

ON TARGET LANGUAGE TYPICALITY
AND LEARNER DIVERGENCES —INCORPORATING
INSIGHTS FROM CORPUS BASED CONTRASTIVE
INTERLANGUAGE ANALYSIS (CIA) IN EFL
VOCABULARY TEACHING AND LEARNING

Yu Yanming
Zhejiang Normal University

Abstract

Compared with Contrastive Analysis (CA) and Error Analysis (EA) , the corpus based Contrastive Interlanguage Analysis (CIA) is a more sound way of analyzing learner language . CIA s focus on target language typicality and learner divergences should be a guide in English vocabulary teaching and the teaching of English as a whole . In view of the focus , three practical ways of vocabulary teaching and learning are suggested . 1) Making good use of corpus based materials ; 2) Incorporating learner divergences into production dictionary compiling ; 3) Applying Data driven Learning (DDL) in and out of classroom .

Key words

Contrastive Interlanguage Analysis (CIA) ; vocabulary teaching and learning ; corpus linguistics ; typicality ; divergences

1. Introduction

English vocabulary teaching to college students is one of the most difficult and ineffective areas of instruction in China . Both teachers and students are in urgent need of more effective and practical ways of dealing with vocabulary . Among various theories and approaches to which they can turn for help , corpus linguistics is one of the latest and most insightful . As pointed out by some scholars , corpus linguistics is mainly a methodology (e. g. Gu 1998) ; contributions from corpus linguistics to vocabulary teaching and learning are most noticeable from a methodological perspective . Its methodological advantages are embodied in the research method of Contrastive Interlanguage Analysis (CIA) . By comparing native speaker language and learner interlanguage , CIA can show not only learners misuse / errors , but also overuse and underuse / avoidance of particular features of a given word / phrase , thus shedding light on the development of learners interlanguage . This paper will take a detailed look at CIA and its implications for vocabulary teaching and learning .

2 . Contrastive Interlanguage Analysis (CIA)

2 .1 Frequency based typicality

Linguistic studies always require sources of evidence for the support of various theories . Such evidence traditionally comes from intuition or introspection , from experimentation or elicitation , or from limited samples of a language (e. g. Yang 2005) . While in the case of corpus based research , the evidence comes directly from authentic texts . Such texts are representative and natural , reflecting how people use a language in the real world . The importance of corpora in language studies in fact lies in the

empirical data that enable linguists to make observation and statements more objectively and comprehensively. And the computer has a role of tremendous importance in achieving this. While corpora enable linguists to obtain representative empirical data as compared to data obtained by intuition or introspection, computer and text analysis software guarantee all the advantages of large storage, incredible speed, accurate replicability and statistical reliability. These advantages make it possible to conduct powerful quantitative analysis in addition to qualitative analysis, a significant advance over traditional research which tended to be dominated by qualitative analysis (Li 1999). Software packages like Oxford Microconcord, Wordsmith with their Concord, Keyword, Wordlist and other tools can greatly aid in efficient use of corpora.

The most salient characteristic of corpus based studies is that they deal with not only what words, structures or uses are possible in a language, but also what is probable and typical, i.e. what is most likely or least likely to occur in language use, typicality in other words. This is achieved mainly by getting statistical distribution of linguistic items in the contexts of use, such as the word frequency lists, so that the probability can be shown clearly. Such typicality oriented corpus based research has given rise to one of the most outstanding discoveries in corpus linguistics: the native speaker of English has a native like competence not because he/she uses whatever is possible, but rather because he/she uses what is probable, i.e. what is frequent: frequent vocabulary, grammatical constructions, etc; and such probable language can also be acquired by learners of English as a foreign language (e.g. Sinclair 1991; Mukherjee 2000). What counts most here is the notion of frequency. Based on frequency analysis, the uneven distribution of many linguistic features can be clearly presented, and what occur most frequently constitute the probable and typical language. This can make a tremendous contribution to both theory development and practices in language teaching and learning. For English teaching and learning, frequency gives some practical hints on what vocabulary and grammar are the most important and thus on what should be tackled first. On these lines Kennedy observed "such (frequency) information is important for curriculum design and for giving teachers and learners part of the basis of information to help decide where pedagogical focus might be directed most profitably" (1998: 93). This holds true for vocabulary as well. Corpus analyses show that different words/phrases, different senses, different colligation patterns and even different collocates are distributed unevenly for a word/phrase. It is thus assumed that students should acquire the most frequent ones supplemented by those required by needs. In a word, the notion of frequency tells us what occur more frequently in real language, and frequency based typicality is convincing. The general implication is that the most frequent items should be the focus of language education. Considering that EFL students can only devote limited time and energy to English learning, it is neither possible nor practical for them to acquire all possible language. What they should acquire is the most probable language, the typical language in other words. As for vocabulary, the most typical lexical features covering senses, colligation, collocates and so on of the most typical vocabulary should be a focus in EFL vocabulary teaching and learning.

