CORPORA IN THE LANGUAGE CLASSROOM: A GUIDE TO USAGE

Nick Canning
Kanda University of International Studies, Japan

Abstract
This article reports on the use of free online corpora in L2 academic literacies classes. The corpus being a record of over 100 million words, phrases and samples of spoken or written English from a variety of sources has long held potential for incorporation in second language learning. The report uses both the Corpus of Contemporary American English and the British National Corpus depending on the online based resource used. The article will examine the rationale for incorporating corpora in the L2 language classroom in terms of developing vocabulary clusters, developing a student’s personal corpus around a topic, examining questions of usage and register as well as error correction. With reference to university level L2 academic literacies classes, the author maintains that the corpus can be an invaluable resource to both learners and practitioners.

INTRODUCTION
The challenge has long existed for practitioners to provide learners in the classroom with authentic examples of how language is used (Reppen, 2010). From the Early studies which reported on the use of the efficacy of concordances (Stevens, 1991: Cobb 1997) to the later studies on the use of corpora in the classroom (Braun, 2007; Reppen, 2010) it has already been concluded that corpora in the classroom can be of benefit to L2 language learners. In addition, the use of a corpus of language for consultation requires students to practice greater cognitive skills. Students have to undertake a process when consulting a corpus from formulating the question, looking at the evidence found and then drawing conclusions to their question (O’Sullivan, 2007). This article will not add evidence based research to further conclude that the use of online corpora should be used in the language classroom; rather it will give practitioners advice on the use of corpora on specific aspects of language learning and show examples of materials for each use.

BACKGROUND
The article will examine how corpora can be used in academic L2 literacies classes. The classes in question are those at a Japanese university specializing in foreign language education. The students are all second year English majors and their proficiency levels range from level 4.5 to 6 on the IELTS band score scale. Learners are required to undertake both academic reading and writing courses which may or may not be conducted by the same lecturer. In the academic reading course students complete both an extensive and intensive reading component. Academic written tasks may take the form of narrative texts, academic essays or reflective journals as well as a lengthy independent research assignment. While the specific uses of corpora detailed in this article have been tailored to fit tertiary academic literacies classes, it is important to note that corpora have been shown to be an effective learner aid in Secondary school language classes (Braun, 2007) and for lower level learners (Ackerley & Coccetta, 2007). Thus the inclusion of corpora may present both practitioners and learners in other contexts with similar benefits.
VOCABULARY ACQUISITION

To foster the ability to learn and use new vocabulary has long presented practitioners with questions over the most effective method not least to those teaching speakers of non-Indo European languages for whom leaning English vocabulary, and in particular academic English vocabulary, can be an area of particular difficulty (Corson, 1997; Mizumoto & Takeuchi, 2009). The use of word lists, for so long a mainstay of an L2 academic course, has been shown to present the learner with little or no context for the vocabulary and as such is a less efficient method of learning than through context (Redounne, 2011). The resulting dilemma means the practitioner has to somehow provide learners with authentic examples of language in context and encourage not only its acquisition, but its usage. The efficacy of the use of corpora in the language classroom has been shown to be beneficial in this regard (Biber et al. 1996; Schmitt & Carter, 2000; Braun, 2007; Liu & Jiang, 2009; Reppen, 2010). Corpora can be used at the very beginning of an academic or language course when outlining study skills. The material (see appendix 1) can be used to introduce the skills of guessing the meaning of new vocabulary from context, vocabulary activation and the concept of collocations, which is necessary to examine vocabulary clusters or ‘chunks’ rather than isolated words (Schmitt & Carter, 2000). The activities guide the learners in recording vocabulary as verb and noun collocations or verb, noun and preposition collocations. The material also shows the ease with which students can then form coherent sentences with the new vocabulary.

The above screen shot from one of the free online corpora shows the parts of speech colour coded allowing the learners a very visual guide to authentic examples language. The vocabulary in question is not necessarily vocabulary that the learners would use. However, what makes this particular corpus even more beneficial is that the context in which a word is most commonly found, in this case from magazines or an academic context rather than from spoken English, is given to the left of the screen shot. This means learners can make an informed choice as to how they wish to use the word, if at all.

