Practical Application of on-Line Corpora in English Classroom Teaching

Wenlian, Zhan¹,², Zhulin, Shen¹,²
¹School of Foreign Languages, Hunan International Economics University
Changsha, China
²e-mail: 513869962@qq.com

Abstract—Nowadays on-line corpora are extremely convenient language learning and research resources, especially those available free to the public are of great value. In recent years how to make full use of the on-line corpora resources to assist English classroom teaching has become what linguists and language teachers concerned. The author in this paper exemplifies how resources of on-line corpora help to solve some problems in teaching practice, which illustrates that some world famous Corpora, as BNC, COBUILD, LDC etc. are applicable in assisting English classroom teaching.

Keywords: on-line corpora, application, English classroom teaching

I. INTRODUCTION

Modern corpus is a real language electronic text database. The emergence of electronic corpus provides a very convenient tool for language research and teaching. The corpus is particularly of significance to foreign language teaching, as it can help narrow the gap between the authentic language in reality and what they have been learned in the classroom. Though corpus application is mainly served as a research tool for some experts and linguists, there is a possibility of exploiting corpora resource on-line in classroom English teaching with the rapid development of the network and people’s deep knowledge of corpus. And what language teachers most concerned in recent years is how to make use of the on-line resources of corpora to assist English classroom teaching, with the hope of innovation the outdated foreign language teaching mode and promoting the foreign language teaching efficiency. This paper illustrates by some examples of application of the on-line corpora in her own teaching practice that on-line corpora are of great value in assisting classroom English teaching.

II. ON-LINE CORPORA: BNC, COBUILD, LDC

There is tremendous difference in the size and quality among the on-line Corpora. After years of trial and comparison, the author find that the following large-scaled corpora available on line can be used directly as foreign language teaching resources. British National Corpus (BNC) is the largest corpus can be directly used in the current network, it created jointly by the Oxford University Press and several dictionary publishers and research institutions, to be completed in 1994, more than one hundred million the BNC corpus word capacity. Of 4124 represent a wide range of modern British English text, including national and local newspapers and magazines, all kinds of journals, theory books, popular fiction, the University of papers, letters, and a taped conversation text, 90% written and 10% spoken. The Co build corpus is the network era earliest large corpus, called the English library (The Bank of English), completed in 1991 by Collins Press, and the University of Birmingham, after the Corpus expanding, in January 2002, the English library Words capacity has reached four hundred fifty million. LDC is the abbreviation of the language data Union (The Linguistic Data Consortium), was established in 1992 by the research departments of universities, companies and government, the University of Pennsylvania Union location.

The above three corpora are rich in language materials, with the wide range of topics, a variety of linguistic phenomena that reflect the contemporary English. Corpus is not only an excellent tool for linguists and lexicographers, language teachers, students and translators infinite source of corpus consultation. The frequency of common phrases and structures in the corpus can be a direct reflection of the breadth of the size and subject matter of the corpus, the following are some common phrases and structure arbitrarily chosen from three corpora:

<table>
<thead>
<tr>
<th>Phrase</th>
<th>BNC</th>
<th>COBUILD</th>
<th>LDC3125.2</th>
<th>LDC57.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>look forward to</td>
<td>3007</td>
<td>&gt;40</td>
<td>&gt;40</td>
<td>&gt;40</td>
</tr>
<tr>
<td>asked him to</td>
<td>220</td>
<td>&gt;40</td>
<td>24</td>
<td>229</td>
</tr>
<tr>
<td>paid for</td>
<td>282</td>
<td>&gt;40</td>
<td>149</td>
<td>&gt;40</td>
</tr>
<tr>
<td>out of the way</td>
<td>286</td>
<td>&gt;40</td>
<td>109</td>
<td>422</td>
</tr>
<tr>
<td>used to get</td>
<td>47</td>
<td>&gt;40</td>
<td>&gt;40</td>
<td>107</td>
</tr>
<tr>
<td>used in doing</td>
<td>94</td>
<td>&gt;40</td>
<td>5</td>
<td>61</td>
</tr>
<tr>
<td>used in getting</td>
<td>36</td>
<td>11</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>made him happy</td>
<td>17</td>
<td>7</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>interested in getting</td>
<td>18</td>
<td>23</td>
<td>25</td>
<td>80</td>
</tr>
<tr>
<td>used in doing</td>
<td>22</td>
<td>7</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>be indifferent to</td>
<td>20</td>
<td>9</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>even been to</td>
<td>67</td>
<td>&gt;40</td>
<td>&gt;40</td>
<td>41</td>
</tr>
</tbody>
</table>

