

# Data-Driven Learning and EAP Materials Development: A Corpus-Based Action Research Project

Brian Rugen  
Hawai'i Pacific University

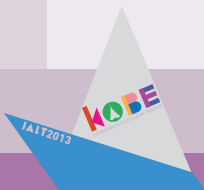
## Reference Data:

Rugen, B. (2014). Data-driven learning and EAP materials development: A corpus-based action research project. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings*. Tokyo: JALT.

The current availability of online corpora resources offers access to countless examples of authentic language, making it increasingly possible to incorporate corpus-based activities, or data-driven learning (DDL), into the language classroom. DDL encourages students to discover language patterning, rules, and frequency details through guided tasks with an online corpus and with teacher-designed materials based on corpus data. In this paper is described a corpus-based action research project that combined a guided, computer-based task with paper-based materials designed to supplement the vocabulary activities in an English for Academic Purposes (EAP) course textbook. The aim of this project was to determine to what extent a combined approach to DDL would affect classroom interaction and student attitudes toward using online corpora resources. Results of the first cycle of action research indicate both increased interaction and favorable attitudes toward DDL. The paper concludes with reflective suggestions for a future plan of action.

現在の電子コーパスが提供する大規模なデータは言語研究のみならず、英語教育の分野にも極めて有効なツールであると考えられている。コーパスのデータを英語教育に用いることによって、学習者は現実の英語圏の世界で頻繁に用いられている表現や文法のパターンを習得することができる。本稿では、授業で使われている一般学術的英語 (English for Academic Purposes) の教科書で学ぶ英単語習得を補うために、電子コーパスデータを用いて作られたペーパーとコンピューターベースの教材の結合教材についての実験が記述されている。これらの結合された教材を使った英語教育がどのように学習者の英語習得に対する態度や教室内でのインタラクションに影響するかを調べた。実験の結果、コーパスのデータを基にした教材を利用した英語教育は、学習者の英語習得に対する好意、及び教室内でのインタラクション上昇に繋がり、習得に貢献したと考えられる。本稿の終わりにまとめと今後の課題を挙げている。

**T**HE CURRENT availability of online corpora resources offers access to countless examples of authentic language, making it increasingly possible to incorporate corpus-based activities, or data-driven learning (DDL), into the language classroom. Reppen (2011) described three ways in which a corpus can be utilized to encourage DDL learning in the classroom. First, teachers can prepare materials for students from their own corpus searches. This can allow for a level of teacher control over the materials. Second, teachers can introduce students to available online corpora. In this case, students interact directly with the online corpora, giving them an opportunity to become language researchers. Third, teachers can create specialized corpora for their classes, such as a collection of course readings or student papers. Regarding the second suggestion, some have expressed concern with the sharp learning curve for tasks that involve students interacting directly with an online corpus (Boulton, 2010; Sun,



2003). This concern points to the importance of scaffolding and the need for some degree of control over the tasks. Teachers need to be careful not to turn students loose on an online corpus resource, where they might get lost because of unfamiliar navigational demands and where the “potential for confusion and frustration is high” (Oghigian & Chujo, 2010, p. 202). Johns (1997) alluded to the importance of scaffolding too, suggesting that first introducing students to prepared, corpus-based materials is a way to prepare them for direct interaction with a concordance (p. 113).

What has not been adequately covered in the literature are studies combining a guided, computer-based task with paper-based materials (Chujo, Anthony, Oghigian, & Uchibori, 2012), essentially a combination of Reppen’s (2011) first two suggestions above. This paper addresses this gap by describing a corpus-based action research project that combines a controlled, computer-based task with teacher-designed, paper-based materials designed to supplement the vocabulary activities in an English for Academic Purposes (EAP) course textbook. The first part of the paper contains a brief discussion of the importance of academic vocabulary and the recently developed *Academic Vocabulary List* (AVL; Gardner & Davies, 2013). In the second part I describe the action research project that combined the guided, computer-based task with paper-based materials. A final discussion highlights the outcomes of the project and suggests future actions to take based upon a critical reflection on the project.

### Academic Vocabulary

For both first and second language speakers, it has been recognized that control of academic vocabulary plays a central role in student success (Gardner & Davies, 2013). For years, the *Academic Word List* (AWL; Coxhead, 2000) has been a standard of reference in English language teaching. In fact, many commercially produced academic vocabulary texts are based on the AWL.

