

Spelling Errors Analysis in College English Writing

Yanyan Liu

College of Foreign Languages, Qingdao University of Science and Technology, China

Abstract—How English learners master and use vocabulary in writing is not only an important criterion for their English writing, but also a crucial aspect to show their ability to express in English. Moreover, the use of English vocabulary in their writing can further represent how well they learn English vocabulary. However, it can also reflect some existing faults in the vocabulary teaching and learning. How to improve vocabulary teaching and how to help learners use words effectively are the main problems to deal with during vocabulary teaching. This paper tries to do an empirical investigation, helping both teachers and language learners locate the difficult points in the target language and get prepared to acquire the language in a more oriented way. This study is carried out by collecting 68 compositions written by non-English majors. The errors are classified into three categories: local syntactic errors, global syntactic errors and semantic errors. As required, each category is further divided according to different criteria. Major causes are offered by illustration. Finally, some suggestions on vocabulary teaching are offered.

Index Terms—error analysis, real-word spelling errors, causes of errors

I. INTRODUCTION

Error analysis is a very important branch in the research field of second language acquisition (SLA) no matter in our country or abroad. For example, there is a whole chapter in the famous work about SLA by Ellis and even an entire book dedicated to error analysis by James, not to mention tons of reference books of grammatical errors for learners and teachers of English as a second or foreign language, such as Turton. In our country, a world famous corpus on English learners' errors (CLEC) was built years ago and plenty of researches have been published since then, including Yang, Gui and Yang.

“Among the grammatical errors made by writers, spelling error occupies the most prominent position, no matter in the writers of native speakers or language learners” (Leacock *et al.*, 2002, p.15-17). Nowadays, along with the fast development of computational techniques in automated spelling error detection, especially the broad using of spelling checker in Microsoft Word, students are being able to eliminate most of the spelling errors by themselves in their daily writing practice. Yet, spelling checkers are usually not good at detecting those real-word spelling errors and it is hard for students to find those errors by themselves.

II. LITERATURE REVIEW

A. *Learner Errors and Error Analysis in SLA*

1. Second language acquisition

In the term second language acquisition, “‘second’ can refer to any language that is learned subsequent to the mother tongue” (Ellis, 1987, p.3). “The activity of acquiring a second language is often referred to as a L2. ‘L2 acquisition’, then, can be defined as the way in which people learn a language other than their mother tongue, inside or outside of a classroom, and ‘Second Language Acquisition’ (SLA) as the study of this” (Ellis, 1987, p.4). “The goals of SLA are to find out what learners actually do by collecting samples of learner language which provide evidence of what the learners know about the language they are trying to learn and describe how learner language changes over time. SLA also focused on the formal features of language the linguists have traditionally concentrated on, such as pronunciation, words, and grammar.” (Ellis, 1987, p.4-5). This paper studies the word spelling errors made by Chinese students in their English-written compositions. Therefore, L2 refers English in this paper.

2. Errors

“It may seem odd to focus on what learners get wrong rather than on what they get right. However, there are good reasons for focusing on errors. First, they are a conspicuous feature of learner language, raising the important question of “why do learners make errors?” Second, it is useful for teachers to know what errors learners make. Third, paradoxically, it is possible that making errors may actually help learners to learn when they self-correct the errors they make (Ellis, 1987, p.15). As to the definition of errors, Rod Ellis wrote errors reflect gaps in a learner’s knowledge. Carl James defines a language error as an unsuccessful bit of language. Brown holds that an error is a noticeable deviation from an adult grammar of a native speaker, resulting from the inter-language competence of the learner.

B. *Spelling Errors and Automated Spelling Error Detection*

1. Spelling errors

According to Kukich, spelling errors can be divided into two parts: typographic errors and cognitive errors “Typographic errors have something to do with keyboard adjacencies (insertion, deletion, substitution, or even transposition)” (Kukich, 1990, p.350), for example, ttheir (their), whre (where). While “cognitive errors usually stem from phonetic similarity” (Kukich, 1990, p.352), for example, tirm (term), acedemy (academy). Other factors include shifting key characters, shape similarity.

Formal misselection occurs when words that looks and sounds similar. For example, *This is the course [cause] of the accident*. The student wants to use *cause*, however, he chooses the wrong word. Unfortunately, the error is a real existent word, which makes the error more serious.

Misformation refers to the words that do not exist in the target language. They can originate either in mother tongue or be created by the learner from the resources of target language itself. Chinese students usually will not make interlingual misformation errors, because Chinese and English belong to quite different language families. But if the Chinese students have learned some other languages, it is possible for them to make such errors. For example, in the sentence *The preise [price] of tomato is high* (preise means price in German). The student uses German word preise. It seems that the student has learned some German words.