2.2 Contrastive Interlanguage Analysis (CIA)

Selinker (1972) in his famous paper *Interlanguage* points out that three psychologically relevant data are involved in SLA: native language, target language and interlanguage. Interlanguage refers to both the internal system that a learner has constructed at a single point in time and to the series of interconnected systems that characterized the learner's progress over time. It is characterized by both native and target language features, yet it is systematic and independent of both the learner's native language and target language. Learners, with the development of their interlanguage, move gradually towards the target language, but they generally do not reach the level of competence of native speakers. Instead, fossilization seems to be a nuisance bound to happen, which is the process by which non-target forms become fixed in interlanguage. Selinker's theory has led to large quantities of learner interlanguage studies since it first appeared. However, before learner corpora became available, learner interlanguage analyses, for lack of large quantities of representative learner data, are mainly qualitative rather than quantitative and not systematic enough (Li 1999). The establishment and development of learner corpora enable far more systematic and quantitative learner language analyses in addition to qualitative ones. One of the most important ways of doing such learner corpus based analyses is the Contrastive Interlanguage Analysis (CIA), which Granger (1998) holds, lies at the heart of all learner corpus based research. By

definition ,CIA means “comparing /contrasting what non native and native speakers of a language do in a comparable situation ” (Pery -Woodley 1990 143). By comparing non native speakers (NNS) data with native speakers (NS) data ,this approach can not only provide evidence of non native characteristics of interlanguage ,but also show systematically their divergences including misuse /errors ,overuse and underuse /avoidance of target language features . While the target language corpus can provide authentic and reliable data for such work as syllabus design and textbook compiling etc . ;to improve English language teaching and learning ,there is a need for reliable and representative learner data as well . For while the former tells one how English is typically used ,the latter tells one where learners typically make errors and other divergences ,in other words ,where their difficulty lies (Granger 1998) .

2.3 CIA in contrast with CA and EA

This section will review and comment on three major approaches to learner language : Contrastive Analysis (CA) , Error Analysis (EA) and Contrastive Interlanguage Analysis (CIA) . Based on a comparison of the different focuses of these approaches ,advantages of CIA can be clearly demonstrated .

2.3.1 Contrastive analysis (CA)

Contrastive analysis (CA) associates cross linguistic differences with difficulties students may face and claims that the difficulties of second /foreign language acquisition can be determined through contrastive analyses . The Contrastive Analysis Hypothesis (CAH) as formulated by Lado was based on the following assumption : “ Those elements that are similar to his native language will be simple for him , and those elements that are different will be difficult . The teacher who has made a comparison of the foreign language with the native language of the students will know better what the real learning problems are and can better provide for teaching them ” (1957 2) . On such a basis it is then assumed that many of the “ distortions ” that arise are traceable to differences in the languages involved . In its strongest form , the CAH claimed that all L2 errors could be predicted by identifying the differences between the learner s native language (NL) and the target language (TL) . Lee ,for example ,stated that “ the prime cause ,or even the sole cause ,of difficulty and error in foreign language learning is interference coming from the learner s native language ” (1968 182) . So the focus of teaching and learning is where TL differs from NL as represented by the dark area in Figure 1 with very little attention paid to learner s own language .

Figure 1 . Focus in CA

As more and more research has demonstrated that many errors are not caused by NL transfer and that many errors predicted by contrastive analysis did not actually occur ,a weaker form of the hypothesis was put forward that admits that only some errors are traceable to NL transfer and contrastive analysis could be used only a posteriori (Ellis 1994) to explain rather than predict . In other words ,it needed to be used hand in hand with error analysis ,which arose later .

2.3.2 Error analysis (EA)

Error analysis (EA) regards NL transfer as one major source of error ;its attention is also directed towards other sources ,these factors being transfer of training ,strategies of second language learning ,strategies of language communication and overgeneralization of target language material (Selinker 1972) . The four plus NL transfer (L1 transfer in Selinker s words) are what Selinker identifies as the five principle cognitive processes responsible for the development of interlanguage (IL) . Figure 2 ,adapted from Yang Caimei (2002 74) ,shows this . Where IL and TL overlap is the correct part of learners interlanguage . Where NL ,TL and IL overlap (the dark area) shows similarities between native language and target language and may be regarded as where positive native language transfer occurs . The horizontal line area and the square area are where learner errors lie . The former represents errors caused by negative native language transfer and the latter represents errors caused by the other four factors

mentioned above . EA tries to gain insights into the process of learner interlanguage development by discovering and describing these errors rather than by predicting possible learner errors . And the aim is to move the learner's interlanguage towards the target language through appropriate treatment of errors , reflected by the reducing size of error areas and the increasing size of the vertical line area . Thus it is concluded that errors as covered in the horizontal line area and the square area are the focus of research in EA (Figure 3) .

Figure 2 . A model of EA

Figure 3 . Focus in EA

2.3.3 Contrastive Interlanguage Analysis (CIA)

As modern corpus linguistics calls for attention to be paid to the notion of typicality , attempts are made to take both native language and typicality of target language into consideration . Figure 4 and Figure 5 below are adapted from Pan Fan and Feng Yuejin (2000 §4) :

Figure 4 . Typicality in TL

Figure 5 . Focus in TL Education

The focus of language education is where TL typicality is not covered in NL as shown in the dark area in Figure 5 . Thus it reduces the daunting task of learning a new language considerably .

