# A Tentative Corpus-based Study of Collocations Acquisition by Chinese English Language Learners

## UNE ETUDE TENTATIVE DE RECUEIL-BASEE SUR L'APPRENTISSAGE DES ACCORDS PAR LES APPRENANTS CHIOIS D'ANGLAIS

Gao Youmei[1]          Zhang Yun[2]

**Abstract**:  Though collocations have drawn much attention in the field of language acquisition, yet difficulties with them have not been investigated in much detail.

This paper reports on a corpus-based exploratory study that analyzes the mistakes learners made when they produced English collocations. The study shows that not only beginners but also advanced learners have difficulties in choosing the right collocates and the difficulties that learners of different levels have are more or less the same. The biggest challenge for them is to choose the appropriate verbs. The L1 influence on the production of L2 collocations exists at every stage of learning though it varies with the learners' L2 competence. Based on this study, a corpus-based approach is advanced in the end to cope with the difficulties in the acquisition of L2 collocations.

**Key words**:  Collocation, second language acquisition, corpus-based, CLEC

**Résumé**:  Ce document fait un bilan sur une étude explorateur de recueil-basée qui fait une analyse des erreurs commis par les apprenants au cas des accords. Cette étude montre que non seulement les débutants mais aussi les apprenants du niveau avancé ont du mal à choisir un bon terme d'accord et que les erreurs y reviennent au même pour tout niveau. Le plus grand défi pour eux est de choisir le mot juste. Le fait que la langue 1 inflence sur la production du choix d'accord existe au niveau quel que ce soit malgré la variation du niveau de langue 2 des apprenants. Basé sur cette étude, une approche recueil-basée est engagée à la fin pour traiter ce problème existant dans l'apprentissage de l'accord en Langue 2.

**Mots clefs**:  Accord , apprentissage de la langue secondaire, recueil-basée , CLEC

## 1. INTRODUCTION

Although a corpus-based approach to SLA research and foreign language teaching is still in its infancy, there has been a growing interest in this new field. There are several reasons for this. Firstly, it is the computer that has introduced incredible speed, total accountability, accurate replicability, statistical reliability and the ability to handle huge amounts of data (Kennedy 2000: 5). Secondly, a growing awareness of the usefulness of quantitative data provides major impetuses to the re-adoption of the corpus-based language study[3] as a

methodology in linguistics (McEnery and Wilson 1996: 18). Many SLA researchers have found it very difficult to simply follow what theoretical linguists or psycholinguists say about an L1 acquisition model and check if the same abstract linguistic principle is still applicable in L2 learning. More and more researchers now prefer to look at real language performance data instead of relying too much on intuitive or introspective data. Thirdly, it is widely accepted that in modern language classroom the teacher should act as a research facilitator rather than the more traditional imparter of knowledge. Under such a student-centered teaching background, corpus-based adaptive learning has gained much attention (曾用强 2001).

Recently there has been a growing awareness that it is necessary to investigate learner language by collecting a large amount of learner performance data on computer. The term 'learner's corpus' was first used for Longman's learners' dictionaries, in which the information on EFL learners' common mistakes was provided. A project called ICLE (International Corpus of Learner English) was launched as a part of ICE (International Corpus of English) project in 1990. Now more than a dozen of projects constructing learner

---

corpora have been underway around the world. (see the web site: http://www.lancs.ac.uk/postgrad/tono/) In China, an available learner corpus is CLEC (Chinese English Learner Corpus), which was built under the lead of Professor Gui Shichun and Professor Yang Huizhong.

Some researches on language acquisition have focused on the phenomenon of collocation. Alexander (1984) suggests that three Cs, namely, collocation, context and connotation, should be emphasized in the teaching process. Bahns (1993) proposes adopting a contrastive approach to teaching collocation. Wong-Fillmore once concluded, "The strategy of acquiring formulaic speech[7] is central to the learning of language" (Kennedy 2000: 110). Verstraten (1992) points out the need for information on fixed phrases to be included in learners' dictionaries. Moreover, Bahns and Eldaw (1993) argue that, for advanced students, collocations present a major problem in the production of correct English. Stubbs (1999) points out that collocations are a notoriously difficult area for language learners. So we can see that collocation, as a common language phenomenon, has received much attention in language acquisition studies.

