CORPUS LINGUISTICS OF MATERIALS AND APPLICATIONS IN TEACHING, LEARNING FOREIGN LANGUAGES

Le Chi Hieu

Foreign Language Faculty, Hanoi University of Industry
Email: *hieulc@haui.edu.vn*

The aim of this paper is to review some relevant factors related to the corpus and its using in english teaching. First we describe some definitions of corpus and its development in linguistics, especially in context of Vietnam (it is called corpus linguistics). Then, we attempt to present some tools for exploiting the corpus called concordance such as Antconc, Corpus, Concordance, etc... They are useful in using the electric corpora. Finally, we outline the implementation of Corpus in english learning and teaching in EFL class.

Keywords: *Corpus; Using corpus in english teaching.*

1. Introduction

Corpus linguistics was born with John Sinclair and the Cobuild project at the University of Birmingham. He defined it as "a collection of naturally occurring language texts, chosen to charactrize a state or variety of a language" (Sinclair, 1991: 171). In general, corpus is a collection of natural language (text, and/or transcription of speech) constructed with a specific purpose. Ideally, a corpus is a set of language production samples designed to be representative of a language through careful selection. Until 90's decade, with the development of ICT, corpus and corpus linguistics are clear outcomes of these technological advances.

Nowaday, linguists apply more and more the corpus in researching and studying of foreign language. It is useful to discover the term in real context, to help them creating the automatic translating tools and to support teachers and students comparing language phenomenons through paraell and compared corpus. Especially, corpus is so necessary in learning new words and collocations.

2. Overview of issues research

The use of real exemples of texts in the study of foreign language is not new issues in EFL. However, using corpora in teaching EFL has developed considerably in the last decades. The availability of computer and electric tools help linguistics and students getting data quickly and easily. Since 90s and 2000s years, some universities in UK and France intergrate corpus linguistics courses in their learning curiculums. For exemple, University Paris 7 develop some tools to help students using corpora in learning english in EILA department. Besides us, some colleagues in Malaysia start to buil corpora serving students and teachers in learning and teaching ESL such as a corpus of English language of Malaysia School Students as EMAS Corpus created by the University Putra Malaysia. The next two corpora were developed by the Universitu Malaya. The first is the MACLE corpus (Malaysia corpus of Learner English) based on students's essays. The second is the COMEL corpus (Corpus of Malay English) based spoken corpus project. In Vietnam, building many corpora for vietnamese language is real and necessary demand. But until now we don't have yet one corpus in national level such as BNC or CFC (Corpus du Francais Comtemporain), etc…

In Vietnam, applying corpus in linguistics is new approach. But it demands the development in future. Theorycally, we look many researches about this with some reports, jornals and books of corpus linguistics. And Dr. Dao Hong Thu is the scientic pioneer as entering it in Vietnam from 2006. And then, some young linguists follow her in applying it automatic translation, teaching and studying English as ESL and partucularly in teaching Vietnamese laguage for foreigners. But, we look that we need more research outcomes and more attractions forward young linguists in this domain. For application, the researchers built up one small community using electric corpus (look website https://corling.wordpress.com/) where we can find out some journals, reports, books and tools serving automatic translation,…with thu cooperation of young linguists such as T.T.M.Dao, D.T.T.Hien,…. So we can conclude that application of corpus linguistics is new scientific path and approach for both linguistic research and learning.

3. Research Methods

Using corpora is one of major method for research. In fact, it is important analysis tools because it provides quantative data to test and examine the text, speech, signs in the authentic contexts. Corpus collection is important for both the buiders and the users. For exemple, Alcohol Languages Corpus (Schiel, Heinrich, 2011) is specialized corpus consisting of speech samples from 162 speakers of German. The samples are recorded in a car. Besides, many corpora are made by the results from individual and collective efforts. And it become the valuable ressources for vocabulary analysis. For exemple: the corpora represent different populations such as The Brown corpus; or the varieties of English such as British National Corpus; or modes of communication such as the British Academic Written English Corpus. In addition to these corpora, there are other corpus tools to exploit the corpora such as concordances, online versions of websites such as AntConc or Corpus Concordance that allows users search concordance in different sources and work with different languages.

For teaching and learning English as ESL, corpus-based and corpus-driven approach is so useful. In the class, students can build up themselves small specialised corpora serving in learning of foreign language. For exemple, in Jounalism school, students can creat up one special corpus of 300.000 words from NewYork Times journal in the year of 2000. Then they use the concordance to discover the terms in press domain and its context to analyse the real texts and understand how the terms work.

Besides, we use some others approaches to discover the corpus such as quantity, quality and descriptive approach. In fact, student can not

only analyse thousand samples of word but they also describe how these samples work and which collocations they goes with. And the mission of users is to conclude the terms are used in which context and how to use them correctly.

4. Research results

Using corpora means showing all exemples of word in context. By observing over the concordance, lexicographers can find all the meanings of the word, and phrases. If it is a big corpus, or a common word (or both), there might be thousands of examples of the word. Then, the computer can go one step further, and prepare a summary of the contexts, collocations and phraseology for the word. Look the exemple:
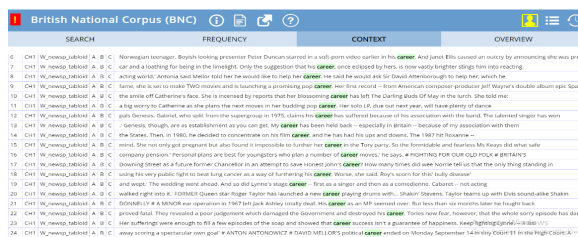

Figure. Sample of concordance for "**career**" from BNC (2020)

However, Google is one "app" that does something like that (with the Internet as its corpus). But it is not designed for people doing language research. One that is widely used for making dictionaries, Dictionary-makers were leaders in corpus use. So, in English language teaching, there is plenty of direct corpus used by teachers, even students? In some cases, the dictionary isn't enough. Using corpora is better to discover the word in the original context.

