Corpora and the Teaching of Languages

JIAO Ran

Ritsumeikan University, Japan

Email: jiaoran55@live.cn

KEYWORDS	ABSTRACT
corpus linguistics, corpora, English as Second Language	This paper explores the relationship between corpus linguistics and language teaching, with a particular focus on how corpora can inform both language pedagogy and classroom practice. It begins by defining the concept of a corpus and outlining the development of corpus linguistics as a methodological approach to language study. The paper then examines two major areas of pedagogical application: indirect and direct uses of corpora in language teaching. Indirect applications mainly involve syllabus design and teaching materials, where corpus-based evidence helps identify frequent vocabulary, typical collocations, and authentic patterns of language use, thereby narrowing the gap between classroom language and real-life language. Direct applications focus on classroom interaction, especially data-driven learning (DDL), which encourages learners to explore corpora themselves, observe language patterns, and induce grammatical and lexical rules. The study also discusses the role of specialized corpora, such as learner corpora and translation corpora, in areas including error analysis, English for Specific Purposes, and translator training. Through practical classroom examples, the paper illustrates how corpora can enhance the teaching of vocabulary, collocation, and grammar. Overall, the paper argues that corpus linguistics offers valuable methodological and pedagogical innovations for language teaching and holds significant potential for future development in materials design and instructional practice.

1. Introduction

In order to understand what the relationship between corpus and language teaching is, we need to know what exactly corpus is. Corpus approaches the study of language through a collection of texts in a computer-based database. Some dictionaries suggest that corpus is a methodological basis for language studying. (Kennedy, 1998)

Corpus linguistics, is a relatively modern term, thus, the use of corpora for pedagogical purposes is generally classified as a relatively new filed of co-operation between corpus linguistics and the teaching profession. (Kennedy, 1998) It is widely accepted that corpus-based approaches to teaching languages, shows different ways of applying corpus-based research to those learning environments. (Astron, 1997) Many linguistics experts agree that there are three main fields in this study which includes teaching a language for specific purpose, pedagogically motivated uses of corpora, and corpora tools for the language learning context. (Aston, 1997)

The first question which will be dealt with here is when the corpus linguistic and language teaching began. We know that the connective between corpus linguistic and language teaching are new and is far from being complete. (Boulton & Wilhelm, 2006) However, there are a number of programs in the processing, and according to John Sinclair, the earliest development in this field took place at the University of Birmingham in the 1980s. (Sinclair, 2004) Researchers who worked at this university emphasized that the heart of this project was to provide a better English database for learners. Vocabulary is the most important part for the students, but, there are about at least 52, 6000 words consist in English language, so how many words do we need to know? To help solve this problem, language researchers created a lexical database which was later referred to corpus database. This computer-based database can easily be used to analyze English language and to calculate frequencies of a single lexical item. Later, it was pointed out that there are two pedagogical corpora, one is named the direct pedagogical corpus and the other is the indirect pedagogical corpus (Fligestone, 1993).

In this article I will examine the relationship between corpora and language teaching and provide an overview of the most important pedagogical applications of corpora. Of course, in the paper we will also discuss how an electronic corpus of English can be used to help both teachers and learners in teaching languages.

2. Literature review

2.1 Indirect Applications of Corpora in Language Teaching

Corpora in teaching languages, has been discussed for the few years now, and the corpora explored in this section will mainly focus on the indirect applications of corpora in teaching languages.

