Contextualizing Language Learning: The Role of a Topic- and Genre-Specific Pedagogic Corpus

Chau Meng Huat
Chau English Language Learning Centre, Malaysia

Introduction

Jackie F.K. Lee has argued for the use of corpora in language learning in the 2003 April issue of the *TESL Reporter*. Among the benefits suggested of using corpora in the classroom are (1) that it helps students to explore for themselves the correct usage of the target language, (2) that it assists in vocabulary building and (3) that it provides evidence for debatable usages. A stimulating context is thus provided in which the learner develops strategies for self-discovery. Corpora then serve as a useful tool to promote learner autonomy.

While acknowledging the various benefits noted and indeed welcoming concordance technology to take place in the language classroom, this article contends that a general corpus such as that of the World Wide Web Access to Corpora Project (W3-Corpora at http://clwww.essex.ac.uk/w3c, a Website address generously provided by Lee) may prove inaccessible in the first instance to many students, especially those of lower level of proficiency. Further, because they are unfamiliar with concordance lines, complete reliance on concordances from a general corpus may lead to the students running the risk of getting confused. As Fox (1998, p. 39) observes, each concordance line "is discrete, often on a completely different subject, and also most of the lines do not even show whole sentences".[1]

This unfamiliarity with the language and message content is a real concern which requires attention from every teacher before wishing to use the general corpus. This is not to suggest, however, that the general corpus should be relegated to the wings while "traditional" classroom pen-and-paper work takes centre stage. Rather, it suggests an alternative role for the corpus: that of an *additional*, *valuable* resource for evidence confirmation and language learning—a point which we will return to in the conclusion section.

This article proposes the use of a topic- and genre-specific pedagogic corpus as the main learning resource to address the concern of language and content unfamiliarity.

We will first consider some design issues of the proposed corpus and then discuss some ways to use it in as well as outside the classroom. As we will see, a topic- and genre-based corpus frames language learning within familiar and meaningful contexts, thereby promoting learning opportunities and offering a very good return for students' learning effort.

Developing and Using Topic- and Genre-specific Pedagogic Corpora

The term "pedagogic corpus" was first introduced in Willis (1993) to refer to any materials, written and spoken, to which the learner is or has been exposed. A topic- and genre-specific pedagogic corpus then, as the name suggests, is assembled through a careful selection of texts on a recurrent topic and genre. If we have, for example, decided to collect a corpus for a unit of work based on the topic "English as an international language", we could select eleven topic-related, factual texts to form a factual pedagogic corpus on that topic. If the syllabus for a course consists of twenty units of work, then we will have twenty topic- and genre-based pedagogic corpora of ten to fifteen texts each.

It is crucial that the developed corpora be exploited in relation to tasks framed within a recursive EAR framework. The "E" stands for Exploration. It is a stage at which the learner carries out a series of well-designed tasks that are derived from a particular topic- and genre-based corpus. These tasks could include, for instance, identifying main points, problem-solving, listening to recorded spoken material, pair or group discussion and public speaking, in order of linguistic and cognitive difficulty.

After the learner has processed the particular corpus for meaning through a progressive series of tasks, the corpus will be exploited for focusing the learner on examining the target language. The processed texts then provide valuable input for the learner to notice recurring linguistic features within familiar contexts. This represents the "A" or Analysis stage in the EAR framework. We will look in detail at a couple of language analysis activity types in the section to follow shortly.

The final stage is the "R" or Reflection stage at which students are encouraged to reflect on the learning process with the aim to help them to take control of their own learning. They are guided to reflect on, among other concerns, the texts they have studied, the linguistic occurrences they have come across and the possibility of other options. Possible reflection questions to be posed to them include: *What are the features of the texts that make them typical or look alike? What do you notice about the words, phrases, and tenses in these texts? Do you notice repeated words and phrases*

across the texts? Could you suggest some alternatives to these repeated words from the texts? (See Chau, 2003, for further, relevant discussion.)