However, difficulties with collocations have not been investigated in much detail. With the goal of shedding some light on the problems of CELL (Chinese English language learners) in the production of collocations, the present research was carried out in both ways, cross-sectional and longitudinal. In the cross-sectional study, we focus on the reasons which lead to the occurrence of the errors while in the longitudinal study, we attempt to find out whether there are any manifestations of similarities or differences among learners of different levels in the production of L2 collocations.

## 2. DEFINING COLLOCATIONS

### 2.1 Definitions by Different Scholars

There is no doubt that collocations have long been studied, yet there is no definition accepted due to the fact that different researchers have used different criteria to define them and delimit them from other types of word combinations. Firth (1957: 197) introduces the notion of collocation as part of his overall theory of meaning. He argues, "You shall know a word by the company it keeps". For example, "one of the meanings of night " is its collocability with dark and of dark, of course, collocation with night. He also distinguishes collocation from colligation. According to his idea, colligation is "the syntactical characteristic of the text", while a collocation is "actual words in habitual company". If we take send for an example, colligation is concerned with the patterns where send can be used, such as send something to somebody and send somebody something, while collocation studies what words can frequently co-occur with send. Firth has made quite clear the difference between collocation and colligation, but his definition of collocation as "actual words in habitual company" is too abstract to be feasible in the present study. It seems intuitively right but empirically problematic, for how could one decide whether a combination of words is habitual or not?

Later, Palmer (1981: 79) reviews the previous studies and works out three kinds of collocational restrictions: some are based wholly on the meaning of the item; some are based on range— a word may be used with a whole set of words that have some semantic features in common; some restrictions are collocational in the strictest sense, involving neither meaning nor range, as *addled* with *eggs* and *brains*. Palmer also airs his view on idioms. He says, "Idioms involve collocation of a special kind."(ibid: 79) He uses *kick the bucket* to illustrate his idea. "For here we not only have the collocation of kick and the bucket, but also the fact that the meaning of the resultant combination is opaque— it is not related to the meaning of the individual words."

Like Palmer, Sinclair (1991) also makes his contribution to our understanding of collocations and their difference from idioms. He advances his two famous principles of the organization of language, namely, the open-choice principle and the idiom principle (ibid: 109-115). According to the idiom principle, the choice of one word affects the choice of others in its vicinity. Collocation is one of the patterns of mutual choice, and idiom is another. He defines collocation as "the occurrence of two or more words within a short space of each other in a text"(ibid: 170). His definition of idiom is quite similar to that of collocation: "a group of two or more words which are chosen together in order to produce a specific meaning or effect in speech or writing"(ibid: 172). But the difference lies in the fact that the individual words in idioms are not reliably meaningful in themselves. According to Sinclair, idioms overlap with collocations and the line between them is not clear. He says, if the co-occurrence of words gives a single unit of meaning, it is called an idiom; while if the occurrence is the selection of two related words and each word keeps some meaning of its own, it is called a collocation.

### 2.2 The Working Definition Used in the Study

The difference between collocation and idiom seems, to both Palmer and Sinclair, to depend on whether the meaning of the resultant combination of words is opaque or apparent. Though this way to distinguish collocation from idiom may not be the best one and to some degree, is based on instinct, we have to admit that it is quite effective and also easy to handle in practice. We will advance the working principle applied in our

study on the basis of Palmer and Sinclair. But before this, we have to introduce another term, namely, free combination. Like idiom, free combination can also be easily confused with collocation.

It is widely accepted that combination of words falls into three major classes: free combination, collocation and idiom (e.g.卫乃兴 2002, Nesselhauf 2003). Let's take read a letter as an example. On the one hand, the occurrence of the verb read does not expect the company of a letter, for read can be followed by any work in written form such as a book, a newspaper or a report. On the other hand, the occurrence of a letter does not expect the company of read, either, for any word can appear before a letter so long as it is semantically and syntactically acceptable. For example, we can write, send or even tear a letter. So the senses in which read and a letter are used are both unrestricted. While in the combination deliver a letter, the sense in which deliver is used is much more restricted than that of read in read a letter. Here, deliver, according to the Cambridge International Dictionary of English (CIDE), means, "to take (goods, letters, parcels etc.) to people's houses or places of work". Therefore, deliver, in this sense, can only collocate with a short range of nouns including goods, letters, and parcels. That is to say, the occurrence of deliver, to a certain degree, expects the co-occurrence of a letter. Thus, we should treat deliver a letter as a collocation while read a letter a free combination.