The material then looks at ‘activating vocabulary’ or using it to formulate questions to ask classmates, with the idea of being able to attach a personal memory or experience to it. The activity in appendix 1 should provide a model for students. Using the corpus in fig 1, students can then find authentic examples of the word in use and, rather than relying on teacher generated discussion prompts, use the examples to create their own questions. Learners should feel greater engagement in a discussion task that they themselves create. These ideas echo those of Vygotsky (1978) in that students will likely learn more effectively if they are actively constructing knowledge. This will then provide a future base
for further knowledge accumulation, a shared understanding of the topic of study and future vocabulary acquisition.

Learners may also examine different uses of vocabulary and in this respect the online corpora can assist in developing the ‘generative use of vocabulary’ (Nation, 2008). The author states that a more advanced stage of vocabulary use than recall or use in a similar context is the generative use of vocabulary: to use the vocabulary in a different context or a different meaning of the word. The corpus in fig.1 also shows how learners can search for the word by type of collocation depending on how the word is used in a sentence, for example as a noun or adjective. This means learners can examine how to use words in a way that suits their immediate needs. In all cases the collocation in question can be clicked to see a list of authentic examples. The corpus also allows the learner to search for alternatives to build up associations of synonyms related to the original word. Such associations or ‘vocabulary depth’ play a significant role in the increase of vocabulary size (Akbarian, 2010). A question formulation or other generative use activity could be used after vocabulary input from any text either in class or, more significantly, out of class.

WRITTEN ACADEMIC DISCOURSE

The difficulties encountered by second language learners in academic writing have long been studied and documented (Raims, 1985; Asaoka & Usui, 2003; Lee, 2008; Fernsten & Reda, 2011). The problems learners may face vary from difficulties with the writing process itself (Raims, 1985; Cumming, 1990; Stapleton, 2010) or difficulties incorporating learners’ identities within the discourse patterns of academic writing (Ha, 2009) to lexico-grammatical accuracy (Asaoka & Usui, 2003). The latter authors make an important distinction between problems students faced. ‘Macro level’ problems refer to the content of a piece of writing or how the ideas are supported or explained, whereas those described as ‘surface level’ refer to lexico-grammatical problems. While the multitude of difficulties experienced by second language writers cannot alone be countered by a single intervention, the introduction of corpora may assist learners with ‘surface level’ problems. In particular, with regard to paraphrasing which learners find problematic (Maas, 2002; Walker, 2008). Difficulties arise from reformulating sentences which may contain complex ideas. Students may not have sufficient language skills to do this or may not realize how much reformulation is required which may lead to instances of unintentional plagiarism (Walker, 2008). While Maas (2002) advocates the use of E-Prime in paraphrasing to prevent plagiarism it seems illogical to restrict students, who are already struggling to express the ideas of others in their own words, from using certain grammatical structures. Rather, the example material (see appendix 2) guides the students through different stages of paraphrasing and uses the same online corpora as in fig. 1 to examine how synonyms can be used to convey the same meaning or different forms of the original words or synonyms. Through an analysis of the different words or word forms from authentic examples, students will be better equipped to re-phrase the words of others in their own writing.

A further application for corpora in accessing written academic discourse for second language learners is in the response to written corrective feedback. The subject of its efficacy is an enormously controversial one (Guenette, 2007; Ellis, 2009). While there have been several studies which cast doubt over whether corrective feedback should be used at all see Truscott, 1996; Ferris, 2003, proponents argue that many of the studies which indicate a limited impact of corrective feedback suffer from flaws in their design (Guenette, 2007). Indeed it would seem few teachers would seriously advocate giving no feedback, rather many practitioners feel that the full potential of corrective feedback is not being utilized (Hyland & Hyland, 2006). In terms of what techniques are available to
the teacher, Ellis (2009) makes an important distinction between the different types of feedback. Direct feedback, according to the author, will involve the teacher providing the learner with a correction to an error. Whereas indirect feedback will only tell the learner, perhaps by way of highlighting or through notes in the margin, that an error is present. Indirect feedback is often preferred by practitioners to guide the learners to solve their own problems (Ellis, 2009). A separate form, and of the most common forms according to the author, is metalinguistic corrective feedback and specifically the type of feedback that involves a code. This is in contrast to metalinguistic feedback which gives the students a number in the margin, with the frequency of the error also given, which corresponds to a grammatical explanation at the end of the piece of writing. The application of online corpora in this context will focus on the use of metalinguistic feedback with a code. The example error correction practice material (see fig. 2) shows the codes above the errors although the same codes could be used in the margin thus not indicating the specific position of the error.