In our view , a more effective approach would be one that synthesizes these ideas , taking native language , various sources of divergences (not only errors) and target language typicality into consideration . This is embodied in the Contrastive Analysis Approach (Figure 6) . We expand the concept of "error" to "divergences" , which includes not only errors but also overuse , and underuse / avoidance of particular features . Thus the horizontal line area and the square area are where learner divergences lie . The former represents divergences caused by negative native language transfer and the latter represents divergences caused by other factors . The inner circle of TL represents typical target language (typicality) , and the dark area represents the yet to be acquired typical target language . Typicality of TL is the main goal of our English teaching and learning and the part that has not been covered by interlanguage (dark area) calls for more attention and effort . And the effectiveness of teaching and learning is ensured by studying the learner divergences (horizontal line area and square area) that are caused by various factors including NL transfer . Thus the process of language teaching and learning is a process of enhancing students' development of interlanguage (IL) to make it approach the target language (TL) in a more efficient way , manifested by the progressively smaller size of the divergences areas and the shrinking size of the dark area . Both typicality and divergences can be best obtained from corpus with typical native use from a native corpus and typical learner divergences from comparisons of a learner corpus with a native corpus . With its focus on target language typicality and learner divergences , CIA incorporates both these two corpora and can offer much more insightful help in this respect .

Figure 6 . A model of CIA

EA has advantages over CA in that EA takes into consideration learner language and is based on comparisons of target language and learner interlanguage . Yet ,CIA has enormous advantages over EA and can ,to a great extent ,overcome limitations inherent in EA . First ,the data of learner language in CIA are larger , more representative , more improved , and more importantly , computerized and quantitative . Second ,a most frequently mentioned limitation of EA is that EA ,focusing exclusively on what learners do ,fails to provide a complete picture of learner language ,as observed by Larsen Freeman and Long “ EA researchers focused on decontextualized errors and disregarded the rest of the learners performances . As a result ,they were denied access to the whole picture ” (1991 §1) . To know our students knowledge of English at any single point of time or their development within a period of time , we need to know what learners do correctly as well as what they do incorrectly . In contrast CIA with its systematic and comprehensive quantitative functions can overcome this limitation of EA . Granger contented that CIA presents “larger ,representative ,computerized ,quantitative frequencies of words / structures to study not errors ,but what differ in quantity :overuse ,underuse ,avoidance ,which are never addressed in the era of EA ” (1998 7) . We use the term “divergences ” to cover all these . By comparing the frequencies of words and structures in learner and native corpora ,we can have access to learners divergences in a systematic and exhaustive way .

Based on the above discussion we thus conclude that with its focus on target language typicality and learner divergences ,corpus based CIA is an extremely important approach to learner interlanguage .

3 . Implications of CIA in EFL vocabulary teaching and learning

Since its first appearance ,CIA has attracted ever increasing attention . Being one of the most effective approaches to learner interlanguage ,it has been adopted by both scholars and teachers in their research and teaching . Most of these studies and also the most outstanding discoveries made are on lexis and lexical grammar . Abroad ,one of the leading figures in this field is Sylviane Granger . Granger began to compile the International Corpus of Learner English (ICLE) in the 1990s ,which is made up of compositions by advanced English learners from more than a dozen countries . She has since then conducted a large number of CIA studies between ICLE and native corpora like the Louvain Corpus of Native English Essays (LOCNESS) . These studies have produced encouraging results . Examples of these studies include “Learner English on computer ” edited by Granger in 1998 , “The grammatical and lexical patterning of MAKE in native and non-native student writing ” by Altenberg & Granger in 2001 , “Computer learner corpora §second language acquisition and foreign language teaching ” by Granger , Hung & Petch-Tyson in 2002 , “Computer Learner Corpora and Monolingual Learners Dictionaries :the Perfect Match ” by Cock and Granger in 2005 , “Learner corpora :the missing link in EAP pedagogy ” by Gilquin ,Granger & Paquot in 2007 , “Learner Corpora in Foreign Language Education ” by Granger in 2008 ,etc . These studies and findings have contributed much to the development of CIA studies in China . In China ,CIA studies have become more accessible and more rewarding with the establishment of the Chinese Learner English Corpus (CLEC) ,which is a joint project by Guangdong University of Foreign Studies ,Shanghai Jiao Tong University and some other universities in China under the leadership of Professor Gui Shichun and Professor Yang Huizhong . It is made up of five subcorpora :middle school students (ST2) ,students for band four examinations (ST3) ,students for band six examinations (ST4) , junior English majors (ST5) and senior English majors (ST6) ,totaling one million words . Its large size , wide coverage of topics and representative subjects make it an authoritative corpus for studies on Chinese

EFL interlanguage. Moreover, learner errors in CLEC have been classified and tagged, which has proved to be of great help and convenience to CLEC based studies. Yet, there is still room for improvement. For example, other divergences in vocabulary including overuse, underuse/avoidance can also be taken into consideration. This is a huge project but it will be of tremendous value for such studies. A large number of CLEC based CIA studies have been published ever since its establishment, mainly on vocabulary and grammar. As for vocabulary, there are a number of CIA studies on individual words/phrases (e.g. Wei 2002; Zhang 2002; Cai 2003; Li 2003; Yu 2004; Zhang 2004; Lu 2005; Wang & Wang 2005; Huang 2007; Luo & Zhou 2007; Shi 2007; Wang & Chen 2007; Zhang 2007; Lei 2008). These studies have yielded quite insightful findings as they showed distinctive Chinese EFL learner interlanguage lexical features as they differ from the corresponding target norms, probed into the factors accounting for the differences and also put forward some helpful pedagogical implications. However, such studies are still far from systematic and exhaustive. In our opinion, it is necessary to make systematic and exhaustive CIA analyses of typical English vocabulary so as to gain a better understanding of how Chinese EFL learners acquire vocabulary and of the factors underlying the whole process and thus to offer more insightful help. Since our focus here is on the implications of CIA in vocabulary teaching and learning, we will concentrate on some very practical ways of incorporating CIA in vocabulary teaching and learning for Chinese EFL learners.

