



A Corpus-Based Study on the Vocabulary Errors in CET-4 Writing and Its Pedagogical Implications

HUO Yanjuan^{[a],*}

^[a]School of Foreign Languages, China West Normal University, Nanchong, China.

*Corresponding author.

Received 15 April 2014; accepted 26 June 2014

Publish online 27 August 2014

Abstract

With the rapid development of computer science today, our teaching has been changed from traditional education with a single means or method to modern classroom teaching with scientific way of imparting knowledge. As the enormous corpora have been introduced, the investigation and analysis of common errors made by Chinese learners seems valuable both to teachers and researchers. CET-4 (College English Test 4), the comprehensive evaluation of target language proficiency of Chinese students, plays a crucial role whose passing rate exerting influence on college English teaching. Therefore, analyzing errors made by students in CET-4 can give a valuable and significant reference to EFL (English Foreign Language) teaching and learning in China.

Key words: CET-4; CLEC; Corpus; EFL

Huo, Y. J. (2014). A Corpus-Based Study on the Vocabulary Errors in CET-4 Writing and Its Pedagogical Implications. *Studies in Literature and Language*, 9(1), 38-42. Available from: <http://www.cscanada.net/index.php/sll/article/view/5238> DOI: <http://dx.doi.org/10.3968/5238>

INTRODUCTION

Writing, one of five fundamental language skills in English learning, is often regarded as an important measurement to examine the target language output. Various errors frequently occurring in compositions of students can remind us of the ineffectiveness of language

input, the poor acquirement of target language, and the inadequate authentic language context. For improving the quality of writing, teachers rack their brains, yet the overall situation of English writing is not so satisfactory. With respect to the most influential English tests in China, CET-4 which is administrated by the National College English Testing Committee together with the Examination Centre of the Ministry of Education is well recommended. CET-4 is a large-scale standardized test aiming to examine the university undergraduates who have finished the curriculum of the 4-level of English. The performance students have made in writing part of CET-4 can give a persuasive and effective feedback to the college English teaching and learning. The present study aims to investigate and analyze the errors students have made in CET-4 examination, to accurately measure college students' ability to use English in well-rounded topics and to explore the practical pedagogical implications for teachers. Thanks to the advanced computer technology and the emergence of enormous corpora, researchers and teachers can give the scientific, objective and convincing investigation on errors based on corpus.

1. THE BASIC KNOWLEDGE OF RELEVANT THEORIES

1.1 Error Analysis

In the 1950s and 1960s, Contrastive Analysis was on its prime time. The theory believed that researcher would be able to identify difficulty in language by comparing the native language and the target language. In the 1967, S.P. Corder, published his master work *The Significance of Learners' Errors* which is regarded as the birth of EA (Error Analysis). As we know, in the process of learning a foreign language, it is natural and unavoidable for students to make a variety of errors. Corder (1974) divided

errors into three categories according to systematicity of error: pre-systematic errors, systematic errors and post-systematic errors (Corder, 1974, pp.19-27). Ellis (1994) gave us his understanding of Corder's three categories. Pre-systematic errors occur when learner cannot give any account of why a particular form is chosen. Systematic errors occur when learner is unable to correct the errors but can explain the mistaken rule used. Post-systematic errors occur when learner can explain the target-language rule that is normally used (Ellis, 1994, p.56). Comparing with CA, EA shows its strong point of describing and analyzing learners' language systematically and becomes the first theory concerning about learners' language system. EA has become a vital part in applied linguistics for it attaches importance to observing the whole process of language development.

1.2 CLEC

In 2003, the largest learner corpus in mainland China, Chinese Learner English Corpus (CLEC) was set up by Shanghai Jiao Tong University and Guangdong Foreign Studies University. CLEC is acknowledged as one of the most representative English learner corpora which contains English compositions written by Chinese students widely ranging from middle schools to the non-English major students and English major students in universities. There are five groups of essay writings tagged from ST2 to ST6 with each group containing 200,000 words and the total words add up to more than 1.1 million words of English compositions. The present paper focuses on CET-4 students representing as ST3 in this corpus which shows the target language proficiency of non-English major freshmen and sophomores. The corpus is valuable both to researchers and teachers for the sampled data collected from the original compositions written by the variety of students covering various topics. By identifying and analyzing various word errors students made, teachers can make clear their teaching objects, find out the cause of errors, and obtain rich and informative resources in compiling textbook or English language tests. As far as EFL learning in China, construction of this particular corpus for learner's English

writing and meanwhile taking it as a source for learner's language study are of great significance both in innovative teaching and academic values.

