

Corpus-based Study on Language Errors in English Writing

Xin Chen

Southwest Jiaotong University, Chengdu, China

Abstract: English writing, as a significant English skill, can directly and efficiently embody English learners' comprehensive abilities. In the English subject of Chinese college entrance examination, the score of writing accounts for as high as 16.7%. English Curriculum Standards for Senior High School Students, promulgated in 2017, puts forward specific requirements of English writing for high students. However, students still make various types of errors in compositions. Guided by Corder's error analysis steps, and collecting 96 English writings of the senior three students from three senior high schools in Nanchong, Sichuan as samples, the author builds a small corpus after tagging errors in samples based on the Chinese Learner English Corpus error annotation principle. Through the research, it is found that the language error rate of senior high school students is about 9.75%; the error types mainly focus on word errors, verb usage errors, word forms errors and collocation errors, namely, local errors, accounting for 75.3%. In this way, it hopes to improve students' language accuracy in writing and promote English writing teaching in senior high schools.

Keywords: English Writing; Language Errors; Self-built Corpus.

1. Introduction

English writing is a key skill for English learning. It can embody students' comprehensive ability to use English and can indirectly examine students' innovative abilities. As one of the evaluation and testing methods, English writing can help consolidate language knowledge and develop the core literacy of English. Nowadays, the writing scores of most provinces in China have reached 16.7% of the total scores of English papers of the English college entrance examination.

However, students' actual writing level leaves much to be desired. According to New English Curriculum Standard for General Senior High School (2017), senior high students should be able to write at least 100-word compositions in English within limited time, without obvious grammatical errors, with coherent semantics, smooth overall structure, and proper cohesion. Many students still consider writing the toughest one among the four basic English skills including listening, speaking, reading, and writing. Therefore, errors in students' English writing demand prompt solution for both English teachers and learners.

This study seeks to find out the major types of language errors that high school English. Through analyzing the possible reasons for errors learners made in writing, it helps English educators better understand the common weakness of senior high students. Besides, it also aims to provide new enlightenments and new ideas for high school English writing teaching.

As for the significance, this study offers some important insights into teaching practice. For English educators, it is hoped that this study contributes to a deeper understanding of learners' learning characteristics and psychology through the analysis of English learners' language errors in writing. So that, teachers can teach English writing more pointedly (mainly in grammar and vocabulary), and can correct students' errors more efficiently. For English learners, the results of this study provide a perspective to recognize their own fade zone of English knowledge, helping them improve learning methods and English writing competence.

The overall structure of the thesis takes the form of five parts.

The first part is the introduction, mainly concerned with the research background, purpose, significance, and layout of the essay. The second part begins by laying out the theoretical dimensions of the research, including the introduction of three theoretical bases: error analysis theory, comparative analysis theory, interlanguage theory, and then gives a review of previous studies on language errors in English writing. The third part deals with the research design, including research questions, research methods, research instruments, and procedures. The fourth part presents lots of examples of errors and analyzes the major types. It also discusses the main reasons for these errors. And the final part is the conclusion, summarizing the main findings of the study and providing some pedagogical implications and suggestions for further research concerning the limitations of the study.

2. Literature Review

At the beginning of the study, it is important to define and classify the error first. Then, the author will introduce three basic theories, respectively error analysis, comparative analysis, and interlanguage theory. Besides, this part also presents the related empirical research both at home and abroad.

2.1. Definitions of Error

Researchers vary from each other about the error definition. Dulay and Krashen (1982) elaborates that errors represent learners' level of the target language and considers them grammatical errors. Corder (1967) considers that errors are systematic, not self-correcting, and caused by the learners' insufficient knowledge of the target language. In this paper, the definition of "error" takes Corder's words. And in the same year, Corder also first classifies errors and mistakes. In his view, different with errors, mistakes are random, self-correcting, which are caused by carelessness, fatigue, or negligence. In fact, it is difficult to distinguish an error from a mistake except that there is evidence to show a learner

commits this error or mistake repeatedly (Shichun, 2004). Gui Shichun uses “shiwu (失误)” instead of “errors” or “mistakes” in his article. However, in English it is difficult to find such a word to incorporate both errors and mistakes. Therefore, in this thesis, the word “error” covers both errors and mistakes.

2.2. Classifications of Error

Linguists have different perspectives on the classifications of errors. Burt and Kipasky (1972) put forward two kinds of errors: local errors and global errors. Allen and Corder (1974) divided errors into pre-systematic errors, systematic errors, and post-systematic errors. Dulay and Krashen (1982) classified learners’ language errors into phonetic errors, lexical errors, grammatical morphological errors, and sentence or discourse errors. Domestic scholars Cai and Dai (2001) got three kinds of errors: cognitive errors, linguistic errors, and behavioral errors.

Therefore, error classification is never an easy task. This thesis adopts the error classification principle from Gui Shichun and Yang Huizhong (2003) in Chinese Learner English Corpus (CLEC). According to that, the error is classified into 11 broad types, that’s, the form of words (fm), word (wd), noun phrase (np), pronoun (pr), verb phrase (vp), adjective phrase (aj), adverb (ad), conjunction (cj), preposition phrase (pp), collection (cc) and sentence (sn). Each broad type has its several subclasses, with the successive number 1, 2, 3... to mark. For example, [wd4] means the fourth sub-type under the broad error classification of word. In total, there are 61 kinds of sub-types. The detail of the codes and their meanings can be seen in the following table.

