

A Corpus-based Study on Definite Article Errors in English Writing

Yuan Yang¹, Huaqing He²

¹ Pan Xi Mian Yang Nan Shan International School, Xichang, Sichuan, China

² College of Foreign Language Education, China West Normal University, Nanchong, Sichuan, China

²Corresponding Author:

Huaqing He

E-mail: masha4567@sina.com

Abstract

Based on corpus approach, the author adopted error analysis, used the sub-corpus St2 of Chinese Learner English Corpus (CLEC), and made a careful study on the misuse of definite article. The following are the findings: (1) There are three categories of misuse of definite article which are omission of the, redundancy of the, indefinite article and definite article confusion. (2) The highest frequency error type is absence of the, the next is redundancy of the. (3) The causes are L1 transfer, the related factors of second language and the article semantics. The causes are concluded from interview on the basis of interlingual and intralingual sources. At last, the results contribute to the English writing teaching.

Key Words: Error Analysis; Corpus; Definite Article

1. Introduction

As everyone knows that one can not learn any language without committing errors, as Dulay and Burt(1974:95) put it” You cannot learn without goofing.” Therefore many linguists and foreign language teachers pay much attention to the errors committed by second language learners, which contribute to the development of Error Analysis. Error Analysis focuses on learners’ inner psychological factors, and supplies a scientific methodology for second language acquisition. It is learner-centered. What’s more, it has changed the version which took errors as pathology to be eradicated and has built up the point that learners’ errors are evidence of learning stages and strategies. Furthermore, there is no article system in Chinese, coupled with the complexity of English article make it much harder for Chinese English learners to use definite article correctly.

2. Literature Review

The definite article has attracted many linguistics from abroad and home for its complex usage. Both oversea and Chinese researchers make contributions to learning and teaching definite article by doing different studies.

Dulay and Burt (1974) regarded articles as morphemes and studied together with the other morphemes to analyze the acquisition order. Huebner (1983) and Pica(1985) used various methods to describe the acquisition order of articles. Mitsuharu Mizuno(1999) made an analysis on the 20 Japanese English learners to study the errors of article usage and causes. Master (1997) and Daniel Robertson (2000) were all analyzing the features of articles acquisition of learners on the basis of theories of interlanguage. Zegarac (2004) pointed out the acquisition of articles not only affected by the pragmatic principle in native language, but individual learners can form their own special assumptions and strategies to use articles variably. Trenkic (2007, 2008) made a further and detailed study on features of article, based on pragmatic principles and strategies.

The empirical study on English articles in domestic is relatively later than those in foreign country. Until 2001, Li Jingquan and Cai Jinting, published the first paper of empirical research on articles, which discussed the misuse of articles on Chinese English learning based on corpus, found the diversity and regularity of misuse of articles. What's more, they divided the the misuse articles into three patterns: omission of articles, redundancy of articles and confusion of articles. Thus, the author classified the errors of definite article in senior high students' writing on the basis of article classification method of Li and Cai , but the author combines the interview and interlanguage to analyze the underlying the cause of the errors.

3. Methodology

Inspired by the former scholars, the author will unit quantitative method with qualitative method to carry out this study. That is a corpus-based error analysis and interview to analyze the usage of definite article. Therefore, in this part, the author will discuss the research design of this study, containing research questions, corpus selections, research instruments, data collection and data classification.

3.1 Research Questions

The author put forward two questions on the basis of purpose of this research.

1. What kinds of errors do the Chinese senior high Students commit when they use the definite article?
2. What are the possible causes that affect the Chinese senior high school students' misuse of definite article?

3.2 Corpus Selection

The CLEC project was compiled by Professor Gui Shichun of Guangdong University of Foreign Studies, and Professor Yang Huizhong of Shanghai Jiaotong Univeristy. It is an important part of Chinese National Social Sciences and Humanities. CLEC is a tagged corpus where the errors of learners are marked and it contains more than one million words. In general, the subjects of CLEC are Chinese English learners, while the subjects come from different phases, that include senior high school students, non-English majors and English majors which stand for five sub-corpus St2, St3, St4, St5 and St6.

