

The Use of Corpora in the Vocabulary Classroom

Yu Hua Chen

yhstella2004 [at] yahoo.com.au

University of Melbourne (Melbourne, Australia)

Introduction

Lexical competence recently has been identified to be the most significant predictor to general language ability (Carter&McCarthy, 1988:97); however, it is also identified by most learners to be one of the biggest challenges of language learning (Coady&Huckin, 1997:1;Cobb, 1999). Fortunately, with the advent of technology, a new view of language learning and teaching has emerged; attempts to integrate computers as tools in language classrooms and facilitate the learning have been made.

This paper suggests language corpora can enhance the quality of vocabulary teaching and learning in second or foreign language classrooms. By presenting benefits of language corpora to the pedagogy, it is hoped that this paper can be helpful to both teachers and learners who are struggling to search for an efficient way of teaching and learning vocabulary.

What are Corpora?

Corpora, plural term of a 'corpus', refer to electronic authentic language databases that can be available via internet or as software installed in desktops (Hasselgård, 2001). Language corpora can be either collections of written or spoken texts; for example, collections of written texts can be extract from newspapers, business letters, popular fictions, books, or magazines, published or unpublished school essays and etc. Collections of spoken texts can be any recorded formal or informal conversations, radio shows, weather broadcasts or even business meetings and etc.

What Can Corpora Do?

Language corpora can be used by anyone who is engaged in language learning, teaching, or research; language learners or even native speakers may find it useful to assist academic writing or lexical knowledge (Qiao, 1995); teachers can utilize the authentic collections of data as classroom materials for ESL, EFL, or EAP (English for Academic Purpose) learners; language researchers or linguists often use corpora as language sources to analyse certain aspects of a certain language.

Usually users of corpora use the searching tool, the concordance, to look for vast number of authentic language contexts analysed from corpora (Witton, 1993); this feature provides users not only better quality of examples but more exposures to an unfamiliar word.

Examples of Corpora

Most software based corpora need to be purchased; however, there are still lots of free online resources, available from internet, for teachers and learners.

BNC (British National Corpus)

The British National Corpus (BNC) is one of the most famous corpora consisting of 100 million collections of written and spoken language samples. Both online service and CD-Rom are available for users. For more information, please visit:

<<http://www.natcorp.ox.ac.uk/>>

VLC (Virtual Language Centre)

The web-concordance is one of the language projects presented by the Virtual Language Centre in Hong Kong. Users can search for language samples from various corpora such like students' academic writings, Time Magazines, the Bible, business and economy and etc. To experience the on-line concordance, please visit <<http://www.edict.com.hk/concordance/>>

ICE (The International Corpus of English)

The International Corpus of English consists of 100 million spoken and written English samples. It is a product presented by the Department of English and Literature in University College London. The ICE provides both authentic written and spoken English samples of various English varieties such like Australia, New Zealand, Hong Kong, India and etc. For more information, please visit: <<http://www.ucl.ac.uk/english-usage/ice/>>

For users who are looking for other language samples rather than English, there are also corpora in many other languages such like Italian, Polish, and Japanese and so on. For more information, please visit: <<http://www.sfb441.uni-tuebingen.de/c1/corpora-engl.html>> (maintained by University Tübingen) or <<http://www.ruf.rice.edu/%7Ebarlow/corpus.html>> (maintained by Rice University).

Pedagogical Benefits of Corpora

Provides High Speed Searching Tool

Traditional language learners rely heavily on dictionaries as the main source to look up word definitions and examples; however, this task is often too laborious and time-consuming (Cobb, 2003). By using the concordance tool of corpora to search for word contexts, learners are involved in a more speedy and efficient language learning experience.

Provides Better Quality of Language Samples

It is suggested that language learning is more likely to happen when adequate examples are noticed and processed by learners (Cobb, 2003); for example, when learning a word, learners need to go through adequate sentence examples or contexts in order to develop more retainable lexical knowledge. Traditional dictionaries often provide unclear, limited, and artificial examples for each definition, which is insufficient for fully understanding an unfamiliar word. Using a concordance to search for word examples enable users to obtain mixed types of authentic language examples including phrases and collocations (group of words that often appear together), rather than simple and clear sentences. Learners' can not only develop their breadth but depth of lexical knowledge.

Encourages Active and Student-Center Learning

Traditional vocabulary learning is usually labeled as a 'passive way of learning'; no matter what approach the teacher uses in the classroom, intentional or incidental learning, the learning process that learners experience is usually inactive: receiving word lists or reading texts given by teachers, looking for word definitions, memorizing words, and luckily they remember them or usually they face the depression that they forget the words and need to go through the same process endlessly. Teachers can do nothing but ask their students to do anything they can to 'memorize the word definitions'.

By designing activities that involved learners in exploring and noticing the target language, teachers can engage learners in a 'content decision making' learning situation (Hadley, 2001). This technique is called Data-Driven Language Learning. It is suggested by Krishnamurthy (2004) that a word often has many meanings; however, the actual meaning should be determined by its surrounding contexts. By using the data-driven technique, teachers involve learners in tasks of exploring, choosing and determining the language from various resources that the computer found. The classroom therefore becomes student-center and learners have active control of their own learning (Nation, 2001; Rüschoff, B)

Suggestions for Incorporating Corpora into Vocabulary Classrooms

Create Concordance Sheets for Young Learners or Beginners

The searching task from a corpus may be too overwhelming for beginners or young learners who do not know much about computers. Learners often become demotivated or frustrated because of the enormous data produced by the concordance (Hadley, 2001). However, teachers can still utilize a suitable corpus and create a concordance sheet for students so they do not need to encounter the enormous data by themselves.

