

# *Study of Corpus and the Development of English Textbook*

Yang HE

Department of Finance and Trade  
Ningbo Dahongying University  
Ningbo, China

**Abstract**—This study investigates the differences between what textbooks are teaching and how native speakers actually use language as evidenced in the corpora. The author first makes a review of the literature on using authentic versus invented examples in the ELT context, followed by a critical account of what is meant by ‘native speaker’ and the implications following this classification a comparison between examples from textbooks and/or dictionaries and evidence you have extracted from a corpus. A critical discussion of possible difference is shown after analysis of the differences. The author also offers some implications on using corpora to inform textbook and dictionary design.

**Keywords**—*Corpora, Traditional Textbooks, Native Speaker, COCA*

## I. INTRODUCTION

In China, English language teaching aims to promote students’ overall language ability, which is composed of five interrelated components, namely, “language skills, language knowledge, affects, learning strategies and cultural understanding” [8]. The ultimate goal of language instruction is to equip the learners with the ability to use the language for communication. As more and more emphasize has been given to communicative competence of students, therefore, textbooks have designed to cater to this need. Textbooks are the fundamental tool of learners. The study found that “98% classroom instructions get the source from textbooks instead of the teachers, while 90% homework also gets guidance from textbook” [6]. However, with the advent of corpus linguistics, the analysis of authentic corpus data makes it possible to shed light on language learning and teaching. “Corpus offers rich materials for foreign language instruction and major source for compiling dictionaries, grammar and various coursebooks and corpus can be regarded as bases for outline design and textbook development” [14]. Therefore, the traditional English textbooks are challenged. For instance, the subjectivity of the contents of textbooks is questioned.

Corpus, however, has not been completely applied to English instruction in China due to some restraints. Consequently, Chinese students can hardly enjoy the authentic materials and environment. Therefore, it is crucial to combine the traditional textbooks or dictionaries with authentic materials from corpora. This essay is going to discuss the comparison between examples from English textbooks used in Cixi, China and their evidences extracted from one corpus:

Corpus of Contemporary American English (COCA). The project will firstly review literature on previous studies on using traditional textbooks versus corpora in English Language Teaching (ELT) context as well as a critical account of “Native Speaker”. Secondly, specific corpus-based analysis will make it clear the possible differences between textbooks and corpora. Finally, using corpora to inform textbook and dictionary design will be discussed.

## II. LITERATURE REVIEW

### A. Review of Textbooks in ELT Context

#### 1) Description of textbooks

Textbooks are tools for teachers and also liken to teachers as media bridging a target language and learners. Textbooks denote “materials”, referring to “anything used by teachers or learners to facilitate the learning of a language” [7]. They can be in the form of “coursebooks, dictionaries, grammar books, videos, teacher’s book, photocopied materials etc.” [7].

#### 2) Previous studies on textbooks application in ELT

Having evaluated various textbooks, Cunnings worth [7] mentions multiple roles of coursebooks in ELT: “A resource for presentation materials (spoken and written), a resource of activities for learners practice and communicative interaction, a reference source of stimulation and ideas for classroom language activities, a syllabus (where they reflect learning objectives which have already been determined), a resource for self-directed learning of self-access work and a support for less experienced who have yet to gain in confidence”.

Murray and Christison [8] proposed another perspective of textbooks. According to them, “textbooks impact learner identity and therefore their investment in learning English. Textbooks often reflect stereotypical views of race, gender, ethnicity or nationality” [8]. Carter [8] also argued that “the language coursebook represents a “can do” society on which interaction is generally smooth and trouble free, the speakers cooperate with each other politely, the conversation is neat, tidy and predictable, utterances are always as complete as sentences and no-one else can interrupt anyone else or speak at the same time as anyone else”. However, not only do such English textbooks provide unrealistic models for learners, but also imply that English as a language, operates differently from what learners are used to in their own language. Consequently, learners, undoubtedly, may be confronted with troubles in using the target language in the real situations.

It can be concluded that some scholars held negative views toward the existing textbooks and pointed out the limitations of the textbooks. However, there are few studies on making comparison between textbooks and other advanced teaching tools, e.g. corpora to develop the textbooks.

### *B. Review of Using Corpora in ELT Context*

Corpus has gained great importance in English language teaching. According to Sinclair (1991) and Stubbs (1996), “corpora lead to new description of a language, so that the content of what the language is teaching is perceived to change in radical ways. Meanwhile, corpora themselves can be exploited to produce language teaching materials, and can lay basis for new approaches to syllabus design and to methodology” [10].