3.3 Research Instrument

Since this present study is based on sub-corpora St2 of CLEC. Therefore, the author chose AntConc3.2.1 as the available corpus software, which the key terms are indexed, and the definite article are marked and numbered. The second research tool is excel which is used to calculate the errors of definite article.

In order to have a better understanding to the underlying causes of misuse of definite article, the author conducted an interview for this present. The participants are 25 senior high school students who were randomly selected in Nan Bu Senior High School.

3.4 Data Collection

In this study, subjects are sub-corpus St2 of CLEC, which includes more than 1000 writings written by senior high school students. Basing on the research question of this present study, the data are selected as follow.

Because the error tags for misuse of definite article are np1(error in combination with other words/grammatical), np7(confusion a/an with definite article), wd4(absence of a word), wd5 (oversuppliance of a word). Then, the author will locate the concordance errors which tagged with np1, np7, wd4, wd5 at the aid of AntCon3.2 sub-corpus St2 of CLEC.

3.5 Data Classification

The author made use of the software AntCon3.2 to search and index the 208088 words, and found that the misuse of definite article could be classified into three types: The first is omission of definite article(OOA) that indicates the situation where native English speakers or writers will use definite article, while Chinese English learners omit it. The second is redundancy of definite article (OSA). It stands for the place where Chinese English Students use definite article, but the native English will not. The last category is confusion definite article with indefinite article (ATC), which means Chinese English Students are not able to make a correct judgment between definite article with indefinite article when they are writing. The next table is about the total number of OOA, OSA, ATC.

Table2. The number and percentage of misuse of definite article.

Type	number	percentage
OOA	238	52.8%
OSA	171	37.9%
ATC	42	9.3%
Total	451	100%

4 . Results and Discussion

In this part, the results will be presented and then the writer will describe the errors with some examples.

4.1 Three Kinds of Misuse of *the* in St2 of CLEC

The author has yielded there are three primary types error of definite article in chapter three. That is omission of *the*, redundancy of *the*, confusion *the* with *a/an*. The author will discuss them one by one with examples.

4.1.1 Omission of *the*

Omission of *the* occupies more than half percentage in the total errors of definite article, and it appears in the following situations:

- **Omission of *the* before adjective of comparative forms, superlative forms and ordinal number.**

Example 1. at [fm3,-] 7:30, we have [wd4,s-] first lesson.

Example 2. it [fm3,-s] has [wd4,2-4] largest number of people in the world.

Example 3. [wd4 , -s] more you stand , [w d4 , -s] far you see.

The examples are the most usual phenomena in the senior high school students' compositions. According to these examples, we can see that the examples 1 omit *the* before ordinal number, and examples 2 is omission definite article before adjective of superlative degree. The example 3 omit *the* before adjective of comparative degree.

- **Omission of *the* before fixed collations, and specific terms**

Example 1. Then we went to [wd4,0-2] North sea Park to go boating

Example 2. When you go into in [wd4,1-1]early morning

Those 2 examples are common situations where students omit definite article before fixed collocations and specific terms. The usage of definite article on fixed collocations is easy to grasp, then absence of definite article at this situation indicates that students do not have a good master of definite article.

- **Omission of *the* in the case of naphoric reference and fixed reference in the context**

Example 1. Mu Zhong was [wd4,s-] emperor of Tang period

Example2. After [wd4,1-1] match, [wd4,-4] head of our class praised [fm1,-] us for it

The definite article is used when one thing or one person was mentioned before. However, the author observed and analyzed sub-corpus St2 of CLEC, found that a great many of students eave out *the*, as they refer to the things or persons which were mentioned before.

