

Using Corpus Linguistics to Teach ESL

Drew S. Fagan

San Diego State University

Proceedings of the CATESOL State Conference, 2005

Using Corpus Linguistics to Teach ESL

ESL instructors strive to bring authentic English interaction into their classrooms to help their students prepare for activity in an English-speaking world. However, when it comes to selecting those materials to help reinforce lesson objectives, teachers may not be satisfied with the appropriateness of the texts chosen to represent authentic English. Another problem that persists is once the teacher is satisfied with the authentic text, how can materials be made to best satisfy the students' varied needs?

The presenters for this poster session presented a corpus-based solution to these problems. Namely, a corpus is a group of authentic texts put together ranging from a couple hundred words to a couple million (for more information on corpus linguistics, refer to Biber, Conrad, and Reppen, 1998). Some sources that have come out of corpus linguistics work have included grammar nooks (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Other sources have focused on how corpus use interacts with applied linguistics and language pedagogy (Hunston, 2002).

The presenters' use of corpus in the classroom, though, is related to helping the individual instructor produce his or her own corpus that matches the course's objectives. This solution is in the form of Paul Nation's vocabulary program (Nation, 2002). This free-to-download program is easily used with all types of texts. The purpose of the program is to allow instructors to gather their own sets of authentic English texts (their "corpus") that complement the topic, theme, grammatical point, or textbook being used and see which words are most prevalent in those authentic texts. The program enables teachers to run up to 32 different texts of varying size at the same time. After downloading the program, there are easy-to-follow instructions that show the teacher how to use the program (refer to page 1 of the instructions manual)

To make it easier to understand these directions, it is important to realize that teachers need to simply save the texts that they want to use as a ".txt" before running them through the program. If the texts are from internet sources, it is simple to copy and paste them into a Word Document and save them as ".txt". If the source(s) are from books, articles, or other materials not found online, the instructor can easily scan the sources onto the computer and then follow the same directions as listed above. Finding

the texts and saving them are the most time-consuming parts of the process. However, running the texts through the program takes only a few seconds and your results are given almost immediately.

At this point in the process of making materials using the teacher's own authentic corpus, there are two main functions with which the program can serve. One function focuses on looking at the word count of all the texts combined to get a more general understanding of how the text, or genre, is being used. Another function is to compare the word counts between the texts to determine which texts have the language which is more suitable to the needs of the students. These two functions are exemplified in the following two scenarios of activities developed by the presenters of this presentation for their own classes.

Scenario 1: Teaching résumé writing to advanced ESL university students

Louhiala-Salminen (1996) explains that many textbooks used in, for example, the business ESP classroom are not providing students with accurate preparation for future professional careers. The textbooks' uses of English are either not authentic or their authenticity is not clearly explained. With this in mind, Teacher 1 decided to incorporate his own corpus of authentic résumés from professionals in business to teach his ESL students how to construct these essential texts for future use. He asked the students to get business professionals' résumés. Thirty résumés were gathered electronically and saved as ".txt" files. Since he wanted to see general patterns for the words used in this genre, the teacher clicked on the "Frequency" file (listed next to the "Instructions") on the opening screen. He then ran all thirty résumés through the program and got his results in the format shown in Figure 1. This figure on the following page is an example of the page that teachers will see once they run their texts through the program. "Total tokens" means the total number of words in all the résumés put together. "Total types" are all the different words in the résumés counted only once. The order of words shown under "Word Type" is the order of the most commonly-used words. "And" in this corpus is the most commonly-used word, being used 683 out of the total 12,415 words (tokens).

Outfile file was: A:\Total Nation Info 30 resumes.txt			
Total tokens:	12415		
Total types:	3760		
... in frequency order ...			
Word Type	Rank	Frequency	Cumulative Percent
AND	1	683	5.50
OF	2	398	8.70
TO	3	238	10.61
THE	4	220	12.38
FOR	5	210	14.08
IN	6	185	15.57
A	7	165	16.89
BUSINESS	8	140	18.02

Figure 1. “Frequency” results screen.

Once the counts were found, the teacher decided to focus on the use of verbs that describe job duties in résumés; thus, he went through the findings and focused solely on verbs. Some examples of the most commonly used verbs from the authentic corpus were: ACHIEVED, ESTABLISHED, DRAFTED, DESIGNED, COLLECTED, and CREATED. However, rather than just telling the students to use these words, he wanted them to understand how to use them in context. Refer to Appendix A for activities describing how the teacher had his students think about the context in which their résumés would be written and how the verbs found in the authentic texts could be used appropriately:

Activity 1

Ask the students what their résumés’ readers will expect from them (talk about different expectations in their different fields). Get a list and write it on the board. Sample: LEADERSHIP SKILLS, RESEARCH SKILLS, OFFICE (CLERICAL) SKILLS, FINANCIAL SKILLS, COMMUNICATION SKILLS, START-UP (ENTREPRENEUR) SKILLS

Activity 2

Using the verbs found using Nation’s Vocabulary Program, have the students put them under the appropriate skill category. This can be done in pairs or groups. NOTE: Some verbs can go into more than one category; have them discuss or explain how.