Thus, what is an indirect application of corpora in language teaching? When we talk about the application of corpora, we have to realize that there are two types of pedagogical corpus applications, one is classified as the use of corpus tools and methods in a language teaching, and language learning context, it influences the teaching syllabus and teaching materials. (Mindt, 1981) As a consequence, indirect applications focus on researchers and materials writers. As Barlow (1996) notes: "the results of a corpus-based investigation can serve as a firm basis for both linguistic description and, on the applied side, as input for language learning." (Hunston, 2002)

2.1.1 Corpora and the Teaching Syllabus

The first question we deal with here is why do we need a corpus for teaching syllabus? To answer this question, we need to know that a teaching syllabus corpus can be systematically analyzing syllabus language. (Willis, 1990) Like any other computer-based corpora, a teaching syllabus corpus provides the frequency used in lexical items and an accurate description of language use. (Willis, 1990) Thus, many scholars are using corpus data to look through the syllabuses and teaching materials. Mindt, for example, noted that to compare the "school English" and "real English" in term of "school English" it does not really exist outside the language classroom, meaning that many language learners have difficulty in communicating with native speakers. (Kennedy, 1992) Therefore, the significant role of corpora is to provide a more realistic example of language usage and natural language.

In addition, corpora may provide language data, and present the frequency words. However, it may not able to provide negative evidence, that means corpora cannot tell us what is correct or incorrect in language. (Kennedy, 1992) Thus, we need to pay attention to the authenticity in language teaching. Basically, the central idea of authenticity is that should we teaching English in natural communicative situations or should we use the specific texts for the purpose of teaching in our language course. It is interesting to analyze what exactly the situation is like at present. Hunston suggests that "the experience of using corpora should lead to rather different views of syllabus design." She classified a "lexical syllabus" which was originally proposed by Sinclair and Renouf (1988). According to Sinclair and Renouf, the idea of lexical syllabus expresses their situations, firstly the commonest word forms in a language; second the central patterns of usage; third, the combinations which they usually form. (Partington, 1998).

In order to continue with lexical syllabus, it is necessary to know that the first

teaching syllabus was probably design of the "Collins COUBUILD English Course" (Willis & Willis, 1989). The contents of lexical syllabus focused on lexis and lexical patterns, it responds to the commonest words and phrases in English teaching. (Willis, 1990). A lexical syllabus covers almost all aspects of language, it is worth mentioning, that Sinclair says that the grammar covered in a lexical grammar, which means "build a grammar and lexis on an equal basis." (Sinclair, 1991) In 1953, Michael West published General "Service List of English Words" (GSL) and has achieved a great influence on syllabus design. West, though, suggests that a teaching syllabus must be based on words rather than on grammatical structures. It is also based on frequently used words rather than rare words. It can be useful for the students to develop their own language skills. (McEnery, Xiao & Tono, 2006)

Another strand of lexical approach to corpus research is teaching collocations. (Willis, 1990) We are often reminded, that the traditional language teaching modals, are learning grammar, language function, and understanding the categories of idiom and phrasal verb. (Wilson, Rayson & McEnery, 2003). The collocation knowledge is important for development in learning foreign languages. (Cowie, 1994) Cowie noted, that "learning a lexical item entails learning what it occurs with and what grammar it tends to have." (p.131) A powerful reason for the employment of collocations is that because language consists of relatively phrases such as noun + verb, adjective + noun. That is, collocation can help both teachers and learners to acquire the frequently lexical items and word combination within a sentence. (Sinclair, 1991) It is important for language teaching, because such sentence knowledge is more important than individual words themselves. As the teacher focuses on word collocation, they cannot only isolate the major grammar patterns and vocabulary meanings, but also will see more than they used to in a text.

In addition to the lexical syllabus, a corpus-based teaching material tries to explain how the second language is used in different contexts and examines language in reality. In Biber's (2002) *Longman Student Grammar of Spoken and Written English*, this focuses on how English is used differently in spoken and written registers.

2.1.2   Corpora and Teaching Materials

As we have mentioned, some textbooks in the previous section do not just inform us of the language teaching syllabus, but also helps with the decision about the teaching materials to use.

The publications in the Collins COBUILD series have got a major achievement, because if we look closer at the English teaching materials, all of them are based on real English and the COBUILD dictionaries, grammars, usage guides, and concordance

samples. (Capel, 1993; Carpenter, 1993; Goodale, 1995; Sinclair et al., 1990) Because these types of language materials provide the information teachers and students need. The advantage of COBUILD is that they incorporate corpus-derived findings and contain genuine. (Capel, 1993). It is the common advantage to other corpus-based materials for example, those books published by Longman, Macmillan and CUP.