The AWL is organized into 570 word families, including the headword, or stem, and the inflections and derivations containing that headword. One concern with the AWL, however, is a lack of frequency information about the different forms of a word in a family. Do students studying academic English vocabulary need to learn all the forms? Which forms are most important for students to learn? For example, in the AWL there is no indication of which of those forms are academic high-frequency words, which are simply general high-frequency words, and which might have a discipline-specific frequency.

The recently developed AVL, on the other hand, provides this valuable information. The list is based on a 120 million-word subcorpus (derived from the 425 million-word Corpus of Contemporary American English [COCA]; see Davies, 2008) of academic texts from nine disciplines. It can be compared to the older AWL, which is based on a 3.5 million-word corpus of academic texts from only four disciplines. The words in the AVL are not only listed by lemmas, but are also grouped into word families, which was useful for this particular project. Unlike the list of word families in the AWL, the AVL provides clarity in distinguishing whether forms (inflections and derivations) are core academic words (those that appear in the majority of the nine, broad academic disciplines) or simply general high-frequency words. In addition, words that appear at a higher than expected rate in a particular discipline are noted as technical- or discipline-specific words. Consider the following two word family examples for the headword *illustrate*:

Illustrate	<i>illustrated</i> , illustrates, illustrating, illustration, illustrations, illustrative
------------	---

Figure 1. Academic Word List (Coxhead, 2000).

Illustrate	9662	<b><u>illustrate</u></b> (v) <sub>9224</sub> illustration (n) <sub>4493</sub> <b><u>illustrative</u></b> (j) <sub>438</sub> <i>illustrated</i> (j) Hum <sub>247</sub> illustrator (n) <sub>194</sub>
------------	------	--

Figure 2. Academic Vocabulary List (Gardner & Davies, 2013).

The AWL (Figure 1) provides a minimal amount of information, showing only the most frequent form from the relatively small corpus in italics. The AVL (Figure 2), on the other hand, provides much more detail. The bolded and underlined **illustrate** (verb) and **illustrative** (adjective) indicate that the two are core academic words. Italics indicate that a word is more technical and occurs mainly in one domain; in this case, *illustrated* appears more frequently than expected in the humanities discipline. A normal font indicates that the word is simply a member of the word family, not necessarily a core academic word. For the online version of this list, these distinctions are color-coded.

For activities requiring students to use an inflectional or derivational form of a word to complete sentences or paragraphs, it makes sense to focus on the forms that are indeed academic, high-frequency words. In other words, by focusing only on the forms of a word that are part of an academic core list, rather than trying to learn all the forms, or perhaps some at random, teachers and students of EAP can get more bang for their buck.

## The Current Project

The aim of action research (AR) is to identify and address a particular classroom problem or dilemma in a systematic way in order to improve an unsatisfactory situation. Burns (2009) noted that “the central idea of the action part of AR is to intervene in a deliberate way in the problematic situation in order to bring about changes and, even better, improvements in practice” (p. 2). Burns called the overlapping phases of action research a

*reflective research cycle* that consists of planning, acting, observing, and reflecting. The planning phase involves identifying a classroom problem or dilemma and developing a plan of action—with ideas gleaned from the literature, discussions with colleagues, or both—to address it. As a result, the teacher should have a specific research question to answer. The action phase involves organizing and implementing, over a set period of time, whatever deliberate intervention was planned in the previous phase. During this intervention, the observation phase involves systematically collecting and analyzing data regarding the effects of the intervention. The final reflection phase involves evaluating the effects of the intervention and formulating a plan for further action.