Finally, to check the students understanding of verb use in résumés, a simple cloze exercise was used, taking job duties directly from the résumés and having students fill in the blanks with the appropriate verb.

Scenario 2: Teaching prepositions to lower-division undergraduate ESL students

As was previously mentioned, many teachers encounter problems of authentic English usage in textbooks. This extends to the teaching of prepositions, where most textbooks have exercises focusing on locative prepositions. Many of these exercises However, most preposition uses in authentic English extend beyond locative uses into the more abstract. To make preposition teaching more authentic, and at the same time more interesting for the students to use in the classroom, teacher 2 chose to use songs as her way of aiding students in understanding preposition use in English. Contrary to teacher 1, who wanted to have a more general understanding of verb use in the genre, résumé, teacher 2 wanted to pick a song that included many of the prepositions that she was currently teaching. Therefore, she needed to find a song that was suitable for her class’s needs. Based on her student population, she felt that using a song by her favorite group, Rush, would be appropriate. However, which song would be best?

As with Teacher 1, Teacher 2 collected a group of 15 Rush song lyrics and saved them as “.txt” files. However, instead of looking at the total word count under “Frequency”, she chose “Range” instead to not only see the overall counts for all the songs, but also to compare how often different words were used in the different songs. She started with the prepositions “of”, “in”, and “on”, and therefore wanted a song that used these prepositions often. The following figure shows the results of five songs and their counts with “of”, “in”, and “on”:

<u>Type</u>	<u>Range</u>	<u>Frequency</u>	<u>F1</u>	<u>F2</u>	<u>F3</u>	<u>F4</u>	<u>F5</u>
The	13	211	19	12	20	28	34
A	12	86	2	10	2	9	16
And	13	67	9	6	3	5	6
Of	11	60	0	4	6	6	15
In	10	51	3	7	8	2	7
On	9	33	0	6	2	1	14

Figure 2. “Range” results screen.

Figure 2 shows the different words and their “range”- how many of the fifteen songs they were in, their total word count for all fifteen songs, and how often they were used in each file (F1, F2, etc...), which in this case is each song. From the figure, it was apparent that F5, or song five (“Dreamline”) had the most uses of the prepositions needed for the class. Therefore, based on the findings from this program, Teacher 2 chose to use this song as a cloze activity to confirm that her students understood how to use these prepositions correctly.

Concluding Remarks

In all, this presentation was given not to tell teachers to throw away the textbook, but to help teachers expand their repertoire of materials development to include more uses of authentic English communication to help support learning objectives in the classroom. ESL students need to be exposed to these genuine uses of English as often as possible to help them prepare for future encounters with the language. Additionally, Nation’s Vocabulary Program enables teachers to use any type of text or genre that they feel is necessary in helping students succeed both in the classroom and beyond. Teachers who saw this presentation remarked at how easy they thought it would be to use and how important it would be for their students to see “real” English in action rather than just use the textbook for support. Other teachers remarked at how easy the program appeared to be and thought that, even though they were not technologically savvy, this program would be overall simple to use and make materials from.

References

- Biber, D., Conrad, S., and Reppen, R. (1998). *Corpus linguistics: Investigating language, structure, and use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, J., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman grammar of spoken and written English*. New York, NY: Longman.
- Coxhead, A. and P. Nation. (2001). The specialized vocabulary of English for Academic Purposes. In J. Flowerdew and M. Peacock (Eds.). *Research perspectives on English for Academic Purposes* (pp. 252-267). Cambridge: Cambridge University Press.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Louhiala-Salminen, L. (1996). The business communication classrooms vs. reality: what should we teach today? *English for Specific Purposes, 15*, 37-51.
- Nation, P. (2001). Using small corpora to investigate learner needs: Two vocabulary tools. In H. Ghadessy & Roseberry (Eds.) *Small Corpus Studies in ELT* (pp. 31-45). Amsterdam: J. Benjamin Publishing.
- Nation, P. (2002). *Vocabulary programme*. Retrieved March 2, 2005, from Victoria University of Wellington, School of Linguistics and Applied Language Studies Web site: <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>.