Distortions refer to the intralingual errors of form created without recourse to first language resources. The outcomes are forms non-existent in the target language. They result from the misapplication of one or more of the processing operations: omission, overinclusion, misselection, misordering and blending. Here are some examples: (1) omission: seris [serious]; (2) overinclusion: polloution [pollution]; (3) misselection: folgen (follow); (4) misordering: soceity [society]; (5) blending: popolution [population + pollution].

2. Automated spelling error detection

The writer has found that the above-mentioned error categories cannot cover most of the spelling errors. Therefore, the paper will focus on whether the form is existent or not in the target language. For those which are not existent in the target language we call them non-word errors. While those which are existent in the target language called real-word errors.

That is to say, in this paper we can classify all the spelling errors into two parts; one is non-word errors, while the other is real-word errors. For those non-word errors, it is easy to check against words list, and flag them if they are not in list. So we do this with the aid of Microsoft Word and correct them. However, those real-word errors are also in list, only unsuitably used in the context. Therefore, it is difficult for writers to notice them. What makes things more complexity is that real-word errors written by learners always have a large deviation from target words, contain multiple errors and misspell the first letter?

In practice, there is no agreed-upon list of confusable real words. It is also hard to predict which words will be confused.

III. RESEARCH METHOD

A. Objects and Data Collection

In this study, more than 600 compositions are collected from non-English majors to be the data of the study. These compositions are the students' writing exercises. The topics in these writings are diverse and the students do not know their compositions will be samples of the research. Therefore, they are all free productions that provide the best research materials. After using Microsoft Word automatic spelling checker to correct all the non-word errors, the writings of two topics were chose as research subjects.

B. Research Procedures

The process of this study extends more than four months and includes the following stages.

The first step is data collection. The writer randomly chose more than 632 copies of compositions, including the following topics: My Experience of Learning English, Family is more/less important than in the past. In the beginning, the writer collected different kinds of compositions of non-English majors. And the total amount of words is more than 120,000. The writer input all the data into one document and corrected all the spelling errors that Microsoft Word automatic spelling checker could detect.

The second step is to identify those real-word spelling errors that Microsoft Word automatic spelling checker cannot find out. At this stage, the writer only analyzed 68 compositions including 11,503 words, which focused on two topics, namely, “My Experience of Learning English” and “Family is more/less important than in the past”.

The third step is to describe errors. There are many different kinds of classification systems that are mentioned in chapter two. In this paper, Atwell and Elliott's error categories will be used. These researchers analyzed errors in four categories: (1) non-word errors; (2) local syntactic errors; (3) global syntactic errors; and (4) semantic errors. This paper only focuses on real-word errors, so the last three categories will be adopted.

The fourth step is to explain errors. Explanation is concerned with establishing the source of the errors.

TABLE 1.
THE TOPICS OF THE SAMPLES

Topic	Amount (68)	Total Words
My Experience of Learning English	51	9,823
Family is more/less important than in the past	17	3,029

C. Qualitative and Quantitative Analysis

Qualitative and quantitative analysis are two basic approaches to language study.

Qualitative analysis is a kind of research that the data are not in the form of numbers. On the contrary, the data in quantitative analysis are in the form of numbers. In detail, the focus of qualitative analysis is to use language to describe the study. This kind of method mainly depends on the researchers' experience, intuition as well as the information and situation of the research object to make a judgment. While quantitative analysis refers to a method, which uses mathematical language to describe the study. It is based on statistical data to set up mathematical models and then make conclusions. Qualitative analysis is the basis of quantitative analysis. Quantitative analysis makes qualitative analysis more accurate and scientific, and it can also promote qualitative analysis to draw an extensive and in-depth conclusion. Therefore, in this paper the writer combines the two methods to do the research. In the aspect of qualitative analysis, the writer will divide all the real-word spelling errors into three parts: local syntactic errors, global syntactic errors and semantic errors. In the aspect of quantitative analysis, the writer will discuss the relative proportions of real-word spelling that fall into each of the three categories.

In this paper, I use quantitative analysis describe the proportion of each category, which can make us easy to know the easiest wrongly written words and the categories they are belong to. The qualitative analysis in this paper can give us a detailed description of different errors, which make us have a deeper understanding of them and help us avoid making such errors.