4.1.2 Redundancy of *the*

Redundancy of *the* occupies a major percentage of the misuse of definite article, thus it is more complicate and confusing. The writer classified the following categories.

- **Redundancy of *the* in the set expressions**

Some fixed phrases must be used with definite article, some phrases needn't. In English learning, there are lot of fixed phrases which will be meaningfulness if they collocate with *the*. Such as:

Example 1. And I will read the text in the [wd5,2-2] spare time.

Example 2. I put on the [wd5,1-] clothing [wd2,5-] and left home to run

- **Redundancy of *the* before abstract nouns, and the nouns with modifiers.**

Example 1. Though the [wd5,1-1] time pass [vp6,4-1] by.

Example 2. When they were taking [wd3,3-1] lessons, the [wd5, 0-2] some teachers of our school will listen.

Definite article is not required before abstract nouns, such as: time, hope, knowledge, friendship. Besides, if there are modifiers before nouns, definite article is also supposed to be omitted.

- **Redundancy of *the* before ordinal numerals**

Example 1. The [wd5, 1-] second, I want to know the words "further"

Example 2. The [wd5,s-] fourth, after the meal, they eat fruit or ice cream

When the ordinal number emphasizes order or it works as nouns, the definite article cannot be omitted. However, if the ordinal numbers wok as adverbs or in some fixed expressions, (at first, first of all, first or last) the definite article cannot be used with it.

- **Redundancy of *the* before three meals of a day, and ball games.**

When we refer to breakfast, lunch, and supper, or ball games, the definite article could not be used before them.

Example 1. And then it was time to prepare the [wd5,6-1] supper

Example 2. He had played the [wd5,3-5] table tennis with his parents for a moment.

- **Redundancy of *the* before a country, geographical names, position, and official title.**

Example 1. Every year, in the [wd5, 1-] Guang Zhou there had some show

Example 2. "flesh [fm3,-] " means "beauty" in the [wd5,1-1] England

Example 3. In 1949, the [wd5,2-5] Chairman Mao ... [sn2,s]China was liberate

“Guangzhou” is a name of city, “England” are the name of country, and Chairman Mao is a name of person, so *the* should not be used before these kind of nouns.

- **Redundancy of *the*, when two nouns or more are used simultaneously**

Example 1. People eat too much food with sugar and fat, like the [wd5,1-] hamburger [fm1,-], fired [fm1,-]chicken, coca.

Example 2. For example: students, peasants, workers, the [wd5,7-2] foreign friends

In these two examples, there are more than two nouns, which are used simultaneously, therefore, “the” before hamburger, foreign friends should be omitted.

- **Redundancy of *the* before seasons, months, festivals, dates, and days**

Example 1. At that the [wd5,1-] New Year, My father gave me one hundreds Yuan

Example 2. In the [wd5,1-1] winter. [sn2,s] The mind [wd3,1-1] blows so heavily.

Example 3. But on the [wd5,1-1] Thursday, I must fill in a member.

New Year”, “winter”, belong to seasons category, and Thursday,are the representatives of days. If definite article is added before those nouns, it will not conform to the grammar rules.

4.1.3 Confusion of *the*

ATC is the lowest one, which only occupies 9.3%. The writer classified confusion definite article into two categories: one is misuse definite article for indefinite article, which takes a proportion of 76.19, the other is misuse indefinite for definite, which is 23.81%. The following are some examples to explain it respectively.

- **Misuse definite article for indefinite article**

When we refer to someone, something, or someplace firstly, and before some descriptive adjectives or when the item is a part of a category, indefinite article is needed. However, many Chinese English learners use *the* instead of a/an.

Example 1. met the [np7,d] family.

Example 2. The night of that day, I went to the [np7,-1] old house with my Grandpa [sn9,-]

Example 3. I'm [np7,1-4] the student of Class 3

- **Misuse indefinite for definite article.**

There are only ten examples in sub-corpus St2 of CLEC which misuse indefinite article for definite article. The writer concluded it into following categories:

When the nouns phrase specifies referent of something, and in some fixed collocations, the definite article is required, rather than omitted or replaced by a/an.