As we have mentioned in the previous section, the problem of language authenticity has been discussed for many years, and at present time, there are a number of publications that dealing with this problem. Due to the corpus revolution, the students can choose the course materials from a range reference works throughout corpus-based material database. Now, almost all language teaching materials are drawing on corpus research and are based on real-life data (Barlow & Burding, 2006).

Also, it is worth paying attention to collocation patterns and phraseology that are found in native speaker language data. (Hill & Lewis, 1997) Sinclair has emphasized that the word combination and prefabricated strings must be more natural, fostering fluency, accuracy and idiomaticity in a pedagogical context (Hill & Lewis, 1997)

2.2 Direct Applications of Corpora in Language Teaching

Part of using direct application of corpora in teaching languages, is being interactive between teachers and students. They can create a corpus database and find out about language patterning by themselves. Leech, pointed out that the three main points with the direct uses of corpora include "teaching about", "teaching to exploit", and "exploiting to teach" (Leech, 1997).

"Teaching about" means the teaching corpus linguistics as an academic subject such as syntax and pragmatics. McEnery has explained that "teaching to exploit" means proving students with "hands-on". Teachers should encourage their students, and once they have acquired knowledge of the corpus-based language study, they can exploit corpora in their own way for their own purpose. "Exploiting to teach" means using non-corpus-based methods to teach corpus-based approach language course and linguistic course.

Here we need to introduce a new method for direct application which is data-driven learning (DDL) or corpus-based exercise activities with the language learner and general corpora. (Mindet, 1997) The example illustrated below:

```
 1    ery? Can I think.  .. Well I 'd like to speak about the gallery I like to speak for m
 2    d morning.   Hello! Yeah,  I 'd like to speak about the the squeeze on the benefits.
 3    ught people might be less willing to to speak at the meeting if they knew it was bein
 4    d   pointing. And er when you get up to speak at the conference,  you have to give yo
 5    re 's not quite as bad as when I had to speak for Amnesty on Radio Essex last year an
 6    ke to speak about the gallery I like to speak for myself and er just the visual arts
 7    't got a word  . you 're not allowed to speak for the rest of the week. he 's hiding
 8    Well football fans? well David I ca n't speak on behalf of Hibs,  all I can say Mm. i
 9    a member of the public I 'm not here to speak on behalf of the theatre at all. I migh
10    rhaps people could as actually come and speak to me afterwards,  if they, if they thi
11    e will if you 'd like to come along and speak to him individually afterwards he will
12    hat, then.   When do you start again? I speak to Stella now Do you speak, do you star
13    e appropriate way of doing it.  Shall I speak to Paula about that then?  Yeah. Paula,
14    er Okay.   I 'll speak to Simon.  I 'll speak to Simon erm about borrowing his P C at
15    ot to go    home and do? Are you sure? Speak to me Yes Okay,  right, one person from
16    rt is to go back one step,  not just to speak to people who are experiencing hurt, bu
17    ikey.  Speak up loud, you 've got ta be speak up loud and clear.  No. Uniform. Unifor
18     I think that they may be frightened to speak up and that they 're scared that if the
19     of talking about.  Excuse me could you speak up just a little bit? Yes yes er Thank
20     first floor.  Oh I 'm sorry!  Can you speak up then?  Oh sorry!  Have I Mm mm. Have
```

```
 1    ney. All of that being said and Mr will talk a bit more about the figures when he com
 2     have an M P here on the phone line and talk about this er proposed pay rise. They ca
 3    ? Mm.  . Erm I just want to go back and talk about a few things we just touched on ea
 4    are?   Erm Without drawing it,  try and talk about a square.  What would happen? Erm
 5     Mm.   Like we had you know,  who could talk about experiences you know,  in a, in a
 6    ey include Gerry Addams and we we could talk about the other side,  we could talk abo
 7    ot  , you know, I 'm not er gon na even talk about the disaster at airport things, yo
 8     Mm.   And basically,  again can we just talk about what we 're trying to achieve? My
 9     on, and Mr   I know in a minute we 'll talk about the number of people who attend co
10     Becky. No, Becky. Becky  .  And we 'll talk about this in committee and let you   kn
11    ople or all about that I do n't want to talk about that I want to talk I want to talk
12     get them around the table and begin to talk about those sort of things.  Show a bit
13     bit, we just have done.  We started to talk about the solar system.  How far have we
14     them   anyway. Right.  What else did we talk about and we need to know, we have n't g
15    ber  of possible pieces there,  Did we talk about this?  We did.  Ah, right, yes. So
16    .   They look tired and worn out.  They talk quietly amongst themselves sometimes fin
17    s boss, the professor to come along and talk to us,  and let,  some of their time is
18    than then, because as a kid they do n't talk to you about them things do they like, y
19     going to talk to somebody who will not talk to you,  who will not s possibly even sm
20    how nice the man was when he started to talk to Jason oh I 'll go and help you how fr
```

