

A Corpus-based Analysis on Word Errors in College English Writing

Huaqing He

School of Foreign Languages
China West Normal University
Nanchong, P.R. China
E-mail: masha4567@sina.com

Abstract—Based on contrastive analysis and error analysis, this paper attempts to explore word errors committed by non-English majors in their writing collected in Chinese Learner English Corpus. The aim is to offer Chinese learners some help concerning the methods to improve their English writing level and yield some suggestions on English language teaching. The main findings are: first, out of total word errors, the most frequently occurring sub-types of word error are substitution error and part of speech error; second, the less frequent sub-types are the redundancy error and absence error; third, interlingual errors account for 72%, intralingual errors 28%. In the end, causes of the above findings were analyzed and some remedial teaching methods were suggested.

Keywords—college learners; English writing; error analysis; word errors

I. INTRODUCTION

During the process of learning foreign languages, learners are potentially making different types of errors. Corder stated that errors could be significant in three ways: (1) they provided the teacher with information about how much the learner had learnt, (2) they provided the researcher with evidence of how language was learnt, and (3) they served as devices by which the learner discovered the rules of the target language [1]. Therefore, it has been common in foreign language teaching to emphasize the role of errors. Researchers and teachers have made great efforts to find causes of errors and their solutions. One of the most influential factors they believe is mother tongue interference—that is, the interfering effects of the native language on the target language. This paper probes into the interfering effect of Chinese on learners' English writings through contrastive analysis (CA) and error analysis (EA). Specifically, the focus is on word errors committed by Chinese Non-English Major Students. The paper attempts to analyze all the word errors in 290 compositions in Chinese Learner English Corpus (CLEC), gain an insight into the causes of the word errors and help students improve their English proficiency.

II. LITERATURE REVIEW

Language learning, like any other human learning, is fundamentally a process that involves the making of errors. Mistakes, misjudgments, miscalculations, and erroneous

assumptions form an important aspect of learning virtually any skill or acquiring information [2]. In order to analyze learner language in an appropriate perspective, it is crucial to make a distinction between mistakes and errors, technically two very different phenomena [2]. "Whereas a mistake is a random performance slip caused by fatigue, excitement, etc, and therefore can be readily self-corrected, an error is a systematic deviation made by learners who have not yet mastered the rules of the second language (L2)" [3]. However, in many situations it is difficult to distinguish errors and mistakes. Therefore, in this paper, "errors" covers both errors and mistakes.

Researchers try to use many methodologies to analyze these errors so as to help teachers teach more efficiently and students learn better. Contrastive analysis, error analysis and language transfer are very important means they resort to. From the 1940s to the 1960s, contrastive analyses were conducted, in which two languages were systematically compared. Researchers at that time were motivated by the prospect of being able to identify points of similarity and difference between native languages and target languages. "Where two languages were similar, positive transfer would occur; where they were different, negative transfer, or interference, would result" [3].

Error Analysis was, of course, not a new development and had long been part of language pedagogy [4]. It was S.P. Corder who first advocated in English Language Teaching the importance of errors in language learning process. Corder explained the procedures for EA. It is as follows: (1) A corpus of language is selected. This procedure includes deciding on the size of the sample, the medium to be sampled, and the homogeneity of the sample (with regard to the learners' ages, L1 background, stage of development, etc.) (2) The errors in the corpus are identified. Corder points out the need to distinguish "lapses" (i.e. deviant sentences that are the result of processing limitations rather than lack of competence) from "errors" (i.e. deviant sentences that are the result of lack of competence). (3) The errors are classified. This involves deciding the criteria of classifying errors and assigning a grammatical description to each error. (4) The errors are explained. In this stage of the procedure, an attempt is made to identify the causes of the errors [1].

Errors can be divided into interlingual errors and intralingual errors. An interlingual error is one, which results from language transfer, that is, which is caused by the learner's native language. An interlingual error is one which results from faulty or partial learning of the target language, rather than from language transfer [5]. Intralingual errors are often further subdivided into: overgeneralization errors, ignorance of rule restrictions, incomplete application of rules and false concepts.