The focus on target language typicality and learner divergences of CIA can shed light on many aspects in EFL vocabulary pedagogy: syllabus design, teaching and learning material compilation including textbooks, reference books and dictionaries, practical classroom teaching and learning, testing, and so on. It is proposed that all these should 1) provide the typical use of typical word/phrase to students; 2) incorporate learner divergences when necessary. We will here suggest three ways, which can be of some practical help.

3.1 Making good use of corpus based materials

What corpus linguistics tells us is that a dictionary should show people not only what real language is but also what the usual way of using a language is like, in other words, authenticity and typicality. This is manifested in many ways: the selection of words, of senses of a particular word, and of colligations of a particular sense; the way the senses and colligations are arranged; the way definitions and examples are presented and other relevant aspects of dictionaries. Though the importance of real and typical language has long been realized in the pre-electronic era, corpora before the advent of computer were a bit far from typical and representative English language as a whole. Today, with the development of computer technology this goal has been achieved to a great extent. Computers and software packages enable the storage of large quantities of data, quick retrieval and sophisticated statistical analysis. Based on careful corpus analysis, the most typical features of any given word/phrase can be identified and incorporated into dictionaries. Such corpus based dictionaries are powerful tools for all language learners, especially for foreign language learners who, living in an environment of limited exposure to the target language, are unlikely to get a complete picture of the use of a word/phrase, or sometimes even the most typical use of common words/phrases.

There are a number of very representative corpora for dictionary compilation in this respect. The most well-known are the Bank of English, Longman Corpus Network and British National Corpus. All of them are large and representative enough. The Bank of English has led to the publishing of the Collins COBUILD dictionary series including Collins COBUILD English Dictionary (CCED). The latter two corpora gave birth to the famous Longman Dictionary of Contemporary English. All the dictionaries based on these two corpora are of great help to EFL teachers and learners. Most of these dictionaries have been introduced into China and some have even been adapted into bilingual dictionaries with Chinese illustrations added to suit Chinese EFL learners. We would like to have a close look at CCED in the Collins COBUILD series to see what advantages it has and how it can help in vocabulary teaching and learning in general.

First, vast representative data. The representativeness of data is well illustrated by Sinclair himself as he declared "For the first time, a dictionary has been compiled by the thorough examination of a representative group of English texts, spoken or written, running to many millions of words" and that

“this dictionary is based on hard measurable evidence” (Sinclair et al., 1987 :XV). Based on such representativeness, the data were carefully analyzed to provide the most accurate information. The central concept involved in the analysis is frequency. Not only frequencies of words themselves, but also relative frequencies of senses and colligations are provided, and even the frequencies of relatively common collocates of words are taken into consideration. Based on the frequency file, statistical analyses were conducted to decide which structures to include in the dictionaries. Thus all senses listed in the dictionaries are the most frequent ones in the Bank of English and are arranged according to their relative frequencies: the most frequent one appearing first, the second most frequent, next, and so forth. The same is true with colligations of any particular sense. This is very important for EFL learners for they can then know the relative importance of different senses and colligations and decide on the ones suited to their needs. For teachers, this provides some criteria for the selection and sequence of the words, senses and colligations in teaching.

Second, authentic, typical examples. All examples in CCED are carefully selected to show how the word is actually used by native speakers of English. They are not invented; rather, they are “natural”. What is more, examples also show the typical grammatical patterns, typical collocates and typical contexts of the usage as well as other information. The elusive phenomenon of semantic prosody is usually best captured in examples. Take the verb *cause* for example. Examples are “Attempts to limit the family size among some minorities are likely to cause problems” and “The damage to Romanian democracy caused by events of the past few days”. In these two examples collocates “problem” and “damage” show the strong negative semantic prosody of *cause*. Such authentic and typical examples can help learners to extract such subtle yet very important information.

Third, full sentence definition. Another important feature is the definition, which is one of the most striking things about CCED and other dictionaries in the series. Being full sentences rather than phrases, the definitions tell us not only what a word means, but also how to use it—in what phrase, in what grammatical structures and in what context. For example, looking at the definition of Sense 1 of *adapt*—“if you adapt to a new situation or adapt yourself to it, you change your ideas or behavior in order to deal with it successfully”, we can know it is usually used in the colligation of “V to n” and “V pron refl to n”. And semantic prosody can also be obtained from the definition. For example, the definition of *cause*—“to cause something, usually something bad, means to make it happen”, clearly shows its negative semantic prosody by highlighting the word “bad”.

