On Corpus-based Autonomous Learning for College English Learners

Lihua Cai

School of International Studies, University of Science and Technology Liaoning, Anshan, 114051, China

Cailihua0412@126.com

Abstract

With the rapid development of interdisciplinary research in computer science and linguistics, corpora of different types are employed in the second language acquisition field. For Chinese English learners, native corpora like BNC will offer comprehensive guidance in lexical acquisition process; Chinese learner corpus is supposed to be used to predict and avoid effectively some typical errors and mistakes produced by college students.

Keywords

Corpus; autonomous learning; college English.

1. Introduction

The official document issued by Education Department of China in 2007, namely "Curriculum teaching requirements for College English", put forward that modern information technology, especially IT, should be fully employed to change traditional teaching model to promote students' integrated application skills in English and cultivate students' autonomous learning ability as well as comprehensive cultural quality. With development of computer technology, online information and corpus resources are expected to be employed efficiently for Chinese college students to learn English in and after class. This study aims to explore the practical way of computer-aided and corpus-based English learning for college students in their autonomous learning process.

2. Autonomous learning and corpus

Autonomous learning can be regarded as a sort of learning attitude as well as some kind of independent learning ability. Cohen (1990) pointed out that the success of language acquisition lies in learners' factors and their ability to make good use of various chances to learn the language. Autonomous learning involves both external environment and internal environment. The former is the prerequisite and physical basis of the latter one which includes learners' attitude and competence (Zhang Dianyu, 2005). Generally speaking, the theory of autonomous learning covers the following aspects (Shu Dingfang, 2004): a) learner's attitude; b) learner's ability; c) environment.

Corpus refers to a collection of naturally occurring samples of language which have been collected and collated for easy access by researchers and materials developers who want to know how words and other linguistic items are actually used. A corpus may vary from a sentence to a set of written texts or recordings. In language analysis corpuses usually consist of a relatively large, planned collection of texts or parts of texts, stored and accessed by computer. A corpus is designed to represent different types of language use, e.g. casual conversation, business letters, ESP texts (Jack R, 2005). Actually a number of different types of corpuses may be distinguished according to different purposes, structures or targets, namely specialized corpus, general corpus or reference corpus, comparable corpora, learner

corpus and so forth. In this study, learner corpus and specialized corpus on line may are employed to promote Chinese learners' English learning.

3. Corpus-based autonomous English learning model

Compared to traditional learning model, corpus-based autonomous English learning model may have some inherent advantages due to its features in nature. The details are as follows in the section below.

3.1 Learner corpus and errors in the model

In recent years some high-quality learner corpora have been built in China to help Chinese English researchers shed new light on second language acquisition process of Chinese English learners. In addition, relevant findings in these studies may be of great value for English teaching and learning activities. Take the Chinese Learner English Corpus (shorten for CLEC) for example, it is the most important and representative learner corpus of written English produced by Chinese English learners in Mainland China. It was established under the supervision of Professor Gui Shichun at Guangdong University of Foreign Studies and Professor Yang Huizhong at Shanghai Jiao Tong University. The compositions in CLEC were produced by Chinese learners at different stages: middle school students (St2), college sophomores (Band IV students, St3), college juniors (Band VI student, St4), English-major juniors (St5) and English-major seniors (St6). CLEC is claimed to be reliable due to its scientific sampling process. First, the samples are all original writings without any correction; second, the sampling proportion is balanced among learners at different levels; third, the samples are from diverse sources so that the corpus covers learners' written output widely enough.

Therefore, it can be used to help Chinese learners to acquire English in autonomous learning style as well as a guide for college English teachers in the teaching process. CLEC has been tagged according to different types of mistakes and errors (see the table below), which is convenient for index engine to search for KWIC (key word in context) information closely related to typical errors produced by Chinese English learners. Both software package like AntConc or self compiled programs can fulfill the target. College English learners will benefit a lot from the study of high-frequency errors extracted from the corpus, hence the promotion of English proficiencies and accuracy in both written and spoken English.

Table 1 Text sample in CLEC

<ST 6> <SEX ?> <Y ?> <SCH GIFL> <AGE ?> <WAY 1> <DIC 2> <TYP 2> Euthanasia should be Legalized in China Euthanasia, or mercy killing, means helping to hasten the death of a person who is badly suffering. According to Joseph Fletcher, it includes people getting incurable dease [fm1,-] and people in a helpless condition, such as trapping in a blazing fire. [sn8,s] In China, suicide is legal, which means, people are legal to kill themselves in a helpless condition, so what we conside [fm1,-] is only whether it is legal to end the life of a [np7,1-] incurable patient. I am in favor of the legalization of euthanasia, though some others against it.[sn8,s] Those who against it agues [fm1,-] that euthanasia is inhumane. It is a false argument. Death, most of the time, is the end of long suffering period. With the advanced medical techinique [fm1,-] and equipment, human life can be extended. On one hand, it is in deed a good thing to provide people health when the diseases are curable; on the other hand, it is rather a bad thing to extend people's suffering when the diseases are incurable.

