Introducing Online Corpora into ELT Classroom

Yang Jiezhi & Liao Jianhua

Jiujiang Institute

Abstract:
Online corpora are convenient resources for language learners and researchers. Some online corpora are publicly available; they are accessible as long as you are connected online. However, many English language teachers are still wondering if online corpora are appropriate to be employed in the ELT classroom. This paper attempts to demonstrate that some large online corpora are the most useful resources that can be utilized in the ELT classroom.

Keywords: online corpora, BNC, Cobuild, LDC, ELT classroom

There has been much talk about how corpora can facilitate language learning, especially foreign language learning and how they can narrow the gap between the English acquired in the classroom and authentic English. Some even declare that corpora will revolutionize language teaching. With the most powerful tools that have ever been created, the concordancer and corpora of natural English, “language learners can short-cut the process of acquiring competence in the target language, because the computer is able to help students organize huge amounts of language data so that patterns are more easily discerned.” (Stevens) While interest is growing in the application of corpora to English language teaching, many English language teachers are just wondering if corpora are appropriate to be employed in the ELT classroom, because most of them either have vague ideas about what corpora can offer to them and the students or they don’t know where to find publicly available corpora online. And many more just feel more comfortable with traditional ways of instruction. So the real challenge is how to get the language teachers to accept the notion that the new teaching technique is the right thing to improve ELT in the classroom. This article will address those questions by examining why it is necessary to use online corpora, how it can be applied to teaching English, what pattern future foreign language teaching may take.

First, let’s take up the question why it is necessary to use corpora in the ELT classroom? As we know, the problem that most English language teachers and students encounter here in China is the shortage of authentic English language resources, like newspapers, magazines, academic books and popular fiction and audio material. There is hardly any up-to-date material available for them to use or to cite examples for classroom reference. That mostly accounts for the fact the students are not adequately motivated and the teachers' efforts don’t seem to pay off, to get the desired results. The best they can do is to act like the only knowledge imparter, trying to spoon-feed the
students with all the grammar rules and invented examples. Though students often feel their vision of the target language is limited by the scope of knowledge of the teachers, they cannot but passively receive whatever is being fed to them. Their enthusiasm for studying English diminishes quickly as there are hardly any language resources written or spoken available for the students themselves to explore. When teaching new words, most teachers are well aware of the awkward situation that presenting the definition is far from enough to get the students to have a clear understanding of a certain word or phrase, but they often feel constrained to provide enough appropriate examples to show the use of that word or phrase. So it is evident that language resources are essential for language learners. A modern English corpus is a large collection of written and spoken samples of authentic English language, made conveniently available by a computer program called a concordancer which can help researchers and learners of the English language find examples of authentic usage so that they can get familiar with features of vocabulary, typical collocations, points of grammar. As you will see, corpora are just what we need to solve our language problems.

The obvious advantage of using online corpora is convenience. Internet connection can save teachers the trouble of gathering many reference materials. Though online corpora are subject to some unpredictable elements like network paralysis, their value should not be overlooked. Size is an important aspect to consider when choosing a corpus. If a corpus is not large enough, it may disappoint an advanced English learner when infrequent words will not show up. Some online corpora are very large because they are designed to include relatively infrequent words and provide for a large and broad sample of language. However, on the other hand, if a corpus is so large that its size reaches beyond a certain limit, say a hundred million words, some people would consider it too large and believe that students will be at a loss what to do with so much data and even drown in all of the data. So the question seems to be how to balance somewhere between too much and not enough data. John Sinclair, founding editor of COBUILD, suggests that a corpus should be as large as possible, and should keep on growing. Another advantage of using online corpora is the availability of many other online tools that can be used as cross references, online dictionaries, encyclopedias, such as Encarta Online Encyclopedia, Cambridge International Dictionary of English, Merriam-Webster's Online Dictionary, LDC American English Spoken Lexicon and so on. No modern language learner can afford to overlook the convenient online resources even though they may have the printed or CD ROM material available.

BNC, Cobuild and LDC are some of the readily available large online corpora. The British National Corpus (BNC) is a 100 million-word sample of contemporary spoken and written British English, which consists of extracts from 4124 texts of all kinds so as to represent as wide a range of modern British English as possible. The Cobuild corpus or The Bank of English with some 450 million words (2002) is said to be the largest corpus in the world, some 50 million publicly accessible on-line. LDC The Linguistic Data Consortium is an open consortium of universities, companies and government research laboratories, which is involved in a number of projects to support language education, research and technology development. LDC-Online corpora
consist of several finished projects, including LDC93T1 with Brown corpus; LDC93T3 with data sources from Wall Street Journal, Associated Press and some other publications; LDC95T6 with data sources from LA Times & Washington Post, New York Times, Reuters News in the 1990s. These online corpora are the most useful resources that the teachers and students can find anywhere.