2. THE CORPUS-BASED STUDY OF VOCABULARY ERRORS IN CET-4 WRITING AND ITS IMPLICATION

Since error can provide valuable information for language pedagogy and second language acquisition, error studies constitute a far-reaching significant part of EFL teaching in China and have become a useful measure. Sponsored by the Ninth Five-year Plan of the Chinese National Foundation of Social Sciences and Humanities, CLEC represents different levels of Chinese English foreign language learners within which English writings of Chinese learners are labeled as ST2, ST3, ST4, ST5, and ST6. The present study will focus on ST3 which in CLEC refers to compositions written by non-English majors in CET4 (Gui, Yang, & Yang, 2005, pp.1-3).

2.1 Formal Errors

Since words constitute the fundamental part of language, lexical study can offer an inspiration into the development of the English vocabulary. Corpus-based lexical study, with the availability of electronic, machine-readable corpora, can save a lot of time and labor. According to Gui Shichun and Yang Huizhong, both of them are prestigious scholars in China, CLEC clarified language errors into 11 categories, they are form, verb phrase, noun phrase, pronoun, adjective phrase, adverb, prep phrase, conjunction, word, collocation and sentence. Among these errors, form errors include spelling, word building and capitalization (Gui & Yang, 2003, pp.7-8). In order to process data, my finder (the program of CLEC searches) will be employed to pick up the given code wherever it is spotted in corpus files and to display the results. The present research will firstly focus on form errors which are taken as the basic part in vocabulary. The following Table 1 can offer some information of form error in CET4 based on CLEC..

Table 1
Form Errors in CET-4

Label	Tokens	Type of vocabulary error	Explanation of errors
fm1	3349	spelling	spelling, coinage, abbreviation, etc.
fm2	522	word building	inflection, derivation, plurality (noun), compounding, irregularity(verb), etc
fm3	851	capitalization	lower initial letter for upper initial letter or upper initial for lower initial letter

Searching from the five sub-corpora of CLEC, the tags related to spelling errors were extracted by running the

finder program. Comparing the distribution of these three types of form error, we can get the following Table 1.

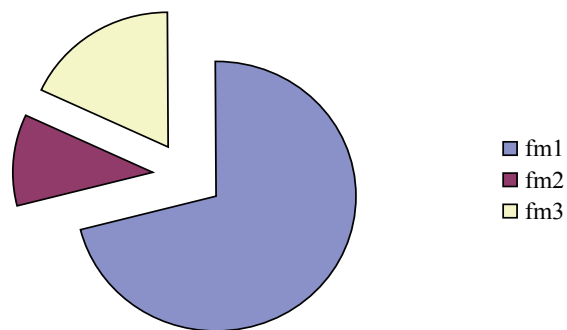


Figure 1
Distribution of Form Errors by Pie Chart

It is obvious that fm1 is the most frequent error students have made. Therefore, let us give more observation and analysis to fm1. The following examples are extracted from CET-4.

Qingdao Daily ect [fm1,-]. I am going to tiried [fm1,-] to be a suitabl Qingdao Daily ect [fm1,-]. I am going to tiried [fm1,-] to be a suitable man in the future job. we can know of the society. The most convinient [fm1,-] is the public media, we can read the world, I think I should form the good habit [fm1,-] of reading newspapers and watching programs of It is necessary for our colledge [fm1,-] students to get to know the world outside they knowing the society we also can widen our rang [fm1,-] of the knowledge. But how to knowate the experience of the society. As a colledge [fm1,-] student, I should go out from the campus, and we should know something about it. The approaches [fm1,-] to know about society are] . For example, we can read newspapers, megines [fm1,-] and watch TV to learn information ety . Also we can practice in society by ourself [fm1,-] , we can participate in free service activitie have learned. Before we go into [wd3,1-] socity [fm1,-] , we must have [vp6,-s] known good use of our knowleges [fm2,-] in the socity [fm1,-] . We must know what kinds of man wd2,1-] if [wd4,-s] have mastered the technologe [fm1,-] that the socity [fm1,-] require

The above examples all come from the sample compositions in which we can easily find the square brackets and their enclosed letters and numbers [fm1] which is the error tag representing the error of spelling. The writer roughly divides spelling errors made by students in CET-4 into three types:

(a) students always miss a letter in a word (the letter in the brackets is omitted), such as knowle(d)ge , soci(e)ty, activit(i)es, laugh(t)er, stren(g)then, fina(n)cial, ar(r)ange.