3.2 Incorporating learner divergences into production dictionary compiling

The notion of typicality should be captured well by textbooks and major reference books. However, it is not easy work for textbooks to cover all the typical uses, for texts are highly limited in number and content. Thus it is strongly suggested that major reference books provide more comprehensive and useful information. To cater to learner needs, especially production needs, we even advise incorporating typical learner divergences as explicit information in the reference books. Of course, one of the most important reference books is the dictionary.

The idea can be found as early as 1936, when T. J. Fitikides published his collections of common errors under the title *Common Mistakes in English*. And *Les Faux Amis des Vocabulaires Anglais et Americain* by Maxime Koessler and J. Derocquigny was published in France in 1975. Then explicit error information in a dictionary was first seen in the form of usage notes, focusing on cases of particularly notorious student errors. By the mid 1980s various researchers pointed out that it was important for ELT dictionary publishers to analyze learners' errors (e.g. Maingay & Rundell 1985). With the development of corpus linguistics, the collection and analysis of learner errors are more representative and objective than ever. In 1987, Longman began to build a learners' corpus—the Longman Learners Corpus (LLC) to help in compiling ELT dictionaries. Thus we see the Longman Language Activator in 1993, which aimed at improving intermediate level students' production ability by presenting typical learner errors in “help boxes”. Other major publishers of dictionaries such as Oxford, Cambridge, and Collins also contain error warnings in various forms. Yet, the fact is, though overseas learner dictionaries have error warnings to help users avoid some errors, they fail to point out the typical errors committed by Chinese learners. These dictionaries are intended for foreign language students in general and not specifically for Chinese

learners . With respect to this particular shortcoming , Yuan Kele observed that “they fail to emphasize points where Chinese students typically make mistakes ”(1994 :281).

In China , much research has been conducted since the recognition of the importance of interlanguage ,and many Chinese scholars and teachers have been working on collecting common Chinese learner errors in specialized dictionaries (Gao & Wang 1986 ;Liang 1987 ;Liu 1987 ;Yu & Wang 1991 ;Ge 1993 ;Zhao 2002). However ,learner dictionaries on the whole haven't paid attention to error warning . Zhao Xueqin (2000) pointed out that current bilingual dictionaries have given few warnings against common errors by Chinese students . Rather ,these dictionaries focus merely on trying to provide learners with equivalents ,near synonym discrimination ,examples ,collocations ,etc . Only a few scholars have directed their attention to error warnings in dictionaries (e. g. Yuan 1994 ;Zhao 2000). All this points to the need to compile a learner's dictionary especially designed for Chinese EFL learners .

Our suggestion here is such a learner's dictionary should not only include typical errors ,but also other typical divergences found in learner use including avoidance ,underuse and overuse in its Warning Columns . Since it would be intended especially to improve students' production ,it would be a production dictionary . Such a dictionary would cover typical learner divergences in colligation ,collocation ,semantic prosody ,sense preference and divergences in other aspects of language use that have been discovered . Take adapt for example again . A CIA analysis of adapt based on LOB¹ and CLEC shows that Chinese learners over used sense 1² and its colligation pattern “ V to n ” ,misused “ V n ” and avoided sense 2 and its two colligation patterns “ V n to n ” and “ V n ” (Yu 2004). Based on such an analysis ,we illustrate below a possible entry for adapt with the italicized words providing information on typical learner divergences .

Adapt :

1 . Become suitable (overuse)

The world will be different ,and we will have to be prepared to adapt to the change .³ (V to n : overuse)

They have been tightening their belts for months ,adapting themselves to a war economy . (V pron refl to n)

Do not use V n in this sense .

2 . Make suitable (underuse)

Shelves were built to adapt the library for use as an office . (V n :underuse)

Their life has been well adapted to the climate there . (V n to n :underuse)

3 . Alter or modify for the use of a film or television

The film has been adapted from a play of the same title . (V n)

Such a dictionary would be of practical use for students as well as for teachers . We suggest the idea be incorporated into other EFL materials ,the efficiency of which ,we believe ,could be “improved if materials designers had access not only to authentic native data but also to authentic learner data ,with the Ns data giving information about what is typical in English ,and the NNs data highlighting what is difficult for learners ...” (Granger 1998 :7).

3.3 Applying DDL in and out of classroom

We recommend here a corpus based method of form focused instruction — data driven learning (DDL) ,which ,according to Granger and Tribble ,is “a particular attractive way ”(1998 :199) to raise accuracy . It can make use of both native speaker corpus and learner corpus . And the focus on typicality and divergences can be well captured by adopting the method .

DDL is defined by Johns and Ring as “the use in classroom of computer generated concordances to get students to explore regularities of patterning in the target language ,and the development of activities and exercises based on concordance output ” (1991 :111). A concordance is a list of occurrences of a particular word ,a part of a word or a combination of words in its contexts drawn from the corpus . A concordance is very powerful in providing information on collocation ,colligation and semantic prosody . By looking at both sides of the keyword ,we will know quite easily how common particular lexical and

grammatical features are associated with it. In addition extended context is available when the original context is not enough to perceive the patterning. By actively engaging students in searching and analyzing concordances of a particular word / phrase, DDL calls for learners' inductive skills, and "favors learning by discovery" and "takes on character of research, rather than spoon feeding or rote learning" (Tribble & Johns 1990: 12). This new method presents not only a shift in the focus of teaching towards form, but also an introduction of a more inductive approach to the project of learning and teaching languages than has been common in the usual way.