In the rest of this article, we will consider two activity types that can be carried out both in- and out-of-class. We will focus exclusively on the Analysis stage here, although the Exploration as well as the Reflection stage deserves equal, if not more, attention in the classroom to help the learner to acquire the target language.

Exploiting Topic- and Genre-specific Pedagogic Corpora for Language Learning

As already indicated, the value of a topic- and genre-specific corpus is that it familiarizes the learner with the language and message content of a particular topic and genre and situates learning in context. This learning condition is further strengthened when the learner has gone through the Exploration stage. Given the learner's past language experience of the input, the language to which the learner has been exposed thus constitutes familiar and meaningful linguistic data for language analysis work to take place in context.

By way of illustration, we will look at two activity types: focusing on patterns and focusing on collocations. The activities are derived from a topic- and genre-based corpus of eleven texts on "English as an international language" from a number of Malaysian course books. We will see how an eleven-text topic- and genre-specific corpus can offer a wealth of learning opportunities to the learner.

Focusing on Patterns

Previous corpus-driven studies, most notably those by the Birmingham team of corpus linguists that draw exclusively upon the data from the Bank of English corpus (e.g., Francis et al., 1996; 1997; 1998), reveal that there are observable patterns that are associated with particular lexical items in English texts. The patterns of a word, as Hunston and Francis (2000) point out, refer to "all the words and structures which are regularly associated with the word and which contribute to its meaning" (p. 36). They give the following examples of the pattern of verbs with introductory *it* and with a *that*-clause which expresses a reaction to a fact or piece of news:

1 *It* amuses *me that every 22-year-old now wants to own property.*

2 *It* frightens *me that kids are now walking around with guns.*

3 *It* puzzles *me that people are willing to pay any taxes at all to this Government.*

(Hunston & Francis, 2000, p. 264)

Like Hunston and Francis, I believe that information about frequent patterns is of great value to learners and should be incorporated into language lessons for learners to notice how the target language works. After all, the central challenge for language teaching is to develop learners' communicative language ability which includes, among others, the ability to use the language naturally and fluently. Let us now turn to examples of language patterns of some selected words.

Nouns and patterns (1): the case of ENGLISH

Since the topic under discussion is "English as an international language", we can draw our students' attention to the patterns of some useful topic-related words such as *ENGLISH, LANGUAGE, INTERNATIONAL, WORLD,* and *STANDARD*. A search for concordance lines containing the word *ENGLISH* reveals that there are 195 matches or occurrences in the corpus; *LANGUAGE 87*; *INTERNATIONAL 13*; *WORLD 29*; and *STANDARD 22*. These numbers indicate sufficient examples for students to examine the use of each word in the context of the topic under discussion. We can either prepare in advance for the students a set of concordance lines containing all the instances of the words to be examined or ask them to find out, depending on the frequency of the words, all or some of the sentences containing those words. (It is important to note that ten to fifteen familiar texts provide manageable data for language analysis.)

If we consider the noun *ENGLISH*, we will notice that this word behaves as any noun does, as in *English can be considered as a world language*. One frequent pattern, however, emerges: the *noun + preposition + English/the English language* structure. There are 62 of them in total, with 48 occurrences of the pattern *noun + of + English/the English language* and 14 occurrences of *noun + in + English/the English language* in the corpus. Here are some recurring examples of the pattern *noun + of + English/the English language* that students can (and should) be guided to notice:

a good knowledge of English
a word of English
a working knowledge of English
our (poor) command of English
the development of English
the growth of English
the importance of English
the knowledge of English
the role of English
the spread of English
the standard of English
the use of English

a good command of the English language
ample command of the English language
the standard of the English language

While this pattern may be too common to qualify for an entry in a reference grammar, it merits learning effort for students to notice and learn the accompanying phrases in chunks which are in their most natural environment. The students can be required, after examining the phrases, to complete the concordance lines (with these phrases omitted and with more than one acceptable answer in certain gaps), first, taken from the original pedagogic corpus and later, from other resources like the W3-Corpora website maintained by the University of Essex recommended by Lee or the Internet through such search engines as Google (at www.google.com). The following lesson would focus on the *noun + in + English/the English language* pattern (e.g. *proficiency in English, sources in English*) and the use of *ENGLISH* serving the function of an adjective (e.g. *the English Week, English programmes*).