Example 1. It was [np6,2-1] a finals of Hong Kong football united match. (St2)

Example 2. he asked to make a [np7,s-] same clothes [fm1,-] (St2)

4.2 Factors Affecting Misuse of *the*

Since the effect of misuse of definite article is hard to measure, so it is indispensable to investigate the underlying causes. In the following, the author will discuss the reasons of definite article error, which

contain first language transfer, second language related factors, and the article semantics, and the interview results.

4.2.1 L1 transfer

Odlin(2001) gave a introduction of transfer, which is the influence resulting from similarities and differences between target language and any other language which has been previously (and perhaps imperfectly) acquired.

According to L1 transfer, the first language namely the habits of language usage will directly affect second language acquisition, and the influence may be negative or positive. So L1 transfer can be divided into positive transfer and negative transfer. The former refers to the similarities between first language and second language, such as custom, patterns and rules, which contributes to second language acquisition. The negative transfer will hinder or impede second language acquisition.

Larsen-Freeman&Long(2000) pointed out when there are some similar components between native language and target language, positive transfer will occur. Otherwise, the negative transfer will take effect, when the two language are different. Because articles are absent in Chinese but present in English, L1 transfer can not be avoided.

4.2.2 L2 Related Factors

The L2 related factors just like second language transfer and incomplete rules may affect Chinese English learners' acquisition of article. In addition, some errors are caused by the rules which transfer from mother tongue. We called these type errors as interlingual errors. While learners also commit other errors by incorrectly using rules, which are called intralingual errors.

There is no doubt that generalization plays an important part in the process of language learning. Only by generalization can we distinguish the difference between words and rules, which are helpful to perform correctly, when we are using language. What's more, it also can predict how the different items will behave in certain circumstance. Sometimes, if we overuse generalization or the predictions are wrong, the overgeneralization will occur.

4.2.3 Articles semantics factors

English articles are functional words and it cannot be used independently. The referential meanings of nouns phrase govern the mapping of the form and meaning of English articles. Therefore, if we want to know how articles are used and how they are acquired by learners in second language learning, we need to consider the proprieties of relevant nouns.

English article system has only three members, but the relation between their form and function are very intractable, which makes it quite difficult for learners to acquire the usage of article and use them correctly. For Chinese senior high school English learners, it is inevitable and common to commit errors, because they are only at a intermediate English level

4.3 An Analysis of Interview Results

As the writer mentioned, the interview has been conducted on 25 senior high school students to explore the causes of misuse of definite article, which works as a supplement of general causes discussion, and make the cause analysis more complete and reliable. The following are the results concluded by the writer.

4.3.1 The causes of omission of *the*

According to the 25 interviewees, more than half of them said that they did not have a good master of article. 12 interviewees said when they are writing, they tend to write Chinese composition firstly, then translate it, which will lead to omitting the definite article, because their teachers always translate sentence by sentence as they teach reading comprehension.

4.3.2 The causes of redundancy of *the*

Almost every interviewee indicated that the rules of definite article are too complicated to master. 21 of them know that definite article should be omitted before ball games, while referring to the use of definite article before meals only 10 interviewees have known it. In addition, the memory discrepancy will also contribute to definite article redundancy especially in some fixed expressions.

4.3.3 The causes of confusion of *the*

According to the interviewees, they sometimes are restricted by grammatical rules and forms. For example, if they used definite article in the first sentence, they would use indefinite article in the second sentence. Furthermore, L1 transfer not only leads to omission of *the*, but also mislead students to make an improper choice between definite article and indefinite article. 16 interviewees expressed that before they write, they would think out the Chinese sentences, which can cause some confusion to choose articles.