Researchers and teachers in China have conducted intensive researches on EA and the mother tongue interference and composed many papers and works. Some of them aim at analyzing learners' writings and point out the role the mother tongue played in the process of foreign language learning. A few of them focus on the errors collected in CLEC [6][7]. All these papers indicate the negative influence of mother tongue on learners' writings objectively.

III. METHODOLOGY AND FINDINGS

This study involves contrastive and error analyses on the word errors collected from written samples of non-English majors in CLEC. This study adopts Corder's procedures: corpus selection, error identification, error classification and error explanation.

A. Corpus Selection

Chinese Learner English Corpus (CLEC) is selected as the data pool of this study. CLEC contains five sub-corpora, including junior and senior English-major corpus, junior and senior non-English major corpus and middle school learners' corpus. The corpus offers basic information of the essays, including types of compositions, personal information, learners' level, and evaluation. Professor Yang Huizhong points out that the information is of great importance because it can reflect the typical differences between learners of different levels and help us understand the process of foreign language learning [6]. Some researches on CLEC have demonstrated that the subjects of this corpus can represent English learners in China [6][7]. This paper chooses 130 compositions in junior non-English-major corpus (st3) and 160 compositions in senior non-English-major corpus (st4) as research subject. These compositions' title is "Health Gains in Developing Countries", made up of 40320 words, excluding all error tags and annotations.

B. Error Identification

CLEC adopts COCOA as its standard of annotations which are being put into "〈 〉" to offer the main information of compositions. The information locates in the first line of each composition so as to facilitate the recognition of computers. For example, <Band ?><ST 3><Female><Y 8><AGE 20><WAY 1><DIC 2><TYP1><SCH ?><SCORE 10><TITLE Health Gains in Developing Countries > indicates that this argumentation composition is written without dictionary by a non-junior English major and the score is 10. The writer is female and at the age of twenty. Where she studies is uncertain.

CLEC tags 61 types of errors, among which 11 types are first-class categories while the other 50 are sub-categories. The first-class categories of errors include word form error(fm), verb phrase error(vb), noun phrase error(np), pronoun(pr), adjective phrase error(aj), adverb error(ad), preposition phrase error(pp), conjunction error(cj), wording error(wd), collocation error(cc), syntax error(sn). In CLEC, errors are tagged as follows: "*and helps them to resolute [wd3 - 4] all kinds of problem*". Wd3 stands for the type of the error, - stands for the position of the error word—resolute and 4 means that if we want to decide whether the error exists, it is necessary to consider four words on the right of "resolute". The search engine used in this study is ConCapp. The result is achieved through opening file "st3" and "st4" respectively, and then inputting "Health Gains in Developing Countries" and "wd" into the search string.

C. Error Classification

Word refers to the choice of words in a given linguistic context. Word errors are those made either by the wrong choice of part of speech or by improper substitution (usually the use of a grammatically acceptable but semantically unacceptable word in a given linguistic context). Word errors tagged in CLEC can be classified on the basis of two dimensions.

The first dimension follows Gui and Yang's division [6], that is, word errors are divided into order error, part of speech error, substitution error, absence error, redundancy error, repetition error and, ambiguity error. Different abbreviations stand for different categories and wd1 represents the first type of wording errors—order, wd2—part of speech, wd3—substitution, wd4—absence, wd5—redundancy, wd6—repetition and, wd7—ambiguity. "Order" errors are defined as "misplacement of any word other than an adverb", for example, **another for example/ ✓for another example*. "Part of speech" errors are defined as "an error in part of speech: right root, but wrong word class", for example, *If we want to be a *succeed/ ✓successful man*. "Substitution" errors are defined as "an error in word choice: right word class but wrong selection", for example, *But some people couldn't really know its *means/ ✓meaning*. "Absence" errors mean an omission of a word, for example, **in his first/ ✓in his first time*. "Redundancy" errors mean an extra word, for example, *It *is refer to that a lot of practice can form an ability of doing something*, word 'is' is redundant. "Repetition" errors refer to unnecessary repeating of a word, for instance, *you should think and *think to find out the best way*, word "think" is unnecessarily repeated. "Ambiguity" error refers to "unclear word meaning". For instance, *Not only reading but also speaking will be better when we practice more*, the meaning of *reading and speaking* is not clear.

The second dimension is interlingual and intralingual errors.