DDL is gaining popularity in language teaching and learning in western countries. However, in China, even though DDL has been frequently discussed (e.g. Li & Pu 2001; Wang 2004; Zhen 2005; Liang 2005; Huang & Sun 2006; Jiang & Yang 2006; Liu 2007), it is rarely practiced with the exception of Lu Yanhui (2007) who conducted DDL in vocabulary teaching to college non-English majors and obtained positive results. Moreover, DDL as discussed by almost all these scholars / teachers made exclusive use of native English data. The advantage of using native English data is the authenticity of the data that ensures that learners are presented with samples of a language which reflect the way people actually speak or write. However, it is our contention that DDL can be more effectively used when the findings from the learner corpora are taken into consideration. Incorporating CIA into DDL can offer both teachers and learners greater help in identifying typical target language use and typical learner divergences. Teachers can accordingly design exercises to draw students' attention towards the misused, underused / avoided features, especially those concerned with typical target language features. For example, as CIA analysis shows, students frequently / usually avoid sense 2 of adapt and its corresponding colligation patterns "V n" and "V n" as mentioned in section 3.2. Guided by this finding, teachers can design the following task to foster students' command of this feature.

Example task 1⁴

The verb "adapt" has a sense meaning changing something to make it suitable for a new purpose or situation, e.g. shelves were built to adapt the library for use as an office. Identify those items which contain an example of this use of "adapt".

The workers have been helped to adapt their old designs.
 For gradually the humans adapted better to the world.
 And it frequently happens that they adapt their act to each venue.
 It is not easy even for those people physically adapted to it.
 An Anglican missionary adapted a system of syllabics to the Eskimo tongue.
 It is not adapted to the teaching of inherently progressive subjects.
 It has, in fact, readily adapted itself to modern tastes and the fine illustrations.
 He is most successful in adapting folk dancing for the chorus of peasants.
 Does the evidence obtained support a rationale for adapting a pre-retirement planning to the needs of the employees?

Native corpus concordances can also be presented to learners in parallel with learner concordances to make learners themselves discover grammatical or lexical features or features of other aspects that distinguish their interlanguage from the target norm. Consider once again the item adapt. Since CIA analysis shows that students often diverge from native speakers by using adapt in the colligation pattern "V n" in place of "V to n" or "V Pron refl to n" in sense 1 as shown in section 3.2, teachers could design tasks like the following one.

Example task 2

Consider the following samples of native and non-native speaker writing.

- 1) What colligation patterns can be found of "adapt"?
- 2) Do any colligation patterns only appear in the non-native speaker examples?

Native speaker writing

For gradually the humans adapted better to the world .
 It has readily adapted itself to modern tastes .
 It is not easy for those people who have adapted themselves to it .
 It is hard for the modern schools to adapt themselves to the new situation .
 The temperament made it impossible for him to adapt himself to the different life there .

Non native speaker writing

We will find it difficult to adapt the society .
 So we can adapt ourselves to the changing world .
 To adapt the new situations is necessary .
 If they want to adapt themselves to the society .
 We should improve our skills and knowledge to adapt the new life .
 They think it easy to adapt the changing surroundings .

Helpful as it is , DDL , being a new methodology , is not a replacement for traditional ways of vocabulary teaching and learning . As many concordances need to be carefully selected and edited to help learners find the relevant features , it puts high demands on teachers . And if vast quantities of concordances are thrown at learners there is a considerable risk that DDL activities can become too time-consuming and frustrating for learners . Another concern is the question of learning style , which presents a problem for almost any methodology . So it is suggested here that concordance based exercises should best be viewed as an additional way , complementary to traditional ways of vocabulary teaching and learning . In order to make DDL more effective , we put forward the following two suggestions . First , teachers should enhance their roles as a guide and a monitor . As students go through the whole process from edited data to raw data , difficulties and confusion will inevitably appear . Teachers should constantly monitor the whole process so as to be able to offer timely guidance and encouragement to students . Second , we suggest designing a DDL vocabulary file packet for Chinese EFL learners . This packet could include software like Wordsmith , native corpora , learner corpora and a large number of exercises / tasks on typical English vocabulary with typical learner divergences highlighted as exemplified above . Learners can thus , under the guidance of their teachers , practice learning and testing at their own convenience and even at their own speed . As a result , learner autonomy may be fostered , which may in turn contribute considerably to English vocabulary teaching and learning .

4 . Conclusion

During the past three decades , the mainstream in language education has been the process of teaching and learning through communication . While the widely adopted communicative approach has undeniably helped improve learners fluency , it has , as increasingly acknowledged by many of its advocates , been accompanied by a decrease in accuracy . “In focusing exclusively on meaning and overall success of communication , we have overlooked the issue of accuracy ” (Williams 1995 13) . As a result , more and more EFL specialists call for a renewal of form focused teaching and learning in the communicative framework . This paper echoes this call . We argue that such a call can be well answered by incorporating insights from corpus linguistics and CIA . With its focus on target language typicality and learner divergences , CIA may help learners acquire the most important and typical English language in a more efficient way . We can , to be practical , reduce the scope of application to words in the syllabus , especially those to be acquired as productive words . And we suggest utilizing everything the machine and the corpus are able to offer , whether typicality or divergences .