As can be seen from Table 1, there are a number of advantages, BNC. This describes the breadth of the BNC the corpus source of the sentence richness, and a direct impact on whether the user as the preferred corpus. The the COBUILD trial Corpus term frequency of more than 40, can still represent quite a large number, but in the lower half of the results to determine, its corpus should be less than the BNC. At the same time, we also notice that, although the LDC mainly by several well-known newspapers constitute extremely rich corpus, the word capacity is quite large, and its value should not be underestimated.

Judging from the query, three corpora have their own characteristics. BNC queries simpler way to directly query word or phrase knock query box, use the dash “_” instead of between words any word, such as: bread _ butter. However, users will find the equal sign “==” and “{}” is almost
irrelevant to what in handy. Cobuild and LDC using wildcards "*" and some other symbols, inquiries can also be carried out as part of speech corpus part-of-speech tagging shows another function, it is not only about the word's part of speech known sentence, but through use POS tag query a variety of sentences with the same structure, this function is much to do for people engaged in English teaching and research. Cobuild and LDC speech tags, this is a great convenience for people who use either two Corpus, but there are still some parts of speech tags.

To illustrate the superiority of the part-of-speech tagging, we can use some common syntactic structure as an example, if worth +ing verb structure, Cobuild and LDC are available "worth VBG structure identified worth different verb ing structure, andBNC can only find out once ing structure of a verb, as worth doing or worth reading. Others, such as with + object + past participle structure, Cobuild and LDC are available "with NN VBN" queries, the BNC but no such a function.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>POS PROPERTY TAGS IN THE COBUILD AND LDC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cobuild</strong></td>
<td><strong>LDC</strong></td>
</tr>
<tr>
<td>N</td>
<td>common noun</td>
</tr>
<tr>
<td>N</td>
<td>plural noun</td>
</tr>
<tr>
<td>J</td>
<td>adjective</td>
</tr>
<tr>
<td>P</td>
<td>preposition</td>
</tr>
<tr>
<td>R</td>
<td>adverb</td>
</tr>
<tr>
<td>PR</td>
<td>personal pronoun</td>
</tr>
<tr>
<td>PP</td>
<td>noun,possessor</td>
</tr>
<tr>
<td>DT</td>
<td>article</td>
</tr>
<tr>
<td>SD</td>
<td>determiner</td>
</tr>
<tr>
<td>MD</td>
<td>adverb is the most recent</td>
</tr>
</tbody>
</table>

Retrieve sampling from the point of view, BNC retrieval results is given directly to the sentence in the form of more than 50 if the sample sentence, click again to search Kin View more diverse sentence. Search results arranged in a sentence is better way of striking keyword centered (KWIC), KWIC gives at a glance the feeling, but the sentence the way to provide more context to the query to enable teachers to select sentences. Query can also enter by registering Sara retrieval procedures syntactic structure or context information in a variety of ways. Cobuild just BNC relatively KWIC way to provide query results can not be extended punctuate sure people feel the vision is limited, but the the query directly diversification and word frequency statistics the query was the Cobuild still rare in the network language learning and research tool, with the analysis of words and the observed syntactic structure aspects Cobuild the role more obvious. LDC retrieval of both BNC and the Cobuild two corpora characteristics can display search results Keywords centered manner, but also to provide a sentence or paragraph, page View 100 after the sentence. Certain network Corpus trends is both convenient query, but also a full-featured, in this regard, the the late American National Corpus ANC will exceed BNC and Cobuild.