Table 1. The CLEC annotation instructions

fm1	word	spelling	spelling, coinage, abbreviation, apostrophe
fm2	word	word building	derivation, inflection, compounding, plurality(noun), irregularity(verb), 3 rd person singular form(verb), syllabification, hyphenation, word division or fusion
fm3	word	capitalization	lower initial letter for upper initial letter or vice versa
vp1	vb phr	pattern	error in transitivity(<i>vi</i> as <i>vt</i> or vice versa), transitive verb pattern/grammatical (cf. <i>Oxford Advanced Learners’ dictionary of Current English</i> , six edition edited by A. S. Hornby)
vp2	vb phr	set phrase	phrasal verb and verbal phrase: error in form or use
vp3	vb phr	agreement	number agreement with its subject(noun or pronoun)
vp4	vb phr	finite/non-finite	finite verb for non-finite verb or vice versa
vp5	vb phr	non-finite	infinite error: form and use infinitive for particular or vice versa/-ed participle for -ing participle or vice versa
vp6	vb phr	tense	error in tense use within a sentence/ the sequence of tenses between sentences
vp7	vb phr	voice	error in the use of voice: active for passive or vice versa
vp8	vb phr	mood	error in the use of mood: imperative, subjunctive/ improper structure of conditional sentences
vp9	vb phr	modal/auxiliary	misuse of modal/auxiliary verbs/ wrong form of modal verb(or auxiliary verb) and verb combination(e. g tense form, voice form, etc.)
np1	nn phr	pattern	error in combination with other words/ grammatical
np2	nn phr	set phrase	omission or replacement of a fixed element that goes after a certain noun
np3	nn phr	agreement	number agreement of a noun, with its determiner or a word that refers to it
np4	nn phr	case	possessive case error: form or use
np5	nn phr	countability	uncountable noun used as countable noun

np6	nn phr	number	countable noun used with no determiner or -s/a or -s with plural noun
np7	nn phr	article	a/an confusion or definite /indefinite confusion
np8	nn phr	quantifiers	misuse or confusion between <i>many/much</i> , (a) <i>few/ (a) little</i> , <i>some/any</i> , etc
np9	nn phr	other determiners	misuse or confusion of demonstratives, wh-determiners, numerals, etc
pr1	pron	reference	incorrect/ ambiguous pronoun reference/ anaphoric
pr2	pron	anticipatory <i>it</i>	improper or wrong use of anticipatory <i>it/ it</i> replaced by a demonstrative, etc
pr3	pron	agreement	number agreement with a noun it refers to
pr4	pron	case	case error of any personal pronoun
pr5	pron	wh-	Misuse or confusion of interrogative, relative and conjunctive pronouns
pr6	pron	indefinite	misuse or confusion of indefinite pronouns such as <i>all/both</i> , <i>few/little</i> , <i>some/any</i> , <i>either/neither</i> , etc
aj1	adj	pattern	error in the combination with other words/ grammatical
aj2	adj	set phrase	error in the idiomatic use of an adjectival phrase/ omission or replacement of a fixed element that goes after a certain adjective
aj3	adj	degree	adjective degree error:form and use
aj4	adj	-ed/-ing confusion	-ed adjective for -ing adjective or vice versa
aj5	adj	predicative /attributive	predicative adjective used as attributive adjective
ad1	adv	order	improper adverb placement/ wrong position
ad2	adv	modification	adjective modifier used as verb modifier/ other kinds of confusion
ad3	adv	degree	adverb degree error: form and use
pp1	prep	pattern	unacceptable combination with other words/ grammatical
pp2	prep	set phrase	error in the formation or use of an idiomatic prepositional phrase

cj1	conj	pattern	unacceptable combination with other words/ grammatical
cj2	conj	set phrase	error in the formation or use of a phrase functioning as a conjunction
wd1	word	order	misplacement of any word other than an adverb
wd2	word	part of speech	error in part of speech: right root but wrong word class
wd3	word	substitution	error in word choice: right word class but wrong selection(any part of speech)
wd4	word	absence	omission of a word (any part of speech)
wd5	word	redundancy	oversuppliance of a word(any part of speech)
wd6	word	repetition	unnecessary repeating of a word
wd7	word	ambiguity	not clear word meaning/ semantic
cc1	collocation	n/n	improper noun (phrase) and noun (phrase) combination/ semantic
cc2	collocation	n/v	improper noun (phrase) and verb (phrase) combination/ semantic
cc3	collocation	v/n	improper verb and noun (phrase) combination/semantic
cc4	collocation	a/n	improper adjective and noun (phrase) combination/ semantic
cc5	collocation	v/ad	improper verb and adverb (or ad/v) combination/ semantic
cc6	collocation	ad/a	improper adverb and adjective combination/ semantic
sn1	sentence	run-on sentence	improper addition of clauses/ fused sentences
sn2	sentence	sentence fragment	subordinate clause as a sentence/ any phrase as a sentence
sn3	sentence	dangling modifier	illogical adverbial modification of a clause
sn4	sentence	illogical comparison	error in the comparison of words or phrases in a sentence which can not be compared
sn5	sentence	topic prominence	the co-occurrence of an initial noun phrase and its equivalent(usually a pronoun) in the same sentence
sn6	sentence	coordination	faulty parallelism of clauses(or words/ phrases) in a sentence
sn7	sentence	subordination	faulty attachment of a subordinate clause to the main clause
sn8	sentence	structural deficiency	error in the grammatical construction of a sentence: improper splitting, pattern shifting, confusing structure, etc
sn9	sentence	punctuation	overuse, absence, choice, apostrophe, comma splice, etc.

2.3. Theoretical Basis

The predecessor of error analysis is comparative analysis whose psychological basis had been doubted and criticized, so that error analysis and interlanguage research came into being. These three theories formed a solid and reliable theoretical basis of this study in turn.

2.3.1. Comparative Analysis

Based on behavioral psychology, comparative analysis theory prevailed in the 1950s and 1960s. Behaviorism theory holds that learning a foreign language means learning a new set of habits. In the process of acquisition, the old habits (mother tongue knowledge) will inevitably influence the learning of new habits (target language knowledge) (Weidong, 1994). This phenomenon of mutual influence between original knowledge and new knowledge learning is called transfer. Comparing the mother tongue with the target language, when the two language systems are similar, the learners will have a positive transfer in the process of second language acquisition. If there are many differences, negative transfers will occur.

2.3.2. Error Analysis

In 1967, Corder published a paper inspired by the error analysis (EA) of mother tongue acquisition, *The Significance of Learner's Errors*. In this book, Corder believes that learners' language errors reflect learners' current language system and are of positive significance to people's cognition of the learners' language development.



Figure 1. Five steps of error analysis

According to Corder, error analysis procedures have five steps: collect learners' materials; identify errors; classify errors; explain errors; evaluate errors. This paper will follow the five steps to study.

2.3.3. Interlanguage Theory

Interlanguage, as a formal language concept for the first time, was proposed by British linguist Selinker in his paper in 1969. Interlanguage refers to a transitional language constructed by second language learners between the mother tongue and the target language, which is in a state of constant change and gradually approaches the target language.

