. Identify the noun phrases and prepositional phrases in at least five titles. Underline or highlight these in a different colour.
3. Identify at least five verb phrases. Place a box around these or highlight in another colour.

PRACTICE 2
1. Load the title files from your own corpus into Anctonc. First look at the titles in FILE VIEW. Can you make any conclusions about the general style of the titles?

Most of them are in NP+PP style

2. Using the CONCORDANCE tool, search *ing. What are some common VPs used in titles? List at least three here.

considering, monitoring, using

3. Using the CONCORDANCE tool, search *tion to find common NP titles. What are some common NPs used in titles? List at least three here.

solution, distribution, generation

PRACTICE 3
Write at least three possible titles for your research paper. Use each style (hanging, NP+PP, VP + PP).

1. An Investigation of Technologies Needed to Establish Smart Cities from a Power Systems Engineering Standpoint

2. Investigating Technologies Needed to Establish Smart Cities from a Power Systems Engineering Standpoint

3. What Technologies are Used in Smart Cities: An Investigation from a Standpoint of Power Systems Engineering

Figure 3. A student example of producing titles based on data from a corpus search.

Data Collection

Of 24 participants, 15 were identified as having previously taken TW1. Because most TW1 students would have been given instruction on NPs, to have more comparable data, the NPs in pre- and post-samples were limited to these 15 students. The first 250 words from the results section from their previous TW1 papers and their final TW2 papers were analyzed for the percentage of correct NPs from the total NPs. The results section was chosen because students generally received a great deal of teacher feedback for the abstract, introduction, and method sections. An NP was incorrect if there was any error and was counted as incorrect only once even if it contained multiple errors. Examples of correct and incorrect NPs are shown in Table 1. Feedback on the use of corpus analysis was also collected from all 24 students, using an anonymous Likert-scale questionnaire given on the final day of the term.

Table 1

Examples of Correct and Incorrect Noun Phrases

	NP Pattern	Incorrect Example	Correct Example
1	[art + adj + n]	[a central ideas]	[the central idea]
2	[prep + art + n]	[in the slope]	[on the slope]
3	[prep + n] [prep + n]	[in the end] [of the year]	[at the end] [of the year]
4	[n] [conj + n]	[Table 1][and table 2]	[Table 1] [and Table 2]
5	wrong word (ww)	[these problems] . . . [it]	[these problems] . . . [they]
6	word form (wf)	[the necessarily]	[the necessity]

Results and Discussion

Analysis of Noun Phrases

Because the range of TOEIC scores was so broad (375-975), the 15 students were divided into two groups: A (700-975) and B (375-699). The results in Table 2 show that the more advanced students in Group A had a very high number of correct NPs (an average of 93%). All students except one made gains, with one student having a 23% gain. Student 6 (S6) had a net loss of 12%. His TW2 paper used very advanced level language compared to his TW1 paper, and was based on an actual lab experiment. Although he had fewer correct NPs, he submitted a highly technical paper with complex journal formatting (Vancouver style). Overall, the average gain for Group A was 5%.

Table 2

Results for NP Analyses of Pre- and Post-Writing Samples from Group A

	Pre-Writing Sample			Post-Writing Sample			Gains
	Correct NPs	Total NPs	% Correct	Correct NPs	Total NPs	% Correct	
S1	53	79	67%	53	59	**90%**	**+23%**
S2	53	60	88%	68	71	**96%**	**+8%**
S3	58	65	89%	61	65	**94%**	**+5%**
S4	64	73	87%	70	76	**92%**	**+5%**
S5	72	73	99%	61	61	**100%**	**+1%**
S6	67	68	99%	69	79	87%	-12%
Average	61.2	69.7	88%	63.7	68.5	93%	+5%

Table 3 shows that the results are mixed for the intermediate level group (Group B). Although the overall average is +8%, five of nine students had losses or a 0% gain. Two students (S14 and S15) had 17% gains, and another (S13) had an 8% gain. The students who showed losses also had poor listening skills and may not have been able to follow all instructions. For example, one student did not understand that she was looking for articles and NPs; she noted only nouns and her sentences omitted articles. The fact that many of these lower level proficiency students were not as successful in improving NP usage suggests that the corpus work may not necessarily have been as useful or appropriate for them; this supposition is supported by the fact that there are many more studies at the advanced level than at lower levels (Boulton, 2008).

