Using Corpus-based Tools in Teaching and Learning Academic Vocabulary

Gözde Durgut

Summary: Drawing on the research literature on EAP teaching practices, this paper outlines a number of corpus tools that are available for teaching academic vocabulary and discusses how they can be used in the process of material design and development.

Keywords: EAP, corpus tools, academic vocabulary, material design and development

Introduction

Academic vocabulary has been a major area of study in corpus-based research. Recently developed web-based tools and software have eased the process of compiling and generating word frequency lists (e.g., Academic Word List by Coxhead (2000), Science Word List by Coxhead and Hirsh (2007), and others). In addition, studies on discipline-specific vocabulary have shed light on how language works across different disciplines, providing teachers and learners in EAP contexts with further insights into various aspects of academic language such as words taking on specific meanings in a particular discipline different than their everyday use (Coxhead, 2010). However, teachers in EAP settings might need more support or guidance in teaching the vocabulary and grammar of academic language. Therefore, it becomes instrumental to take a fine-grained, research-practice oriented approach in developing a materials-based agenda for the teaching of EAP at tertiary level.

Which corpus tools should be used for teaching academic vocabulary?

Written academic language has long been the focus of corpus-based research in an attempt to improve teaching and learning practices in EAP contexts. To achieve that end, a number of corpora that focus on academic written English have been developed to fulfill the needs of L2 writers studying in different disciplines in academic settings.

The Academic Word List (AWL), developed by Coxhead and Byrd (2007), consists of around 3.5 million words, and includes texts from a wide variety of disciplines such as commerce, science, law, and arts. It has long been used by researchers, materials developers, EAP teachers and students since it includes core academic vocabulary. Notwithstanding its widespread use, some researchers suggest that AWL may not be as general as it is assumed to be since frequency and usage patterns show variation across disciplines (Hyland, 2007).

With 6.5 million words of proficient student writing, The British Academic Written English (BAWE) (Nesi et al., 2005) covers four disciplinary areas including social sciences, life sciences, arts and humanities, and physical sciences. Although teachers may not have direct access to data for the purposes of materials development, Nesi et al. (2018) have recently addressed this concern by introducing a new resource: BAWE Quicklinks: links for EAP teachers. The website includes concordances retrieved from BAWE and makes it possible for teachers to give feedback by creating hyperlinks that could be pasted into student papers. These concordance lines in the form of "quicklinks" are aimed at providing EAP students with more felicitous uses of language while at the same time allowing them direct access to data. Imagine, for example, that a student writes *Skiba (2016) describes that…* Clearly, 'describes' is awkward in this context. With BAWE quicklinks, a teacher can provide this reusable link: https://ske.li/pointsout_that which includes the following examples of concordances indicating the words that could be used instead of the word "describe" (see Figure 1).

How can corpora be used in developing EAP materials for teaching vocabulary?

There exist a number of web-based tools used in analyzing language features in teaching EAP vocabulary. One of them is the AWL Highlighter (Haywood, 2007), which