The features of interlanguage can be induced as four points: creative; unstable; incomplete; systematic. And Selinker (1972) believes that the following five factors play a major role in the formation of an interlanguage system: language transfer; transfer of training; strategy of the second language (SL) learning; strategy of SL communication; over-generalization.

2.4. A Review of Previous Studies on Writing Errors

In this part, the author will present a review of previous studies on writing error. According to the contents of the research, the author will make a discussion from three perspectives.

2.4.1. Research on the Proportion of Language Errors

According to Jin-In Seo(2014), "the number of interlanguage errors was more than for intralingual errors."

Mohammed Nuruzzaman (2018) analyzed the types, characteristics, and reasons for errors made by Saudi non-English majors in English fragment writing through a comparative study. He found that the amount of grammatical errors (different from lexical errors, semantic errors and mechanic errors) are always the largest.

2.4.2. Research on a Certain Kind of Language Errors

Hamzah (2012) analyzed the grammatical errors in freshmen's English writing, which were divided into fifteen types from serious errors to minor errors. Helen Forsyth (2014) conducted a study on syntactic errors in writing, pointing out the influence of negative language transfer on multilingual learners' language acquisition. Wang Rui (2015) made a study on lexical collocation errors. Li Yu, He Huaqing (2011) classified and analyzed verb phrase errors in compositions and discussed their causes. Xu Mengqing (2011) through qualitative research method, after investigating the sub-corpus ST3 of Chinese Learner English Corpus (CLEC), explained the wrong classification of the preposition in and the reasons. In terms of grammatical errors, Zhao Xiaodong and Liu Yang (2012) divided all temporal errors in the ST5 and ST6 sub-corpora of the Chinese Learner English Corpus into 16 categories and explained them.

2.4.3. Research on Common Errors in Chinese Learners' Writing

Wu Xiaoping (2014) analyzed the types and causes of English writing errors in the context of EFL and found that interlingual errors are more common than intralingual errors. He Huaqing (2009) conducted a study on language errors in English compositions of non-English major college Chinese students, finding that the lexical error including the word form error (fm) and the word error (wd) accounted for 59.33% of language errors. Jia Guanjie (2014) found that the highest frequency of error is the article error in master's theses for English majors in China, followed by incomplete sentences. Jiang Xiaowen (2014) found that local errors (different from global errors) predominated in Chinese senior high school students' ordinary and exam writing papers.

2.5. Summary

To sum up, the research on the pure theory of writing error analysis started earlier in foreign countries and later in China, while the number of empirical research abroad is less than that at home. Besides, most of the research participants fasten on college students and graduate students, while few of them are middle or high school students. A large proportion of research focus on a certain type of errors, such as vocabulary, grammar, collocation, etc., and a small number of which makes a general statistical study on the types and frequencies of errors made by students in English compositions. Therefore, this paper is going to make general research on the errors made by Chinese senior high school English learners in English practical writing. Through analyzing the major error types and conjecturing the deep reasons behind the specific errors, this paper aims to provide enlightenments and inspiration for English teachers to understand students' writing level and psychology and to innovate English writing teaching approaches.

3. Research Methodology

This part includes research questions, research methods, research instruments and research procures, explaining how this study was carried out.

3.1. Research Questions

Through the analysis of errors in senior high school students' English compositions, this paper aims to answer the following two questions:

- (1) What are the major types of language errors in senior high school English writings?
- (2) What are the possible causes of these language errors?

3.2. Research Instruments

(1) A small self-built corpus: It collects 96 compositions of the grade 12th from three senior high schools in Nanchong, Sichuan. The 96 compositions are tagged according to CLEC annotation instructions. The whole corpus contains 11129 words.

(2) BFSU PowerConc 1.0 (Xu, Jiajin, Liang & Jia, 2012): It is a retrieval software for corpus. With the student's composition corpus tagged with error types into this software, it can help more accurately count the number and frequency of different types of error.

3.3. Research Procedures

(1) Corpus collection: In two senior high school united examinations in Nanchong, Sichuan, 48 compositions of the Grade 12th were selected each time (16 compositions were from Nanchong High School; 16 from Baita Middle School; 16 from Nanchong No.1 Middle School) with the same topic, totally 96. The topics of the two compositions are respectively an invitation letter and a speech draft. The invitation letter asks students to write a letter to invite a foreign teacher Jack to attend the Mid-Autumn Festival party hosted by classmates. The speech draft is set in an International Forum on Environmental Protection for high school students to call on people to protect the environment. These two compositions are both practical style. The detailed information on the two topics can be seen in the following figure.

Topic one:
假如你是李华，中秋节来临之际，请代表你班同学给外教 Jack 写封信，邀请他参加你班组织的中秋晚会。要点如下：时间和地点；活动安排（品尝月饼、赏月）；休闲着装，词数 100 左右，可适当增加细节，以使行文连贯。

Topic two:
假如你是李华，你将代表学校参加国际高中生环境保护论坛，请你根据以下要点写一篇发言稿。1. 环境保护的重要性；2. 如何开展环境保护；3. 呼吁大家积极参与，词数 100 左右，可适当增加细节，以使行文连贯。

Figure 2. The two writing tasks

Then, here are two collected samples for the two topics especially for a more clear illustration.

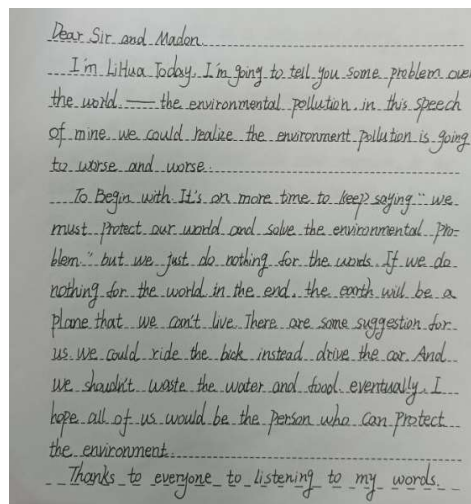
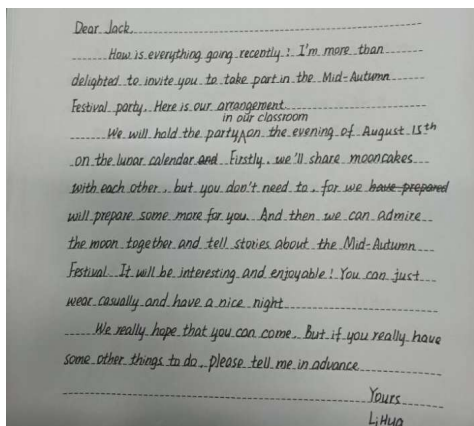


Figure 3. The two samples of writing tasks

(2) Corpus building: 96 compositions are entered in the form of text to constitute a small, written English writing corpus, possessing 11129 words. Each corpus document in the folders is encoded.