III. APPLICATION OF ON-LINE CORPORA IN ENGLISH CLASSROOM TEACHING

The large network Corpus has a considerable number of applications, they can help researchers study language phenomenon, edit dictionaries, grammar books, teaching reference books, translators can take advantage of network Corpus check in terms of the translation is accurate, the teachers can use the network Corpus lesson planning, the choice of sentences, writing exercises, and the questions. Some experts is very important corpus application in foreign language teaching in the classroom, however, corpus used in foreign language teaching is still yet to be reclaimed field. Network practical Corpus mainly from several aspects for foreign language teaching reflected

1) by the volume of sentences to learn words, to understand the variety of the meaning of the word or phrase;
2) to master phrases with familiar words;
3) help to distinguish between synonyms;
4) observation of word order, master syntactic structure;

A. Vocabulary studying

For English learners to overcome the word disorder seems that apart from rote What better way, the emergence of the corpus will completely change our understanding of the misunderstanding. Learn a new word is not just to remember the pronunciation, spelling and meaning, but also familiar with its context and sentence, so as to truly grasp the usage of the word, firmly remember the word. Corpus is such a tool that can provide a large number of instances, so we focus on a word usage, the locale in the notice that the words and sentences, for example, a variety of meanings: adjective delicate, can be translated into: exquisite, delicate, sickly, fragile, delicate, tricky, sensitive and precise. But understood as "sophisticated" under what circumstances, and under what circumstances understood as "delicate" or "fragile" you need practical examples to speak clearly, more clearly. From the LDC corpus, we can sort out the following results:

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>EXAMPLE OF VBG AND WHEN VBN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>worth using...</strong></td>
<td><strong>with legislation passed...</strong></td>
</tr>
<tr>
<td><strong>worth highlighting...</strong></td>
<td><strong>with hat down...</strong></td>
</tr>
<tr>
<td><strong>worth preserving...</strong></td>
<td><strong>with passion spent...</strong></td>
</tr>
<tr>
<td><strong>worth noting...</strong></td>
<td><strong>with information related...</strong></td>
</tr>
<tr>
<td><strong>worth asking...</strong></td>
<td><strong>with maintenance provided...</strong></td>
</tr>
</tbody>
</table>

B. Collocation

Ability to grasp the commonly used phrase, skilled use of collocations related to learning English whether authentic, whether close to the tribe's language. However, with any one dictionary that can be related to the phenomenon are quite limited, it is difficult to meet English learners to further understanding of the relationship needs collocation. Collocations include not only the common words used in conjunction with the structure of the phrasal verbs,
compound nouns, prepositional phrases, fixed phrases, idioms, and also includes numerous habits.

**TABLE V. VARIETY OF COLLOCATIONS**

<table>
<thead>
<tr>
<th>Phrasal verbs</th>
<th>Prepositional phrase</th>
<th>Fixed phrases</th>
<th>Verb + noun</th>
<th>Adverb + noun</th>
<th>Adverb + adjectively + noun</th>
<th>Adverb + adjectively + preposition</th>
<th>Nouns + noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take off/step down</td>
<td>away from / regardless of</td>
<td>in the light of / open to suggestion</td>
<td>throw a party / accept responsibility (useful)</td>
<td>square meal / gain determination</td>
<td>take vigorous exercise / make steady progress</td>
<td>strongly suspect / briefly see</td>
<td>pay pal / money bag</td>
</tr>
</tbody>
</table>

**TABLE VI. "GO ABOUT" WITH PREPOSITION**

- go about the job of finding...
- go about their business...
- go about our business...
- go about our work all right...
- go about doing their business...
- go about getting a job like...
- go about their activities...
- go about their daily chores...

