Application of Corpus Linguistics to EFL Teacher Education in

South China Normal University

Anping He

South China Normal University, China

Abstract: This paper reports the application of corpus linguistics to EFL teacher education in South China Normal University in the past 10 years. Focus is laid on solving problems including: where to get relevant corpus for learners, how to design corpus-aided activity for daily teaching goals, how to make corpus manageable in classroom teaching and how to enhance teaching practice by corpus research.The practice is featured in four aspects: 1) Constructing and sharing EFL pedagogical corpora by both learners and teachers, making it as a component in teacher education courses. 2) Starting from textbook-corpus analysis and resulting in corpus-aided exercise design, taking it as one of the goals in teacher education. This includes investigating quantity and quality of textbook input and salient features in different types of exercise design. Linguistic forms and patterns retrieved from the corpus are further associated with pedagogical ideology embedded. 3) Implementing data-driven learning approach in classroom teaching by improving the teaching environment in terms of both soft ware and hard ware. 4) Reflecting teaching effect by corpus-based research.Some examples of the above are presented and problems are discussed. All this indicates that extension from corpus linguistic research to teaching practice has to be initiated in language teacher education.

Keywords: EFL corpora, joint construction, course book analysis, exercise design, implementation

Introduction

"Teaching is a natural extend of research" (Leech 1997: 3).Yet the application of corpus linguistics to langue education, classroom teaching in particular, is not to an extent as is expected. One reason is that teachers themselves have not had the ideology and technique of corpora. In China, there is a population of 50 millions of learning English as a foreign language. We are now undergoing a new round of curriculum reform in English education, which is aiming at a 'big leap forward' development but based on a low level English proficiency as a whole throughout the country and poor teacher resources. South China Normal University is to educate pre-service and in-service EFL teachers for secondary and tertiary schools in Mainland China and we therefore should play an initial role to bring corpus linguistics into EFL teaching.

In the past 10 years of applying corpus linguistic to EFL teacher education, we have been trying to solve four problems: 1) where to get relevant corpus for teacher

trainees, 2) how to design corpus-aided activity matching the ongoing daily teaching goals, 3) how to make huge corpus data manageable in classroom teaching, and 4) how to evaluate teaching effect. This paper is reporting our considerations and practice.

Construction of corpora for EFL teacher education

Together with importing huge amount of English corpora abroad as a reference corpus, we also build our own pedagogical corpora for routine course learning and teaching practice. This is a joint effort of teacher educator and teacher trainees. We make it as part of course practice by requiring in-service and pre-service teachers to bring in their small portion of data when they attend our courses such as "*Basic English Phonetics and Phonology', "EFL Syllabus Design and Teaching Material Development", "Corpus Linguistics & EFL Teaching" and "Discourse Analysis*'. The data includes scanned textbook, classroom teaching video transcription and students' written and spoken performance in every year's examinations. All this is pooled into our *Corpora of EFL Education in China* (CEEC) and open to the trainees to do corpus analysis during the processing of the above courses. The corpora keep expanding every year and it now comes to a size as follows:

Structure of CEEC (He 2007)

- Imported Corpora (300 million)
 – Native English speaker' spoken & written data
- Adults / teenagers / children
- Literature: 3000 classic works …
- Self-built Corpora (9 million）
 – EFL teaching materials：(2.88 million)
- 120 course books at tertiary, secondary & primary level at home & abroad
 – EFL classroom teaching (0.8 million)
- 222 classes at tertiary, secondary & primary level at home & abroad
 – EFL learners' inter-language: (5.54 million)
- Spoken & written data at tertiary, secondary & primary level at home & abroad ）

The imported corpora offer us information on the centrality and typicality of the target language in use, while the self-built corpora connect corpus research findings with daily teaching objective and exercise design. Both contribute to fine resources to EFL education which we did not have ever before.