Figure 1: Concordance samples of 'speak' and 'talk', based on the spoken part of the British National Corpus

How to use corpora in the classroom has been discussed for many years, data-drive learning encourages the students to discover about languages for themselves. Actually, data-driven learning has been proven by "three I's" includes illustration, interaction and induction. (McCarthy, 1995, p.155)

One benefit from the data-driven learning is "corpora will clarify, give priorities, reduce exceptions and liberate the creative spirit." (Sinclair, 1997, p. 38).

2.2.1 Direct Applications of Specialized Corpora

Data-driven learning can be analyzed by teacher-directed or learner discovery, basically the central role is learner. However, it is necessary to note that teachers have to help the learner to do some research on their languages and to make individual contribution. (Leech, 1997) Although data-driven learning is student-central, but teacher also has a prominent role in this process, they have to guidance the students' in different level such as learner's age, language level.

It is widely accepted that people learning a second language prefer to using the general words in special ways in certain genres (Brodine, 2001). As Gavioli (Brodine 2001) stated, "ESP students become familiar with a productive idea of idiomatic language

features, they learn to use and adapt language patterns to their own needs." Therefore, in general language teaching, the small corpus database can be used for data-driven learning materials.

Johns (1991) has noted that there are three stages in the data-driven learning approach: observation, classification and generalization. Observation means the concordance evidence, classification means the salient features, and generalization means the rules of corpus. These three stages are totally in accordance with the McCarthy "three I's".

Seidlhofer suggests that by using corpora for learning, would be crucial in finding out how best to deal with student's errors. In terms of error analysis in a learner corpus, that is, the learners will not only deal with their paper errors but also will be able to analysis their fellow students' papers. On the other hand, the learner will be able to see their own mistakes and the analysis from their classmates (Mukherjee & Rohrbach 2006).

The other type of specialized corpus is translation corpus. Translation corpus is considered as a corpus database that consists of original text and their translations. Aston (1995) notes there are two main perspectives within translation corpus, descriptive and practical. That is to say, descriptive research can compare between the original text and standard translated texts.

Aston distinguishes different types of corpora: monolingual corpora, this corpus only consists of text in a single language; comparable corpora consists of original text and their translations.

Here is an example to help us to understand what parallel corpus is. Johansson (2007) provides an example of using the English-Norwegian Parallel Corpus with a group of Norwegian students in solving the translation problems. This experiment aims to help the younger translator and students find the most acceptable translation. As a result, the translation corpus has considerable application in the area of translator trainee.

3. In the Classroom

If we wish to build a computer-based corpus that for language classroom, we would have to create a corpus database in teaching vocabulary, grammar and language usage. According to previous research from the past few years, many language teachers found out that it is a challenge to use corpus-based activities in their classrooms, moreover, some teachers avoid using corpora because after a few years in the classroom, they felt that teaching in corpus-based materials is a very difficult thing to do (Farr, 2002).

