

A Corpus-based Approach to Vocabulary Instruction in the CBI Classroom

Jesse Conway

Northern Arizona University

Abstract

In the content-based instruction (CBI) classroom, students' vocabulary knowledge is essential for their success in class. Without adequate vocabulary acquisition, students will be less likely, or even unable, to understand and work with the content-specific information introduced in class. Following the idea of more frequent exposure to words leading to increased acquisition, it stands to reason that words occurring more frequently in class texts are those that should be chosen for vocabulary instruction in the classroom. However, in the CBI classroom, it is often the case that a variety of texts are utilized, making the selection of important vocabulary a daunting task for a teacher. This study proposes that through the use of corpus analysis, the most frequently occurring terms in class texts can be identified for use as class vocabulary. This study compared the frequency and accuracy of use of vocabulary terms between two groups: a control group learning instructor-selected terms and a treatment group learning terms identified through corpus analysis. Results show that the treatment group not only utilized vocabulary more frequently in their writing, but also utilized terms more accurately than the control group.

Keywords: vocabulary, content-based instruction, corpus analysis, L2 writing

A Corpus-based Approach to Vocabulary Instruction in the CBI Classroom

Background

The CBI classroom provides teachers with many chances to tailor classroom materials to student needs, often utilizing more than one source to instruct a unit. While this allows a teacher to make the classroom a much more relevant, and therefore motivating environment (Grabe & Stoller, 1997), it also provides the teacher with a challenge; how to make sure students are able to comprehend all of the information that they encounter? Vocabulary is a key area for any comprehension of language (Nation, 2001) and corpus analysis can be utilized in order to determine the words that students may encounter most frequently in texts. Research has shown that students benefit from instruction of higher frequency terms in order to work with content more effectively (Cobb, 2007; Huang & Liou, 2007). This runs counter to the traditional approach to CBI vocabulary selection, which often relies on teacher intuition. Nation (2001) states that word knowledge begins with receptive knowledge, which through repeated exposure becomes productive knowledge. Therefore, if the vocabulary chosen for instruction are high frequency terms, it stands to reason that students are likely to internalize them at a greater rate than low frequency, but highly-content specific words that an instructor may select.

While there are a wide variety of proven vocabulary learning methods (Zimmerman, 2009), studies report positive effects in English for Academic Purposes (EAP) student writing from examination of collocations with the use of concordancing software (Kaur & Hegelheimer, 2005; Cobb, 1997; Cobb & Boulton, 2015). Examination of collocations can not only provide accurate examples of vocabulary use, but may also push more motivated students to examine other words encountered in collocation identification activities, resulting in more lexical diversity in student writing (Laufer & Nation, 1995).

Research Questions

- Does the vocabulary instructed in class appear in student writing without prompting? If so, how frequently do they occur?
- When vocabulary words appear in writing, are they being used correctly?

Methods

A corpus was created from selected chapters of *Harriet Tubman* (Sawyer, 2010), a text utilized in a level 4 CBI unit. The most frequently occurring words were identified utilizing AntConc (Anthony, 2014) and then compared against an instructor-created vocabulary list against three criteria: frequency of the word in the text, number of forms of the word in text, and frequency of a word's occurrence in BNC and COCA corpora as identified with Vocabprofile (Cobb, 2015). Following this, a corpus-influenced vocabulary list was created utilizing the most frequently occurring, multi-form words in the text.

Research was conducted on with two level 4 classes. The control group (N=14) was a mix of male and female students, with ages ranging between 19 – 24 years of age. First language background for the control group consisted of ten speakers of Chinese and four speakers of Arabic. The treatment group (N=13) was a mix of male and female students, with ages ranging between 19 – 24 years of age. First language background for the treatment group consisted of seven speakers of Arabic, four speakers of Chinese, and two speakers of Japanese.

Two testing instruments were devised following the creation of the corpus-influenced list. One measured receptive vocabulary ability, while the other measured productive vocabulary ability. Each instrument featured the 31 terms from the corpus-influenced list and was given to

the control group and the treatment group before instruction of the unit began. Tests were scored by two raters, achieving an inter-rater reliability rate of 0.96 for the initial test.

Vocabulary instruction methods in the control group were varied, but did not include the use of collocation analysis or concordancing software. Vocabulary instruction for the treatment group utilized identification of vocabulary collocates through text analysis. Results were then confirmed and expanded upon using the concordancing website WordandPhrase.info (Davies, 2015).