(b) add an additional letter (the letter in the brackets is added), such as colle(d)ge , ha(b)bit, knowle(a)dge, tir(i)ed

(c) select the wrong letter to replace the right one (the word in the brackets is the right form), such as *technologe* (technology), fund (fond), perpose (purpose)

As is known to all, college students in China have studied English for many years, yet they still repeatedly make many errors in spelling which can be easily identified in CET-4 examinations. By the observation and analysis of examples, the writer summarizes three possible reasons students made so many spelling errors in their writings. Firstly, from phonetic perspective, the phone does not always correspond with the morpheme. Some phones are silent such as: *h* in hour. It is more challenging to students that one phone will relate to more than one morpheme. The phone /k/ can be represented as the morpheme c, ck, k. Secondly, deep rooted by their mother tongue, students are easily interfered by Chinese which phenomenon is called negative transfer. Influenced by the Chinese pronunciation system, students tend to pronounce each letter in a word, while English phonetic system is complicated and the pronunciation of English words are illogical and inconsistent. Thirdly, it is related to the attitude of non-English majors in learning English. Some students study English only with the aim of passing the examinations so they are lack of motivation and interesting in learning. Students are always complaining that they learn English just because teachers or parents require them to learn. Students may ignore the spelling because sometimes they spare more effort in practicing pronunciation and remembering Chinese meaning corresponding to the English word.

2.2 Collocation Errors

Firth, the father of collocation, claimed that the collocation of a word or ‘piece’ is not to be regarded as mere juxtapositions but an order of mutual expectancy. Collocations are actual words in habitual company (1957). CLEC attaches its importance to the lexical collocation which deals with the word choice and does not include preposition, infinitive or clause. In CLEC, collocation errors are tagged with the code cc1, cc2, cc3, cc4, cc5, cc6. The brief introduction of these six types of collocation errors can be found in the following Table 2.

Table 2
Six Types of Lexical Collocations

Type	Collocation pattern	Explanation	Examples
cc1	noun1 + noun2	Improper noun(phrase) and noun(phrase) combination/semantic	Today morning
cc2	noun + verb	Improper noun(phrase) and verb(phrase) combination/semantic	My heart sang
cc3	verb + noun	Improper verb and noun(phrase) combination/semantic	Eat Chinese tea take jokes
cc4	adjective + noun	Improper adjective and noun(phrase) combination/semantic	Wide sky senseless day
cc5	verb + adverb	Improper verb and adverb (or ad/v) combination/semantic	Know me deeply
cc6	adverb + adjective	Improper adverb and adjective combination/semantic	Strictly correct

After careful observation of these six types of lexical collocations in CET-4, we can find that the distribution of them is not distributed evenly. The following Table 3 can

present the distribution of each category of collocation errors in ST3.

Table 3
Six Types of Lexical Collocation Error in ST3

Type	cc1	cc2	cc3	cc4	cc5	cc6	Total
Concordance hits	88	207	599	114	47	14	1069
Percent	8%	19%	56%	11%	4%	2%	100%

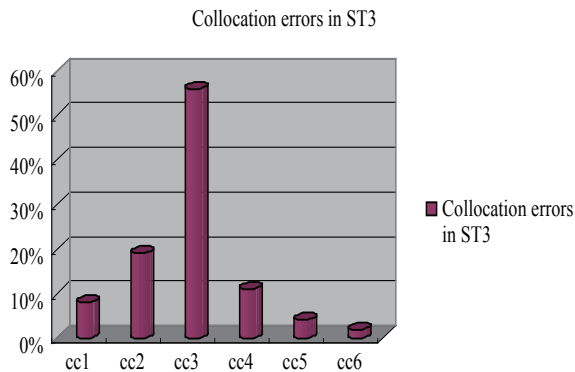


Figure 2
Collocation Error in ST3

From the above Table 3 and Figure 2, we can infer that among six types of lexical collocations errors in ST3, Chinese learners make more errors in the area of verb-noun collocation (cc3) in their CET-4 examination writings. The writer uses cc3 as the search word, and subsequently hundreds of instances of verb-noun collocation can be found in ST3.

er when we practice usually. When we play sports [cc3,1-] , for example, play basket-ball] don't be afraid of failure, just make practice [cc3,1-] . And one day you'll find that you are very people know how to live and take activities [cc3,1-] , how to prolong their lives by [wd3,-2] scie be 200 infans [fm1,-] [wd4,s-] went to death [cc3,2-] every 1000 infans [fm1,-] , while in 1990 the [sn1,s-] With those [fm1,-] protect problems [cc3,-1] the health of the people surely will bout people's living conditions. The states does [cc3,-2] their infrustructure [fm1,-] continually omes 60 years old in 1990. It increased [vp6,s-] [cc3,1-3] 20 years old during thirty years. On s this change cause? [sn8,s-] I think it lies in [cc3,1-2] three reasons. First, economy has been e the [wd4,2-3] republic of China, the people used [cc3,-4] their minds and bodies to get rid of the socialist country. We also builds [vp3,2-] hope [cc3,1-] for the future and do our best to make our are also created. People can relax their spirit [cc3,2-] in great joyment [fm2,-] . This improved ncy and the infant mortality above, we concluded [cc3,-2] the reasons of rising [vp1,-2] the] it is a [np7,-1] error. Play [vp4,-s] computer [cc3,1-] is necessary for us to practice it [pr1,s-] .rtaility [fm1,-] also took on a new variation [cc3,4-] since 1960. In 1960, it was ople [fm3,1-1] have plenty of money to take food [cc3,1-] and see doctors. Secondly, in 1960's most of people know an idea [cc3,2-] that [sn6,s-] is [wd5,-s] "Practice Makes we mastered the skill of driving [cc3,-2] the bikes. The words "Practice Makes [fm3,-1]