Nouns and patterns (2): the case of LANGUAGE

Let us turn to another example, the noun *LANGUAGE*, which occurs 87 times in the pedagogic corpus. Here is a sample of 25 random concordance lines from the corpus:

makes English merit the status of a world	language. Already English has been accorded this
and learn about the complexities of the	language and the many errors commonly made by
such, the importance of English as a world	language cannot be denied. As the nation strides
in the percentage of passes in the English	language. Concern was raised in the mass media
foreign language if we are to look to another	language for sources. Although many people
must have "output" in order to improve our	language further. What he meant is that we must
The chances of a pupil's exposure to the	language have been greatly diminished or
English than to start off with a foreign	language if we are to look to another language
lingua franca increasingly used as a second	language in important areas of the world', says
realise that English is a very important	language in the world and it is definitely important.
colleges and universities, and is also the	language of administration. The importance of English
who are bilingual. English is also the	language of art and literature. Many great literary
books are all in English. English is the main	language of communication of commercial firms and
the United Nations. It is the language - the	language of commerce and industry, science, arts and
there. Not only has English enriched the	language of these countries, but it has also added a new
is an international language, it is also the	language of science and technology. Proficiency in
role of the national language as the official	language of the country remains unquestionable
who feel that English is an important	language that can propel the nation towards develop

in Bahasa Melayu except for the English
speaking countries. It is an ideal second
natural result of the emphasis on the national
the store of knowledge in the national
complexities the many different parts of the
of English. It is the most widely spoken
and origins. If you are good at the English
language. Thus the opportunity for using.
language to master because it is used in business
language to create a national identity for Malaysians
language. What should really be of concern to
language which make it difficult to understand by
language. You need not worry that you will not be
language, you will find the doors of English, Canadian

Given these familiar, concordance lines, students can be required to identify the patterns of *LANGUAGE*. A simple question like "What most frequently comes after language?" or "Is there any same category of words that often come before language?" can elicit encouraging responses. As we can see, *LANGUAGE* is frequently followed by the preposition *of* and a noun. This is the pattern **N *of* n** (noun followed by *of* and noun), as follows:

the language of administration
the language of art and literature
the (main) language of communication
the language of commerce and industry
the language of these countries
the language of science and technology
the (official) language of the country

Like most other nouns, *LANGUAGE* frequently follows an adjective, forming the pattern **adj N** as follows:

foreign language
important language
main language
national language
official language
second language
spoken language

The introductory it + adjective + that-clause pattern

As illustrated earlier, the introductory *it + verb + that-clause* pattern is a common structure in English. This structure also includes the introductory *it + adjective/noun + that-clause* pattern (see Hunston & Francis, 2000, pp. 264-65). We will now look at the introductory *it + adjective + that-clause* pattern that is found in the eleven-text pedagogic corpus. There are eight of them in the corpus, as follows:

4 *It is* good *that you have realised the importance of English and are willing to work hard to improve it.*

5 *Owing to Malaysia's growing participation in international affairs, it is* imperative t*hat those who represent our country are effective in the language in which meetings are conducted.*

6 *English is a very important language in the world and it is definitely* important *that we become proficient in it.*

7 *However, considering that Malaysians study English from their first year in primary school, it is only* logical *that it would be easier to promote English than to start off with a foreign language if we are to look to another language for sources.*

8 *It is only* natural *that Malaysians who wish to be employed must have a working knowledge of English.*