Through the ConCapp's searching, 1372 word errors are found in the sub-corpus, including 32 order errors (wd1), 350 part of speech errors (wd2), 541 substitution errors (wd3),

184 absence errors(wd4), 189 redundancy errors (wd5), 14 repetition errors (wd6) and, 62 ambiguity errors(wd7). In Figure 1, substitution error (wd3) and part of speech error (wd2) are the top two word error sub-types which can respectively explain 39% and 26% of the total number of word errors. Most students made these two sub-type errors and the average number of errors by each student is more than one. The less frequent sub-types are the redundancy error (wd5) and absence error (wd4), which explain 14% and 13% of the total number of word errors respectively. Only a small proportion of students made order errors (wd1), repetition errors (wd6) and ambiguity errors (wd7). Out of 1372 errors recorded, 987 errors are classified as interlingual errors, 385 as intralingual errors. Interlingual errors account for 72%, intralingual errors 28%.

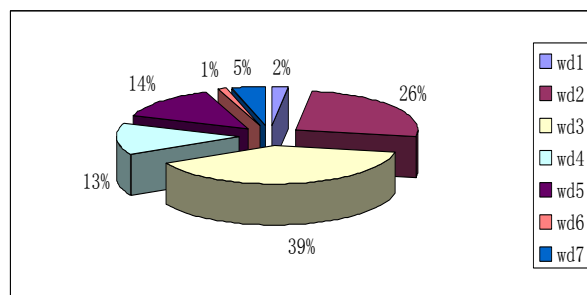


Fig. 1. Rate of the subtypes of word errors

D. Error Explanation

Error explanation is the fourth step of error analysis. There are only 32 order errors and 14 repetition errors in the sub-corps, and ambiguity errors are difficult to be judged, so these 3 types of word errors are not considered here. The other 4 types of word errors are analyzed mainly from the mother tongue interference's perspective because the amount of interlingual errors overwhelms that of intralingual errors. As the four errors show a higher percentage of occurrences and therefore more traces of L1 influence, it is beneficial to have a brief view of the basic features of English and Chinese related to the occurrence of word errors at the very outset. Part of speech is discussed first.

1) Part of speech errors

Part of speech errors account for 26%. Among 350 part of speech errors, errors in the use of adjective as adverb are 70 (example1), 12 errors in the use of adverb as adjective (example2), 83 errors in the use of verb as noun (example3), 40 errors in the use of noun as verb (example4), 27 errors in the use of noun as adjective (example5), 24 errors in the use of adjective as noun (example6), 15 errors in the use of verb as adjective (example7), 91 errors in the use of verbs' different forms (example8). These examples are:

Example1. You can write the words **correct* / $\sqrt{\text{correctly}}$.

Example2. Newspaper is the most **simply* / $\sqrt{\text{simple}}$ and efficient way.

Example3. It will lead to **succeed* / $\sqrt{\text{success}}$ easily.

Example4. During the study of English, we will get an easy method to **memory* / $\sqrt{\text{memorize}}$.

Example5. When you begin to do something first, you can feel **difficulty* / $\sqrt{\text{difficult}}$.

Example6. It needs **patient* / $\sqrt{\text{patience}}$.

Example7. Most people want to be a **succeed* / $\sqrt{\text{successful}}$ man.

Example8. During the **studying* / $\sqrt{\text{study}}$ of English,

All these error words are content words, which have incorrect suffixes. There are two reasons causing the errors. On the one hand, English has many complex suffixes which may change parts of speech. A noun, a verb, an adjective and an adverb may trace to the same root. For example:

Noun: Her *carefulness* wins over her boss.

Verb: I don't *care* a bit what critics think.

Adjective: He was *careful* enough to check up every detail.

Adverb: Please read this article *carefully*.

It is very difficult for learners to remember these suffixes. On the other hand, English and Chinese are two unrelated languages they differ much in lexicology and morphology [8]. No morphological changes happen to Chinese words. Many Chinese words such as “容易, 正确, 小心” can be used as an adverb and adjective. However, In English, suffix “ly” should be added to many adjectives to form adverbs, for instance, “easy and easily”, “correct and correctly”. Influenced by Chinese, Many Chinese learners ignored different derivations of English words and committed these part of speech errors.