We observe an other exemple to illustrate the corpora. The situation like that: Students compare the learners's use of some words with the native use of those words using the Santiago University Learner of English Corpus and the British National Corpus. Corpora allow them to detect the differences in the use of the same word by native speakers and by learners and examine how much they differ. They propose to analyse the noun "*career*" both in Santiago University Learner of English Corpus (SULEC) and in the British National Corpus (BNC) in order to determine the differences and different associations that native and learner speakers assign to this word.

Description of the activity: Students are presented with two examples of the noun "*career*"; one taken from SULEC, the other from the BNC.

Learners have to work out any differences in the use of this noun by native speakers of English and by EFL learners.

Table. Examples of *career* from SULEC and from the BNC (2012)

SULEC: Most of the students that finish their __career__ are not prepared to assume that they probably do not get a job related to the university degree that they have been studying for 3-5 years of their live. (code: SULEC-AE-19)	*BNC: He became a trader in Nigeria and when this __career__ failed, worked for a time as a clerk to Richard Beale Blaize, publisher of the shortlived Lagos Times (code: CDU 52)*

As seen in these examples, learners use the noun "*career*" to denote "*university course or degree*". However, native speakers use this word when referring to a "*position or job*". Data reveal that the noun *career* is used in connection with students by EFL learners; thus, students can finish a career while in the BNC, career is associated with professions, that is, being a trader is a career.

This activity allows students to reflect on the different connotations and uses of this word by first language speakers of English and non-native speakers thus allowing teachers to see that learners use this word in the wrong way. On the other hand, students learn that they are mis-using this word and that they should use it in a different sense and context.

5. Discussion

Corpus linguistics is not same as obtaining data through the use of computer. This is the study and analysis of data obtained from a corpus. The main task of linguistics is not to find the data but to analyse it. Computer are so useful tools used in this process. So how to help students using well this tools is not easy for teachers in ESL because it demands not only knowledge in linguistics but also in linguistic computer. The second question is if electric copora is research tools or research method. So It depend to academic aims of linguist.

Otherwise, teaching-oriented corpora are so useful in teaching LSP. In fact, linguists often choose small corpora based on specific needs such as Hongkong financial services corpus, Hongkong engineering corpus,… So LSP corpora is so necessary for student and teachers in the context of Hanoi University of Industry where many faculties such as Electric, Automobile Technology, Textile,… need specialised English teaching as ESL. Perhaps, the deploiement of this is not easy and demand the

time and cooperation between them.

In some Foreign Language Schools, students of last year in specialisation of interprating and translating branch base mostly in experience to learn the verbal terms, verbal collocations,…For them, google is useful tool. But we notice that it is search engine not corpus sample. If they want to discover the word in real context, they need use corpus and concordance. In fact, they can creat up themselves one smal specialised corpus of 300.000-500.000 words for their learning objectives. Otherwise, students can use some open big corpus such as BNC, Lancaster Corpus, Lemonde Corpus,…Finally, we hope that students and teachers in foreign language apply more and more the corpus in learning English as ESL.

6. Conclusion

Corpora is useful and suitable for vocabulary study. Infact, its contribution to study of vocabulary is remarkable. It brings the real English to classroom's activities. They allow students to analyse the meaning, context in which the words typically occur. This help them understand more and more the realistic picture of how the language and its vocabulary work. So the using of corpus in foreign language teaching and learning need to be researched seriously, totally and applied step by step in the universities of foreign language.

References

Chong, A. S. (2013). *5 ways to use the corpora for classroom activities*, English Teaching Professional, August.

Kilgarriff, A. (2014). *Corpora in English language teaching*. British Council.

Kübler, N. (2014). *Mettre en œuvre la linguistique de corpus à l'université, Vers une compétence utile pour l'enseignement/apprentissage des langues?* Les cahiers de l'Acedle.

Oxford University Press. (2020). *British National Corpora (BNC)*, https://www.english-corpora.org/bnc/.

Rawal, H. (2014). *Using Corpora in English Language Teaching*. Global thought, Scholarly Article, Technology.

NGÔN NGỮ HỌC KHỐI LIỆU
VÀ ỨNG DỤNG TRONG DẠY, HỌC NGOẠI NGỮ

Lê Chí Hiếu

Đại học Công nghiệp Hà Nội
Email: *hieulc@haui.edu.vn*

Mục đích của bài báo đề cập đến các vấn đề liên quan đến ngôn ngữ học khối liệu (corpus) và cách sử dụng chúng trong giảng dạy tiếng Anh. Tác giả mô tả các khái niệm về corpus và quá trình phát triển của nó đối với ngôn ngữ học. Sau đó, giới thiệu một số công cụ để khai thác khối liệu ví dụ như Antconc, Concordance,… Đây là các công cụ rất hữu ích để khai thác được các corpus điện tử. Nội dung ứng dụng trong dạy và học ngoại ngữ của bài viết được phân tích thông qua một số tình huống sử dụng corpus trong giảng dạy và học tập tiếng Anh về lĩnh vực từ vựng .

Từ khóa: Ngôn ngữ học khối liệu (Corpus); Ứng dụng trong giảng dạy và học ngoại ngữ.