#### *1) Previous studies on using corpus in ELT*

John (1986) was “one of the pioneers who advocated the direct use of corpora in language teaching and learning” [5]. As Li [5] mentioned, John (1986) proposed “data-driven learning (DDL)”, which means “the use of computer-generated concordances in the classroom to get the students to explore the regularities of patterning in the target language”. According to Sinclair and Renouf [5], “the primary task of foreign language teaching is the words of high frequency, essential meaning and the common collocation based on language use frequency”. By using a frequent list, we can see clear core vocabulary.

It can be drawn from the above studies that corpus linguistics in language pedagogy is selecting examples from authentic language used in real contexts which is collected in corpora. However, it is still disputable over the so-called “Native Speaker” model. This will be discussed through a critical account of what is meant by “Native Speaker”.

### *C. Native Speaker (NS)*

Historically speaking, Native Speaker is defined as “the first language a human being learns to speak is his native language; he is a Native Speaker of this language” by Leonard Bloomfield [1]. Later on, Native Speaker is “an ideal speaker-listener, in a completely homogenous speech community”, as defined by Noam Chomsky [3], whose position was that “non-native Englishes are inadequately learned versions of “correct” native English forms and therefore not valid as teaching and models”. However, Many scholars have also claimed that “the NS is myth or fallacy” [2]. According to Murray & Christison [9], Paikeday (1985) was the earliest who stated that “Native Speaker is dead, in response to Chomsky’s idea of speaker-hearer” and “it doesn’t allow to assign each individual to native and non-native speaker, because many factors contribute to a person’s language use”. As a result, there is still no consensus on NS. However, as is claimed by Davis [9], “We need native speaker as model, a goal almost as an inspiration”.

There are many existing corpora which contain the authentic languages by “Native Speaker”, such as BNC, COCA, and BROWN etc. Due to the notion of “Native Speaker”, it is possible to examine how a certain word or sentence is expressed or shown differently in our textbooks

and corpora. Then it may offer some implications on making some developments on textbooks design.

## III. METHODOLOGY

### *A. Research Questions*

1) What are the differences between examples from textbook *Go for It* and their evidences extracted from COCA?

2) Based on the differences in the previous question, what are the problems and limitations of the design of textbook *Go for It*?

### *B. Subject*

In this study, the research mainly investigates the textbook “*Go for it, Grade 8, Volume One*” used in high schools of Cixi, China, which was edited by People’s Education Press and American CENGAGE Learning in 2013.

### *C. Instruments*

COCA online will be applied to verify the evidence of examples from the textbook. It was created by Professor Mark Davis of Brigham Young University. This corpus contains 520 million words of materials published in America from 1990 to 2015. It covers spoken corpus, fiction corpus, magazine corpus, newspaper corpus and academic corpus. The online interface allows users to search for specific words or phrases, lemmas, part of speech by frequency.

In this project, both qualitative and quantitative methods will be applied in data collection. Quantitative data is going to answer how the examples in the textbook and the evidences in COCA differ. Meanwhile, qualitative method will be used to analyze and make conclusion about the limitation of the textbook design, which can also offer some implications on the development of textbooks.

## IV. ANALYSIS

In this part, the researcher is going to illustrate some examples from the textbook “*Go for It*” and their evidences in COCA from different aspects to show the possible differences between textbook instruction and corpus. In this respect, the existing problems of textbook design are presented. According to Hunston [10], “the use of corpora has changed dictionaries and other reference books through a series of new emphases: an emphasis on frequency, an emphasis on collocation, an emphasis on variation, an emphasis on lexis in grammar and an emphasis on authenticity”. The differences between the invented textbook and authentic language in COCA will be analyzed through these aspects.

### *A. Frequency*

According to Hunston [10], “one area in which speaker intuition is acknowledged to be of very little use is in the assessment of relative frequency between words, meanings and usages”. Compared with textbooks, corpus has the inclusion of information about relative frequencies.

1) The choice of words

According to Zhang [14], “when learners are using a second language, it is often the case that they show their hesitation as to the appropriate choice of a certain word, e.g. making a choice among people, person, individual or one”. Neither the dictionary nor the coursebooks can display a true picture. Besides, Biber [14] mentioned, “coursebooks lay emphasis on concrete uses instead of abstract one or rhetorical cases. Abstract uses, however, show high frequency of occurrence in authentic language, like ‘way’ to show the way to do some activity as the way of cooking fish or the only way out of the difficult situations”. “The word ‘Thing’ is another case of preferring abstractness” [14]. Take “Thing” as an example, in this textbook “Go for It”, the researcher has counted that the total number of “Thing”, which is shown below.