**Look at the following paragraph with your partners and decide how to correct the mistakes:**

T.

Almost _ef^ students in this class _ave a good experience at High School. The main
pl       sv       pl
reason for this _ elevation activities, talking with friends and school _ip. First and
T.       T.
foremost, lots of students said they _elop to ^ club activity. They _joy playing with
col
their _rrends in ^ club activity and even though it was hard they could _ake a good
w,f+pl.   T.       pl.   T.
relate with their friends. The next reason is talking with friends. Many learner _ike
v,f

*fig. 2 an extract from a collaborative correction code practice activity*

![fig. 3 a screenshot from www.netspeak.org](https://www.netspeak.org)
The image above in fig. 3 is of an online corpus which allows students to check for a missing word or up to three words, compare two different options of word choices, find synonyms and also check word order. This particular site allows students to correct missing word errors such as incorrect dependent prepositions or incorrect collocations by searching for them and examining examples of authentic usage, for all entries examples sentences can be seen by clicking the plus sign. This allows students to respond to corrective feedback accurately rather than making a guess at what the teacher may want the student to write. While metalinguistic feedback which includes a summary of usage has been shown to be effective, (Ellis, 2009), the type of dependent preposition or collocation error which are not based on a grammatical rule will require the teacher to guide the student to the answer differently.

The following image in fig.4 shows how a student would find out which word was missing between almost and people in a sentence. Here the student will notice that almost every entry is a quantifier or numerical figure. Through reflection the student will be able to determine that almost can be used together with quantifiers and numbers when talking about the amount of a noun and in this case it is different to the word ‘most’ despite both words sounding similar. Similar searches could be conducted for word order errors with similar reflection perhaps providing students with grammatical rules which can be used in future writing.

![fig. 4 a screenshot from www.netspeak.org showing search results](image)

In this sense online corpora may allow learners to engage in metalinguistic reflection, identify the gaps in their language learning and foster second language acquisition as well as accuracy development, all of which are the primary goals of written corrective feedback (Van Beuningen, 2010).

With the development of metalinguistic reflection and greater learner autonomy in second language writing, the application of corpora in peer feedback may also be of benefit to practitioners. Although generally not given the same weighting as teacher feedback, see Yang et al. (2006), peer feedback has been shown to improve learner autonomy in second language writing classes (Vilamil & Deguerro, 1998; Tsui & Ng, 2000) and critical reasoning (Berg, 1999) the latter stating that students would have to compare the peer feedback with their own knowledge and evaluate it based on what they know and in turn, whether to apply the feedback to their writing. It is with this stage that the application of corpora could be used. The student has the opportunity not just to evaluate the peer feedback against their own knowledge but against corpora. This application may afford the learner a greater degree of autonomy in the decision to adopt the feedback or not. In this regard perhaps, peer feedback of word choice or word pairing may provide the learner with the most useful application of corpora.
An additional use of corpora to facilitate the access of written academic discourse is in increasing the size of learners’ topic specific vocabulary. All learners, whether non-native speakers or otherwise, will need to increase their knowledge of specific terminology related to the topic of study. This terminology will be necessary for a variety of written academic texts on the topic, this requires learners to not only be familiar with the terminology but also be able to use it in academic discourse. The application of corpora here can benefit the learner, through analysis of how terminology is used in authentic examples, to develop their own corpus on a specific topic (Charles, 2012). The example below, see fig. 5, shows an example of a resource for second language learners. The example shows verb noun collocations for the word war. The reasons this word was chosen were that it is difficult to find verb noun collocations with a simple search of authentic texts as the word is most often used in conjunction with an adjective indicating the name of a war in question. A student also chose to examine an area of this topic for a research assignment and found the terminology difficult to use. In a case where the most common use of a word is not the required use, a learner can use different search strategies such as the online corpus in fig. 1 to find word pairs.