**C. Discriminating Synonyms**

The large corpus Collocation Discriminating Synonyms Dictionary irreplaceable role plays Discrimination. Let's look at another teaching examples, the second volume in Advanced English Lesson Love is a Fallacy in the fourth paragraph of the second sentence consecutive five synonyms, keen, calculating, perspicacious, acute and astute. While teaching reference made a distinction between these five words in the definition and Discrimination Dictionary exposition of; but due to lack of instances, teachers will find it difficult to grasp. The corpus can help students to fill the gaps in the process of understanding phrases and sentence structure, deepen learning more instance impression enhanced learning effect. The corpus fragment below can help students understand the "go about" with the object, to know that it contains a mean to engage in a transaction, chores or activities.

By contrast, it is easy to notice three synonyms tendency when paired with the noun, keen in understanding, the ability to observe and feel sharp. acute addition to keen observation and the ability to feel, the more used to a certain seriousness of the state, such as illness, problems. Astute mainly refers to people engaged in some kind of industry savvy and keen. Of course, not only provided by the Corpus contrast, intuitive language materials, through the use of these corpora, students observe and judge the ability of language materials will be greatly improved.

**D. Observing word order and syntactic structure**

Mastered words and collocations based on, if you want to create the perfect sentence, we must also be familiar with the basic word order and syntactic structure of the foreign language. In this regard, the network Corpus same as we provide avenues to explore. For example, in the word order, so word order is often plagued some students such + a / an + adjective + noun "so + adjective + a / an + noun" the use of network Corpus we can easily solve this problem, as long as in the Cobuild or LDC query box, type "such DT JJ NN" or "so JJ DT (a | an) NN" we can get a lot of the kind of sentence that contains this structure, each kind of sentence can release a real vivid language information, is such a variety of authentic language information coupled with repeated sentence structure to play a role to enhance learning. Seen in this light, alone summarized sentence does not completely solve the problem, we need enough sentence containing the desired real sentences, the students left a deep impression on the sentence structure.

**TABLE VI. THE LOCATION OF THE TWO STRUCTURES**

- **such a few memory lacks...**
- **such a good Trace...**
- **such a hard time...**
- **such a bad day...**
- **such a big payoff...**
- **such a short time...**
- **such a large offer...**
- **such a low percentage of...**
- **such a large amount...**

- **a large state...**
- **to complete an organization...**
- **to go to attention...**
- **to fine a pen...**
- **to keep a quiet...**
- **to email an essay...**
- **to receive a subject...**
- **we write a communication...**
- **to typical job...**

**TABLE VII. DISTRIBUTION OF SYNONYMS**

- keenment
- keen perception
- keen pleasure
- keen observer
- keen competition
- keen interest
- keen observation
- keen intellect
- keen eyes
- keen investors
- keen sensitivity
- keen sense
- keen enjoyment

- acute pain
- acute headache
- acute burden
- acute problem
- acute question
- acute deficiency
- acute sense of order
- acute urgency
- acute shortage
- acute behavior
- acute operation
- acute concern
- acute chief executive

- acute injury
- acute damage
- acute solution
- acute heartache
- acute remorse
- acute sense
- acute disappointment
- acute suffering
- acute anxiety
- acute care
- acute assistance

- acute reaction
- acute distress
- acute danger
- acute persuasion
- acute defense
- acute conflict
- acute condition
- acute intervention
- acute need
- acute violence

- acute change
- acute stress
- acute relief
- acute assistance
- acute need
- acute decision
- acute condition
- acute intervention
- acute need
- acute violence

- acute change
- acute stress
- acute relief
- acute assistance
- acute need
- acute decision
- acute condition
- acute intervention
- acute need
- acute violence

- acute change
- acute stress
- acute relief
- acute assistance
- acute need
- acute decision
- acute condition
- acute intervention
- acute need
- acute violence
IV. INNOVATION OF TEACHING INSTRUCTIONS