After all terms had been instructed, students completed three-part essay, the final assessment for the unit. The essay underwent two rounds of content and grammar revision before the submission of a final draft. Both the control and treatment groups' final drafts (N=13 for each) were then placed in separate corpora and analyzed for vocabulary frequency and accuracy. Vocabulary use was manually checked in order to determine grammatical accuracy. Finally, the vocabulary testing instrument was delivered again to both groups. This was again scored by two raters, resulting in an inter-rater reliability of 0.95. These results were then analyzed to determine vocabulary acquisition of the two groups.

Results

- Does the vocabulary instructed in class appear in student writing without prompting? If so, how frequently do they occur?

It was found that in both the control and treatment groups, class vocabulary did appear in writing without prompting. The control group utilized 26 of 33 words on their vocabulary list (79%) while the treatment group utilized 24 of 31 words on their list (77%), indicating a near

equal amount of term usage between the two groups. The most frequently utilized words can be seen in Table 1, with words shared between control and treatment list bolded.

Table 1

Most frequently occurring terms from each group, per 1000 words

Term	Rank of Occurrence (group)	Frequency - Control	Frequency - Treatment
Slave	1 (both)	35	41
Escape	2 (control)/3 (treatment)	8	6
Underground railroad	2 (treatment)	n/a	12
Fugitive	3 (control)	6	n/a
Freedom	3 (treatment)	n/a	6
Own (slave owner; owned)	4 (both)	4	5.2
Abolitionist	5 (control)	2	n/a
Hide	5 (treatment)	n/a	3

Analysis of these results show that the treatment group were more likely to utilize two out of three shared terms (*slave* and *own*) than the control group. Furthermore, analysis of less frequently used terms found the treatment group utilizing vocabulary terms more frequently in their writing than the control group.

- When vocabulary words appear in writing, are they being used correctly?

The control group was seen to use vocabulary accurately only 52% of the time they utilized a vocabulary word in writing. In comparison, the treatment group utilized vocabulary accurately 71% of the time. While the control group had some instances of vocabulary being used inaccurately 100% of the time, this was not the case with the treatment group, who had a

minimum level of 84% accuracy with any vocabulary word used, and never had an inaccurate use of a vocabulary word. These results can be seen in Figures 1 and 2, below.

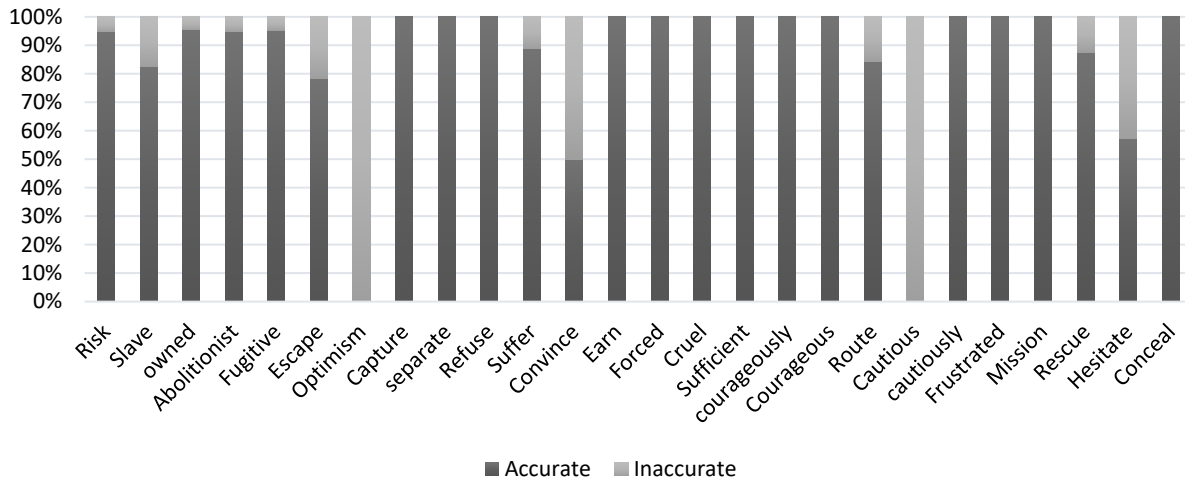


Figure 1. Control group accuracy.