Design corpus-aided exercises

As information and communication technology (ICT) "can not longer be an added extra but rather an intrinsic part of teacher's methodological repertoire" (O'Keeffe & Farr 2003: 389), we teacher educators took a lead in cooperating corpus resource and techniques in the above courses and then require trainees to conduct corpus based research and exercise design with principles as follows:

1) Select specific teaching goals from current textbooks and further elaborate them as a problem to be solved;

2) Retrieve authentic information from the two types of corpora above either in form of concordances, collocation list, cluster list, wordlist or keyword list;

3) Edit the data and make it acceptable by students in terms of vocabulary and grammar structures;

4) Provide practical guiding instructions, i.e., questions to be answered step by step by different observation focus;

5) Write in teaching notes about possible answers and reasons.

Enlightened by concepts of corpus linguistic such as "frequency driven", "co-text meaning construction" and "lexicalgrammar" (Sinclair, 2004), trainees conduct corpus based research and design corpus-aided exercises with topics include but not limit to the following: "association between pronunciation and spelling", "word constructions", "lexical item with features of collocation, colligation and semantic preference", "micro-language skill training of guessing, catching gist, plotting and discourse structure", "cross-cultural language expressions comparison", "genre analysis", "literature style study". The following are a few examples of corpus-aided exercises based on a unit named *Nelson Mandela* from a popular EFL textbook for middle school in China, ranging from word study to skill training.

Case 1   Word construction

The word structure of "Adj. + *ness* → noun" is a learning target of the unit. By presenting a list of words ending with *ness* which is retrieved from the textbook corpus,students are to observe and answer questions including:

1) What words are not Adj. in their original?   (→word class identification)

2) What change has to make to the Adj. in spelling after adding *ness* ? (→spelling regulation)

3) Is there any change in the word stress pattern after changing the part of speech? (→pronunciation regulation)

4) Is this type of Adj. sharing some meaning in common? (→ semantic preference)

Case 2   Word collocation

"Came to power" is a verb phrase highlighted in this unit, but the teaching can be extended to the delexicalization of *COME* , i.e., *COME to* + (none physical places). For more details, see those lexical items highlighted in Version 1 below. Guidance for observation includes:

1)   Observe nouns just after "*come to*" and think if they refer to some actual places or to certain situations? Name some of the later.

2)   Find how many of these nouns are actually coming from verbs (e.g., "*conclusion")* is from the verb "*conclude")*?

3)   Paraphrase the sentences by using the verb forms instead (e.g., change "*came to a conclusion*" into "*concluded that …*". Try at least 3 of them and think about the differences.

4)   Translate some of the sentences into Chinese, such as "*come to hand" "come to my mind", "come to life"*…

Case 3 Retell the life story of Mandela

A keyword list of this unit in relates to the whole textbook reveals the 'aboutness' of the reading materials in the unit referring to Nelson Mandela. The concordances of *Mandela* and *he* demonstrate the major content points and ways of expressions which are useful in retelling this hero's life story. This can be obtained by attending to those action verbs and verbal verbs around the two node words, telling students about what Mandela has done and what he has said (as can be seen below).

No.	Token	keyness	word
1	53	402.352	I
2	27	222.242	Mandela
3	22	181.086	Nelson
4	19	156.393	Elias
5	14	115.237	Africa
6	20	80.038	black
8	14	76.881	south
9	16	74.938	prison
10	8	65.850	ANC
11	9	48.123	guards
13	39	43.252	he
14	5	41.156	league
16	9	41.098	workers
18	10	37.259	government
20	5	35.782	youth

Extract of Keyword list of *Story of Nelson Mandela*

```
Mandela was a very difficult period of my life.
he had opened a black law firm to advice poor bl
Mandela told me what to do and helped me was one
He told me how to get the correct papers so I co
he was and when he organized the ANC Youth Leagu
he organized the ANC Youth League, I joined it a
He said: "The last thirty years have seen the gr
Mandela said: "We were put in a position in whic
Mandela was also there and in one way it helped
Mandela began a school for those of us who had s
He taught us during the lunch breaks and the eve
Mandela allowed the prison guards to join us. He
He said they should not be stopped from studying
Mandela and the ANC came to power in 1993. All t
Mandela remembered me and gave me a job taking t
```

Extract of Concordances of *Mandela* and *he*

Case 4 Pattern and meaning
"--- *because* + (clause)" and "*because of* + (noun phrase)" are grammar patterns in the unit. After comparing the content of the clauses and noun phrases after the two patterns in the format of cluster lists retrieved from both textbook corpus and reference corpus, the designer of the exercise comments:
... we always assume that <u>because</u> + a clause and <u>because of</u> + a phrase can be used

alternatively without any change of the flavor, the corpora has helped us to bring out the hidden knowledge of these two: excerpt for their similarities, <u>because of</u> has a negative semantic prosody, and it may be used to 1) express uncertainty of judgment, 2) to bring out undesirable consequences, or 3) to excuse oneself.