As we have already discussed in the previous sections, all students' English dictionaries, are based on word databases. This corpus-based database provides the basis for checking intuitions about language. In terms of how the teacher should teach,

numerous studies have shown that the teacher should help the leaner to create a frequency list for single words, because this basic corpus teaching model will help students to investigate the vocabulary building and ascertain which words the native speaker use.

4. Methodology and Analysis

4.1 Choosing Key Words

Let's consider the word "work", this word can be understood by dictionary definition and the student will understand its meaning, until they come across the word "job". From this standpoint, a question may be arising, what is the difference between work and job? The differences between these two words not only depend on the dictionary definition, but also on the ways the words are used.

As far as the classroom is concerned, the corpora collocation seems to be important. Teachers are supposed to show word collocates in the classroom, and lecture to the students about how these words should actually be used, and how they collocate rather than explaining what they simple mean.

A corpus database includes almost ten million words but there are numerous words just for terminology, and have a special purpose. As we move along these issues, and the degree of words, it will lead us to a list of the words that are most frequently used in real life. If we want to examine the frequency of words, we need to use the corpus of language. Table 1 shows the most frequent vocabularies in a ten-million-word corpus database (Granger, 1998).

Table 1: Most frequent words: 10-million-word corpus (CIC)

	word	frequency		word	frequency
1	the	439,723	26	as	49,697
2	and	256,879	27	at	49,578
3	to	230,431	28	we	46,025
4	a	210,178	29	her	45,574
5	of	194,659	30	had	45,524
6	I	192,961	31	not	44,977
7	you	164,021	32	no	44,541
8	it	150,707	33	what	44,125
9	in	142,812	34	this	43,024
10	that	124,250	35	like	42,297
11	was	107,245	36	all	41,790
12	yeah	86,092	37	mm	41,639
13	he	78,932	38	er	40,923
14	is	75,687	39	there	39,883
15	on	71,797	40	do	39,744
16	for	69,392	41	his	38,420
17	but	64,561	42	well	37,671
18	she	61,406	43	one	36,889
19	they	58,021	44	just	36,275
20	have	55,892	45	if	36,007
21	with	54,994	46	are	35,279
22	be	52,008	47	oh	35,026
23	It's	50,585	48	right	33,598
24	so	50,531	49	or	32,686
25	know	50,307	50	from	31,444

In selecting vocabulary items from corpus, there are differences between the spoken and written words, for example "well", and "right", are frequently used in the spoken language, and are used in daily conversations. In contrast, the written vocabularies are more academic and formal.

In general, teachers should be able to present the basic nouns to their students. These types of words convey general, non-create words, such as, life, age, problem, room, car, school, door, cat, dog, water, house, accident, cheese, cake, red, along with the basic day, name, terminology, places, basic activities (Carpenter, 1993).

Most frequency corpus words are the par of lexical chunks (Baigent, 1999, p.23). The noun "time" for example,

expression	frequency
all the time	1,019
the first time	834
at the time	733
a long time	657
by the time	583
at the same time	460
in time	323
the last time	238
at a time	216
a good time	127

Table2: Frequency of expressions with "time"

Table 2 shows the frequency of the most common expressions with "time", as teacher, they have to develop their students' investigation and sensitivity to words in general. Once students have learnt the basic nouns, they will be able to involve the different words form the basic forms.

Basic nouns are important for learning language, and they cover everyday things, ideas and frequency words. They are the core of the language.

4.2 Collocation

To address the issue of how vocabulary input can be made more easily available to language teaching in the classroom, we should start with the word collocation.

Collocation is important from a pedagogical point of view, because the ways that words combine in collocation is probably fundamental to all language use. (Jimmie, 1997)

4.2.1 Teaching Collocation

Generally, teacher should help students to identify the collocation in texts. Most of the time, they guide students to read a text and help them to understand the vocabulary within the text. When student asks what the word "views" means in the following sentence: *She holds very strong views on marriage. She thinks everyone should married in a church.* The definition of the word "word" would be in the following ways:

Synonymy: view = opinion

Paraphrase: view = what you think of something

Contextualization: I think it is wrong to kill animals. What are your views?