Based on the discussion above, three working principles for the classification of word combinations are advanced:

1st. If we cannot guess the meaning of a combination from the meanings of its member words, it will be treated as an idiom (e.g. to rain cats and dogs). Otherwise, it will be treated as a free combination or a collocation.

2nd. If we can guess the meaning of a combination from the meanings of its member words, and the senses in which its member words are used are restricted or the verb and noun in the combination are mutually expected, the combination will be treated as a collocation (e.g. break the record).

3rd. If we can guess the meaning of a combination from the meanings of its member words and the senses in which its member words are used are unrestricted or the verb and noun in the combination are not mutually expected, the combination will be treated as a free combination (e.g. look at the picture).

## 3. THE RESEARCH DESIGN

### 3.1 Research Questions and Hypotheses

This study is designed to answer the following questions:

**3.1.1** Is collocation competence related to second language proficiency?

1st. To what extent does the use of collocations by different-level learners differ from each other?

2nd. Is there any difference in using collocations between college students who are English majors and those who are not?

**3.1.2** Is there a significant influence of learners' first language on their production of second language collocation?

3rd. Does the influence diminish with the raise of second language proficiency?

4th. Is the influence significantly different between college students who are English majors and those who are not?

Based on the four questions, we advanced four hypotheses:

1st. Higher-level learners have a better command of L2 collocations.

2nd. English majors have a better command of L2 collocations than non-English majors.

3rd. The influence of learners'L1 on the production of L2 collocations will diminish but still exist even at the advanced level.

4th. When producing L2 collocations, English majors are not so much influenced by their L1 as non-English majors.

### 3.2 Materials and the Research Scope

The materials we used were drawn from CLEC (Chinese Learner English Corpus). In CLEC, learners are in five different developmental phases:

1st. ST2 stands for high school students.

2nd. ST3 stands for the first and second-year college students who are non-English majors and will attend CET4.

3rd. ST4 stands for the third and fourth-year college students who are non-English majors and will attend CET6.

4th. ST5 stands for the first and second-year English majors.

5th. ST6 stands for the third and fourth-year English majors.

**Table1: The types of combination**

| Marker | CC1 | CC2 | CC3 | CC4 | CC5 | CC6 |
|--------|-----|-----|-----|-----|-----|-----|
| Type | Noun/Noun | Noun/Verb | Verb/Noun | Adj/Noun | Verb/Adv | Adv/Adj |

CLEC is an annotated corpus. The errors of combination were marked as CC (for detail, see Table One). The distribution of the errors about combinations is shown in Table Two. After comparison we decided to choose verb-noun combination as our research target.

**Table2 : The number of different types of combination errors in each sub-corpus**

|  | CC1 | CC2 | CC3 | CC4 | CC5 | CC6 | Total number |
|--|-----|-----|-----|-----|-----|-----|-------------|
| St2 | 95 | 44 | 211 | 85 | 31 | 22 | 488 |
| St3 | 88 | 207 | 599 | 112 | 47 | 14 | 1,067 |
| St4 | 93 | 60 | 505 | 163 | 36 | 8 | 865 |
| St5 | 38 | 8 | 95 | 53 | 6 | 3 | 203 |
| St6 | 60 | 26 | 162 | 60 | 8 | 1 | 317 |
| Total number | 374 | 345 | 1,572 | 473 | 128 | 48 | 2,940 |

## 3. RESEARCH METHOD

### 3.1 Concordance

Firstly concordance was made with cc3 as the headword.

### 3.2 Correcting the Errors

To ensure the reliability of our correction, both native speakers and dictionaries were consulted during the whole process.