9 *It is* true *that English is the mother tongue of the people of Great Britain.*

10 *Of course, it is* true *that the government has sent engineers for training in technologically advanced countries like Germany, Japan and Korea, where English is not the main or even a major language.*

11 *It is* undeniable *that the standard of English is declining in almost all, if not all, Malaysian schools.*

Again, students can be required to identify these sentences after the teacher has highlighted the introductory *it* + *adjective* + *that-clause* pattern in the classroom. Working with a personal computer with a word processing programme would prove very efficient in locating these sentences in the electronic version of the texts, but it is worth repeating that ten to fifteen printed texts which have been processed before for meaning at the Exploration stage present little difficulty to students in searching for this structure.

For further examples on this pattern, any curious students, as suggested by Hunston and Francis (2000), can always be directed to relevant reference books such as Francis *et al.,* (1998). With more practice through different units of work that involve a wider range of topic- and genre-specific corpora, students are likely to develop a greater sensitivity to various patterns of the language as well as the necessary skills to work out the correct usages by themselves.

Focusing on Collocations

A closely related notion of "patterns" is collocation, that is, the tendency of two or more words to co-occur to form a piece of natural-sounding language in speech or in writing. For example, the collocates of *LANGUAGE*, as shown earlier, include *foreign, national, official, second* and *spoken.* The significance of collocation is realized when one considers these stored sequences of words as the essential bases of language learning and use (Ellis, 2001). Here are some examples of the range of collocation for *LIGHT* shown in the Oxford Collocations Dictionary for Students of English (2002, p. ix):

adjective + noun:	*bright/harsh/intense/strong* **light**
quantifier + noun:	*a beam/ray of* **light**
verb + noun:	*cast/emit/give/provide/shed* **light**
noun + verb:	**light** *gleams/glows/shines*
noun + noun:	*a* **light** *source*
preposition + noun:	*by the* **light** *of the moon*
noun + preposition:	*the* **light** *from the window*

In the classroom, we should initially focus on topic-related collocations that consist of very frequent words since it is these words that learners often meet and are likely to employ for productive use. Very frequent words are generally also useful words especially if they fall within the first most frequent 1,000 word families in English. Focusing learners' attention on these first 1,000 words gives a very good return for learning effort because they cover "around 75 percent of the running words in academic texts and newspapers, over 80 percent of the running words in novels, and about 85 percent of the running words in conversation" (Nation, 2003, p. 136).

For illustration purposes, we will look at the collocates of one such word, *KNOWLEDGE,* that occurs twelve times in the present pedagogic corpus (*ENGLISH* and *LANGUAGE*, incidentally, also fall into the group of the first most frequent 1,000 words).[2] Here are the twelve sentences that contain *KNOWLEDGE:*

12 *A very important reason for the use of English is that the world's* knowledge *is enshrined in it.*
13 *Countries in Asia and Africa that were till recently under British rule get their scientific* knowledge *and technological know-how from English books.*
14 *It is the* knowledge *of English that helps these countries maintain their high level of intellectual and scientific training and achievement.*
15 *English is a key which opens doors to scientific and technical* knowledge *indispensable to the economic and political development of vast areas of the world.*

16 *It opens doors to scientific and technical* knowledge.
17 *It is only natural that Malaysians who wish to be employed must have a working* knowledge *of English.*
18 *If you wish to go abroad, be it for further studies or for a tour, you need to have a working* knowledge *of English.*
19 *If you have a good* knowledge of English, *you will be able to read and enjoy these great literary works.*
20 *Proficiency in English enables students to have access to current scientific and technological* knowledge, *much needed to help Malaysia become a fully developed nation by 2020.*
21 *It also improves our knowledge of current affairs.*
22 *Proficiency in English may thus be used as a vehicle to progress as it can increase the store of* knowledge *in the national language.*
23 *Since this has now been achieved, Malaysians can once again look to English in order to keep up with the latest developments in the world and acquire a greater store of* knowledge.