Table 3
Results for NP Analyses of Pre- and Post-Writing Samples from Group B

	Pre-Writing Sample			Post-Writing Sample			Gains
	Correct NPs	Total NPs	% Correct	Correct NPs	Total NPs	% Correct	
S7	59	87	68%	44	75	59%	-9%
S8	54	80	68%	48	82	59%	-9%
S9	60	73	82%	52	70	74%	-8%
S10	63	78	81%	55	73	75%	-6%
S11	59	76	78%	70	90	78%	0%
S12	73	112	65%	49	73	**67%**	**+2%**
S13	63	82	77%	70	82	**85%**	**+8%**
S14	52	72	72%	58	65	**89%**	**+17%**
S15	55	85	65%	59	72	**82%**	**+17%**
Average	59.8	82.8	73%	61.6	75.8	81%	**+8%**

Analysis of Questionnaires

Previous studies looking at student attitudes toward the use of corpus linguistics in the L2 classroom show overall positive responses (Chujo & Oghigian, 2008; Thurston & Candlin, 1998; Tribble, 2002); this study is no exception. Of 24 questionnaire responses, an overwhelming majority was positive (*yes* and *mostly* responses and *not really* and *no* responses were combined to calculate percentages). Students reported that they had improved their writing skills (83%) and that the following tasks were useful: learning to use corpora and corpus tools (80%), text analysis of the Tesla paper (75%), looking at paper-based concordance lines (83%), looking at computer-based concordance lines (88%), looking at paper-based statements of intent (75%), searching a corpus for technical words (80%), searching for NPs (80%) and VPs (83%) in a corpus, and searching a corpus for aspects of titles (80%). Encouragingly, 90% of students indicated they would use a corpus in the future to improve their writing. Only 67% reported that they had explored a corpus beyond the assigned tasks and 63% felt that creating their own corpora was useful. A little more than half of the students (58%) expressed a preference for doing the corpus searches in class rather than at home. For the open-ended questions, six students reported that using corpora was the "best aspect of the course" and two reported this was the "most difficult aspect." There were no significant differences in questionnaire responses between the advanced level (Group A) and the intermediate level (Group B), nor between the students who had and who had not taken TW1.

Table 4

Questionnaire Results

	Yes	Mostly	So-so	Not really	No
1. I improved my writing skills.	12	8	4	0	0
2. Learning how to use corpora and corpus tools was useful.	11	8	5	0	0
3. Doing text analysis on the Tesla paper (such as examining logical connectors and verb phrases) was useful.	10	8	3	3	0
4. Looking a paper-based concordance lines (such as a list of titles or list of statements of intent) was useful.	6	14	3	1	0
5. Looking at computer-based concordance lines (such as searching for "Figure" and "Fig." or for articles for NPs) was useful.	9	12	2	1	0
6. Looking at a corpus of statements of intent helped me to write my statement of intent.	10	8	5	1	0
7. Searching a corpus for technical words in my field was useful.	8	11	5	0	0
8. Searching a corpus for NPs using technical words in my field and writing practice sentences was useful.	10	9	5	0	0
9. Searching a corpus for VPs using technical words in my field and writing practice sentences was useful.	13	7	3	1	0
10. Looking at a list of titles and searching a corpus for aspects of titles helped me to write my title.	8	11	4	1	0
11. Creating my own corpus was useful.	8	7	6	3	0
12. I prefer to do corpus searches in class and not at home.	5	9	5	4	1
13. I used a corpus to look for additional writing aspects I was curious about.	6	10	6	1	1
14. I might use a corpus in the future to improve my writing.	11	9	3	1	0