This project came about as the result of dissatisfaction with a particular section in each chapter of an assigned academic vocabulary textbook in a pre-university EAP course. The course is part of an EAP preparation program at a private university (approximately 7,000 undergraduate and graduate students) in Hawaii. Courses in the program provide support in academic English for international students whose scores on a recognized English proficiency test fall just below the minimum requirement for admission to the university (i.e., TOEFL iBT: 70-79). Most of students in the program are from Asia (Japan, Korea, Taiwan, and China), and most matriculate to degree-seeking, undergraduate status after one or two semesters in the pre-university EAP program. The vocabulary textbook is based on the AWL, with each chapter targeting between 30 and 40 words from the AWL. In each chapter, a *word forms* section consists of a two-part activity. First, students are required to complete a table with the noun, verb, adjective, and adverb forms of approximately 10-15 of that chapter’s targeted AWL words. The targeted AWL word appears in the chart, in the appropriate column depending on its part of speech (e.g., noun), and students are required to write in the other forms of the word (in this example, verb, adjective, and adverb). In the second part of

the activity, students are required to complete a set of fill-in-the-blank sentences with the correct form of one of the words from the chart.

I was dissatisfied with this activity for two main reasons. First, the chart-completion part of the activity seemed tedious. Despite students being encouraged to interact as they worked on the activity in pairs, this particular activity really discouraged student interaction. Students discovered that it was easier to simply consult their dictionaries and complete the chart with little to no interaction. It has been noted that because of the importance of classroom interaction for second-language development, a main objective of a language course should be to maximize the opportunities that students have for target language interaction (Van Lier, 1996). Yet, I felt that the first part of the activity discouraged interaction, instead encouraging the kind of *yakudoku* (a Japanese-style grammar translation method, traditionally the dominant approach in foreign language learning in Japan) pedagogical approach that I questioned philosophically.

Second, for the fill-in-the-blank sentences, the correct form required for many of the items was not necessarily a core academic high-frequency word, according to the more current and more reliable AVL. In fact, there seemed to be no logic behind which form was required to complete the items. For example, in an item asking for a form of the word *illustrate* to complete the sentence, the answer was *illustration*. However, according to the AVL (see Figure 2), *illustration* is not a core academic word. It is considered a general, high-frequency word. This is not to say that the word is not worth learning, but perhaps should be learned in another context. However, if a curricular focus is on learning academic vocabulary, then a principled approach that is focused first on core academic words seems more appropriate.

This classroom dissatisfaction and a subsequent review of the literature on an updated academic vocabulary list—the

AVL—and the increasing role in DDL for academic vocabulary development led to the formulation of two research questions that guided the project:

1. To what extent will the integration of materials designed for a combined approach to DDL affect student interaction in classroom academic vocabulary activities?
2. How does the integration of a combined approach to DDL affect student attitudes toward using online, corpus-based tools for learning?
  - a. What attitudes do students hold toward the first part of the supplemental activity, in which they are asked to interact directly with the online corpus resource?
  - b. What attitudes do students hold toward the second part of the supplemental activity, in which they are asked to complete a fill-in-the-blank handout using concordance lines?

To answer these questions, I designed materials for a two-part activity designed to supplement the textbook's two-part word forms activity.

### Part One (Direct Interaction With Online Corpus)

First, to replace the problematic word forms chart in the textbook, for each chapter students were given a handout (see Appendix A for a simplified version) and asked to work in pairs and simply copy down the targeted 10-15 words given in the chart in the textbook onto the handout. Then, working in pairs in front of a computer, students were required to locate the word family for each of those words on the COCA's online interface for the AVL word families (Davies, 2013). Students accessed the site and simply used the mouse to scroll down the list until they located the appropriate word family. Then,

for each word family, students wrote down which forms of the word were academic, high-frequency words. Before this first part of the activity, students were introduced to the color-coding system used in the online version of the word family list. Words in yellow indicate core academic words; words in red indicate discipline-specific words; words in blue indicate general, high-frequency words (not necessarily academic per se); and words in grey indicate low-frequency words. Thus, for this part of the activity, the words in yellow—core academic words—were the ones students wrote down. When finished, each pair of students submitted one handout to the teacher.

### Part Two (Fill-in-the-Blank Worksheet)

The second part of the activity involved using the handout the students submitted and creating fill-in-the-blank sentences for the academic word forms they identified. The new fill-in-the-blank items would then be based on academic, high-frequency word forms, rather than random forms of a word that may or may not be core academic forms. Creating the fill-in-the-blank sentences was a simple process that involved first doing a list search for each of the academic high-frequency forms that the students identified. For example, for the *illustrate* word family, students would have identified *illustrate* and *illustrative* as academic high-frequency words. The teacher would then enter each of those words into the search box on the front page of COCA, check the list search feature, and highlight the academic register section in the drop-down list of registers. Clicking “Search” would then result in a list of concordance lines containing the target vocabulary word. Second, the teacher would copy a few of those concordance lines, paste them on a handout, and blank out the target vocabulary word. An example of a simplified handout with the academic high-frequency word forms from the *illustrate* and *coherent* word families is included in Appendix B. Part two of this activity is also an improvement over the

textbook items in that the fill-in-the-blank sentences are actual concordance lines taken from authentic academic discourse.