### 3.3 Tagging

Those combinations which did not make any sense and those which had a problematic sentence structure were tagged with NS and STR respectively. They will not be studied due to their irrelevance with collocation. Others were tagged with ID (idiom), CL (collocation) and FC (free combination) respectively according to the working definition. If the same error occurred more than once in a composition, only one of them would be chosen and the others deleted. But, if the same error occurred in different compositions, it would be counted repeatedly.

To find out whether a verb or a noun was used in a restricted sense or whether the verb and the noun in a given combination were mutually expected, the most feasible and practical solution that occurred to me was the use of dictionaries, combined with some native speaker judgments[4]. The procedures were as follows: If

there were no possible restrictions on the use of both the noun and the verb, such combinations as *going back to his hometown* were classified as free combinations. If there were possible restrictions, we would look it up in dictionaries. If we could find the combination or its variations in *NCDCEU* (*The New*

*Century Dictionary of Current English Usage)* both under the entry of its verb and under the entry of its noun, we would consider the verb and the noun to be mutually expected and thus the resultant combination would be classified as a collocation (marked by CL). Lets take *acquire knowledge* as an example. If we consult *NCDCEU*, we will get the following information:

*acquire*: acquire a working knowledge of English;

acquire knowledge through experience

*knowledge*: We acquire knowledge step by step.

With this evidence, it is safe for us to say that *acquire* and *knowledge* are mutually expected to some extent. Thus, the combination *acquire knowledge* should be treated as a collocation.

If one combination only occurs once in *NCDCEU,* we would also look it up in *CIDE.* If we could find the combination or its variations, we would make our decision that it is also a collocation. For example, it was according to this yardstick that *grant one's wish* was classified as a collocation. We found the following in these two dictionaries:

---

[4] If a combination does exist but was not used correctly, it would be tagged by ug1. While if a combination does not exist and both the noun and the verb should be replaced, it would be tagged by ug2. Those tagged by ug1 would be classified into collocations and free combinations according to their original forms; however, those combinations tagged by ug2 would firstly be replaced according to their intended meanings and then tagged again.

*NCDCEU:* wish: grant one's wish

*CIDE:* wish: It's that bit in the story where the fairy grants the little girl three wishes.

grant: She granted their request/wish.

If we could only find one occurrence of a given combination in these two dictionaries, we would ask native speakers for help.

We also analyzed the original combinations and tagged them by NN (stand for the wrong choice of noun), VB (the wrong choice of verb), PP (the wrong choice of preposition) and AR (the wrong choice of article) respectively.

## 4. RESULTS AND DISCUSSIONS

### 4.1 Types of Mistakes

Altogether, 1572 verb-noun combinations were extracted from CLEC, of which 573 were classified as collocations, 492 as free combinations, 16 as colligations, 71 as being problematic in structure, 124 as being meaningless (see Table Three).

**Table 3 Distributions of problematic combinations**

|  | st2 | St3 | st4 | st5 | st6 | total |
|---|---|---|---|---|---|---|
| Collocations | 76 | 210 | 203 | 30 | 54 | 573 |
| Free combinations | 88 | 163 | 146 | 31 | 64 | 492 |
| Colligations | 5 | 3 | 6 | 0 | 2 | 16 |
| Idioms | 0 | 0 | 0 | 0 | 0 | 0 |
| No sense | 7 | 38 | 62 | 6 | 11 | 124 |
| Structure | 8 | 31 | 25 | 3 | 4 | 71 |

Table4 shows the types of mistakes in collocations.

**Table4:Types of mistakes in collocations[2]**

|  | St2 | St3 | St4 | St5 | St6 |
|---|---|---|---|---|---|
| Verb | 54 | 172 | 178 | 27 | 50 |
| Usage 2 | 11 | 15 | 8 | 0 | 1 |
| Noun | 7 | 13 | 10 | 1 | 2 |
| Preposition | 2 | 9 | 0 | 1 | 1 |
| Usage 1 | 1 | 0 | 1 | 1 | 0 |
| Article | 0 | 1 | 0 | 0 | 0 |

### 4.2 The Similarities and Differences in the Use of Collocations Among Different-level Learners

There is an obvious similarity among different-level learners in the production of L2 collocations. They have similar difficulty in the choice of words. It seems that all learners found it difficult to choose the right verbs. Table Four also shows that the types of mistakes in collocations learners made are consistent from st2 to st6. If we arrange different types of mistakes according to the frequency, the sequences we will get are almost the same, with verb on the one end and grammatical words (including preposition and article) on the other.