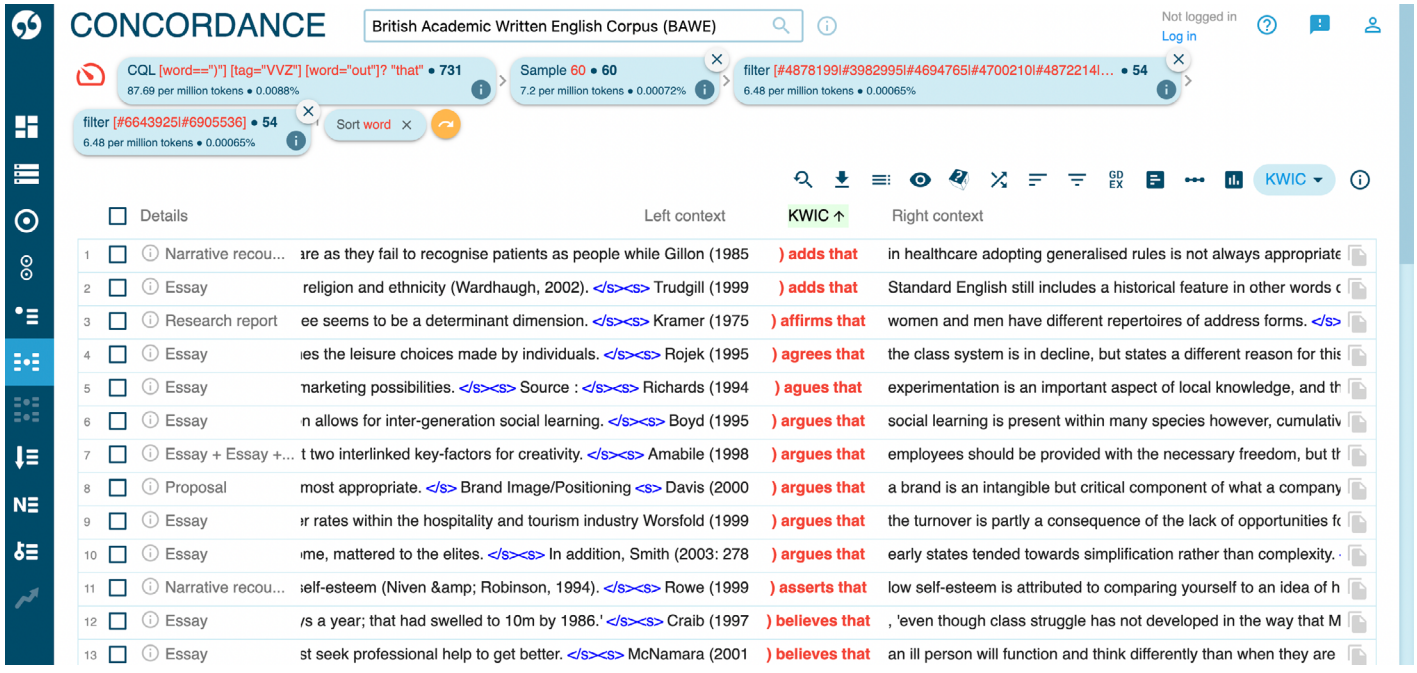

Figure 1: British American Written English Corpus search; (Nesi et al., 2005)

is a computer-based tool that enables users to highlight academic vocabulary in a text. Teachers can cut and paste the texts of their choice and create exercises for their students either as a simple gap fill, a headword gap fill, or a word family gap fill, all of which help learners focus on the target words in a contextualized form.

The Compleat Lexical Tutor (Cobb, 2007), another web-based tool for language teachers and learners, offers a wide variety of tools that could be used in the process of vocabulary development. The website itself includes short texts that teachers can use for developing analysis tasks; alternatively, they can input their own texts and use the concordancing tool available on the website for analysis. The website also includes a section for academic vocabulary lists developed by different researchers in the field, including the Academic Word List (AWL) (Coxhead, 2000), General Service List (West, 1953), and the University Word List (UWL) (Nation and Xue, 1984). Among these numerous tools offered by the website is the Concord Writer which allows users to insert their writing samples and evaluate the accuracy of their vocabulary usage. This could be particularly useful for teachers to encourage learners to experiment with their own writing and engage in independent analysis tasks (Coxhead and Byrd, 2007).

The same tools can be employed by EAP teachers in the process of developing materials or designing in-class activities for teaching academic vocabulary. Some of these activities, for example, can include concordance-based analysis tasks along with some guiding questions that would help learners notice patterns in a particular linguistic environment. See, for example, the sample task based on the concordances of the word "benefit" (Figure 2).

Any activity of this sort helps direct students' attention to some of the lexico-grammatical patterns of the target word in the data set. Other similar questions might include:

1. If the word is used as a noun form, what adjectives or verbs does it collocate with?
2. If the word is a verb, what adverbs or nouns does it co-occur with?
3. Have you noticed any other patterns that are salient in the data? If the word is a verb, for example, what comes after it? e.g., that-clause, infinitive, prepositions etc.