Once the network Corpus enters the classroom, is bound to have an impact on the traditional teaching mode. Traditional teaching mode is easier in the lack of teaching materials in the environment to form, after language experts generally the Foreign Language refined information by teachers to convey to the students, some students mistakenly think the only serious lectures will be able to learn a foreign language. In fact, the effect of foreign language learning student initiative and take advantage of the ability of foreign language information. Therefore, awareness of the school of modern education are all trying to provide as many foreign language learning materials for students, so that students ability to play the active use of existing resources. Network Corpus is an extension of the existing foreign language teaching resources. Network corpus into the foreign language classroom, students direct contact corpus, direct contact with the language of modern real information, students will feel the wide field of vision, like a dip in the foreign language ocean. In this environment, learning, students no longer need others summarized some of the basic concepts ready, they need to go and found that the language of the law, summarized the need personally find language information. Therefore, teachers need to update the traditional mode of teaching, guide students to take full advantage of corpus resources to make a fuss, and mobilize students to take the initiative to find and recognize the language laws.

As a foreign language learning tools, allowing the network the corpus to enter classrooms should be beyond reproach. But how to manage such a huge amount of information tools, efficient use of network corpus resources as classroom teaching services want to apply the corpus of the problems to be faced in teaching teachers. In the case of network classroom hardware facilities, the key to solve the problem is to have the guidance of the use of network the Corpus capacity of teachers and teaching a matching program. Specifically, the guidance of teachers should include: to enter selected corpus query methods, screening useful information. To consider the content of the implementation network Corpus teaching program:

1) To determine the problem to be solved in the textbook before class and use the program to solve the problem of the corpus.

2) Various spare Corpus shortcut on the desktop and tools dictionary.

3) To guide students to master the basic corpus use, so they gradually adapt quickly to inquiries and participate in classroom activities.

4) The independent design use corpus classroom activities, and guide students to find answers to questions and participate in class discussions.

5) To arrange the corpus job, purpose utilization corpus resources for students in the after-school review and preview process.

The implementation of the core of the network the Corpus teaching program is to give students in the corpus exploration and found that the language of the law the opportunity to discuss and exchange, and organized classroom. To achieve smooth and effective network Corpus teaching process also requires teachers to adjust their roles, teachers not just imparting knowledge, but also to become a wizard and his companions to help students explore knowledge. Teachers should be prepared to deal with the various problems that arise in the process of exploring, to guide students searching for answers in the different network corpus. Under proper guidance, the students will gradually adapt to the new mode of learning, from the the passive role the various definitions and concepts to absorb natural language becomes the ocean of the initiative to the corpus.

V. CONCLUSIONS AND OUTLOOK

Obviously the application of corpora on-line in English classroom teaching is of great significance. The attempts of exploiting on-line corpus resources will fundamentally change our idea of existing foreign language teaching resources. Especially for those remote country schools, on-line corpus resources are particularly valuable in making up for the deficiency of their teaching and learning materials. It is no doubt that exploitation of network corpus resources can effectively enhance the CAI ability, and improve the language teaching efficiency. Such available directly on-line corpora as BNC, Cobuild and LDC, with abundant linguistic data, convenient and efficient inquiry mode, meet the teachers and students’ demands of acquainting themselves with various linguistic phenomena, figuring out various kinds of teaching problems, thus surely will become an indispensable English classroom teaching and learning tool. To promote on-line corpus assisted instruction, we appeal to a large number of front-line teaching staff, who are enthusiastic in innovating the traditional foreign language teaching mode, to boldly explore on-line corpus assisted teaching methodologies, and by extensive exchanges experience of exploitation of on-line corpus to make it give service to China’s modern foreign language classroom teaching.

This paper is sponsored by the Hunan Province General Colleges and Universities Teaching Reform Project (No. 2012-401-482).

REFERENCES