Given the sentences above and provided with the range of collocation to be matched with the relevant words, students can be required to identify all the collocates of *KNOWLEDGE* and match the collocational patterns with the identified collocates, as follows:

adjective + noun:	*current scientific and technological/good/ scientific/scientific and technical/working* **knowledge**
quantifier + noun:	*a greater store/the store of* **knowledge**
verb + noun:	*acquire/get/have/improve/increase* **knowledge**
noun + preposition:	**knowledge** *of English/current affairs*

It is useful to encourage students to consult collocation dictionaries as a follow-up activity for further examples and collocations. A look at the entry, *KNOWLEDGE*, in the *Oxford Collocations Dictionary*, for instance, reveals the following additional examples and one new collocational pattern:

adjective + noun:	*considerable/great/vast/sound/thorough/deep/detailed/ extensive/wide/local/direct/first-hand/up-to-date/general/factual/practical/professional/academic* **knowledge**
verb + noun:	*gain/demonstrate/apply/broaden/extend* **knowledge**
preposition + noun:	*in the/to somebody's/with somebody's/without somebody's* **knowledge**

Having completed this initial, "warm-up" in-class exercise, students can then be required, following the same process and repeating the same type of language analysis work as performed earlier, to identify as homework the collocates of other, selected high frequency words. If the class consists of mainly lower proficiency students, the teacher may need to provide them in advance with a framework of the range of collocational patterns (as shown in the entries of a collocations dictionary) of the selected words. Here is an example of such tasks focusing on the word *STANDARD*:

Which words do you think frequently occur with the word standard? *Write them down by searching through the eleven passages and group them into the following:*

adjective + *standard:*
noun + *standard:*
verb + *standard:*
preposition + *standard:*
standard + preposition:

Now add three more words to each group (the more, the better!) by checking in the dictionary. You may want to consider only the first three groups here.

In addition to the language analysis work suggested above, a class that consists of intermediate students or above can be further encouraged to identify, list and keep in a notebook topic-related collocations and classify them as homework according to categories central to the topic. This more challenging task may look like the following:

Complete the following exercise with all the relevant words that you have come across in the eleven texts on English as an International Language*:*

<u>Words/phrases connected with the reasons we need to be proficient in English</u>:

adjective + noun:
quantifier + noun:
noun + noun:
noun + verb:
verb + noun:
verb + verb:
verb + preposition:
adjective + preposition:
adverb + adjective
adverb + verb:
verb + adverb:

Words/phrases connected with the current situation of the standard of English in the country:

adjective + noun:
quantifier + noun:
noun + noun:
noun + verb:
verb + noun:
verb + verb:
verb + preposition:
adjective + preposition:
adverb + adjective
adverb + verb:
verb + adverb:

Words/phrases connected with the possible measures to improve the standard of English in the country:

adjective + noun:
quantifier + noun:
noun + noun:
noun + verb:
verb + noun:
verb + verb:
verb + preposition:
adjective + preposition:
adverb + adjective
adverb + verb:
verb + adverb:

It should be noted that this task may at first prove time-consuming and cognitively demanding. Personal teaching experience of assigning this task suggests that many students (including some highly motivated ones) may, upon the completion of the task, grumble about having had to carry out the task. After a discussion and feedback session of the work completed, however, most report that they appreciate the value of the task and that it has been a worthwhile and rewarding experience.

Many students point out that through identifying and classifying topic-related vocabulary (a process in which they consciously attend to, among other things, sets of words with similar meanings and/or functions), their vocabulary is much enriched. Some even see this vocabulary enrichment activity as being essential to providing them with the necessary vocabulary to produce a good summary and to effectively express

their ideas when writing on a related topic. Perhaps more importantly, they welcome similar tasks in the following units of learning which draw upon different sets of topic- and genre-specific pedagogic corpora.