5. Pedagogical Implications

As we know that the absence of article system in Chinese could lead to the difficulty in definite article acquisition, while the problems existing in teaching are also the important factors related to the misuse of definite article. Consequently, the writer comes up with several implications related to the learning and teaching of definite article.

Firstly, when the teacher explains the articles, they may discuss it with examples and contexts. And the articles in set phrases should be paid more special attention. Moreover, teachers are supposed to give some proper exercises to consolidate the acquired knowledge of articles. What's more, some effective measures should be taken to reduce the effect of language transfer.

Secondly, the rules of definite article are complex and scattered, hence it is improper to present all for once. Teachers could explain the rules of definite article, indefinite article, and zero article respectively. What's more, every kind of category requires some examples to interpret the general rules, and improve the usage of definite article.

Thirdly, when referring to the relation between general rules and special rules, teachers could instruct general rules before special rules. Emphasis on special rules and cannot ignore the general rules. In addition, teachers can make use of comparison method to avoid the confusion special rules with general rules.

Fourthly, since English definite article usage is associated with nouns and noun phrases in most conditions, so it is indispensable to separate definite article acquisition with the learner's overall language knowledge, especially, the knowledge of nouns and noun phrases. Thus, when the teachers instruct the rules of definite article, it should consolidate the relevant acquired knowledge.

6. Conclusion

From the above findings and discussions, we can draw some conclusions:

1) There are three types of misuse of definite article, which are omission of *the*, redundancy of *the*, and confusion of *the*. Chinese senior high school English learners are prone to omit definite article most frequently, which occupies 52%. Redundancy of definite article takes a proportion of 37.9%, and the

confusion definite article with indefinite article ranks the last with a proportion of 9.3%. Therefore, absence of definite article is the most frequent category of misuse of definite article for Chinese senior high school students.

2) The major factors connected with the misuse of definite article which are L1 transfer, L2 related factors and article semantics. The first language transfer is mainly referring to negative transfer that is the primary causes for the errors of definite article, since article system is absence in Chinese language. Some L2 related factors may cause the errors of definite article as well, which mainly contains over-generalization and incomplete application of grammar rules. Besides, article semantics can affect the usage of definite article.

Although, corpus-based approach was adopted in this present study. It dose not used wildly in Chinese studies, We hope there are more and more researchers will adopt corpus-based approach in subsequent studies.

All in all, the author expects the above advice could give some hints for the future on the misuse of definite article. We also hope the present study will provide some help to the teaching of English articles in China.

References

- [1]Corder, S. P. 1981. *Error Analysis and Interlanguage* [M]. Oxford: Oxford University Press.
- [2]Dulay, H. and M. Burt. 1974. 'You can't learn without goofing'. In J. C. Richards (ed.) 1974. *Error Analysis---Perspectives on Second Language Acquisition* 95-124. Longman.
- [3]James, C. 2001. *Errors in Language Learning and Use: Exploring Error Analysis* [M]. Beijing: Foreign Language Teaching and Research Press.
- [4]Master, P. 1997. Teaching the Article system to foreign technical writing students. *The Technical Writing Teacher*, 13 (3): 203-210.
- [5]Robertson, D. 2000. Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, 162, 135-172.
- [6]Trenkic, D. 2007. Variability in L2 Article Production: Beyond the Representational Deficit vs. Processing Constraints Debate. *Second Language Research*, 23(3):289~327.
- [7]Zegarac, V. 2004. *Relevance Theory and the in Second Language Acquisition*. *Second Language Research*, 20(3):193~211.
- [8]桂诗春, 2005. 中国英语学者言语失误分析, 基于语料库的中国英语学习者英语分析 [M]. 上海: 上海外语教育出版社.
- [9]桂诗春、杨惠中, 2003. 中国英语学者英语语料库 [M]. 上海: 上海外语教学出版社.
- [10]李景泉、蔡金亭, 2001. 中国学生英语写作中的冠词误用现象究 [J]. 解放军外国语学院学报, (6) :58-62.