Vocabulary log:
For your research project you will need to use specialised vocabulary in your essay. For example, if you are researching a topic related to war you will need to know which verbs collocate with war.

To do this you can either look at www.wordandphrase.info, www.just-the-word.com, wordneighbors.ust.hk or www.netspeak.org to find example sentences.

<table>
<thead>
<tr>
<th>Word or Expression</th>
<th>type of word</th>
<th>Notes (definition, examples and collocations etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>war</td>
<td>noun</td>
<td>an armed conflict usually between countries</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Many young men went to war.</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>War broke out in 1914 in Europe.</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Many countries entered into war.</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>The country was torn by war.</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>France declared war on Germany</em></td>
</tr>
</tbody>
</table>

fig.5 an example of a record of topic specific terminology

The application of corpora here has implications not just for English majors in tertiary education but for those learning at International High Schools where English is the language of instruction or similarly an international branch campus where the assignments will be at undergraduate or post graduate level and require extensive self study of both the topic and terminology.

LIMITATIONS

While although the application of corpora in second language learning presents several benefits to learners and practitioners alike, the introduction of corpora alone will likely be insufficient to foster usage among learners. Practitioners will need to model usage for learners frequently to display the benefits. The most easily adopted use may be that of responding to metalinguistic feedback. Students will require scaffolded introductions and activities such as the ones in fig. 2 or in appendix 2 to see the immediate benefits of corpora. Whereas in the classroom the vocabulary activation activity in
appendix 1 can be used in class and used to foster corpus usage as part of a literacies lesson. The vocabulary activation activity can also lead to greater skills of linguistic enquiry. Learners will often consult their dictionaries for a word to match their meaning yet the translation given may not be used in the same way in English. If learners are shown that finding out how a new word is used will save them time when correcting work, for example a student may just receive an error code above the translated word as feedback for a re-drafting exercise, this will then necessitate consultation of corpora which could have been done previously. While it is unlikely learners will immediately develop sophisticated linguistic research skills, through emphasizing that the use of corpora can save a student time and be a more efficient method of language learning, greater usage can be fostered.

CONCLUSION

In the context of university second language literacy classes, this report has outlined several applications of different free online corpora. Practitioners may find that learners develop greater vocabulary acquisition from examining how vocabulary is used in authentic contexts which will help to foster not only the students’ use of new vocabulary but generative use and also word associations with new vocabulary, thus developing their vocabulary depth. Greater engagement may also be fostered through having more student generated discussion tasks, rather than teacher generated ones, to experiment with vocabulary usage. Corpora have applications with the access of written academic literacy with regard to paraphrasing and written corrective feedback, from teachers and possibly form peers. The use of corpora with regard to metalinguistic feedback may also develop greater linguistic reflection skills and critical reasoning. Corpora can also be applied in developing a learners own corpus of topic specific terminology. While certain limitations remain with regard to fostering learner usage, the different uses of corpora provide an invaluable resource for practitioners and second language learners alike in a variety of contexts.

References:


APPENDIX 1

Learning and recording vocabulary

Understanding meaning from context

Look at the following sentences and try to guess the meaning of the words in bold without using a dictionary.

1) "I don’t know, I don’t think she has forgiven me. I tried to talk to her yesterday and she completely blanked me. I was absolutely distraught.”
2) Many of the buildings in this town have been long since abandoned. No-one has lived in them for years.
3) When he asked her about her boyfriend she looked quite flushed and couldn’t answer.
4) Many holiday commercials show idyllic scenery, but actually the places are not so beautiful in real life.
5) A: Really? You don’t like 1D?
   B: No, they’re annoying.
   A: Actually, they’re quite talented and lots of people get the wrong impression about them. They’re quite misunderstood.
6) Workers who have to do large amounts of overtime are more likely to succumb to illness or stress.
7) It’s always difficult to strike up a conversation with people on public transport; they just think you are weird.
8) Why do some students make such a minimal effort with their homework? It really sends teachers round the bend.