The above materials retrieved from ST3 are taken directly from CET4 compositions, which are limited in time and the topic is fixed. However, they can reflect students' real target language proficiency. By reviewing verb-noun collocation errors extracting from ST3, some tentative reasons for collocation error can be given in the following:

(a) Over-use of delexical verbs. Whenever students feel it is difficult to find a proper verb, they will seek help from the delexical verbs, like: do, make, take, etc. Such as: make practice, take food, do crime, do shop assistance, make the knowledge, do teaching, have some methods, etc..

(b) Directly translate Chinese into English. When students do not know how to express in English they will turn to their mother tongue and translate them directly neglecting the differences between Chinese and English languages. Such as: use their minds, eat medicine, go to death, build hopes, etc..

(c) Overgeneralization. Students create a deviant collocation according to the previous knowledge in the target language. Such as: drive the bikes, know an idea, conclude the reasons, relax their spirit, etc..

2.3 The Implications for Teaching

On the basis of the present research, pedagogical implications have been put forward for improving EFL writing skills, reinforcing the quality of input, reducing cultural differences, cultivating English ways of thinking and helping students massively avoid committing errors in their target language output. With the scientific guidance of the corpus-based study conducted by the author, four causes accounting for students' poor performance in vocabulary are identified showing in the following. Firstly, students are lack of motivation and interest in writing English compositions. As long time influenced by traditional teacher-oriented teaching style, students are used to play a role of listener and have little chances to do output. Secondly, some students are poor in vocabulary because they hold unserious attitude towards remembering words. For these students, when they are required to remember words and enlarge their vocabulary, they only try to bear in mind the word and its one or two corresponding Chinese meaning while neglecting IPA (International Phonetic Alphabet), synonyms, antonyms of the word, and the context within which the word is used. Thirdly, most students are greatly influenced by negative transfer. As Gui told us that mother tongue interference is inevitable in foreign language learning, and generally speaking, it occurs when two languages seem similar but not exactly the same (Gui, 1985, p.215). Fourthly, numerous collocation errors remind us that it is necessary for us to raise learners' awareness of the importance of collocations and lexical collocations should deserve a place in the process of vocabulary acquisition. Facing the problems of English teaching and learning in China, teachers should try to improve students' interest in learning English, encourage students to explore their potential ability in seeking their own errors by learning the correct forms and picking out their wrong expression and shed importance on English practical writing.

CONCLUSION

In the process of English language learning in China, making errors is inevitable. Based on this error-tagged CELC corpus and the theories of Error Analysis, the present study focuses on investigating errors in the compositions of Chinese college English learners with the purpose of improving the quality of the English teaching and ensuring the high quality English input. Through the observation and analysis of the retrieved concordance lines, vocabulary error is discussed. Especially spelling in form error and verb-noun collocation errors are given priority to analyze because students commit more errors in these two categories. By identification, description, and discussion of all these errors in CET-4 writings, the writer intends to provide some pedagogical implications. First and foremost, teachers should understand that with the scientific guidance of corpus, they can easily identify errors, analyze errors and find out the difficulties as well as the distinctive and shared features of a wide variety of interlanguage of Chinese learners in the process of learning English. Then, error analysis based on corpus can provide teachers a huge number of lively and authentic languages used by students, tell teachers the difficulties

confronted by students and provide insights into the process of acquisition. For the purpose of improvement in writing and vocabulary teaching, the application of corpus is therefore recommended.

REFERENCES

- Corder, S. P. (1974). The study of learner errors. In Jack C. Richards (Ed., pp.19-27), *Error analysis: Perspectives on second language acquisition*. London: Longman Group Limited.
- Ellis, R. (1994). *The study of second language acquisition*. Shanghai, China: Shanghai Foreign Language Education Press.
- Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Gui, S. C. (1985). *Psycholinguistics*. Shanghai, China: Shanghai Foreign Language Education Press.
- Gui, S. C., & Yang, H. Z. (2003). *Chinese Learner English Corpus*. Shanghai, China: Shanghai Foreign Language Education Press.
- Gui, S. C., Yang, H. Z., & Yang, D. F. (2005). *Corpus-based analysis of Chinese learner English*. Shanghai, China: Shanghai Foreign Language Education Press.