Conclusion

This study is a follow-up to a similar study conducted the previous year (Oghigian & Chujo, 2012), which found that the use of corpora overall was generally successful based on NP gains and questionnaire responses. However, the data was taken from a smaller sample and students indicated a preference for paper-based tasks rather than computer-based tasks. In that study, students were taught to use corpora at the beginning of the term, but had no additional instruction or in-class supervision; it is not surprising that they preferred paper-based concordance lines. In response, a far greater number of in-class computer-based demonstrations were done in this current study. Numerous paper-based examples were provided as well. Kennedy & Miceli (2001) refer to these guided hands-on tasks as an "apprenticeship approach" and in their study noted a positive overall response from students. They also noted this approach is time consuming; however, teaching students this important skill is well worth the effort. Just as there is value in teaching students how to use a dictionary or thesaurus, or even tools such as word processors and computers, so is there value in teaching corpus analysis. Although the sample for this study is also small, it is believed this approach could be adopted for English for Academic Purposes (EAP) or ESP courses, since students can build their own specific corpora for use with Antconc.

Author Note

Kathryn Oghigian, Center for English Language Education in Science and Engineering (CELESE), Waseda University, Tokyo, Japan, and Kiyomi Chujo, College of Industrial Technology, Nihon University, Chiba, Japan.

Correspondence concerning this article should be addressed to Kathryn Oghigian, CELESE, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan, and Kiyomi Chujo, College of Industrial Technology, Nihon University, 2-11-1 Shin'ei, Narashino-shi, Chiba 275-8576, Japan. E-mail: k_oghigian@aoni.waseda.jp, chuujou.kiyomi@nihon-u.ac.jp

References

Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly, 35*(2), 331-336. http://dx.doi.org/10.2307/3587653

Boulton, A. (2008, March). Evaluating corpus use in language learning: State of play and future directions. Paper presented at the American Association of Corpus Linguistics, Provo, Utah.

Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL, 17*(1), 47-64. http://dx.doi.org/10.1017/S0958344005000510

Chujo, K., & Oghigian, K. (2008). A DDL approach to learning noun and verb phrases in the beginner level EFL classroom. In A. Frankenberg-Garcia, I. Rkibi, M. R. Cruz, R. Carvalho, C. Direito, & D. Santos-Rosa (Eds.), Proceedings of TaLC 8 – Lisbon, 8th Teaching and Language Corpora Conference (pp. 65-71). Lisbon, Portugal: Associação de Estudos e de Investigação Cientifíca do ISLA-Lisbo Fichier.

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, *34*(3), 548-560. http://dx.doi.org/10.2307/3587743

Coxhead, A. (2011). The Academic Word List 10 years on: Research and teaching implications. *TESOL Quarterly*, *45*(2), 355-362. http://dx.doi.org/10.5054/tq.2011.254528

Flowerdew, L. (2002). Corpus-based analyses in EAP. In J. Flowerdew (Ed.), *Academic discourse* (pp. 95-114). Harlow, England: Longman.

Huang, L.-S. (2008). Using guided, corpus-aided discovery to generate active learning. *English Teaching Forum, 46*(4), 20-27. Available at http://exchanges.state.gov/englishteaching/forum/archives/docs/08-46-4-d.pdf

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom Concordancing: English Language Research Journal*, *Volume 4* (pp. 1-16). Birmingham, England: University of Birmingham Centre for English Language Studies.

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology, 5*(3), 77-90. Available at http://llt.msu.edu/vol5num3/kennedy/default.html

Milton, J., & Tsang, E. S. C. (1993). A corpus-based study of logical connectors in EFL students' writing: Directions for future research. In R. Pemberton & E. S. C. Tsang (Eds.), *Studies in lexis* (pp. 215-246). Hong Kong: University of Science and Technology Language Centre.

Oghigian, K., & Chujo, K. (2012). Corpus-informed writing for science and engineering. *Journal of the College of Industrial Technology, Nihon University, 45,* 1-13.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, England: Oxford University Press.

Thurston, J., & Candlin, C. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes, 17*(3), 267-280. http://dx.doi.org/10.1016/S0889-4906(97)00013-6

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131-149). Harlow, England: Longman.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257-283. http://dx.doi.org/10.1016/j.jslw.2004.06.002