TABLE I. THE DISTRIBUTION OF THE NUMBER OF “THING” IN EACH UNIT OF THE TEXTBOOK (GO FOR IT, GRADE 8, VOLUME ONE, 2013)

Unit	Topics	Number of “Thing”
1	Holidays and vacations	0
2	Free time activity	0
3	Personal traits	1
4	Your town	2
5	Entertainment	0
6	Life goals	0
7	Life in the future	1
8	Cooking	0
9	Invitations	0
10	Decision Making	2

It can be drawn from TABLE 1 that this textbook seldom uses the abstract word “Thing”. However, at COCA interface, it shows that “Thing” has a high frequency of 245054 items.



Fig. 1. the Top 20 of “Thing” in the Context

From Fig.1, the Top 20 of “Thing” in the context is displayed as KWIC. With the help of frequency in corpus, we can conclude that Native Speakers have frequent use of “Thing”, while textbooks seldom apply these kinds of abstract words.

2) The choice of certain phrases

Since the English textbook “Go for It” is published by national authoritative publisher, it is believed that the choice of the content, especially the designed vocabulary is reasonable and scientific. In this part, the researchers chose “Go for it, Grade 8, Volume One” as an example and found that the choice of some phrases had its limitations.

The Words in a corpus can be arranged in order of their frequency in that corpus. “This is most interesting when corpora are compared in terms of their frequency lists” [10]. For example, the sentence “Everyone should play a part in saving the earth” (Go for it, Grade 8, Volume One, Unit 7, p.50) was extracted from a role-play conversation. In this unit, the phrase “play a part in” is taught as a new phrase. The researcher searched “play a part in” in COCA. It shows the frequency of 269. However, if you search “take part in”, the frequency reaches 2013. It can be seen that “take part in” is more commonly used in American English as phrases.

For another example, it is shown in the textbook that two phrases “at the weekend” and “at school” (Go for It, Grade 8,

Volume One) are applied. The research results appear in COCA show the frequency of the following phrases:

TABLE II. THE FREQUENCY OF FOUR PHRASES IN COCA

Phrases	Frequency
At school	7131
In school	12594
At the weekend	104
On the weekend	702

As is shown in TABLE I, American people tend to use “in school” and “on the weekend” more than “at school” and “on the weekend” (Go for it, Grade 8, Volume One, 2013) used in the textbook.

To sum up, compared with the invented textbook, corpus interface offers the frequency of lexical items and shows a true picture of use of vocabularies of Native Speakers. It can help learners to master the natural forms of Native Speakers. In this respect, corpus data can be directly applied to coursebooks, especially in the area of vocabulary choice.

**B. Variation**

Conrad [10] argued that “reference books must cease to be ‘monolithic’, that is, must cease to treat English as a single entity”. Instead, corpora present lexical items in terms of a comparison of frequency between different registers. This, to great extent, plays a significant part in synonym differentiation.

*1) Synonym differentiation*

Synonym differentiation is an important part in English language teaching. It is undoubted that students are confronted with some synonyms when teacher deliver new words. However, as Zhang [14] claimed, “linguistic intuition or dictionary illustration is far from the semantic nuance (there are many synonyms in English, little and small, start, begin and commence, etc. which is apt to error ratio)”.

Take two synonyms “Though” and “Cross” as an example which are extracted from the textbook (Go for It, Grade 8, Volume One, 2013, Unit Two, p.13). According to the textbook (Go for It, Grade 8, Volume One, 2013, Unit Two, p.13), “these two synonyms only differ in their collocations, as “Though” usually collocates with park, forest, hall, etc., while “Cross” usually collocates with bridge, river, desert, etc.”.

However, corpus-based analysis will make it clear how Native Speakers use the two synonymous words not only in the perspective of collocations, but also in the aspects of register distribution, semantic preference and semantic prosody. Take register distribution as an example.

CONTEXT	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
THROUGH	500349	78017	131618	101315	90620	98779

Fig. 2. The Register Distribution of “Through”

CONTEXT	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
CROSS	29765	4622	6728	6938	6553	4924

Fig. 3. The Register Distribution of “Cross”

The two Figs. show the distribution of synonyms based on COCA mainly in the perspective of five registers: SPOKEN, FICTION, MAGAZINE, NEWSPAPER and ACADEMIC. Totally speaking, the proportion of “Through” used by NS is much larger than “Cross”. The majority of “Through” appear in fiction, while “Cross” is mostly used in magazine.