IV. RESULTS AND DISCUSSION

Some researches in China have shown that the biggest difficulty in the process of learning of a foreign language is lexis. Language is a sequence of words. People from different culture have different systems of languages. They differ greatly in way of thinking and ways of expression. Therefore, the fundamental difference between native language learning and target language learning is how the second language learners use the target language to express the same idea in the native language context. It is known that Chinese and English language belong to different language systems. They differ much in pronunciation, spelling, grammar, lexis and usage. That is why Chinese learners make errors in such aspects.

A. Quantitative Analysis of Real-word Spelling Errors

The following table describes the quantitative features of the three kinds of errors.

From the table below, local syntactic errors take up the biggest proportion.

TABLE.2
DISTRIBUTION OF REAL-WORD SPELLING ERRORS

Total	Local Syntactic Errors	Global Syntactic Errors	Semantic Errors
55	32	6	17
100%	58%	11%	31%

B. Qualitative Analysis of Real-word Spelling Errors

1. Local syntactic errors

The local syntactic violations involve local incongruities within 1 or 2 words, which can be divided into three kinds: grammatically misspelled errors, errors in word class and errors in lexical relations.

TABLE.3
DISTRIBUTION OF LOCAL SYNTACTIC ERRORS

Total	Grammatically Misspelled Errors	Errors in Word Class	Errors in Lexical Collocation
32	13	8	11
100%	41%	25%	34%

(1) Grammatically misspelled errors

TABLE.4
DISTRIBUTION OF GRAMMATICALLY MISSPELLED ERRORS

Type of Errors in Grammatically Misspelled Errors	Number	Percent (100%)
Singular and Plural Forms of Nouns	3	23.3
Superlative of Adjectives	2	15
Infinitives	3	23.3
Past Participle of Lexical Verbs	2	15
Subject-verb Agreement	3	23.3
Total	13	100

Errors in singular and plural forms of nouns

- 1) He recommends me to see some interesting English [film] <films> and listen to some cool music.
- 2) Now, I am an undergraduate, many [exam] <exams> of English level will be challenge to me.
- 3) As the saying goes, all [road] <roads> lead to Rome.

In the above three sentences, they all wrongly use the singular forms of countable nouns to describe the concept of many. Therefore, they all should be replaced by the plural forms of these countable nouns.

Errors in superlative of adjectives

4) The teacher didn't teacher us patiently and the interesting has been missing, this time is the [worse] <worst> time although I learn English.

5) But this is just the [worse] <worst> part for me.

Instead of comparative forms of adjectives, superlative of adjectives should be used here.

Errors in infinitives

6) Although my English is not good now, I believe that nothing is impossible if you want to do it. I will try my best to [learning] <learn> English in my university.

7) Because in my heart, English is a mysterious language so that so many people learn it and I wanted to [learning] <learn> English well was a hard work.

8) At that time, I thought that it is unnecessary for me to [learning] <learn> a Chinese course.

Here we should use infinitives instead of present participle of verbs.

Errors in past participle of lexical verbs

9) From then on, I have [make] <made> up my mind to learn English well.

10) Come to the college, everything has been [change] <changed>.

The past participle of lexical verbs should be used in the above two sentences to replace the past tense verbs.

Errors in subject-verb agreement

11) I lost my interest of English, because I thought I [were] <was> the top of our class.

12) Firstly, competitions that children face are more and fiercer nowadays, so parents' support and encouragement for their kids [is] <are> more vital.

13) When I was failed on study, I never given up because of my dad's encouragement, he help <helps> me find back my confidence and cheer me up.

This kind of errors in examples 11) through 13) is subject-verb number disagreement, which is more commonly happened in Chinese students writings.

(2) Errors in word class

Sometimes nouns are replaced by verbs, sometimes adjectives by nouns. Chinese characters don't have any obvious markers to show their word class. That's a reason why Chinese learners are not often aware of word class of English words when using them.

TABLE.5
DISTRIBUTION OF ERRORS IN WORD CLASS

Type of Errors in Word Class	Number	Percent (100%)
Noun as adjective	3	38
Adjective as noun	1	12.5
Noun as verb	1	12.5
Verb as noun	1	12.5
Personal Pronoun as Adjective Possessive Pronoun	2	25
Total	8	100

Nouns misused as adjectives (or vice versa)

14) Now, I do not afraid to write and speak English, I just try hard to do it. I will never forget her who showed me that learning English could be [interest] <interesting> and rewarding.

15) What's more, she told me an [interest] <interesting> method to remember the English words.

16) To begin with, when we as [youth] <young> as a baby, it is our parents who teach us how to step our first step and to speak first word.

17) When I told him I didn't like English and I think English is no [useful] <use> for having a deep conversation with me.

18) In senior middle school, I am very [interest] <interested> in English as much as I'm in junior school .So I always

get the bad scores.