EAP practitioners might also prefer a more genre-based approach and compile a list of discipline-specific vocabulary to use in their teaching practices. There are a number of criteria that teachers should consider. Some of these factors include but are not limited to:

- Frequency of use in corpus data (across disciplines),
- Words frequently used in both corpus data and in sample texts that were previously studied in class,
- Words which are frequent either in all or at least two of these resources: corpus data, the AWL (Coxhead, 2000) and sample texts (Coxhead and Byrd, 2007).

It should also be noted that if students do not have any previous experience in using corpus tools, items selected for the initial phase of analysis should be words with which students are more familiar in order to make the first encounter less overwhelming or intimidating for the students (Coxhead and Byrd, 2007).

Below is a short excerpt from a civil engineering textbook (Lieuw and Chen, 2003) which could be used to develop students' knowledge base of AWL vocabulary as well as commonly used collocations in a civil-engineering

Sample task based on the concordances of the word "benefit"

The questions below will help you recognize patterns in the left/right context of the word 'benefit'. Analyze the concordances of "benefit" from the BAWE corpus and answer the following questions:

1. The word "benefit" can be used both as a noun or verb. Look at the right context. Which one is more frequently used in the data?
2. Look at the right context and find examples of the following uses:
   a) an advantage gained from something.
   b) a payment or gift provided by the state, an employer, or an insurance company.
3. Look at the right context. Sometimes 'benefit' occurs as part of a prepositional phrase. Which prepositional phrase(s) can you find in this set of data?
4. The verb form of "benefit" can be used both as a transitive and transitive verb. Find examples of intransitive and transitive forms in the concordances. Which preposition is used after the verb "benefit" when it is used as an intransitive verb?

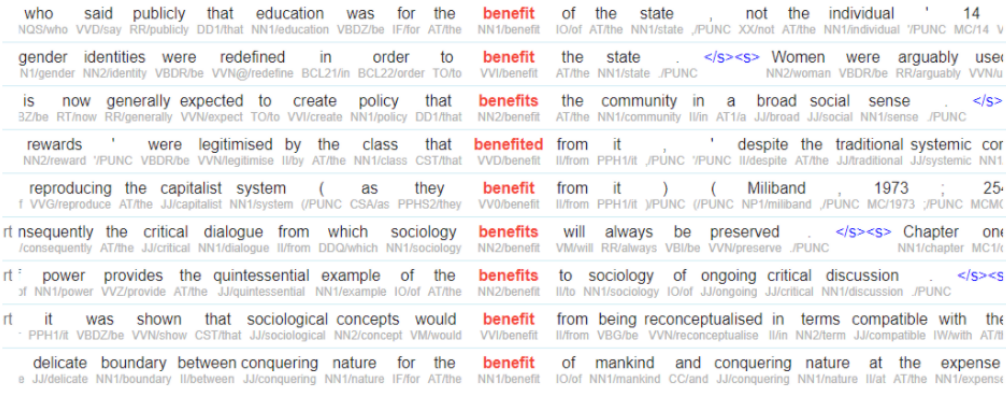

Figure 2: Sample task based on the concordances of the word "benefit"

One of the most important responsibilities of construction project management is the planning and scheduling of construction projects. The key to successful profit making in any construction company is to have successful projects. Therefore, for many years, efforts have been made to plan, direct, and control the numerous project activities to obtain optimum project performance. Because every construction project is a unique undertaking, project managers must plan and schedule their work utilizing their experience with similar projects and applying their judgment to the particular conditions of the current project. Until just a few years ago, there was no generally accepted formal procedure to aid in the management of construction projects. Each project manager had a different system, which usually included the use of the Gantt chart, or bar chart. The bar chart was, and still is, quite useful for illustrating the various items of work, their estimated time durations, and their positions in the work schedule as of the report date represented by the bar chart. However, the relationship that exists between the identified work items is by implication only. On projects of any complexity, it is difficult, if not virtually impossible, to identify the interrelationships between the work items, and there is no indication of the criticality of the various activities in controlling the project duration. A sample bar chart for a construction project is shown in Fig. 2.1. The development of the critical path method (CPM) in the late 1950s provided the basis for a more formal and systematic approach to project management. Critical path methods involve a graphical display (network diagram) of the activities on a project and their interrelationships and an arithmetic procedure that identifies the relative importance of each activity in the overall project schedule.