Notes

- 1 . Lancaster -Oslo /Bergen (LOB) is a one -million -word written BE corpus , covering many registers . It is one of the most representative English corpora .
- 2 . For sense and colligation patterns in this study , we took reference to Collins COBUILD English Dictionary (2000) .
- 3 . Examples of “ Adapt ” here are taken from Cobuild .
- 4 . Sample concordances of learner interlanguage /non native speaker in the two example tasks here are taken from ST3 and ST4 in CLEC and those of native speaker from LOB .

References

- Altenberg ,B . &Granger ,S .2001 . The gram matical and lexical patterning of make in native and non native student writing . Applied Linguistics 22 :173 194 .
- De Cock ,S . & Granger ,S .2005 . Computer learner corpora and monolingual learners dictionaries :The perfect match .In W . Teubert &M . Mahlberg (eds .) . The Corpus Approach to Lexicography (Special issue of Lexicographica) 20 :72 86 .
- Ellis ,R .1994 . The Study of Second Language Acquisition . Oxford :Oxford University Press .
- Fitikides ,T .J .1936 . Common Mistakes in English . London :Longman .
- Gilquin G . ,Granger ,S . &Paquot ,M .2007 . Learner corpora :The missing link in EAP pedagogy .In P . Thompson (ed .) . Corpus based EAP Pedagogy (Special issue of Journal of English for Academic Purposes) 6 / 4 :319 335 .
- Granger ,S .1998 . The computer learner corpus :A versatile new source of data for SLA research .In S . Granger (ed .) . Learner English on Computer . New York :Longman . 3 18 .
- Granger ,S .2008 . Learner corpora in foreign language education . In N . Van Deusen Scholl & N .H . Hornberger (eds .) . Encyclopedia of Language and Education , Second and Foreign Language Education , Springer /4 :337 351 .
- Granger ,S . ,Hung ,J . &Petch -Tyson ,S .2002 . Computer Learner Corpora , Second Language Acquisition and Foreign Language Acquisition . Amsterdam :John Benjamins Publishing Company .
- Granger ,S . & Tribble ,C .1998 . Learner corpus data in the foreign language classroom :Form focused instruction and data driven learning .In S . Granger (ed .) . Learner English on Computer . New York : Longman . 119 131 .
- Johns ,T . &Ring ,P .(eds .) .1991 . Classroom Concordancing . ELRJournl (New Series) 4 .
- Kennedy ,G .1998 . An Introduction to Corpus Linguistics . Harlow :Addison Wesley .
- Koessler ,M . &Derocquigny ,J .1975 . Les Faux Amis des Vocabulaires Anglais et Americain . Paris :Vribert .
- Lado ,R .1957 . Linguistic Across Cultures :Applied Linguistics for Language Teachers . Ann Arbor :University of Michigan Press .
- Lee ,W .1968 . Thoughts on contrastive linguistics in the context of language teaching . In Alatis (ed .) . Reports of the Nineteenth Annual Round Table Meeting on Linguistics and Language Studies . Georgetown University , Washington D . C . :Georgetown University Press . 185 194 .
- Maingay ,S . &Rundell ,M .1987 . Anticipating learners errors implications for dictionary writers . In A . Cowie (ed .) . The Dictionary and the Language Learner — Papers from the EURALEX Seminar at the University of Leeds , 1 3 April 1985 . Max Niemeyer Verlag ,Tubingen . 128 135 .
- Mukherjee ,J .2000 . Disc :Review Green . Available online <http://www.linguistlist.org/Issues/A1/A1-2092.html>
- Pery -Woodley ,M .1990 . Contrasting discourses :Contrastive analysis and a discourse approach to writing . Language Teaching 23 /3 :143 51 .
- Selinker ,L .1972 . Interlanguage . International Review of Applied Linguistics 10 :209 231 .
- Sinclair ,J .1987 . Collins COBUILD English Language Dictionary . London and Glasgow :Collins Publishers .
- Sinclair ,J .1991 . Corpus , Concordance , Collocation . Oxford :Oxford University Press .
- Sinclair ,J . et al .1995 . The Collins COBUILD English Dictionary (2nd ed .) . London :Harper Collins .
- Tribble ,C . &Jones ,G .1990 . Concordances in the Classroom . London :Longman .
- Willams ,J .1995 . Focus on form in communicative language teaching :Research findings and the classroom teacher . TESOL Journal 4 /4 :12 16 .
- 蔡虹 ,2002 , 英语学生写作中的词类混用现象 《解放军外国语学院学报》第3 期。
- 高文志、王庆云 1986 《英语常见错误分析》。北京 :知识出版社。
- 葛传二等 1993 ,《简明英语惯用法》。上海 :上海译文出版社。
- 顾曰国 ,1998 , 语料库与语言研究 ,《现代语言学》第1 期。
- 黄瑞红 ,2007 , 中国英语学习者形容词增强语的语义韵研究 ,《外语教学》第4 期。
- 黄万武、孙炳文 ,2006 , 基于语料库的数据驱动学习模式创新 《武汉科技学院学报》第11 期。
- 蒋婷、杨梅 ,2006 , 数据驱动学习与专业英语写作创新。《重庆大学学报》第6 期。
- 雷春仪 ,2008 , 中国英语学习者“so”使用的失误特点分析 ,《牡丹江大学学报》第2 期。
- 李文中 ,1999 , 语料库、学习者语料库与外语教学 ,《外语界》第1 期。
- 李文中 ,2003 , 基于英语学习者语料库的主题词研究 ,《现代外语》第3 期。