19) I needed someone’s [encouraging] <encouragement> but nobody did so. Finally, I gave up.

The word ‘interest’ is a noun. It should be used as subject or object, but it wrongly used as predictor. The word ‘encouraging’ in example 19 is a verb. It should be added an object. However, students falsely used it as a noun.

Personal pronoun as adjective possessive pronoun

20) Form [they] <their> words, I understand life not only myself but also my parents, I should value it two times.

21) They will do [them] <their> best to take good care to us.

The above two sentences misuse personal pronoun and adjective possessive pronoun.

The use of word class needs a special attention in vocabulary teaching. Chinese has a negative effect on their English language learning. When writing in Chinese, learners don’t need to consider so much of the use of word class because Chinese language lays more emphasis on expression of meaning. Therefore, the word class of words is probably neglected when they are writing in English.

(3)Errors in lexical collocation

In Wikipedia, collocation is defined as a sequence of words or terms which co-occur more often than would be expected by chance, within the area of corpus linguistics. The question is not which word could substitute for this word, but which words typically precede or follow it.

22) Now, I can write down my experience [using] <in> English.

23) Since then, I am interested [with] <in > English.

24) As time [going] <goes> by, I found English was [hard] <harder> than before.

25) [The] <At> last I came to college with a good score of English.

26) [Thank] <Thanks> to their teaching, let us grew-up and successful.

27) [Late] <Later> on, we could touch English a little more. My mark is down at the sane tine. Also I lose interest in English slowly.

28) When I first met English I fell in love [in] <with> it.

29) I love listening to English and English songs, because [on] <in> my opinion, English sounded beautiful .But it’s so sad that my English is very poor.

30) [On] <At> that time, I was good at English that was based on my enthusiasm.

31) But one year [late] <later>, my first English teacher was gone, so I had to be taught by another teacher, who was a beautiful girl.

32) [In] <At> that time, I was a grade 4 student, and I didn’t begin to learn English in Wuxi, but other students have learnt English for 2 years, my English ability is the worst at class, so I was afraid of learning English.

In the above sentences, errors of lexical collocation occur. Seen from these eleven sentences, errors are made especially in the fixed phrases.

2. Global syntactic errors

The global syntactic errors in the following sentence are more problematic.

33) At first, we [know] <knew> little about English and didn’t understand what the teacher said if she didn’t translate the words into Chinese.

34) Because of her teaching, I like to answer the questions and never [worrying] <worry> about other people’s derision. Of course, primary school’s question was very easy.

35) She know I haven’t learnt English ,so she would self teach my English after class, I know that I learnt the first English word is “zoo”, because I think that it is likely 2 hundred, I remember [I] <it> clearly.

36) [Beside] <Besides>, family members seem to care about being together.

37) Parents should not only teach the children how to study, but also taught to something [beside] <besides> the study.

38) [One] <Once> I forget to finish homework, the English teacher very angry.

In sentence 34), readers have to finish reading the whole sentence to realize the exact tense of the sentence. When reading the last three sentences, readers can understand the meaning by considering the meaning of the whole sentence. Therefore, this kind of errors is an obstacle to people’s daily reading.

3. Semantic errors

TABLE.6
DISTRIBUTION OF SEMANTIC ERRORS

Total	Similar in Spelling	Similar in Pronunciation	Similar in Meaning
17	12	3	2
100%	71%	17%	12%

Similar in spelling

39) I had to pay more attention to my poor math. Then my English [proved] <improved> slowly.

40) But [generally] <gradually>, laziness made me not study hard

41) When I [stopped] <stepped> into the senior high school, I began my grammar learning.

42) At that time, I only learned the [sample] <simple> English.

43) Now, English is my favorite [project] <subject>, I learn it well and I always at the top of my class.

44) As we all know that family plays an important role in Chinese mind, but as the advent of modern society and progress of the age, family now [seen] <seems> to be less important than before, according to someone's opinion.

45) On the one hand, they know [must] <more> about us, high efficiency schedule, balance diet and so on.

46) During this society, with the development of [completive] <competition>, the work make us tired, the stress make us illness.

47) Nowadays, most of the people think that family is less important because children can live normally [event] <even> with the high divorce rate.

48) At last, she often instructed us to say oral English as well as [baby] <body> English.

49) She is the most responsible teacher I have [never] <ever> seen.

50) The teacher always [published] <punished> me in class.

In the above sentences, learners may confuse words in similar spelling and even in similar sounding.

Similar in pronunciation

51) In middle school, My English was very [pool] <poor> but I never give up.

52) There are many questions, why I don't like to speak English, why my English [cause] <course> was so poor.