### Data Collection and Analysis

Three methods of data collection were used to answer the project’s two research questions. First, I kept written observations in a research journal while students worked on the two-part activity. Reflecting on the notes, the first thing I noticed was the obvious increase in target-language interaction during the activity, especially the first part. Working on the computer in pairs led to increased interaction, some of which involved the logistics of accessing, navigating, and understanding the online resource, and some of which involved searching through the online word family list to complete the required handout. Much of the interaction across all pairs, for example, involved comprehension checks regarding the list’s color-coding scheme.

To answer the second question, a 10-item Likert-scale survey was administered at the end of the course. The first seven items addressed part one of the activity; the last three items addressed part two of the activity. Table 1 shows the results of the survey.

Results of the survey indicated favorable attitudes toward interacting directly with the online corpus and toward the fill-in-the-blank activity using concordance lines. Most students also indicated a willingness to use online corpora resources at home in the future and a willingness to recommend the resources to other learners of English. One remaining concern is whether to simplify concordance lines or to leave them in their original form. Over twice as many students either somewhat or strongly disagreed that the concordance lines were easy to read. In the next iteration of this project, it may be necessary to be more careful in the concordance lines chosen for the student handout or to simplify the concordance lines altogether.

Table 1. Results of Student Questionnaire Evaluating the Two-Part Activity (N = 17)

Survey question	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
It was easy to use the online corpus resource for this activity.	0	1	1	12	3
I would like to use online corpus resources in English classes in the future.	1	0	2	13	1
I intend to explore online corpus resources at home in the future.	1	0	5	10	1
I would recommend using online corpus resources to other English language learners.	0	0	5	11	4
Interacting with the online corpus resource motivated me to want to learn more about academic vocabulary.	0	1	5	9	2
Using the online corpus resource took up too much time in class.	7	6	2	2	0
Exploring online corpus resources is appropriate for advanced-level students only.	1	7	3	4	2
The concordance lines for the fill-in-the-blank items in the materials were easy to read and understand.	3	8	1	4	1
Using concordance lines for fill-in-the-blank materials is a good way of learning vocabulary.	0	3	6	6	2
Using concordance lines for fill-in-the-blank materials is appropriate for advanced-level students only.	0	8	2	5	2

Finally, exit interviews with students in groups of three to four were held at the end of the course. In the interviews, which resembled focus group discussions, a series of open-ended questions was asked about various aspects of the course, including the two-part activity. Comments indicated that the activity was challenging but useful. Students seemed to enjoy working in

pairs and appreciate being introduced to an online vocabulary list that would help them learn the academic words they needed to learn. Students concurred that the two-part activity led to increased interaction, mainly because it was a new type of task for them and the process involved collaborating with a partner as well as “sometimes needing to ask another group for help.”

Students also expressed interest in not only referring to the on-line interface of the AVL again on their own, but also in exploring other features of the COCA site more.

## Conclusion

The aim of this corpus-based action research project was to determine to what extent a combined approach to DDL would affect classroom interaction and student attitudes toward using online corpora resources. Analysis of the data indicates both increased interaction and favorable attitudes toward DDL. In addition, as a result of the project, students were able to focus on only the relevant forms (i.e., core academic words) of the words targeted in the assigned textbook. Whether the increased focus on the relevant forms leads to actual learning is another follow-up step worth addressing in the next iteration of the project.

Scaffolding was a major consideration in the project. In part one of the activity, scaffolding students' direct interaction with the online resource lowered the learning curve and also served as an introduction to the tools of online corpora resources. In part two of the activity, using concordance lines for the fill-in-the-blank sentences not only provided authenticity but also functioned as a way of scaffolding further explorations with concordance lines to encourage the discovery of particular lexical patterns and collocations (see Reppen, 2010). As mentioned above, however, another follow-up step worth addressing is the extent (if any) to which the concordance lines should be simplified for the students.