**C. Lexis and Grammar**

According to Sinclair [5], “grammar and vocabulary are mixed together”. However, in the traditional textbooks, grammar and vocabulary are separated. In the textbook Go for It, there is an independent section called “Grammar Focus”, which mainly deals with the required grammar instruction of a certain unit and is separated from speech activities and lexis delivery. This textbook also includes an individual chapter called “Vocabulary” at the end of books.

On the contrary, corpora make lexis together with their grammatical structure in form of co-occurrence. Learners are able to generalize their usages from the actual cases and combine lexis, grammar and context. As is concluded by Kennedy [12], “Occurrence frequency in corpus data can shoulder the job competently”. “The Collins Cobuild English Grammar (Sinclair et al 1990) was a pioneer in this respect” [10]. For instance, “describing imperative, it notes the use of this verb-form in sentences such as Consider, for example which focus the reader’s attention on a particular aspect or example of the topic being explained” [10]. In conclusion, corpora make a connection between meaning and pattern in which words are tied with grammar.

**V. DISCUSSION**

The combination of examples from the chosen textbook and their evidences in COCA have shown the differences as well as the problems of traditional textbooks. We have also found that corpus presents quantitative description of language based on the collected statistics by native speaker, including frequency, distribution, context, etc. Therefore, it offers some implications on using corpora to inform textbook design.

First, since corpus shows the frequency of words by native speakers, it is crucial to adopt words with high frequency. The final goal of English teaching is to cultivate learners’ communicative ability. Having a good command of native-like expressions can possibly make learners easier to blend in the authentic situations.

Second, provided with quantitative description of language, the grammar in textbooks can also be designed in descriptive languages instead of abstract rules. Meanwhile, corpus reflects the combination of lexis and grammar. It is not necessary to divide vocabulary teaching and grammar teaching separately. For instance, the section of “Grammar focus” (Go for it, Grade 8, Volume One, 2013) can be blended into other parts.

Third, the chosen textbook has shown some classroom activities, such as “Listen and Practice”, “Role-play the Conversation” etc. in each unit (Go for it, Grade 8, Volume One, 2013). However, these activities enjoy the same purpose as to make learners practice. With the aid of corpus, textbooks can designed with more activities which take places in

different registers. In addition, activities can be in various forms with the cooperation of textbooks and corpus in the classroom, e.g. operating synonym differentiation in the corpus interface.

#### VI. CONCLUSION

Corpus has not been completely applied to textbooks design in China. Consequently, the textbooks applied in China are far from satisfaction. English learners can hardly enjoy the authentic languages. However, "Foreign language instruction needs a new change, a focus change from grammatical feature to the communicative attributes, and realize the English communication with actual meanings" [6]. With authentic materials presented in different aspects, corpus equips students with the ability to make actual linguistic types. It is necessary to combine teaching with authentic language by apply corpus to compile our textbooks.

#### REFERENCES

- [1] Bloomfield, L. (2002). *Language*. Beijing: Foreign Language T&R Press.
- [2] Canagarajah, A. (1999). *Resisting Linguistic Imperialism in English Teaching*. Oxford: Oxford University Press.
- [3] Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- [4] Davis, A. (2003) *the Native Speaker: Myth and Reality*. Toronto and Sydney: Multilingual Matters LTD.
- [5] Li, X. (2012). *A Corpus-based Approach to College English Teaching and Research*, Master thesis, Zhejiang University: Hangzhou.
- [6] Liu, H.Y. (2014). 'The Application of Corpora in the Compilation of English Textbooks taking COCA as an example', in *International Conference on Education, Language, Art and Intercultural Communication*. Available at: [www.cnki.cn](http://www.cnki.cn).
- [7] Liu, Y. (2009). *College English Coursebooks and Cultivation of Intercultural Communicative Ability*, Master thesis, Foreign Language University of Shanghai: Shanghai.
- [8] Ministry of Education of People's Republic of China. (2011). *English Curriculum Standards*. Beijing: Beijing normal University Press.
- [9] Murray, D.E. and Christison, M. (2011). *What English Language Teachers Need to Know Volumn I*. New York and London: Routledge.
- [10] Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- [11] People's Education Press Ltd. (2013) *Go for it*. Beijing: People's Education Press.
- [12] Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman, p.164.
- [13] Sun, Q.B. (2009). 'Evaluation of College English Integrated Course', *Journal of Hefei University of Technology*, 23(5): 131-134.
- [14] Zhang H.Q. (2012). 'Corpus and the Research of English Textbooks Development', in *International Conference on Social Science and Education*. Available at: [www.cnki.cn](http://www.cnki.cn).