53) I think it's the most difficult project to learn .And the thing [whish] <which> is the most difficult thing to stop me learning well is the vocabulary.

In the above three sentences, the errors are caused by phonological rules. When they write, learners may feel it difficult to tell the difference between words which show no difference in pronunciation or similar in pronunciation. Or they may not realize that English words don't always follow the pronunciation rules.

Similar in meaning

54) So I spend more time to learn English, sometimes I even learn English [until] <till> twelve o'clock.

55) Miss Deng is a beautiful and young lady, as long as she came to our class, everyone would quiet with smile, [although] <even> those naughty boys, they would stop talking about their things.

In the above two sentences, learners wrongly choose words to present their ideas and meanings.

The above sentences have shown that familiarity with pronunciation of a word and its phonological combination indeed affects accuracy in perceiving, pronouncing and remembering the word.

C. Analysis of Causes of Real-word Spelling Errors

Since a detail description about the real-word spelling errors committed by students has been demonstrated in the previous section. In the following part, I will hunt for the potential causes of real-word spelling errors.

1. Interlingual causes

Interlingual cause is mainly due to the interference of learners' mother language. Mother language can both have positive and negative influences on the target language. Chinese and English belong to different systems. There are great differences between their vocabulary, including pronunciation, spelling, word class and so on. The differences trigger negative transfer. Therefore, the big difficulty for Chinese learners is English vocabulary.

Taylor has argued that less proficient learners will rely more on mother language. "More advanced learners know more about the target language and more often make analogies between mother language and target language. Since less proficient learners have less such kind of knowledge, they will tend to pay more attention on their native language. For instance, as intermediate-low learners, when they want to express their complex ideas, they often create some expressions by translating Chinese phrases and collocation word for word. But not all L1 equivalents are absolutely the same with L2 words" (Taylor 103-104). That is to say, some sentences produced by the language learners do not make sense in English, but if translated into Chinese, they really do.

2. Intralingual causes

"Intralingual Causes are by definition not predicted on the basis of contrastive analysis (CA), they can't be traced back to differences between L1 and L2, but they relate to a specific interpretation of the target language and manifest themselves as universal phenomena in any language learning process. Technically, they are caused by the faulty or partial learning of one item of the target language upon another" (Huang 307). Therefore, this kind of errors has little connections with L1. It can be sub-divided as follows: false analogy, overgeneralization, ignorance of rule restrictions and incomplete rule application.

REFERENCES

- [1] Brown H.Douglas. (1987). Principles of Language Learning and Teaching. Prentice: Prentice-Hall.
- [2] Ellis R. (1999). The Study of Second Language Acquisition. Shanghai: Shanghai Education Press.
- [3] James C. (2001). Errors in Language Learning and Use: Exploring Error Analysis. Beijing: Foreign Language Teaching and Research Press.
- [4] Huang N. (2000). Lexical Representation and Development in a Second Language. *Applied Linguistics*, 12.7, 305-308.
- [5] Kukich K. (1992). Techniques for Automatically Correcting Words in Text. *ACM Computing Survey*, 14.4, 377-439.
- [6] Leacock, Claudia et al. (2010). Automated Grammatical Error Detection for Language Learners. CA: Morgan & Claypool Publishers.
- [7] Mitton R. (1985). A collection of computer-readable corpora of English spelling errors. *Cog. Neuropsychol*, 2.3, 275-279.
- [8] Richards, Jack, Platt, John and Heidi, Weber. (1985). Longman Dictionary of Applied Linguistics. New York: Longman

Publishing Group.

- [9] Taylor B. (1975). The use of overgeneralization and transfer learning strategies by elementary and intermediate students in ESL. *Language Learning*, 25. 1, 73-107.
- [10] Turton. (1995). ABC of Common Grammatical Errors. London and Basingstoke: Macmillan Education Ltd.
- [11] Young C.M., Eastman C.M., Oakman R.L. (1982). An Analysis of Ill-formed Input in Natural Language Queries to Document Retrieval Systems. *Inf. Process. Manage*, 27. 6, 615-622.
- [12] Yue Meiyun. (1999). Applied Linguistics. Nanjing: Nanjing Normal University.

Yanyan Liu was born in Linyi, China in 1979. She received her master degree in translation from Qingdao University of Science and Technology, China in 2013.

She is currently a lecturer in the School of Foreign Languages, Qingdao University of Science and Technology, Qingdao, China. Her research interests include English teaching methodology and English public speaking.

Ms Liu is a member of Toastmasters Club. She has won the first prize in FLTRP National College English Teaching Contest (2014).