Now that corpora are of great value to ELT, why should we hesitate to exploit them in the classroom? Here are some of the concerns some educators may have about getting started with online corpora. Apart from the basic skills needed to access the web corpora, other factors may also cause them to hesitate. One of them is that introducing corpora into the classroom will involve changing the methods of teaching. And the teaching process may become unpredictable; teachers must be ready to deal with any problems that may crop up. Another concern is that a corpus may appear to be too large to manage, and the number of the hits in the corpus can be overwhelming for students, some even predict students could drown in all of the data. While that scenario sounds quite true, the real situation will prove that the concern is baseless. Above all, our main concern should be about how much students can benefit from using corpora in the classroom. As to the size of a corpus, it does not cause any problems for students as long as it reveals attractive features of authentic English language. To avoid “being drowned” in the data, we should never divert from the focus of our study, for example, to identify some collocations. Don’t bother with irrelevant information. It’s like consulting a very big dictionary. However big it may be, you can’t possibly not know how to get out of it and drown yourself in it.

Introducing corpora into English language teaching will make the old teaching pattern obsolete. The spoon-feeding or rote learning pattern will be replaced by a new one that promotes discovery learning or Data Driven Learning a concept developed by Tim Johns, in which students actively work with “raw” information taken directly from corpora. Students will change from passive recipients to active participants within the classroom. They can effectively acquire language when they draw their own conclusions about word/phrase meanings and grammar rules based on examination of authentic linguistic evidence. The teacher would act as a research facilitator rather than the more traditional imparter of knowledge. The learners can benefit a lot in the way they learn from using corpora since they take an active part in the learning process. As we all know, attitude determines the outcome.

In the classroom, a corpus and concordancer can be used mainly to discover and determine meanings of words and phrases; to identify useful phrases and unfamiliar collocations; to analyze usage and determine the context; to discriminate words with similar meanings; to examine word order; to encourage questions and generate discussion

Let’s just examine how corpora can help students learn different meanings of common words. Take the word “delicate” for example, you will find many definitions under the entry word in a dictionary, like “very subtle in difference or distinction”,

3
“frail in constitution or health” or “requiring tactful treatment” and etc, but presenting the definitions won’t help students get any further in understanding its meanings. With an online corpus available, it is just like playing games, we can ask students to type “delicate” into the query box of a certain corpus like LDC, and see what can be retrieved from the corpus, then students are instructed to take a close look at the concordances of a certain page, formulate their own conclusions about the meanings of the word, there should be a discussion before students are allowed to compare what they’ve discovered with the definitions in a dictionary. In this way students are left with a deep impression of the various meanings of the word.

s confronted with a delicate problem. <s> Short appreciated Miller’s delicate literary style that Jackson discuss the delicate situation created by organization and delicate diplomatic relations. Social idiom, was a delicate and perilous one, p of this master -- delicate, flexible and gem-like -- though this is a delicate matter -- that ever a. <s> Men of more delicate sensibilities refer cherished rare and delicate plants like oleander. <s> Being somewhat delicate in health, at the d to the sight of a delicate butterfly and two s. <s> His wife was in delicate health and nursing every change in the delicate balance of European t be related to its delicate, long alveolar duc e remained only the delicate task of maneuvering to hear.; ; <s> his delicate hand, as veined as

Anyone who wishes to master idiomatic English has to pay enough attention to word collocations. Collocations refer to the association between two or more words that are typically or frequently used together and they challenge one’s effort to learn them. The examples of collocations that one can find in any dictionary are rather limited. So we have to resort to corpora to solve the problem. Concordances will greatly help students identify and get acquainted with unfamiliar collocations. Following is a section of the concordances showing the collocations of past participles with the preposition “upon”.

public since it was agreed upon in open session sho ss seems to have descended upon us,” “John Paul s United States has imposed upon us.” “<s> He is i that can not be infringed upon without due process bilateral ties was built upon a “solid structure” of ours is called upon to guarantee strict m, said he “d been called upon to “navigate the tice departments” relied upon the immunized testi e and energy can be looked upon as a certain phase executive branch is called upon to investigate its adopted, or even improved upon, the mere words of
much damage was inflicted upon the plant under conditions are generally agreed upon in negotiations between defendants reasonably relied upon the proclamation. scenes power and is looked upon by Thais as a fact if something wasn't agreed upon and neither did, '