These definitions are highlighting the useful relations in collocation: adjective + noun – strong views; verb + noun – hold views. (Ter-Minasova, 1996) Traditionally, teacher has to design a syllabus for their teaching in the classroom. In case of the word

"views", teacher may be wanting to add supplementary exercises:

Look at this part of the text:

She holds very strong views on marriage. She thinks everybody should be married in a church.

Most people hold strong views on something. What about you? Write some sentences about yourself following the pattern,

Most people hold strong views on .... Personally, I think ....

This exercise helps students to form their own opinions such as: most people hold strong views on smoking. Personally, I think cigarettes should be banned (Ter-Minasova, 1996).

Teachers have a prominent role to play in teaching collocation in the classroom. While they have to look at their course books and adding exercises, which focus on the word collocation and language creativity skills.

Vocabularies are abstract and cannot take care of themselves, therefore, it is necessary to add some collocation definition and exercise to help students to understand those words. Collocation must become a part of that learning in the classroom.

4.2.2 Collocations and Grammar

Sinclair suggests: "Grammar is part of the management of text rather than the focus of meaning-creation" (Sinclair, 1991, p. 223) This emphasizes that language is not only the study of meaning but also the study of words, collocations and expressions. A corpus can tell us different things about grammar, also there is a general agreement that corpus grammar will help us to understand the traditional grammar and categories (Sinclair, 1991). Nevertheless, corpus would create a links between grammar and lexis for example:

Don't He
She's just
If only you hadn't   *let the cat out of the bag.*
Why did you

Corpus also provides more information for lexico-grammatical, the study of "yet" for example demonstrate below:

```
<$2> Yeah. We haven't got any answer yet We'd like it trimming. <$E> laughs
the wedding.   <$2> I haven't got any yet. Em <$O69> Janet looked lovely <\$06>
but we haven't made er any arrangements yet it's sort of er a bit too early yet
?   <$1> Sorry?   <$2> Has FX arrived yet?   <$1> Who is this?   <$2> MX's f
be in.   <$2> They haven't arrived as yet.   <$1> <$=> It is a whole <\$=> it
yet?   <$1> No not a price breaker as yet. Just their own winter programme.
ame in. <$E> laughs <\$E> Erm but er as yet it's not available in every store.
ll over the place. Em we haven't got as yet a timetable to show you as to what's
 haven't come have they?   <$2> Not as yet. No. Normally about two weeks before
. Well I said I don't know the story as yet.   <$2> Mm.   <$1> <$=> I said But
. But they're not putting anybody up as yet because they have an appeal launch r
ms. Er that's still not p= er set up as yet though. Erm we're gonna do something
n't managed to mark any of your work as yet but I I promise I'll have it back to
Manda are you ready for your assessment yet?   <$F> I think so yeah.   <$1> I'
 Anyway you obviously haven't gone back yet so <$=> erm I won't be er <\$=> you
t know. <$G?>   <$1> Oh he's not back yet.   <$2> No.   <$1> Oh right.   <
eeks ago. And he he hasn't written back yet. So <$E> laughs </$E>   <$1> No. Mm
G?>.   <$4> Have you changed your bank yet?   <$3> My turn. <$E> sighs <\$E
   <$1> Bye. Cheers.   <$3> Won't be yet until I've <$O13> lost <\$O13> a lit
<$2> Have you seen Beauty And The Beast yet?   <$1> No I was wanting to go.
p to see me every year. She hasn't been yet. And she and I like to trip out on a
tomorrow.   <$6> No. No. Not for a bit yet.   <$3> Good.   <$6> We we thought
```

Figure 2 Concordance list for yet.