(3) Error annotating: A teacher and two students form an error tagging group, which annotates the corpus according to the Chinese learner English corpus error annotating method (Gui, Yang, 2003). It includes 11 broad classifications, 61 subtypes. The detailed classifications of errors are presented in these two figures.

A sample writing text of the invitation letter

Dear Jack,

As the Mid-Autumn Festival is approaching, I'm writing to invite you to take part in the party with us. As scheduled, the party is held on our students [fm2,-] own. It is supposed to talk [fm1,-] place at 9 am in our classroom on September fifteenth [fm1,-] in [pp1,-] the lunar calendar. The party mainly covers the two parts: have [vp4,-] a taste [wd4,-] the moon cakes and take [wd4,-] a look at [cc3,-] the moon. Besides, what makes the party appealing is to have a good time with us. [sn7,-] Surely, you will enjoy it. What's more, you'd better wear some casual clothes, which will be convenient for the activity [wd7,-].

What do you think of it? You [np4,-] presence is surely appreciated. Looking [vp5,-] forward to your reply.

Yours,
Li Hua

A sample writing text of the speech draft

Hello, everyone! I'm Li Hua. I will be represented [vp7,-] school visit [vp4,-] International [fm3,-] high Students [fm3,-] protect [wd2,-] environment theme [wd3,-]. I'm writing to tell you about [sn8] protect [vp4] [wd4,-] environment [wd4,-] very important.

First of all, we should have [np9,-] protect [wd2,-] awake [wd2]. Now, a number of water around [sn8,-] badly, and more [sn4,-] trees cut [vp7,-] down. Native [cc4,-] area [np6,-] decrease [vp6,-] [vp3,-]. Not only animals [wd4,-] homeless but also human being often life [wd2,-] [wd7,-], breath [wd2,-] bad air, lead to body bad. [sn2,-]

Then, we should protect home [np6,-]. Most volunteers take [vp6,-] part in activities, whom [cj1,-] growth [wd2,-] [cc3,-] trees and pick up [vp6] [fm1,-] [wd7,-] water [wd7,-] rubbish. And lost water, electrical [sn2,-] Don't waste Food, go out should on feet [np2,-] or by cycle [sn2]

Finally, I hope everyone join volunteer [np6,-]. [sn9,-] protect [vp4,-] [wd4,-] environment [sn2,-]

Figure 4. The detail annotation interface

I am Li Hua, who is a high-school [fm2,-] student from XinXin school. It would be my honor to be here to share with you my opinions on how to protect the environment.

As is known to everybody, our environment would be worse and worse if we don't take some measures. So our task is great [wd3,-] and the time which [cj1,-] we can do something is less and less.

There are some ways [wd4,-] how to protect the environment. First of all, the most important thing is what [cj1,-] people should have a mind to protect the environment. On the other hand, we should make people know some basic ways to protect the environment, like using used water and no wasting.

Finally, I sincerely invite you to join [cc3,-] the action to protect the environment. And I believe we can do a good job.

Thank you and [pp2,-] your time.

Figure 5. A composition sample after error annotation

The whole process of errors annotation takes three rounds. After comparing the annotation results of three rounds and combined with the examples of "Juku Correction Network", the final result is obtained. In this way, the results reduce as much as possible the inaccuracy caused by human subjectivity.

And, the collected corpus data is simplistically annotated with error types, just like [fm1,-]. The following picture is one corpus document sample after error annotation.

(4) Error classification: use the corpus software (BFSU PowerConc 1.0) to retrieve the target label, count the frequency proportion of each error, and further analyze and classify it. For example, input and retrieve “wd4” (the omission of a word), you can get the result as the following picture.

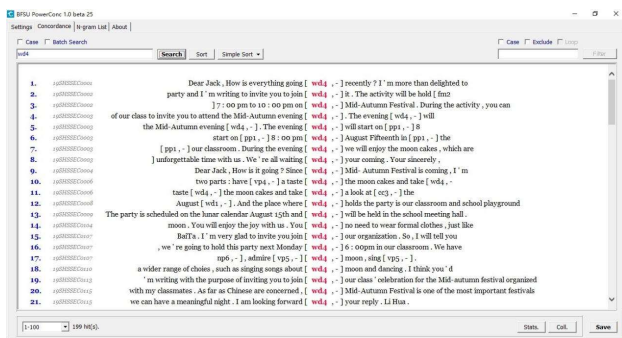


Figure 6. Retrieval result of “wd4”

Error attribution: Consult the literature, communicate with the teachers in Grade 12th, and analyze the possible reasons for different errors made by students in their compositions.

4. Results and Discussion

This part presents some specific error examples of various reasons from collected samples. Through the correction of them, the author tries to summarize the possible reasons for the occurrence of these errors.

4.1. Error Types and Frequency Distribution

In this research, the author elicits the topic requirements in a clear and obvious way, so most students can easily distinguish the writing genre which is practical writing from the given context. Based on this, discourse errors in this research will not be taken account into. This thesis mainly focuses on the language errors appearing in high students' English compositions.

By the corpus software tool, BFSU PowerConc 1.0, a total of 1085 errors have been counted. The error rate is about 9.75%. On average, there are 11 errors in one composition. The following table shows eleven broad error types and frequency distribution.

Table 2. Error Types and Frequency Distribution

Types	Frequency	Proportion
Word errors (wd)	370	34.1%
Errors of verb phrase (vp)	166	15.3%
Word form errors (fm)	165	15.2%
Errors of collocation (cc)	116	10.7%
Errors of sentence (sn)	88	8.1%
Errors of preposition (pp)	64	5.9%
Errors of verb phrase (np)	59	5.4%
Errors of conjunction (cj)	17	1.6%
Errors of pronoun (pr)	17	1.5%
Errors of adjective (aj)	14	1.3%
Errors of adverb (ad)	9	0.8%
Total	1085	100%

The following pie chart illustrates the error types and

frequency distribution more clearly.