2) Substitution errors

Figure 1 clearly indicates that the substitution errors are more noticeable, accounting for 39% of the total word errors. The substitution errors are random in parts of speech, i.e., evidence of substitution concerns most of the word classes. The following are examples:

Substitution errors in the use of nouns:

Example 9: We are lacking in **insistence* / $\sqrt{\text{perseverance}}$.

Substitution errors in the use of verbs:

Example 10 : We must join their society and **study* / $\sqrt{\text{learn}}$ from the people.

Substitution errors in the use of adjectives:

Example 11: That ' s our **respectful* / $\sqrt{\text{respectable}}$ father.

Substitution errors in the use of adverbs:

Example 12: The moon is always **specially* / $\sqrt{\text{especially}}$ round on that day.

Substitution errors in the use of prepositions:

Example 13: We can read, write and so on **under/√in* the light.

Substitution errors in the use of conjunctions:

Example 14: I couldn't hold my tears **but/√and* just cried.

Substitution errors in the use of pronoun:

Example 15: I would like to tell you **some/√something* about this.

Substitution errors in the use of auxiliaries:

Example 16: But the beautiful sense always **be/√is* kept in people's mind.

Substitution errors in the use of articles:

Example 17: People in **a/√the* same village will sit together.

Through observation, it can be found that substitution of notional words, which is slightly more than that of structural words, usually results in an intralingual error, while that of structural words tends to imply L1 influence. For intralingual errors in substitution, mispronunciation, adjacent influence, confusion or false friends in conception seem to be responsible for the occurrence. Examples of such errors are the use of a for an, protest for protect, respectful for respectable, etc. Among the errors of substitution, the improper choice of prepositions is rather obvious. This has much to do with the fact that prepositions not only perform syntactic function in a language, but convey various lexical meanings as well [9]. Prepositions in English and Chinese share such syntactic functions but are in many cases different in the semantic features and co-occurrence with other elements in linguistic constructions. Thus if a Chinese learner's knowledge of L2 prepositions is not sufficient enough to tell the differences in their choice between the two languages, specific L1 conflation would exert certain influence on their L2 production. Example 21 indicates the improper choice of under for in.

Word meaning in the two languages may be different in each type mentioned above. As a result, the difficulty of selecting words comes from the lexical transfer of the mother tongue; and this lexical transfer comes from presupposed synonymy and literal translation. According to Hu Zhuanglin [10], English is said to be rich in synonyms, such as study and learn, world and universe. But total synonymy is rare. The so-called synonyms are all context dependent. They all differ in one way or another. In addition, some synonyms having identical lexical meaning in Chinese have to be used in different contexts. If a foreign language learner cannot distinguish one word from another, they either use improper words or happen to choose right words on the basis of their intuition.

3) Absence errors

Absence errors account for 13% of total word errors. The absence of the articles is the most frequent in all parts of speech, especially the definite article the (example 18). Because the Chinese language has no definite article, it is

very difficult for Chinese learners to master. In the English word class known as articles there are only two linguistic elements: the indefinite a/an and the definite the. Though this part of speech contains the smallest number of elements, their occurrences are always the highest in percentage, e.g. the is the first of the 100 words with the highest frequency of occurrence in both the Brown corpus and LOB corpus while a comes fifth, their meanings are implied and diverse, and their use is definite in certain contexts but optional in others [11]. All these features and the total absence of their equivalents in Chinese make it a problematic aspect of L2 learning for the Chinese learners. Thus adequate knowledge of their usage is very necessary for EFL learning. Auxiliaries, which are placed before adjectives, are easily deleted (example 19). In Chinese, adjectives can be directly used as predicative, however, in English, auxiliaries should be put before adjectives. For example, ni meili, you are beautiful. The conjunctions and pronouns are also deleted (example 20). The omission of content words occurred least frequently, and these deleted content words are usually from expressions and phrases (example 21).

Example 18: Little by little, (*√the*) more practice you do, (*√the*) more perfect you are.

Example 19: we (*√are*) sure that constant practices will help us to succeed.

Example 20: Not only in the study of English, (*√but also*) in many other aspects, the proverb proves quite right.

Example 21: *For another* (*√example*), a carpenter can make the tables or chairs while we can't.