Bearing the similarity in mind, we are going to analyze the differences? However, only considering the mistakes learners made will definitely lead to bias. We also studied the difficulty co-efficiency of the words used by different learner groups, trying to find out whether there was any difference in the choice of words. The tool we used is VocabProfile[5] provided by the Compleat Lexical Tutor website. This tool can be used to analyze the distribution of words along a scale from the most frequently used 1000 words to those of much lower frequency (Off-list Words). The databases we used are what we got when we made concordances of the problematic collocations (CL as the headword) with the spelling mistakes corrected and all the markers deleted. Table Five shows that from st2 to st6, the percentage that the most frequently used 1000 words take falls from 84.88% to 79.58%, while the percentage the Off-list Words take rises from 3.47% to 6.81%. This indicates that learners tend to use 'bigger' words with the increase of their L2 proficiency. That may account for why the advanced learners still make a lot of mistakes when producing collocations. Thus, making more mistakes does not equal lower collocation competence, yet on the contrary it may indicate the progress learners have made on the way to arriving at the native speaker-like competence. The more collocations they acquire and the more difficult the newly acquired collocations are, the more mistakes they are likely to make in collocation production.

---

[5] The tool can be used online, free of charge.

Table 5: Word frequency distribution

| | St2 | St3 | St4 | St5 | St6 |
|---|---|---|---|---|---|
| K1 words (1-1000) | 84.88% | 85.92% | 81.96% | 81.79% | 79.58% |
| K2 words (1001-2000) | 8.18% | 4.99% | 6.33% | 5.84% | 6.24% |
| AWL words (academic) | 3.47% | 5.37% | 8.16% | 7.22% | 7.37% |
| Off-list Words | 3.47% | 3.72% | 3.56% | 5.15% | 6.81% |

Another factor we analyzed is the collocations that earners used correctly. As we have described, the biggest difficulty for learners is to choose the right verbs. Therefore, we chose a noun and then studied how many verbs were used correctly with it and how many not.

The noun we chose is *knowledge* which not only was used requently and but also seemed difficult for all learner groups. Of all the 577 problematic collocations we found in the previous study, 124 are related to the use of *knowledge*.

Table 6: *Verb+ knowledge* collocations

| | St2 | St3 | St4 | St5 | St6 |
|---|---|---|---|---|---|
| Wrong collocations (WG) | 30 | 99 | 68 | 21 | 13 |
| Correct collocations (RW) | 2 | 60 | 40 | 53 | 20 |
| RG/(RG+WG) | 6.25% | 37.73% | 37.03% | 71.62% | 60.6% |

From Table 6, we can see that college students non-English majors have a much better command of the verb + knowledge collocations than those high school students, but they still have a long way to go to match the English majors. Thus our second hypothesis can be accepted now. However, our first hypothesis can only be partially acceptable. Though the third and fourth-year students are supposed to be more capable than the first and second-year students, the percentage the correct collocations take falls from 37.73% to 37.03% and from 71.62% to 60.6%.

However, this may not indicate that the learners' collocation competence decreases from st3 to st4 and from st5 to st6. Compared with the words used by the first or second-year students, those chosen by the third or fourth-year students are more difficult (see Table Five), thus increasing the possibility of the occurrence of mistakes. So now we can only say we have not found

any significant difference between st3 and st4 and etween st5 and st6.

## 5. THE INFLUENCE OF LEARNERS' L1 ON COLLOCATION PRODUCTION

Now we will discuss the second question. The L1 influence was considered likely if the Chinese equivalent of what the learners apparently attempted to produce was similar to what was actually produced. For example, many learners produced *learn knowledge*, the fact that Chinese often say *xuexi zhishi* and *xuexi* is related to *learn* in meaning led to the assumption of L1 influence.