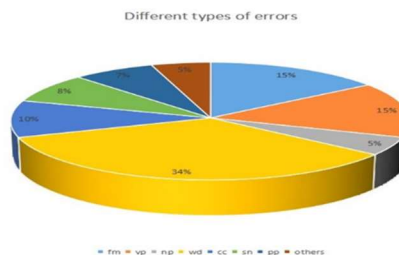


Figure 7. Different types of errors

From the statistics of the table and pie chart above, it can be seen that among the 11 broad error types, the top four with the largest proportion of errors respectively are the word errors (wd), accounting for about 34.1%, the errors of the verb phrase (vp), accounting for about 15.3%, the word form errors (fm), accounting for about 15.2%, and the errors of collocation (cc), accounting for about 10.7%. As mentioned above, each broad type includes several sub-types. Further statistics shows that among these sub-types, the error type according to the frequencies from high to low are word missing error (wd4), 199 times; spelling errors (fm1), 91 times; verb/noun collocation errors (cc3), 86 times; case errors (fm3), 59 times; part-of-speech selection errors (wd2), 57 times; verb tense errors (vp6), 56 times; wrong prepositions (pp1), 53 times. Other sub-types possess the frequencies lower than 50 times. They are errors within a sentence component or a clause in a compound sentence, namely, local errors according to Burt & Kipasky (1972). The data indicates local errors account for 75.3%.

He Huaqing (2012) has made a study about writing errors of college students, finding that the error types in descending order of frequency are word errors (wd), word form errors (fm), errors of verb phrase (vp), errors of sentence (sn), errors of noun phrases (np), errors of collocation (cc) and other errors. From the results, this study also supports her research and views. Alzamil (2020) also found that “50 percent of the errors fell into three categories (capitalization, spelling, and article use)”. It confirms that the English learners' common error types in writing focus on the word form error (fm) and the word error (wd).

In this paper, the author selects these four error classifications as major language errors of high students to demonstrate. Here is the detailed analysis with specific examples and their corrections.

4.1.1. The Word Errors

This error category refers to the word errors. It is the most numerous, including seven subgroups: word order (wd1); part of speech (wd2); substitution (wd3); absence (wd4); redundancy (wd5); repetition (wd6); ambiguity (wd7), as the table below illustrates:

Table 3. The Word Errors and Frequency Distribution

Types	Frequency
Omission of a word (wd4)	199
Errors in part of speech (wd2)	57
Oversuppliance of a word (wd5)	42
Substitution (wd3)	30
Ambiguity (wd7)	27
Word order (wd1)	13
Unnecessary repeating of a word (wd6)	2
Total	370

It can be seen that absence (wd4) accounts for over half of the total seven subgroups, so the author will focus on the analysis of it. “Wd4” refers to the omission of a word and this word can be any part of speech. After retrieving in PowerConc 1.0, 199 results were obtained. Observing these 199 items, the author found that the omission of articles, prepositions, nouns/pronouns, and verbs consist of a big chunk. Here are some examples:

Absence of articles

E.g1: It must be [wd4, -] fun time for...

It must be a fun time for... (√)

g2: Since [wd4, -] Mid- Autumn Festival is coming, ...

Since the Mid-Autumn Festival is coming, ... (√)

E.g3: What is [wd4, -] most important is to teach people how...

What is the most important is to teach people how... (√)

E.g4: ...hope more and more people like [wd4, -] earth...

...hope more and more people like the earth... (√)

In E.g1, when used to express “a period of time”, “time” should be expressed as “a time”. In E.g2 and E.g3, a definite article “the” must be added before Chinese traditional festivals and superlative adjectives. For planets, the sun, the moon, and the earth must contain “the”. So E.g4 lacks a “the” before “earth”.

Absence of prepositions

E.g5: ...to hold this party next Monday [wd4, -] 6: 00 pm in our classroom.

...to hold this party at 6: 00 pm next Monday in our classroom. (√)

E.g6: I am looking forward [wd4, -] your reply.

I am looking forward to your reply. (√)

E.g7: ...this letter to invite you [wd4, -] attend this Mid-Autumn Festival performance.

...this letter to invite you to attend this Mid-Autumn Festival performance. (√)

In E.g5, 6,7, they all miss a preposition. The exact time is preceded by the preposition “at”; “look forward to something” and “invite somebody to do something” are set phrases.

Absence of nouns/pronouns

E.g8: ...of our class to invite you to attend the Mid-Autumn evening [wd4, -].

...of our class to invite you to attend the Mid-Autumn evening party. (√)

E.g9: I’d appreciate it if you could take my tips into account and I hope [wd4, -] will make a difference to...

I’d appreciate it if you could take my tips into account and I hope it will make a difference to... (√)

In E.g8, “attend” should be collected with a party or a meeting but not “evening”, so it should add a “party” after “evening”. In E.g9, the subordinate clause after “hope” lacks a subject, so the author adds an “it” to serve as the subject.

Absence of verbs

E.g10: You [wd4, -] no need to wear formal clothes.

You have no need to wear formal clothes. (√)

E.g11: ...the Mid-Autumn Festival party, that [wd4, -] organized by us.

...the Mid-Autumn Festival party, that is organized by us. (√)

These last two examples both lack predicate verbs. The sentences have faulty construction.

These absences of words mainly indicate that students have a weak English grammar foundation (E.g10, 11), and they are unfamiliar with the usage of set phrases and collocations (E.g5, 6, 7), certain grammatical rules (E.g1, 2, 3, 4) in

English. Students tend to write English compositions in Chinese thinking mode (E.g8, 9).

In addition to the absence of a word (wd4), there are other subclasses of word errors. Some instances are as follows:

Word order (wd1) :

E.g12: ...I suggest that we should keep in mind the protection [wd1, -] and do some things.

...I suggest that we should keep the protection in mind and do some things. (√)

Part of speech (wd2):

E.g13: ...so we have enough time to celebration [wd2, -] the Festival [fm3, -].

...so we have enough time to celebrate the festival. (√)

Substitution (wd3):

E.g14: I sincerely appear [wd3, -] to everyone.