This is an in-depth learning of grammar pattern, which can be further extended to the study of *owing to*, *due to* and *thanks to*.

Implementation in classroom teaching

To make the corpus-informed or corpus-based language input manageable in classroom teaching, we have tried to improve the presentation format and technique by:

1) Use mini-file as input file in class, which is a practice modifying Sinclair's tasks design in his book *Reading Concordances* (2003)

2) Use *Antconc* (Anthony 2006) to justify font size, colour highlighting and techniques of cut-past print-screen on to doc. file or PPT file for class presentation.

3) Use school inner-net to store all corpora above and offer free access to all users with only output files downloadable but no copy of the whole corpus for copy rights projection.

For example, the following version 1 is proved to be more acceptable than version 2 by middle school students.

Version 1: Extract of edited mini-file

Version 2: Extract of original mini-file retrieved by Antconc.

Reflection and research

After designing exercises or trying them in teaching practicum, trainees are to write reflection reports as course papers. This is a stage to enhance the course practice to a theoretical consideration of teaching materials development and classroom teaching practice. The following are some of the key phrase in the titles of their research papers or reflection reports, covering various aspects of EFL education.

On textbook analysis：

-- Ideology of humanism / gender equality / cross-culture awareness / globalization embedded in textbooks

28

-- Cognitive demanding in exercise design before and after curriculum reform

-- Basic vocabulary investigation in its frequency, central meaning and typical pattern

-- Features of orality in textbook dialogues

--"3-dimension grammar teaching" and grammar exercise design

-- "Lexical grammar" & lexical teaching design

-- Schema theory, corpus & reading skill training

On theme & stylistic features in literature reading:

-- Color words in *Sons and Lovers*

-- Beauty of color in *Wilde's fairy tales*

-- Relationship between nature and humanity in *Walden*

-- Shakespeare's view point on the four seasons

On learners' inter-language:

-- "Small words" in LINDSEI-Chinese corpus

-- Spelling errors / connective devices / attitudinal adverbials in learners English compositions

-- Text structures in abstract writing

On classroom discourse:

-- Negotiation sequence in classroom conversation

-- Questioning / feedback giving/ repairing / code switching / dis-fluency / in teacher talk

These are all corpus-based or corpus-aided papers and most of them have been published in books and journals in China.

Conclusion

The above description indicates a smooth circle from corpus building and research to practical teaching and again going back to corpus expanding and researching further. However, as a new way of teaching and learning language, corpus-using teachers have to "maintain control of a potentially large quantity of evidence while trying out generalizations and this requires intellectual skills that have not traditionally been taught (Sinclair 2003: vii). Such a demand challenges our corpus application. We still have many problems, including: 1) how to guide trainees to make corpus analysis, especially from simply identifying repeated forms to categorizing similar semantic or functional groups; 2) how to select and edit corpus examples to meet learners' current proficiency; and 3) how to keep offering relevant corpora to the trainees when they graduate from our school. All this drives us to further study and practice.

References

Anthony, Laurence. 2006 *AntConc 3.2 Ow.bet a3(windows)*, a free software developed in School of Science and Engineering , Wesada University, Japan.

He, Anping. 2007 "Corpus-aided analysis of EFL course books." In *Curriculum. Teaching Material. Methodology.* Beijing: People's Education Press, 44-49.

Leech, G. 1998. "Preface." In *Learner English on Computer*, S. Granger (ed.). London: Longman.

O'Keeffe, MA. & F. Farr. 2003. "Using language corpora in initial teacher education: Pedagogic issues and practical applications." *TESOL Quarterly* 37/3:389-418.

Sinclair, John. 2004 *Trust the Text*, Routledge: London.