3.2 BNC and Online corpora in the model

Large quantities of native speakers' authentic texts in English can be collected and processed to build up a corpus for reference. The British National Corpus (BNC) is a 100 million-word collection of

samples of written and spoken English from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century. The written part which accounts for 90 percent, includes extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fictions, published and unpublished letters and memoranda, school and university essays, among many other kinds of texts. In addition, BNC is encoded according to the Guidelines of the Text Encoding Initiative (TEI) to represent both the output from CLAWS (automatic part-of-speech tagger) and a variety of other structural properties of texts (e.g. headings, paragraphs, lists etc.). Full classification, contextual and bibliographic information is also included with each text in the form of a TEI-conformant header.

BNC can offer authentic guide for Chinese learners in the process of autonomous learning, especially for the collocation identification and distinction of synonyms in terms of semantic prosody and idiomatic level (see the figure below). Furthermore, other native corpora are available for free online such as COCA, COHA, BASE and so forth.

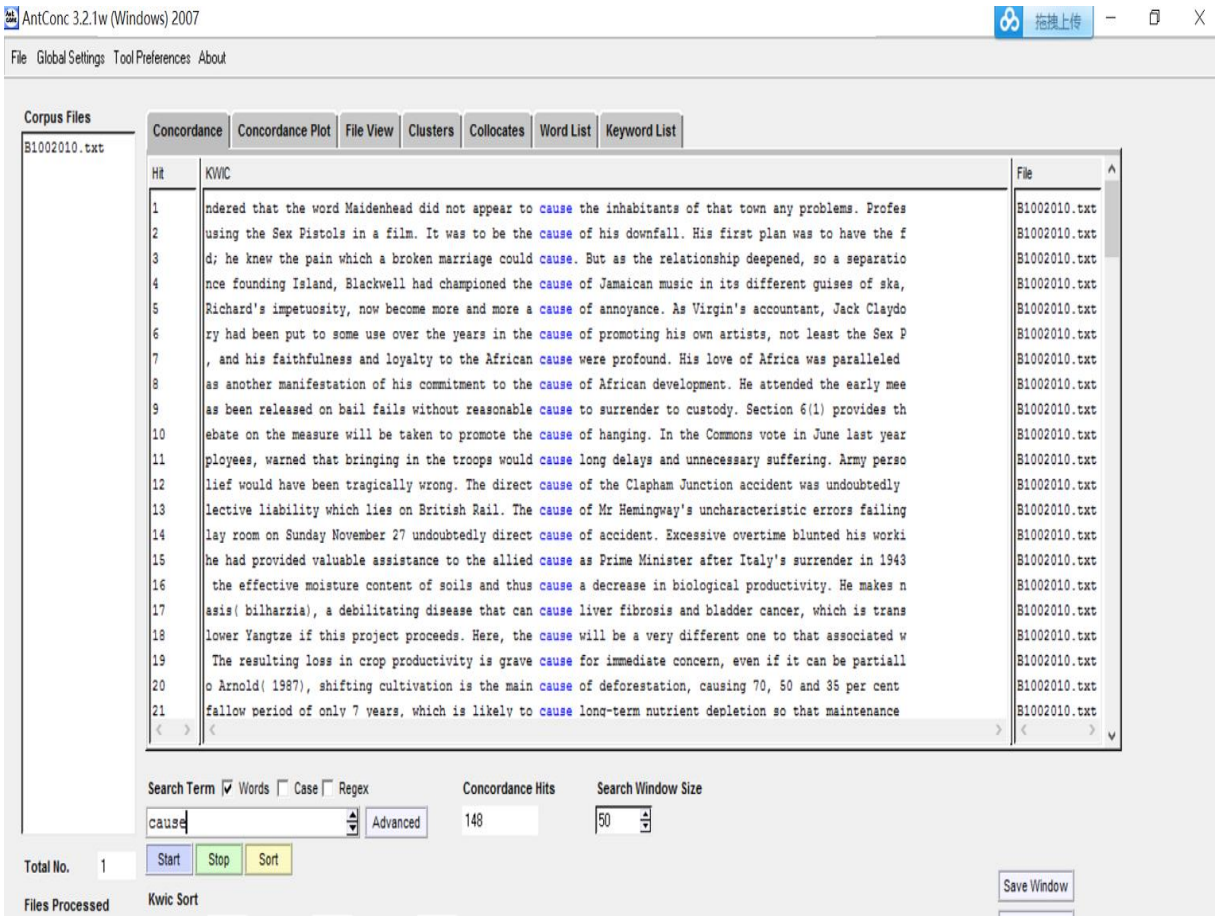

Figure 1. Collocation comparison of "cause" and "reason"

4. Conclusion

To sum up, corpora of different types can be regarded as a valuable resource of language data in both native language and interlanguage. The former one will offer guidance and reference for language learners and the latter may help predict and avoid typical errors and mistakes that Chinese English learners tend to produce in the process of second language acquisition. Therefore, corpus-based model is supposed to be efficiently employed in autonomous learning process.

References

[1] Cohen A. Language Learning: Insight for Learners, Teachers, and Researchers[M]. New York: Newbury House,1990.

[2] Zhang Dianyu. Strategy and Autonomous Learning of English Learning[J]. Foreign Language Teaching. 2005(1).

[3] Shu Dingfang. Foreign Language Teaching and Reform: Problems and Strategies[M]. Shanghai: Shanghai Foreign Language Education Press. 2004.

[4] ack R. Longman Dictionary of Language Teaching and Applied Linguistics[M]. Beijing: Foreign Language Teaching and Research Press.2005.