- 李文中、濮建忠,2001,语料库在外语教学中的应用,《解放军外国语学院学报》第2期。
- 梁三云,2005,语料库与词汇教学策略的研究,《外语电化教学》第5期。
- 梁实秋,1987,《英文正误大辞典》。香港:华视文化事业股份有限公司。
- 刘红梅,2007,基于语料库的英语数据驱动学习,《安徽师范大学学报》第6期。
- 刘毅,1987,《英文正误辞典》。香港:学习出版公司。
- 卢鹿,2005,语料库对同义词研究的贡献,《中国英语教学》第6期。
- 鲁艳辉,2008,数据驱动学习在大学英语教学中的应用研究,《中国英语教学》第2期。
- 罗琴琴、周江林,2007,基于语料库的中国英语专业学生 HAVE 搭配行为调查研究,《外语教育》。
- 潘、冯跃进,2000,语料库语言学与外语教学思想的转变《外语学刊》第4期。
- 史逢阳,2007,基于 CLEC 语料库动/名搭配错误探究词汇学习,《天津外国语学院学报》第4期。
- 孙海燕,2004,基于语料库的学生英语形容词搭配语义特征研究,《现代外语》第4期。
- 王海华、陈国华,2007,中国学习者使用英语强势词搭配的发展特点,《外国语》第1期。
- 王海华、王同顺,2005, CAUSE 语义韵的对比研究,《现代外语》第3期。
- 王金铨,2004,数据驱动学习——英语教学模式的新动向,《扬州大学学报》第6期。
- 卫乃兴,2002,语义韵研究的一般方法,《外语教学与研究》第4期。
- 杨彩梅、宁春岩,2002,本族语、目标语和中介语三语相交的二语习得理论模型《湖南大学学报》第4期。
- 杨连瑞,2005,第二语言习得研究的学科定位和进展,《外语与外语教学》第5期。
- 于国治、王福银(主编),1999,《学生英语正误辞典》。沈阳:东北工学院出版社。
- 俞燕明,2004,中介语对比分析法及其对词汇教学的启示,《国外外语教学》第2期。
- 源可乐,1999,他山之石,可以攻玉——论英日教学型词典给我们的启示,《辞书研究》第4期。
- 张淑静,2002,中国英语专业学生 make 的使用特点调查报告,《解放军外国语学院学报》第4期。
- 张雪梅,2007,基于语料库的中国英语学习者连接副词研究,《中国英语教学》第2期。
- 赵雪琴,2000,学习者错误分析与双语词典编撰,《辞书研究》第4期。
- 赵振才,2002,《英语常见错误解答大词典》。哈尔滨:黑龙江人民出版社。
- 甄凤超,2005,语料库数据驱动的外语学习:思想、方法和技术,《外语界》第4期。

(...continued from p .41)

- Long , M . H . & Sato , C . 1983 . Classroom foreigner talk discourse : Forms and functions of teachers questions . In H . W . Selinger & M . H . Long (eds .) . Classroom Oriented Research in Language Learning . Rowley , Mass : Newbury House .
- Mackey , A . & Philp , J . 1998 . Conversational interaction and second language development : Recasts , responses , and red herrings . The Modern Language Journal 82 / 3 : 338 - 356 .
- Nunan , D . 1987 . Communicative language teaching : Making it work . ELT Journal 41 / 2 : 136 - 145 .
- Nunan , D . 1991 . Language Teaching Methodology : A Textbook for Teachers . New York : Prentice Hall .
- Richards , J . & Lockhart , C . 1996 . 《第二语言课堂的反思性教学》。北京:外语教学与研究出版社;北京:人民教育出版社;伦敦:剑桥大学出版社。
- White , L . & Lightbown , P . 1984 . Asking and answering in ESL classes . Canadian Modern Language Review 40 : 288 - 344 .
- Williams , M . & Burden , R . L . 2000 . 《语言教师心理学初探》。北京:外语教学与研究出版社;北京:人民教育出版社;伦敦:剑桥大学出版社。
- 何苗,2007,《大学英语教师课堂提问研究》(硕士学位论文)。长沙:湖南大学。
- 胡青球、埃德·尼可森、陈炜,2004,大学英语教师课堂提问模式调查分析,《外语界》第6期。
- 孔文、李清华,2007,关于 EFL 课堂中教师提问的对比研究,《国外外语教学》第3期。
- 李敏,2006,《大学英语教师课堂提问多维研究》(硕士学位论文)。桂林:广西师范大学。
- 许峰,2003,大学英语课堂提问的调查与分析,《国外外语教学》第3期。
- 郑佩芸,2003,课堂提问与焦虑控制,《外语界》第3期。
- 支景嫦,2007,《课堂提问与中职英语教学》(硕士学位论文)。济南:山东师范大学。