This example shows that "yet" is commonly used in a negative environment, such as, "I haven't got any yet." And "as yet" is the common pattern in this environment.

Another example of grammar we will be discussed here is verb "be":

form	frequency	form	frequency
he's not	704	he isn't	18
she's not	476	she isn't	15

Table 3: frequencies of verb be

The simple verb "be" is a particular tense marker, based on this table it is not difficult to see that corpus provides useful information for both teachers and students, and this information will help learner to avoid mistakes. (Lewis, 1993) For example:

Exercise

Write yes-no questions. Then write true answers.

  1.   *you/shy ?*

  Are you shy?   *Yes, I am.*

  2.   *this class/easy ?*

  Is this class easy?   *No it isn't.*

The examples above present no such particular pattern which is important to acquire a new language. Studying corpus helps us see how grammar structure and lexis work.

4.3 Classroom Activities

The teacher has the responsibility to guide the students and to draw the students' attention to the most frequent corpora. They should discourage students from re-coding every collocation they meet, this means they must avoid re-coding very weak items, such as nice house, good vocation, because not all vocabularies are useful in daily conversation, students only have to know the majority of the words.

In some sentences, it is necessary to bring collocations into the classroom. The teacher has to make students aware of collocations, like the students' native language, the English language has some words that go together. Once the teacher explained collocation in the class, they can adapt to collocation books. If they are using vocabulary books based on frequent word collocation, students can acquire the English lexicon as quickly as possible (Romer, 2006).

Isolating the frequency words in the classroom, there must be some questions that relate to the meaning of vocabulary, for instance, when students ask a question such as "what is the difference between point and view?" the teacher may want to provide a contextualized example with each word rather than spend a great deal of time by explaining each word. The reason is that students need to explore the language rather than learn an individual word. Below is a classroom scenario:

Student: What does the word "point" mean?

Teacher: We use point in different ways, here are some typical ways we use it.

Why do you want me to do that? I can't see the point.

I know you want to come but, the point is, you're not old enough.

For the teacher, it is not easy to explain exactly what "point" means, but when the students are provided with some different sentences with this word, the student will be able to explore the word in different situations, and realize how to use this it in such different ways (Romer, 2006).

It is becoming clear that the students make mistakes when trying to talk discuss something. Do not just correct the mistake, provide some examples and extra sentences, for example:

Student1: I have to make an exam.

Teacher: (Writes "exam" on the board)

    What verb do we usually use with "exam"?

Student2: Take.

Teacher: Yes, that's right. (Writes "take" on the board)

    What other verbs do we use with "exam"?

Stuent3: Pass.

Teacher: And the opposite?

Student2: Fail.

Teacher: Yes. (Writes "pass" and "fail" on the board)

According to the example above it is clear to see that the teacher has not only corrected the mistake but also provide other words for the students to use.

After studying corpus-based language learning, for a couple of months, students will know many more English words, however a problem may arise, they do not really know how to use those words in the real life. Therefore, the teacher should design teaching materials for what students already know. As Skehan suggested that "very often the pedagogic challenge is not to focus on the brand new, but instead to make accessible the relatively new" (Sinclair, 2004, p. 300). For example the teacher can take such words, usually make a short list of nouns, adjectives and verbs which students think go with those nouns.

Remember that, it is better to ask questions with "or" rather than "and" because as a teacher you will get a full sentence in responds.

5. Conclusion

This paper has focused on the relationship between corpus linguistics and the teaching of language. From many teachers, corpus is another useful way of teaching languages, and indeed, corpus can be used to design the teaching materials and syllabus. I hope I have shown you some basic concepts and pedagogical methods used in the classroom. Teachers of course, have to support their students, and teach them in the best way possible. The relationship between corpus linguistics and language teaching seems very close, because when the teacher brings corpus linguistics into the classroom it is important to realize that they are also adapting the methodological innovation and the theoretical innovation. Bothe innovations will entail the new way of teaching. Considering the future, there is still a lot of scope for the development, of materials, exercises, course books, so we should look forward to studying the field of corpus linguistics and of teaching languages.