Figure 3: A short excerpt from a civil engineering textbook (Lieuw and Chen, 2003) with highlighted words from Sublist 1 of the academic corpus from AWL Highlighter (Haywood, 2007)

context. The highlighted words are from the Sublist 1 of the AWL, which includes the most commonly occurring words in the academic corpus. While teaching these words, teachers could prepare tasks to help students identify patterns that these items frequently occur in and discuss some of the lexico-grammatical features that they notice in the data set (see Figure 3).

Another example is an excerpt from a finance and economics textbook (Kürthy et al., 2018). In this short extract (Figure 4), words included in Sublist 1 and 2 in the AWL are highlighted. Based on this data set, teachers can design activities in which students specifically look at how some of these words take on new meanings across different disciplines such as finance. According to this data set, for instance, *bond*, *maturity*, and *principal* are some of the words that stand out. Learners can look more closely into the linguistic environment these items occur in and discuss how they operate within the boundaries of this particular context. A teacher could ask, for example,

- How do the meanings of these words differ from their everyday usage?
- What collocational behaviors/patterns do we observe?
- How do they operate on a phraseological level?

Teachers do not have to rely on existing corpora only in the process of vocabulary development. Another option for them is to compile

their own corpus that would include sample student writings as well as a selection of journal articles from native-speaker corpora that could be representative of the type of genre that they want to examine. This could help students compare their language use to the native speakers of a particular discourse community and vary their use of vocabulary or lexico-grammatical features (Lee and Swales, 2006).

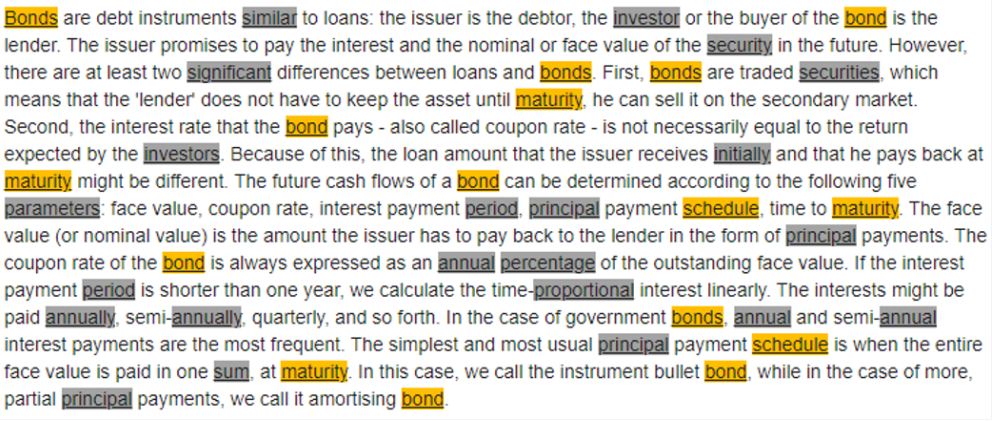
Bonds are debt instruments similar to loans: the issuer is the debtor, the investor or the buyer of the bond is the lender. The issuer promises to pay the interest and the nominal or face value of the security in the future. However, there are at least two significant differences between loans and bonds. First, bonds are traded securities, which means that the 'lender' does not have to keep the asset until maturity, he can sell it on the secondary market. Second, the interest rate that the bond pays - also called coupon rate - is not necessarily equal to the return expected by the investors. Because of this, the loan amount that the issuer receives initially and that he pays back at maturity might be different. The future cash flows of a bond can be determined according to the following five parameters: face value, coupon rate, interest payment period, principal payment schedule, time to maturity. The face value (or nominal value) is the amount the issuer has to pay back to the lender in the form of principal payments. The coupon rate of the bond is always expressed as an annual percentage of the outstanding face value. If the interest payment period is shorter than one year, we calculate the time-proportional interest linearly. The interests might be paid annually, semi-annually, quarterly, and so forth. In the case of government bonds, annual and semi-annual interest payments are the most frequent. The simplest and most usual principal payment schedule is when the entire face value is paid in one sum, at maturity. In this case, we call the instrument bullet bond, while in the case of more, partial principal payments, we call it amortising bond.