Conclusion

It is now well recognized that language learning does not take place in a vacuum. Contextualization is the key word here and it is realized, as proposed in this article, through familiarity with the language and with the message content in known, meaningful contexts. Through sustained exposure to, exploration with, analysis of, and reflection on topic- and genre-specific pedagogic corpora, this familiarity is promoted, developed, and enhanced. Not only is the learner's schema of text structures and central ideas of the selected topics repeatedly reinforced and strengthened, frequent recycling of the common lexicogrammatical features is also ensured.

At the outset of this article, I noted that a general corpus may contain patches of language beyond what the learner can process. This is aggravated by the fact that concordance lines which illustrate such instances of language show only partial or incomplete sentences. Yet at the same time I suggested that a general corpus can serve as an additional, valuable resource for evidence confirmation and language learning. The "trick" here is knowing when to exploit a general corpus to complement the proposed pedagogic corpus.

I have argued that a pedagogic corpus on a recurrent topic and genre provides for learners familiar texts to process for meaning and accessible linguistic data for the study of the target language. I have illustrated the importance of the proposed corpus for contextualizing language learning by suggesting as examples two activity types: focusing on patterns and focusing on collocations. I also contend that because they are familiar with the content as well as the recurring language features of the corpus, learners have little difficulty in carrying out language analysis work even in the form of concordance lines. This past experience with concordances is useful to facilitate the transition to analyzing language data based on concordances from a general corpus.

Given that topic- and genre-based pedagogic corpora consist of only ten to fifteen texts each and that they have been designed as or intended to be *pedagogic* corpora, not *research* corpora, there are admittedly some aspects of the target language that they are unable to reveal to the learner. For instance, debatable usages whose empirical foundation can be traced in general corpora, as shown in Lee (2003), are not available for examination in pedagogic corpora. Also, there will be items that require more exemplification due to their limited instances. This is when general or research corpus is called for as a valuable tool to complement the proposed pedagogic corpus, the core learning resources that promote language learning in context.

Notes

1. To address this problem, Fox suggests spending some time with students to get them used to looking at concordances.

2. Alex Heatley, Paul Nation and Averil Coxhead of Victoria University of Wellington have devised a program known as RANGE which enables users to identify, among other things, the first and second most frequent 1,000 words in English. A free distribution or download of RANGE is available at http://www.vuw.ac.nz/lals.

References

Chau, M. H. *Text Matters in Language Learning*. Manuscript submitted for publication.

Ellis, N. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.

Fox, G. (1998). Using corpus data in the classroom. In B. Tomlinson (Ed.), *Materials Development in Language Teaching*. Cambridge: Cambridge University Press.

Francis, G., Hunston, S., & Manning, E. (1996). *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.

Francis, G., Hunston, S., & Manning, E. (1998). *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.

Francis, G., Manning, E., & Hunston, S. (1997). *Verbs: Patterns and Practice*. London: HarperCollins.

Hunston, S., & Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.

Lee, J. F. K. (2003). Use of corpora in language learning. *TESL Reporter, 36*, 23-29.

Nation, I. S. P. (2003). Vocabulary. In D. Nunan (Ed.), *Practical English Language Teaching* (pp. 129-152). New York: McGraw-Hill.

Willis, D. (1993). Syllabus, corpus and data-driven learning. *IATEFL Conference Report: Plenaries*.

About the Author

Chau Meng Huat (chaumenghuat@hotmail.com) teaches students of ESL/EFL from elementary to high schools in a small language school he runs in Malaysia. His work on the use of texts and corpus analysis techniques to raise ESL/EFL learners' consciousness about the target language was initially inspired by the work of Susan Hunston of the University of Birmingham on corpus linguistics and, thereafter, by the students in his classroom. Meng Huat is currently developing ways of extending access to texts as learning resources for students and looking at ways of providing supporting frameworks within which language learning can be promoted.