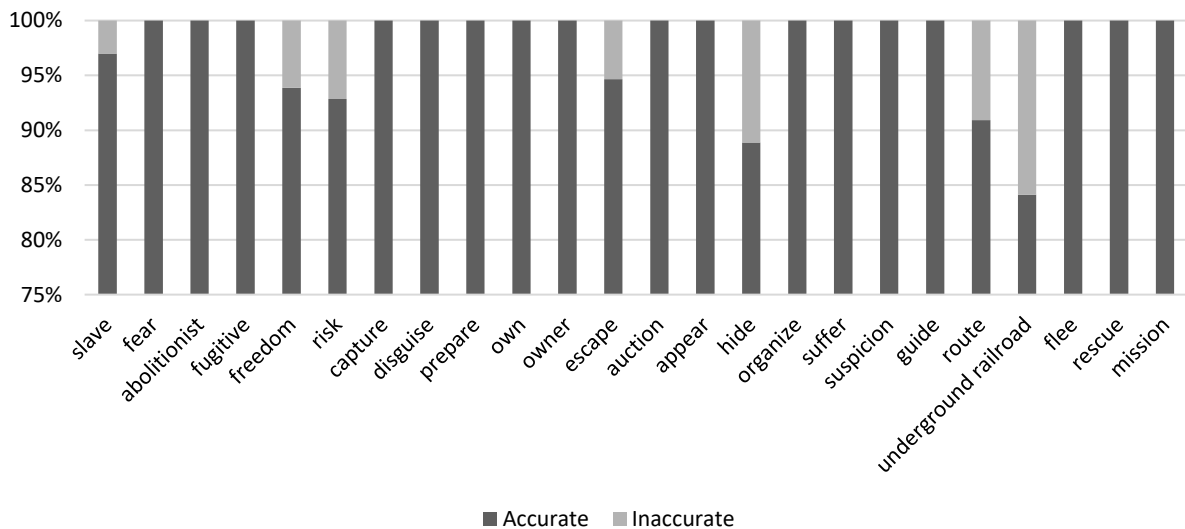


Figure 2. Treatment group accuracy

These accuracy results were then compared to the receptive and productive vocabulary knowledge tests delivered to students before and after vocabulary instruction. Despite not receiving instruction on the corpus-influenced vocabulary list, the control group were more

likely to pass both tests delivered after instruction than the treatment group. However, the treatment group was more frequent and accurate in their usage of the corpus-influenced list terms in their writing. This is a result that may be explained by affective issues associated with testing compared to writing, or the presence of revision focusing on different vocabulary terms in classes. Regardless, this aspect of the study requires further analysis.

Relevance to PIE and Second Language Learning

As CBI courses are a mainstay of levels 3 and 4 at PIE, I believe that this study has great relevance to PIE in not only the way we approach unit planning, but also in how we address vocabulary instruction, especially when students are being assessed with a written task. This study has shown that through some careful planning before beginning a unit, the most frequently occurring vocabulary terms can be identified and used for classroom instruction. Utilizing more frequently occurring words as vocabulary compared to instructor-selected words benefits CBI students in a number of ways:

- It sets students up for success by providing them with multiple encounters with vocabulary, leading to increased recognition and productive use.
- It saves time in the classroom by focusing on the most salient words for content comprehension as opposed to spending time teaching a word that students will encounter only once in their text and may not find useful in their lexicon.
- It leads to increased productive use of these terms in student writing without explicit prompting, leading to more natural usage of vocabulary in student language.

Utilizing collocation-based vocabulary teaching also provides benefits for students such as:

- Increased accuracy of vocabulary use.

- Greater lexical variety thanks to the number of collocations that can be associated with one word.
- An engaging approach to vocabulary that encourages targeted re-reading of text and provides students with an interactive manner of checking their predictions.

This study has shown that identifying key vocabulary requires little more than a few hours of an instructor's preparation time and pays off great dividends for both teachers and students with highly accurate usage of a variety of essential terms in a CBI course.

References

- Anthony, L. (2014). AntConc (Version 3.4.4w) [computer software]. Tokyo, Japan: Waseda University.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301-315.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3), 38-63.
- Cobb, T. (2015). *Vocabprofile* [computer program]. Accessed 12 Dec 2015 at <http://www.lex tutor.ca/vp/> .
- Cobb, T. & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478-497). Cambridge, UK: Cambridge University Press.
- Davies, M. (2015). *WordandPhrase* [computer program]. Accessed 12 Dec 2015 at <http://www.wordandphrase.info> .
- Grabe, W. & Stoller, F.L. (1997). Content-based instruction: Research foundations. In M.A. Snow & D.M. Brinton (Eds.), *The content-based classroom: Perspectives on integrating language and content* (pp. 5-21). White Plains, NY: Longman.
- Huang, H. & Liou, H. (2007). Vocabulary learning in an automated graded reading program. *Language Learning & Technology*, 11(3), 64-82.

- Kaur, J. & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, 18(4), 287-310.
- Laufer, B. & Nation, P. (1995). Vocabulary size and use – lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Sawyer, K.K. (2010). *Harriet Tubman*. New York: DK Publishing.
- Zimmerman, C.B. (2009). *Word knowledge*. New York: Oxford University Press.