After distributing the concordance sheet, teachers can ask students to choose several examples that are meaningful to them and keep those examples in their language diaries. The purpose is to engage learners in exploring and noticing the language contexts; during the process of writing down those examples, learners are expected to undergo a cognitive process of digesting the language input. In Hadley's study (2001), by using this teaching skill, he found his students not only developed their lexical knowledge but improved their writing skill.

Classroom Project-10 Words a Week

Teachers can involve learners in a classroom project. They can choose some interesting stories for students as homework; learners are asked to use markers to write down unknown words on vocabulary cards (20cmx8cm) during the process of reading. Teachers can prepare two boxes in the classroom: unknown word box and learned word box. Students can put their colourful word cards in the unknown word box anytime when they come to class and teachers can choose at least 10 words from the unknown word box as a classroom weekly project, in which students are asked to search for examples from corpora and share the search result with the whole class.

As soon as students learn and retain the target words, teachers can put the word cards into the learned word box or paste it on the classroom wall. As time passes by, students will be motivated to learn more words since the learning has become a shared goal for the whole class.

Task-based Learning-Filling Gaps

Teachers can design a gap filling task for students to do in pairs. By giving students a learning task, teachers can efficiently control and monitor students' learning. Teachers can use the authentic contexts searched from corpora to compose a gap-filling sheet for students to work on; for example, if the target word for the class is 'medicine', teachers may want to present several collocations such like use medicine, take medicine, and prescribed medicine. The gap-filling task may look like this:

- If you wear contact lenses, ask your health care provider if you should wear them while you are _____ the eye medicine. Also ask if the medicine you have been _____ (prescribe) will stain your contact lenses.
- If your doctor _____ a new antibiotic, take all of the medicine as prescribed.
- Many antihistamines cause drowsiness, so you may want to _____ the medicine only at bedtime.

(Search results of 'medicine' from VLC online concordance: the health corpus:

<http://www.edict.com.hk/concordance/WWWConcappE.htm>)

By engaging students in searching for answers from a corpus to solve problems, teachers can encourage students to devote more time to the learning activity (Bracewell &Laferriere, 1996). Moreover, the task-based learning can also develop learners' confidence in their ability once they accomplish the task.

Conclusion

Due to insufficient learning time and inefficient word searching tools, lexical learning has always been one of the main language learning problems that learners pointed out. The advent of technology is about to present a different view toward language learning and teaching; several studies have shown positive learning outcomes by engaging students in activities of decision

making and information retrieving. Integration of corpora into vocabulary classrooms not only provides learners faster searching tools and better quality of contexts that traditional dictionaries are not likely to achieve but enhance their learning motivation.

References

- Bracewell, R., Laferriere, T. (1996). The Contribution of New Technologies to Learning and Teaching in Elementary and Secondary Schools. (Documentary Review). Retrieved from <<http://www.fse.ulaval.ca/fac/tact/fr/html/apport/impact96.html>>
- Carter R. & McCarthy, M. (1988). Vocabulary and Language Teaching. Longman:New York.
- Coady, J. & Huckin, T. (eds.) (1997). Second Language Vocabulary Acquisition. A Rationale for Pedagogy. Cambridge University:USA
- Cobb, T. (1999). Breadth and Depth of Lexical Acquisition with Hands-on Concordancing. CALL Journal 12 (4), pp. 345-360
- Cobb, T. (2003). Do Corpus-Based Electronic Dictionaries Replace Concordancers? In Morrison, B., Green, G.. & Motteram, G. (Eds.) Directions in CALL: Experience, Experiments, Evaluation. Polytechnic University:Hong Kong.
- Hadley, G..(2001). Concordancing in Japanese TEFL: Unlocking the power of data-driven learning. Retrieved from <<http://www.nuis.ac.jp/%7Ehadley/publication/jlearner/jlearner.htm>>
- Hasselgård, H. (2001). Corpora and their use in research and teaching. Retrieved from <http://folk.uio.no/hasselg/UV-corpora.htm>
- Krishnamurthy, R. (2004). Learning and Teaching through Context-A Data-Driven Approach. Retrieved from: <http://www.developingteachers.com/articles_tchtraining/corpora3_ramesh.htm>
- Nation I.S.P. (2001). Learning Vocabulary in Another Language. Cambridge University Press:Cambridge
- Qiao, H. (1995). 'Navigating in corpus space with the Mini-Concordancer: an autonomous language learning perspective'. On-CALL Vol. 9. No.3
- Ruschoff, B. Data-Driven Learning (DDL): the idea. Retrieved from: <http://www.ecml.at/projects/voll/rationale_and_help/booklets/resources/menu_booklet_ddl.htm>
- Witton, N. (1993). A Simple Concordancer. From ON-CALL Vol. 8. No. 1