I sincerely appeal to everyone. (√)

Redundancy (wd5):

E.g15: By the way, you needn’t dressed in [wd5, -] formally.

By the way, you needn’t dressed formally. (√)

Repetition (wd6):

E.g16: ...something that I provide [wd5, -] way maby [fm1, -] [wd6, -] can provide [cc3, -][wd6,-] help in [pp1, -] it.

...something that I provide may offer help for it. (√)

Ambiguity (wd7):

E.g17: I appeal [wd4, -] everyone to know it [wd7, -] and do it [wd7, -].

I appeal to everyone to know the importance of protecting the environment and do action. (√)

4.1.2. The Errors of Verb Phrase

The errors of verb phrase contain seven subclasses, which are: verb agreement (vp3); finite/non-finite (vp4); non-finite (vp5); tense (vp6); voice (vp7); mood (vp8); modal/auxiliary (vp9).

Table 4. The VP Errors and Frequency Distribution

Types	Frequency
Verb tense (vp6)	56
Finite verb for non-finite verb or vice versa (vp4)	49
Infinitive error (vp5)	28
Verb agreement (vp3)	11
Modal/ auxiliary (vp9)	10
Verb voice (vp7)	6
Verb mood (vp8)	6
Total	166

As is shown in the table, tense error (vp6) and finite/non-finite error (vp4) account for a large portion. Errors of “vp4” refer that a non-finite verb is used in a finite form or vice versa. Tense error (vp6) means error in tense use within a sentence/ the sequence of tenses between sentences. Here are some examples of the two subclass error types:

Finite/non-finite error:

E.g1: ...we’ll have supper together allow [vp4, -] us to catch up...

...we’ll have supper together allowing us to catch up... (√)

E.g2: My speech theme is protect [vp4, -] the environment.

My speech theme is to protect the environment. (√)

Tense error:

E.g3: In addition to these traditional activities, we [vp6, -] have a wider range of choices...

In addition to these traditional activities, we will have a wider range of choices... (√)

E.g4: My mum says [vp6, -] “The moon cake is round, so it marks to get together with your families”.

My mum said, “The moon cake is round, so it marks to get together with your families”. (√)

In E.g1, the phrase leading by “allow” acts to modify “a supper”, so “allow” should be used in the non-infinite form, that’s, “allow” should be used in the non-infinite form, “allowing”. In E.g2, “protect the environment” serves as predicative following the copula “is”, so “protect” should be changed into “to protect”. In E.g3, the sentence is talking about a scheduled plan, so the verb should be used as future tense. In E.g4, the saying of Mum happened when “I” was a child, so it should use past tense.

From the errors listed, it can be told that high students still are under the influence of Chinese, unconscious of the importance of verb tense. And they often ignore a vital principle in English syntax, that’s, every English sentence has and only has one predicate verb. Any other verbs which not serve as the predicate verb should exist in a non-infinite form (to do /-ing/-ed).

Beyond those, the following instances represent other types of errors in the verb phrase.

Verb agreement (vp3):

E.g5: All in all, everyone in this lively planet are [vp3,-] supposed to pull together to...

All in all, everyone in this lively planet is supposed to pull together to... (√)

Non-infinite (vp5):

E.g6: What I [wd4,-] most delighted about was eating [vp5,-] mooncake [np6,-] in [wd4,-] Mid-autume [fm1,-] [fm3,-] Festival, when I was a child.

What I was most delighted about was to eat moon cakes in the Mid-Autumn Festival, when I was a child. (√)

Voice (vp7):

E.g7: I will be represented [vp7, -] school...

I will represent the school... (√)

Mood (vp8):

E.g8: It is also high time that we humans took [vp8, -] proper measures to...

It is also high time that we humans take proper measures to... (√)

Modal/auxiliary (vp9):

E.g9: Be relaxing, you mustn't [vp9, -] wear clothes which...

Be relaxing, you needn't wear clothes which... (√)

4.1.3. The Word Form Errors

The third numerous error type is the word form error. It includes three subclasses: the error of spelling (fm1), word building (fm2), and capitalization (fm3).

Table 5. The Lexical Errors and Frequency Distribution

Types	Frequency
Spelling error (fm1)	91
Capitalization error(fm3)	59
Word building error (fm2)	15
Total	165

Among these three sub-types, the spelling errors are the largest proportion. Actually, the great majority of misspelled words should probably be defined as mistakes, caused by slips of the pen. Students can self-correct them if teachers point the mistakes out. However, if it cannot be modified by students

themselves, or is made by amounts of students in similar ways, one misspelled word is exactly an error. Therefore, in this paper, it still is zoned as errors. Observing the misspelled words, in most cases, they are caused by students’ defects in morphology. Here are some examples of spelling errors:

E.g1: ...hoping to have an unforgettable [fm1, -] time with you.

...hoping to have an unforgettable time with you. (√)

E.g2: ...Festival is on the fiftieth [fm1, -] day of...

...Festival is on the fifteenth day of ... (√)

In E.g1, the student wants to write “unforgettable”. This word’s root is “forget”, with a prefix “un-” to express the negative meaning and a suffix “-able” meaning “being able to”. In the meantime, “t” should be diploid to fit the morphological rules. In E.g2, the student confuses the word “fifteenth” with “fiftieth”. Such error is common for many English beginners, but it shouldn’t be for a high student.

The errors in the capitalization of words also reflect students’ poor foundation in the specification of English writing. In English, the initial letter of festivals should be capitalized. Here is an example:

E.g3: The Mid-autumn [fm3, -] Festival is coming.

The Mid-Autumn Festival is coming. (√)

Besides the two kinds of errors, the errors of word-building also constitute word form errors. Through observation, the author finds that it usually happens when a verb have a particular past tense form. Look at this example:

E.g4: The activity will be hold [fm2, -] in our classroom...

The activity will be held in our classroom...

4.1.4. The Errors of Collocation

The collocation errors are also very significant, consisting of six types of subclass: n/n (cc1); n/v (cc2); v/n (cc3); a/n (cc4); v/ad (cc5); ad/a (cc6). Their respective frequency distributions are shown in the table below.