Overall, the project is significant in that it is a move toward academic vocabulary study, based on an improved academic vocabulary list, that encourages students to both interact with a computer in a controlled and scaffolded manner and engage with paper-based materials based on authentic academic language.

## Bio Data

**Brian Rugen** is an assistant professor of applied linguistics in the Department of Languages and Applied Linguistics at Hawai'i Pacific University. His research interests include second language writing and TESOL materials and curriculum development. <brugen@hpu.edu>

## References

- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60, 534-572.
- Burns, A. (2009). *Doing action research in English language teaching: A guide for practitioners*. New York: Routledge.
- Chujo, K., Anthony, L., Oghigian, K., & Uchibori, A. (2012). Paper-based, computer-based, and combined data-driven learning using a web-based concordancer. *Language Education in Asia*, 3, 132-145.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- Davies, M. (2008). *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available from <http://corpus.byu.edu/coca/>
- Davies, M. (2013). *Academic Vocabulary Lists*. Available from <http://www.academicvocabulary.info/>
- Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*. Advance online publication. Available from <http://dx.doi.org/10.1093/applin/amt015>
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). Harlow, UK: Addison Wesley Longman.
- Oghigian, K., & Chujo, K. (2010). An effective way to use corpus exercises to learn grammar basics in English. *Language Education in Asia*, 1, 200-214.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.

Reppen, R. (2011). Using corpora in the language classroom. In B. Tomlinson (Ed.), *Materials development in language teaching* (2nd ed., pp. 35-50). Cambridge: Cambridge University Press.

Sun, Y. C. (2003). Learning process, strategies and web-based concordancers: A case study. *British Journal of Educational Technology*, 34, 601-613.

Van Lier, L. (1996). *Interaction in the language classroom: Awareness, autonomy and authenticity*. London: Longman.

## Appendix A

### Handout: Part One

Academic Vocabulary: *Word Forms*

The academic vocabulary list is a useful resource for developing your knowledge of academic vocabulary. The list can be found at the following address: <http://www.wordandphrase.info/academic/>

#### Step one

In the left column of the following chart, copy the words that are provided in the chart from Exercise 4 of your textbook.



#### Step two

For each word, scroll down the word families of the AVL to find the word's family. In the second column, write down the words from the family that are in yellow. The forms of the words in yellow are academic, high-frequency words and are important for you to learn.

## Appendix B

### Handout: Part Two

Academic Vocabulary: *Word Forms in Sentences*

Directions: Choose the correct word form in brackets to complete the corresponding concordance lines below.

<i>[illustrate, illustrative]</i>	
1	serious and widespread. The following example, from the Horn of Africa, is _____. # Famines are extreme health crises that cause people to move to avoid hunger
2	excitation. Intermodulation distortion # A test signal consisting of two tones was used to _____ the ability of the time-domain model to simulate intermodulation distortion effects. The test signal
3	to things that should not be said during IEP meetings. # Following are six _____ case vignettes, which include statements that parents and professionals should not hear at IEP



4	particularly underexplored. Two examples - one from the litigation context, one legislative - _____ the point. A. Litigation # In Alexander v. Sandoval, 50 the Supreme Court
5	the immediate, intermediate tasks. The technological manifestations and their complicity in the Holocaust _____ the inability of instrumental rationality to adequately incorporate the requisite ethical and moral dimensions,
6	process whereby our memories can be distorted or drastically altered by suggestion. In one _____ experiment, researchers created a fake picture by transposing a childhood snapshot of a boy

*[coherent, coherently]*

7	Standard # The communication standard proposes that students organize and consolidate their mathematics thinking, _____ communicate their mathematical ideas to others, analyze and evaluate the mathematical thinking and strategies
8	integrate these various methods of describing spillover and identify the most likely neural pathway to _____ explain the recorded observations. Finally, the possible advantages and disadvantages of each spillover
9	decades since anyone on the left has been able to articulate, first, a _____ analysis of what happens to the structure of advanced societies as they undergo economic change

10	started changing its posture and strategy in late 2006, arguably did not implement a _____ counterinsurgency campaign until 2009. It would be myopic and irresponsible to conclude that the
----	---