### 5.1 L1 Influence: Group Comparison

Table 7 The L1 influence

| | St2 | St3 | St4 | St5 | St6 |
|---|---|---|---|---|---|
| Number | 41 | 89 | 55 | 12 | 17 |
| Percentage | 53.94% | 42.39% | 27.09% | 40% | 31.48% |
| Average | 53.94% | | 34.74% | | 35.74% |

We found that L1 influence on the production of L2 collocations varied largely among different learner groups (from 53.94% to 27.09%, see Table Ten). On the whole, the L1 influence descended from beginners to advanced learners, but the change was not smooth.

Instead of the English majors, the third or fourth-year college students were least influenced by the L1. High school students received much more influence than the others. However, there is no significant difference between college students non-English majors and

English majors. This means that college students and English majors are equally influenced by the L1. Thus, our fourth hypothesis proves to be invalid. The third or fourth-year students were comparatively less influenced than the first or second-year students. This proves our third hypothesis.

## 5.2 L1 Influence: Different Types of Mistakes

It is quite clear that, of all the mistakes possibly influenced by the L1, the most frequent type is the wrong choice of verbs (see Table 8)

Table 8 The L1 influence on different types of mistakes

|  | St2 | | St3 | | St4 | | St5 | | St6 | |
|---|---|---|---|---|---|---|---|---|---|---|
|  | Number | % | N | % | N | % | N | % | N | % |
| Verb | 33 | 61% | 80 | 47% | 49 | 28% | 11 | 41% | 17 | 34% |
| Usage 2 | 6 | 55% | 2 | 13% | 3 | 25% |  |  |  |  |
| Noun |  |  | 2 | 15% | 2 | 20% | 1 | 100% |  |  |
| Preposition | 2 | 100% | 7 | 78% |  |  |  |  |  |  |
| Usage 1 |  |  |  |  | 1 |  |  |  |  |  |

Although the L1 seems to have an influence on all types of mistakes, unlike the conclusion reached by Nesselhauf (2003),[6] our study shows the influence on different types of collocational mistakes varies dramatically. As shown in Table Eight, the percentage drops from 100% to 13%. This indicates that the L1 influence is universal but not equally distributed.

## 6. L1 INFLUENCE: POSSIBLE EXPLANATION

Here, three possible causes are summarized according to our understanding:

1st. Learners do not know the differences between their L1 and the L2 they are learning. For example, many high school students produced such collocations as *learn knowledge* or *study knowledge*. One possible cause for this is learners are not aware that *learn* or *study* cannot collocate with *knowledge* in English. So one important task for teachers to fulfill is to arouse students' awareness about the difference between their L1 and the L2 they are learning.

2nd. Learners may know the difference between their L1 and the L2 they are learning, but if producing one collocation goes beyond their existing capacity, they have to borrow expressions from their mother tongue temporarily to facilitate their communication.

3rd . It is believed that there is mismatch between what is taught and what is expected from students (Tognini-Bonell 2001). What students have learned in the classroom may not be sufficient enough to guarantee

---

[6] According to Nesselhauf (2003: 235), the L1 influence on all types of collocational mistakes is of similar strength.

a good communication in daily life. That explains why so many students did not know how to produce such simple collocations as *to acquire knowledge* and *to get into the society*. So, in order to reduce the L1 influence, textbooks should be adapted to the needs of learners or some additional materials should be provided.

## 7. SOME IMPLICATIONS FOR TEACHING

As shown in the study, not only beginners but also advanced learners have difficulties in the production of collocations. Thus, collocation does deserve a place in L2 learning and teaching. Based on this study, a corpus-based approach is suggested.

Firstly, appropriate teaching materials selected should be based on two criteria. The first criterion is that teaching materials should contain collocations which are both undoubtedly acceptable and highly frequent in a neutral register. With the appearance of large corpora and powerful concordancers, it is now possible to arrange the collocations used by native speakers according to their frequency. Such a frequency list will further influence the selection of teaching materials. The more frequently used a collocation is, the earlier it should be involved in L2 teaching and learning. The second criterion is that teaching materials should contain those collocations often used incorrectly by CELL. Teachers, while selecting materials, can get some hints directly from the results of those related studies. If possible, they can also do some empirical work themselves to find out which collocations are most frequently used wrongly by different learner groups respectively. Thus, when teaching learners of a certain level, they will know which collocations deserve more attention from learners and which collocations only need to be mentioned briefly.

Corpus can be used to provide extra information.