Table 6. Collocation Errors Types and Frequency Distribution

Types	Frequency
v/n collocation (cc3)	86
n/n collocation (cc1)	11
a/n collocation (cc4)	8
n/v collocation (cc2)	5
v/ad collocation (cc5)	5
n/n collocation (cc1)	1
Total	116

From table 6, it can be seen that the third type (cc3) is the most frequent error type. This type refers to the improper verb and noun (phrase) combination/ semantic. Here are some examples:

E.g1: ...and take a look at [cc3, -] the moon.

...and admire the moon. (√)

E.g2: And the clothes isn't controlled [cc3, -]...

And the dressing isn't restricted ... (√)

E.g3: If you want to come, just dress [cc3, -] casual clothes.

If you want to come, just wear casual clothes. (√)

In E.g1, “look at” just means to watch, but not admire or appreciate the moon. In E.g2, the writer is to express that the types of clothes are not restricted. But the original sentence expression is not so authentic. In E.g3, when followed by something but not somebody, the word “dress” is an intransitive verb. So it should be changed to “wear”.

Examples of other subclass errors are as follows:

n/n (cc1):

E.g4: It's a traditional festival, which plays an important part [cc1, -] to [pp1, -] all of us Chinese.

It's a traditional festival, which plays an important part in all of us Chinese life. (√)

n/v (cc2):

E.g5: I would like to express my things [cc2, -].

I would like to express my opinions. (√)

a/n (cc4):

E.g6: To begin with, we should do it from little [cc4, -] things...

To begin with, we should do it from small things... (√)

v/ad (cc5):

E.g7: ... (litter) and never throw it away [cc5, -]...

... (litter) and never throw it about... (√)

These instances show that students tend to ignore the fixity of some collocations and take wrong collocations for granted.

4.1.5. Other Types of Errors

Apart from the above error types, there still exist other classifications of errors, though they are not in the same large portion as the four broad categories mentioned above.

Here are some examples selected containing error types that are different from the above-mentioned classifications.

E.g1: Above all [sn3,-], join us.

Finally, join us. (√)

E.g2: ...will start on [pp1, -] 8:00 pm.

...will start at 8:00 pm. (√)

E.g3: There are many different types of pollutions [np5,-] in this world.

There are many different types of pollution in this world. (√)

E.g4: First of all, the most important thing is what [cj1,-] people should have a mind to protect the environment...

First of all, the most important thing is that people should have a mind to protect the environment... (√)

E.g5: In the end, I hope everyone will take part in it [pr1,-]...

In the end, I hope everyone will take part in this activity... (√)

E.g6: ...and appreciate the most bright [aj3, -] moon.

...and appreciate the brightest moon. (√)

E.g7: Next, you would [vp9,-] dress casually [ad2,-] clothes.

Next, you can dress casual clothes. (√)

In the next part, the author is going to analyze the main factors in the occurrence of language errors by searching the literature, consulting with the front-line teachers, and reviewing her own learning experiences.

4.2. Attribution for Language Errors in Writing

Brown (2001) once divided the causes of errors into intralingual transfer, interlingual transfer, and other factors. Therefore, according to the research results of Brown and the data analysis results of this study, the author sums up the causes of errors as interlingual transfer, intralingual transfer, and learning habits.

4.2.1. Interlingual Transfer

In the process of studying composition errors, the author finds that the language errors in writing, especially vocabulary and grammar errors are mainly due to the influence of the mother tongue, that is, interlingual transfer. Take an error as an example, "The Mid-Autumn Festival with

a long history in china [sn2,-]." From a semantic perspective, it has no problem, while in syntax, it lacks a predicate verb. In Chinese, one syntactically right sentence can have no predicate verbs, such as "中秋节历史悠久". However, every syntactically right English sentence must contain one verb. This "sentence" uses with as a predicate verb, but with only is a preposition. Besides, in English, there are many ways to express the meaning of "拥有" in Chinese, such as "have(v.)" "with(prep.)" "possession(n.) of". When using the latter two words, a predicate verb must be added in the sentence, to satisfy the need for syntactic correctness. During the practical writing process, high students are usually apt to directly transfer Chinese expression schema to English, which will smother good English compositions.

4.2.2. Intralingual Transfer

English learners' errors in writing are not all caused by mother tongue interference, but by the negative transfer of the target language, that is, intralingual transfer errors. According to the observation and analysis, the errors in the spelling of words, confusion of parts of speech, incorrect use of verbs, conjunctions, and nouns are mainly caused by the intralingual transfer. And high students usually confuse the use of words with similar meanings. Actually, every word has its specific usage scenarios, especially some nouns, verbs, and adjectives. For example, "Secondly, we will assign [cc3,-] some activities..." "Assign" means "giving others the work to do". In this context, it should be changed into "arrange", which means "planning".

4.2.3. Students' Bad Learning Habits

According to the statistical results of writing data, we can see that quite a number of students have bad learning habits, including memorizing English words by rote, ignoring the basic grammatical structures, and not applying what have learned to practice. Actually, many English words are formed in a certain morphological rule, especially many adjectives. But quantities of misspelled words indicate the loss of these rules by high students, such as "unforgettable" "Honourable" "vailable". Besides, many errors in syntax show high students have a weak foundation of the grammatical structure: every single English sentence must have and only has one predicate verb. Here is a negative example, "And lost water, electrical. [sn2,-]". And many errors about capitalization are because few students have a strong sensitivity to standard writing in English. For instance, "this saturday" "chinese traditional festival" "the Mid-autumn Festival".

5. Conclusion

This part makes a conclusion about the study, elaborating on the main findings of the study. Based on that, the author puts forth several enlightenments for English writing teaching. At last, some suggestions are provided for further research.

According to the statistics, it can be seen that high students' errors have numerous types, 11 broad types, 56 sub-types; errors are mostly local errors; errors mainly focus on word (wd), verb phrase (vp), word form (fm), collocation (cc). This statistics indicate that senior high school students' English writing proficiency is still keeping a relatively low level in syntax, morphology, collocation, diction, and thinking mode.

Through the concrete analysis of errors, three reasons for errors are preliminary summarized: intralingual transfer, interference from the target language, interlingual transfer, interference from the mother tongue, and students' bad learning habits.

The analysis of high school students' language errors in writing shows that in order to reduce the number of errors in high school students' English writing and to improve students' writing ability, teachers need to take some effective teaching measures.