4) Redundancy Errors

Redundancy errors account for 14 % of total word errors. More than half the errors, resulting from mother tongue interference, are literally translated. In example 22, structure "If, then" is influenced by Chinese structure "如果, 那么". The conjunction "so", which is redundant, occurred times. Example 23 is illustration. It shows that Chinese learners' compositions have characteristics of oral English. In examples 24, 25, the articles "the" and "a" are redundant, which indicate that the article is so difficult that Chinese learners can't completely master it. They may make both redundancy errors and absence errors in the use of articles. In example 26, the phrase "I'm like" is an intralingual error, based on a blend of the English structures I am and I like.

Example 22: *If* you take this medicine, **then* all your symptoms will disappear.

Example 23: If you only remember their mean, but not practice them, **so* you have difficulties when you speak and write sentences or passages.

Example 24: I usually talk about **the* English with my classmates.

Example 25: For **a* instance, how can we improve our life quality?

Example 26: **I'm like* plying sports.

IV. CONCLUSION

Word errors committed by non-English major students have been carefully analyzed. Word errors are divided into 7 types, that are order error, part of speech error, substitution error, absence error redundancy error, repetition error and, ambiguity error. This study focuses on part of speech error, substitution error, absence error and redundancy error, because these 4 errors occurred more frequently. The amount of interlingual errors overwhelms that of intralingual errors. That is to say the mother tongue interference is the major cause of word errors. Therefore, The 4 types of word errors are analyzed mainly from the mother tongue interference's aspect.

According to the analysis, part of speech errors are mainly resulted from complicated suffixes that change parts of speech. The learners in China sometimes overuse the rules of changing forms of words or forget to change the forms of words when it is necessary. Substitutions errors are chiefly caused by English synonyms, which have the same lexical meaning in Chinese but they are different either in one way or another in English. Absence errors and redundancy errors in the use of the definite article frequently occurred since the definite article does not exist in Chinese.

In a word, the study on errors gives insight into both teachers and learners. For teachers, errors can tell them the difficulties confronted by students and the key points of the classes. For students, error correction can tell them the right usage of a particular word or sentence. However, this study has its limitations. First, one of the weaknesses of error analysis is it is difficult to identify the unitary source of an error [3]. The focus of this study is mother tongue interference, and some other causes have not yet been analyzed. Second, the subject of this research consists of only 290 compositions; whether it can represent the English level of non-English majors, it needs being further researched.

ACKNOWLEDGMENT

This work was founded in part by the Department of Education of Sichuan Province, China, under Grant No. 13SA0015.

REFERENCES

- [1] S.P. Corder, "The significance of learner's errors," *International Review of Applied Linguistics*, No.4, pp: 161-169, 1967.
- [2] H.D. Brown, *Principles of Language Learning and Teaching*, 3rd ed., Beijing: Foreign Language Teaching and Research Press, pp.216-217, 2002.
- [3] L.F. Diane and M. H. Long, *An Introduction to Second Language Acquisition Research*. Beijing: Foreign Language Teaching and Research Press, pp. 53-61, 2000.
- [4] R. Ellis, *The Study of Second Language Acquisition*. Shanghai: Shanghai Foreign Language Teaching Press, P.48, 1999.
- [5] J.C. Richards, *Longman Dictionary of Language Teaching & Applied Linguistics*. Beijing : Foreign Language Teaching and Research Press, p.233, 2002.
- [6] S.C. Gui, and H.Z. Yang, *Chinese Learner English Corpus*. Shanghai, Shanghai Foreign Language Education Press, 2003.

- [7] H.Z. Yang, S. C Gui, and D. F. Yang, *Corpus-based Analysis of Chinese Learner English*. Shanghai, Shanghai Foreign Language Education Press, pp.110-125, 2005.
- [8] L.M. Xiao, *English-Chinese Comparative Studies and Translation*. Shanghai: Shanghai Foreign Language Education Press, 2002.
- [9] R. Quirk and S. Greenbaum, *A University Grammar of English*. London: Longman, p.306, 1973.
- [10] Z. L. Hu, *Linguistics A course Book*. Beijing: Beijing University Press, p.163, 2001.
- [11] R.H. Li, *Language and Culture: Contrastive Studies between English and Chinese*. Shanghai: Shanghai Foreign Language Education Press, p.236, 2002.