References

Aston, G. (1995). Corpora in language pedagogy: Matching theory and practice. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 257–270). Oxford University Press.

Aston, G. (1997). Enriching the learning environment: Corpora in ELT. In P. Wichmann

et al. (Eds.), *Teaching and language corpora* (pp. 51–64). Longman.

Baigent, M. (1999). Teaching in chunks: Integrating a lexical approach. *Modern English Teacher, 8*(2).

Barlow, M., & Burdine, S. (2006). *American phrasal verbs* (CorpusLAB Series). Athelstan. http://www.corpuslab.com

Biber, D., Leech, G., & Conrad, S. (2002). *Longman student grammar of spoken and written English*. Longman.

Boulton, A., & Wilhelm, S. (2006). *Habeant corpus: Should learners have the body? Tools learners have the right to use*. Paper presented at the 27e Congrès du GERAS "Cours et Corpus", Université de Bretagne Sud, Lorient, France, March 23–25, 2006.

Brodine, R. (2001). Integrating corpus work into an academic reading course. In G. Aston (Ed.), *Learning with corpora* (pp. 138–176). Athelstan.

Capel, A. (1993). *Collins COBUILD concordance samplers 1: Prepositions*. HarperCollins.

Carpenter, E. (1993). *Collins COBUILD English guides 4: Confusable words*. HarperCollins.

Carter, R., Hughes, R., & McCarthy, M. (2000). *Exploring grammar in context: Grammar reference and practice*. Cambridge University Press.

Cowie, A. P. (1994). Stable and creative aspects of vocabulary use. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. xx–xx). Longman.

Farr, F. (2002). Classroom interrogations – How productive? *The Teacher Trainer*, 19–23.

Fligelstone, S. (1993). Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal, 17*, 97–109.

Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. xx–xx). Longman.

Hill, J., & Lewis, M. (1997). *LTP dictionary of selected collocations*. Language Teaching Publications.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Johansson, S. (2007). Using corpora: From learning to research. In J. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpus linguistics in use*. Rodopi.

Johns, T. (1991). Should you be persuaded? Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom concordancing* (*ELR Journal*, No. 4, pp. 1–16).

Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.

Leech, G. (1997). Teaching and language corpora: A convergence. In P. Wichmann et al. (Eds.), *Teaching and language corpora* (pp. 1–23). Longman.

Lewis, M. (1993). *The lexical approach*. Language Teaching Publications.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

Mindt, D. (1981). *An empirical grammar of the English verb: Modal verbs*. Cornelsen.

Mindt, D. (1997). Corpora and the teaching of English in Germany. In P. Wichmann et al. (Eds.), *Teaching and language corpora* (pp. 40–50). Longman.

Mukherjee, J., & Rohrbach, J.-M. (2006). Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. In B. Kettemann & G. Marko (Eds.), *Planning, gluing and painting corpora: Inside the applied corpus linguist's workshop* (pp. 205–232). Peter Lang.

Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. John Benjamins.

Römer, U. (2006). Looking at *looking*: Functions and contexts of progressives in spoken English and "school" English. In A. Renouf & A. Kehoe (Eds.), *The changing face of corpus linguistics* (pp. 231–242). Rodopi.

Sinclair, J. McH. (1991). *Corpus concordance collocation*. Oxford University Press.

Sinclair, J. McH. (Ed.). (2004). *How to use corpora in language teaching*. John Benjamins.

Ter-Minasova, S. (1996). *Language, linguistics and life: A view from Russia*. Moscow State University Association.

Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. HarperCollins.

Wilson, A., Rayson, P., & McEnery, T. (Eds.). (1997). *A rainbow of corpora: Corpus linguistics and the language of the world*. Lincom Europa.

Author Bio:
RAN JIAO is a PhD candidate at the Graduate School of International Relations, Japan. Her primary research interests lie in the areas of international education, higher education, and culture sociology. Email: jiaoran55@live.cn