Corpus, as "a collection of naturally occurring language text, chosen to characterize a state or variety of a language" (Sinclair, 1991:171), is undoubtedly a source of authentic language usages. Not only teachers can find useful materials from the corpora, but also L2 learners can be instructed to make use of the corpora to assist their study. Teachers can demonstrate to students how to make a concordance of a certain word with the help of computer and how to read the display vertically to find their needed information about collocations related to the headword. Once students become familiar with the corpus analysis, they will find it a great help not only to the acquisition of collocations but also to their English learning as a whole.

Secondly, a corpus which contains the written assignments done by each student in different periods should be compiled. For learners, they can recall those mistakes they made before and at the same time get to know what mistakes they are likely to make at present. By comparison, they may find what mistakes have been corrected and what mistakes still cling to them. Thus, they will be aware that what collocations need to be paid more attention to. Teachers can also benefit from such a corpus, for knowing learners' difficulties means knowing what to stress in the classroom.

Thirdly, teachers can also use parallel corpora to illustrate the use of collocations which might be used incorrectly due to the L1 influence. He can ask a computer to search out all the occurrences of *knowledge* in an English corpus, and then those of *zhishi* in a Chinese corpus. After that they can ask students to pick out the collocations respectively, compare the results and discuss the difference between them. During the discussion, students' awareness of the L1-L2 differences will be raised. Keeping the difference in mind, they will try to avoid being influenced by their L1 while producing L2 collocations. Otherwise, despite having learned the correct collocations, they are still likely to produce the L1 equivalents. Of course, teachers' interference is necessary during the process. They may guide students where to focus their attention. For example, in the learning of verb-noun collocations, students should be told to focus on the comparison between the choice of verbs in English and that in Chinese, for as shown in the study, it is the verb that presents the greatest challenge. Students should be made aware that verbs cannot be used freely, and that collocations acceptable in one language may not be directly transferred into another. However, to ask students to focus on the choice of verbs does not mean to ask them to neglect that of other elements. As we have discussed before, it is not sufficient to merely teach the lexical elements that go together, but the non-lexical elements such as prepositions and articles should also be involved in teaching.

## REFERENCES

Alexander, R.L. (1984). 'Fixed expressions in English: reference books and the teacher'. E*LT Journal* 38(2): 127-134.

Bahns, J. (1993). 'Lexical collocations: a contrast view'. E*LT Journal* 47(1): 56-63.

Bahns, J., & Eldaw, M. (1993). 'Sould we teach EFL students collocations?'*System, 21(1)*: 101-114.

Firth, J.R. (1957). *Papers in linguistics*, *1934-1951*. London Oxford University Press.

http://www.lancs.ac.uk/postgrad/tono/

Kennedy, Graeme. (2000). *An introduction to corpus linguistics*. Beijing: Foreign Language Teaching and Research Press.

McEnery, T. and Wilson, A. (1996). *Corpus Linguistics*. Edinburgh University Press.

Nesselhauf, Nadja. (2003). 'The use of collocations by advanced learners of English and some implications for teaching' *Applied Linguistics,24/2*

Palmer, F.R. (1981). *Semantics*. Cambridge University Press. P75-79.

Sinclair, John. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stubbs, Michael. (1999). Corpus evidence for norms of lexical collocations. In Guy Cook & Barbara Seidlhofer (ed.) *Principles & practice in applied linguistics*. Shanghai: Shanghai Foreign Language Education Press.

Tognini-Bonelli, Elena. (2001). *Corpus linguistics at work*. Benjamins.

Verstraten, L. (1992). Fixed phrases in monolingual learners' dictionaries. In P. Arnaud & H. Bejoint (Eds.) *Vocabulary and applied linguistics* (pp.28-40). Basingstoke: Macmillan.

卫乃兴. (2002). 基于语料库和语料库驱动的词语搭配研究. *现代外语(2)*

曾用强. (2001). 基于语料库的适应性学习模式. *现代外语(3)*: 267.

## THE AUTHORS

**Gao Youmei**, Tianjin Foreign Studies University, Tianjin, 300204, P.R. china.

**Zhang Yun**, Tianjin Commerce University , Tianjin, 300134, P.R. China.