When learning a word or phrase, we’ve got to understand in what context it should be used. Let’s just take an idiom “to rub shoulders with” for example, and see how the concordances can help students be aware what kind of people that one would like to rub shoulders with. Following are the text concordances for rub shoulders with:
rub shoulders with exotic Hollywood types like Goldie Hawn
rub shoulders with any stars, but he did get a lot of real
rub shoulders with its earthen sides and so low one must a
rub shoulders with, white supremacy. <s> The same Clinton rub shoulders with him and George Bush, the details I ci rub shoulders with the people, ' adviser Nabil Shaath to rub shoulders with the future king of England. </TEXT> <T rub shoulders with the Prince of Wales, next in line for rub shoulders with commoners admitted on scholarships. <p rub shoulders with bankers Sir Evelyn de Rothschild and Lo rub shoulders with the widow of Oliver Tambo, Mandela's

source: Corpus NAN

To define “rub shoulders with” as either to meet and spend time with or to mix or socialize closely won’t be adequate for students to understand under what circumstances it should be used. However, the concordances can provide enough information for learners to detect those circumstances and arrive at the conclusion that the person whom common people would like to rub shoulders with is most likely to be a celebrity. It’s interesting to know some celebrities would also like to rub shoulders with commoners.

In discriminating words with similar meanings, concordances can play a unique role since no description of differences between words in any dictionary can take the place of the evidence in authentic language. So we’d better encourage students to find answers, to generalize from raw material and arrive at their own conclusions, the process really helps them in learning. Let’s take a look at the synonymous words keen and acute, they both mean having unusual mental agility or perceptiveness. Without enough examples one can hardly draw a clear distinction between the two. However, concordances make the discrimination much easier. Learners cannot fail to notice the general patterns of collocation or colligation. Only after we observe the concordances can we understand their definitions better. (Keen implies clear-headedness and acuity: a journalist with a keen mind and quick wits. Acute suggests penetrating perception or discernment: an acute observer of national politics.)
keen wit | acute problem
keen perception | acute pain
keen pleasure | acute boredom
keen observer | acute embarrassment
keen competition | acute questions
keen interest | acute leukemia
keen understanding | acute poverty
keen intellect | acute kidney damage
keen eye | acute sense of satire
keen investors | acute shortage
keen sensitivity | acute radiation sickness
keen sense | acute worries
keen enjoyment | acute concern
keen ability | acute hearing
keen interests | acute observation
keen intelligence | acute sense

Or we may just be interested in whether to use a keen sense of humor or a acute sense of humor. Concordances will help us notice the general pattern of their usage.

<s> She also had a keen sense of humor. </p> <s>
> '' Husseini has a keen sense of realism, and in
> '' Husseini has a keen sense of realism, and in
hey all have a very keen sense of focus, although
play a particularly acute sense of smell. At the o
ine musician with a keen sense of humor. Tipton ' ine musician with a keen sense of humor. Tipton ' ns said Starr has a keen sense of humor and would
ings moving with a keen sense of dramatic tension
of good intellect, keen sense of humor, great co
painful. He has an acute sense of right and wrong.
power, who have a keen sense of the need for maj
t a ready smile and keen sense of humor. On one o
ded. Wolves have a keen sense for weakness. Once
d. ''Bears have an acute sense of smell and they f
r al character and a keen sense of responsibilities

Concordances can help learners in many ways. And there are many ways that a corpus can be used. Before learners get started with concordances, they got to have some ideas and tips on how to use the concordancer to explore the corpus. As they get familiar with it, they will find some corpus may serve one particular purpose better than another. For example, BNC concordances are displayed in sentence mode; they
are more likely to be used to determine the context, to cite examples or to generate exercises. Cobuild concordances are displayed in KWIC mode; they are more likely to be used to identify useful phrases, check unfamiliar collocations or to discriminate words with similar meanings. LDC concordances can be displayed in both ways, and therefore, merits of LDC are obvious.

The value of online corpora is real and definite. But whether an ELT class is successful or not is not determined by how much data is used but by careful planning of class activities and work assignments. However valuable a corpus may be, it is still a tool, and it should not draw our attention away from the lesson we study. How to integrate corpora into English language teaching will challenge our thinking. When corpora are used appropriately in class, we can see the advantages of the lesson over traditional one. There are many challenges lying ahead before we can accomplish our goal of introducing corpora into English language teaching. But the question is when instructors are ready to adapt themselves to the role of a research facilitator and help students make use of authentic language resources and discover more by themselves.

References