Helpful hints:

1. Look at the situation (the context) we know how people feel when they are asked about their boyfriend or girlfriend. We know how they look.

2. What kind of word is it? Is it an adjective, noun etc? This will also help you find out the meaning because you know the function of the word in the sentence.

3. Does the word sound like or look like other words? Maybe you can guess the meaning from this.

4. Is the word part of a word family? Are there suffixes (..er/...ment) or prefixes (mis.../re....) that help you to understand the word?

ACTIVATING VOCABULARY

When you find new words or phrases that you want to use, a good idea is to use them in your written or spoken English. By experimenting with different ways of using the words you begin to attach memories or experiences to them. When you are able to share an experience of being flushed, then you are more likely to remember the word when you want to retrieve it (use it) or when you see it, because the word now has more meaning for you.

Look at the following questions with in small groups and try to think of experiences in your life to answer them.

1. Have you ever been blanked by someone? What happened? Would you ever blank someone?
2. Can you think of other reasons why someone might look flushed?
3. Do you find it difficult to strike up a conversation with new classmates? Why, why not?
4. Can you think of any synonyms for the word idyllic? What kind of place seems idyllic to you?
5. Think of different reasons why you might feel distraught.
6. How might someone send you round the bend?
Recording vocabulary:

When recording vocabulary it is important to not only make a note of the **type of word, the pronunciation** and the meaning but also other words which go together with it or **collocate** with it.

Look at the following example:

It’s always difficult to **strike up a conversation** with people on public transport, they just think you are weird.

We know we have strike up (phrasal verb) + conversation (noun) + with (prep) + object.

Normally you cannot change the preposition, but you can change the verb. E.g.

**have**

<table>
<thead>
<tr>
<th>verb</th>
<th>a conversation +</th>
<th>with</th>
<th>+ object (in this case usually a person)</th>
</tr>
</thead>
</table>

**hold**

By looking at websites such as [www.wordandphrase.info](http://www.wordandphrase.info) you can find more words that collocate with your new vocabulary. Why? Well, this makes it much easier to use it in a sentence. If you have the words **have** + **a conversation** + **with** and you know the following word must be an object then it’s very easy to make a full and correct sentence E.g.

I had a conversation with Misaki yesterday about the new course. Easy!
Simply input the word and click on the word to the right to see the results.

Find other collocations for the following words:

<table>
<thead>
<tr>
<th>verb</th>
<th>noun</th>
<th>verb</th>
<th>noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>succumb to</td>
<td></td>
<td>succumb to</td>
<td>stress</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ok, you might not use the words ‘succumb to something’ yourself but, this is also important for building up your academic vocabulary which you will need for reading and also writing (especially in semester 2 when you write your research report).
APPENDIX 2: Paraphrasing review and practice

In the following activity we look at three separate paraphrasing techniques.

1. Using synonyms
2. Using different word forms
3. Using different sentence structure

The goal of paraphrasing is to use the ideas of others, and make reference to them, to support your own points.

Why do you think it is important to tell the reader where the information comes from?

Why is it also very important to not change the meaning of the original author?

For the following activities you will just be paraphrasing the original statements.

Here is what you do:

Step 1) Make sure you understand the original sentence, read it several times and look up any new words.

Step 2) Note down the main ideas

Step 3) Put the original away and work with a partner to create a paraphrase.

Look at the following website www.wordandphrase.info to help you find synonyms and show you examples of how they are used. Do the synonyms reflect the same meaning as the original?

You could also look at other verbs which collocate with the verbs in your sentence on www.just-the-word.com (remember to choose common collocations which match the original meaning) or change the word forms and look at how they are used.

Step 4) Check that you have not copied vocabulary from the original sentence, or followed sentence structure too closely.
Paraphrase the following sentences using the above techniques as a guide

1. Few residents travel to work by bus.

2. Only Sony suffered loses that year.

3. Members of staff are allowed to wear casual clothes on Friday.

4. It is not always easy to predict how well a product will be received by the public.

5. It has been said that learning lists of words is not an effective use of time.

6. After only a nominal pay increase, the staff made a formal complaint to management.

7. Cycling is becoming increasingly popular in Japan.

8. There doesn’t seem any point in linking these two towns by rail.