Figure 4:A short excerpt from a finance and economics textbook (Kürthy et al., 2018) with highlighted words from Sublist 1& 2 of the academic corpus from AWL Highlighter (Haywood, 2007)

Conclusion

This paper has looked at how corpus tools can be used to inform teaching and learning practices in academic vocabulary acquisition by addressing two main questions: which corpus tools are available in this space, and how they can be utilized in material design and development. It was aimed at providing some practical ways for integrating corpus-based tools into language classrooms while at the same time highlighting some of the issues or factors to be considered in the process of choosing what vocabulary to teach and how.

References

Chen, W. and Liew J. (2003). The Civil Engineering Handbook. CRC Press.

Cobb, T. (2007). The Compleat Lexical Tutor. University of Montreal; available at www.lextutor.ca/.

Coxhead, A. (2010). What can corpora tell us about English for Academic Purposes? In A. O'Keefe & M. McCarthy (Eds.), The Routledge handbook of Corpus Linguistics, (pp. 458-470). Routledge.

Coxhead, A. and Byrd, P. (2007) Preparing writing teachers to teach the vocabulary and grammar of academic prose, Journal of Second Language Writing 16, 129-47.

Coxhead, A. and Hirsh, D. (2007). A pilot science word List for EAP, Revue Française de Linguistique Appliquée XII 2, 65-78.

Coxhead, A. (2000) 'A New Academic Word list', TESOL Quarterly 34(2): 213-38.

Haywood, S. (2007). The AWL Highlighter. https://www.nottingham.ac.uk/alzsh3/acvocab/awlhighlighter.htm

Hyland, K. and Tse, P. (2007) Is there an academic vocabulary?, TESOL Quarterly 41(2): 235-53.

Kürthy, G. (Ed.). (2018). Basics of Finance. Corvinus University of Budapest.

Lee, D. and Swales, J. (2006). A Corpus-based EAP course for NNS doctoral students: moving from available specialized corpora to self-compiled corpora', English for Specific Purposes 25, 56-75.

Nelson, M. (2007). Semantic associations in Business English: a corpus-based analysis, English for Specific Purposes 25(2), 217-34.

Nesi, H., Gardner, S., Forsyth, R., Hindle, D., Wickens, P., Ebeling, S., Leedham, M., Thompson, P. and Heuboeck, A. (2005) 'Towards the Compilation of a Corpus of Assessed Student Writing: An Account of Work in Progress', paper presented at Corpus Linguistics, University of Birmingham, published in the Proceedings from the Corpus Linguistics Conference Series, 1, available at www.corpus.bham.ac.uk/PCLC

Nesi, H. and Gardner, S. (2018). The BAWE corpus and genre families classification of assessed student writing, Assessing Writing 38, 51-55.

West, M. (1953). A general service list of English words: With semantic frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology. Longman.

Xue, G. Y., & Nation, I. S. P. (1984). A university word list. Language Learning and Communication, 3, 215-229.

Gözde Durgut has taught EFL for six years and is currently a graduate student in the Applied Linguistics/TESOL program at the University of Alabama. Her research interests include teaching academic writing, corpus linguistics, ESL/EFL pedagogy, and curriculum development. She serves as a member of Fulbright Alumni Association and American Association of Teachers of Turkic Languages (AATT). You can contact Gözde at gdurgut@crimson.ua.edu.