Firstly, teachers are supposed to pay attention to the differences in expression between English and Chinese, including diction, word order, the degree of context dependence, discourse style and so on. Actually, most interlingual errors are caused by this neglecting, including the omission of articles (wd4), and possessive case forms (np4/pr4); additional prepositions (pp2); and the misformation of prepositions (pp1), verb tenses (vp6), word choice (wd3), and so on. Therefore, teachers should guide students to regularly use English-Chinese dictionaries, to read a large number of authentic local English materials, to watch English movies, etc., so that students can increase the amount of native language input. In this way, students can be less affected by negative interlingual transfer and improve the language accuracy.

Secondly, teachers should improve students' English learning habits and help them master some basic English rules. These habits cover the ways of memorizing, understanding syntax rules, cognition, etc. For example, when asking students to memory the dictionaries, teachers can teach them to use syllables to memorize words. And also, teachers should ask students to check the composition to reduce the occurrence of tiny mistakes such as punctuation (sn9) or capitalization (fm3). Besides, to reduce the global errors, teachers can instruct students that every single English sentence only has one predicate verb.

Thirdly, teachers ought to adopt corresponding error strategies to different students and different errors. In other words, taking students' overall language level as a reference, teachers can take tolerance policy or strict policy. For some errors that are beyond students' current language level, teachers need to employ the tolerant policy and let them go temporarily. For those obvious and important errors especially vocabulary errors or syntactical errors, teachers should strictly control and correct them immediately. Besides, teachers should also take every student as an individual. For those students with high language literacy and strong writing skills, teachers can be relatively strict with errors they make in compositions; while for those students with weak English writing capacity, teachers can adopt tolerant policy on a presumption that these students have made some progress.

There are many limitations in this research. Firstly, the author's research participants are senior three students of three high schools in Nanchong City, which cannot reflect high students' conditions in a wide range. Secondly, it is difficult for the author to grasp the students' psychological state in writing in the process of analyzing. Some students made errors or mistakes, probably not because of the lack of knowledge, but psychological factors, such as carelessness and lack of time, so the difference between mistakes and errors can't be clearly displayed. Therefore, the results of erroneous attribution may be biased.

Concerning the limitations of the present study mentioned above, the author provides two according suggestions for further research. First of all, increase the sample size. Senior high school students, as research participants, can be from three grades: Grade Ten, Grade Eleven, and Grade Twelve, which can better and more generally represent the senior high school students' average level of English writing. Besides, the

attribution of language errors in English writing should combine with a questionnaire or case studies about high students, so that the error attribution can be more reliable and valid.

References

- [1] Allen, J., & Corder, S. (1974). *The Edinburgh Course in Applied Linguistics*. Oxford University Press.
- [2] Alzamil, A. (2020). An investigation of writing errors made by Saudi English-major students. *International Journal of English Linguistics*, 10 (2), 92-97.
- [3] Amp, C. L. , & Weidong, D. . (2001). The integration of error classification. *Foreign Language World*.
- [4] Brown, H. (2001). *Principle of language learning and teaching* (3rd Edition). Foreign Language Teaching and Research Press.
- [5] Burt, M. K., & Kipasky, C. (1972). *Global and local mistakes*. Second Language Learning, Newbury House.
- [6] Corder, S. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5 (1-4), 161-170.
- [7] Dai, W. D. (1994). Some issues in contrastive analysis, error analysis and interlanguage study. *Foreign Language World*. (5):1-7.
- [8] Dulay, H., Burt, M., & Krashen, S. (1982). *Language Two*. Oxford University Press.
- [9] Forsyth, H. (2014). The influence of L2 transfer on L3 English written production in a bilingual German/Italian population: A study of syntactic errors. *Open Journal of Modern Linguistics*, (4), 429-456.
- [10] Gui S. C. (2004). A corpus-based cognitive model of Chinese EFL Learners' error analysis. *Modern Foreign Languages*. 27(2):129-216.
- [11] Gui S. C., & Yang H.Z. (2003). *Chinese Learners English corpus*. Shanghai Foreign Language Education Press.
- [12] Hamzah. (2012). An analysis of the writing grammatical errors produced by freshmen students in English writing. *Lingua*, 6 (1), 17-25.
- [13] He, H. Q. (2009). An analysis of lexical errors in non-English-majors' writing: A corpus-based study. *Foreign Language World*, 3, 2-9.
- [14] He, H. Q., & Chen, W. C. (2012). A corpus-based study on language errors in non-English-majors' writing. *Journal of China West Normal University*, 3, 97-101.
- [15] Jia, G., & Qiao, W. (2014). Study of language errors in the MA theses of English majors. *Foreign Language World*, 3, 63-69.
- [16] Jiang X. W. (2014). The study the error analysis and strategies in English writing of senior high school students. Northwest Forestry University.
- [17] Li Y., He H.Q. (2011). Error analysis of verb phrases in college English writing: A corpus-based study. *Journal of Southwest Jiaotong University(Social Sciences)*. 9 (5): 102-105
- [18] Ministry of Education of the People's Republic of China. (2017). *English curriculum standards for senior high schools*. People's Education Press.
- [19] Nuruzzaman, M. (2018). An analysis of errors committed by Saudi non-English major students in the English paragraph writing: A study of comparisons. *Advances in Language and Literary Studies*, 9 (1), 31-39.
- [20] Selinker, L. (1972). *Interlanguage*. Julius GroosVerlag.

- [21] Seo, J.-I. (2014). The impact of cognitive style on errors in English writing through extensive reading: A case study. *English Teaching*, 69 (2), 175-198.
- [22] Wang, R. (2015). A Corpus-based Study on Verb-noun Collocation Errors by Chinese Non-English Majors from the Perspective of Conceptual Transfer. *Foreign Language Research*, 3, 120-124.
- [23] Wu Xiaoping, E. V. (2014). Types and attributes of English writing errors in the EFL context - A study of error analysis. *Journal of Language Teaching and Research*, 5 (6), 1256-1262.
- [24] Xu M.Q. (2011). A corpus-based study by error analysis of prepositions In. *Journal of Ningxia University (Social science)*. 33 (6): 181-184.
- [25] Zhao X.D. , Liu Y. (2012). A corpus-based study on tense errors in English majors' writing. *Journal of Liaoning Normal University (Social science)*. 35 (4):540-545.