

Corpus-Based Contrastive Study on the Use of the Definite Article “the” between Chinese EFL learners and English Native Speakers

Shuhan Dong

School of Foreign Languages, Quzhou University, Quzhou 324000, China.

Abstract: “The” is a functional word of high-frequency in English. Due to its simple form, Chinese students use it extensively after they have learnt its semantic and grammatical rules, and may show different characteristics from those of native speakers. Based on the use of corpus, this paper compares texts extracted from CLEC and BNC, and finds that both Chinese English learners and native English speakers use “the” most frequently to express “specific reference”, followed by “generic reference”. There is also a feature that Chinese English learners use fixed collocations containing “the” more frequently than native English speakers do, but “the” in these fixed collocations don’t belong to its ideographic function. The enlightenment for English teaching is discussed, as well as reasons for the differences: cultural differences, mother tongue and grammar teaching.

Keywords: Corpus, Contrastive Study, Definite Article, “The”.

1. Introduction

As a function word, “the” is the most commonly used word class. It almost has no lexical meaning and cannot be used independently. However, it can be widely used with nouns, pronouns, adjectives, numerals, verbs, prepositions, etc., especially before nouns. When teaching this word, teachers often directly tell students the semantics and grammatical rules of “the” in general[1].

Due to its simple form, “the” seems easy for learners to use, but in fact, the definite article covers various aspects such as grammar, semantics, discourse structure, and cross-language comparisons in linguistic research. And this study takes the cross-language comparisons as a research perspective. A large number of Chinese students learn English as a second language and there have been previous researches on this perspective at home and abroad. In addition, the comparison between English native speakers and Chinese EFL learners can lead to the development of cross-cultural communication and translation studies. Therefore, there are sufficient theory basis and promising progress to support the study.

To truly master a word, one must not only master its form, but also its function and meaning. Only through the comparative study of the definite article “the” in differences, can we reveal its rules in use and provide more in-depth insights in application. Therefore, it is necessary to study the uniqueness of definite articles used by Chinese learners.

In view of this, based on the British National Corpus (BNC) and the Chinese Learner English Corpus (CLEC), this study aims to explore the similarities and differences between Chinese EFL learners and English native speakers from the perspective of its ideographic function given by Zhang Zhenbang. The difference between the use of the definite article by English speakers and native speakers can be used to help reveal the unique characteristics and corresponding reasons, thereby providing some inspiration for English teaching.

In this study, a corpus-based approach was used to investigate the use of definite articles in written language by Chinese English learners and English native speakers in order

to find out the similarities and differences between them. At the same time, the focus of this study is to investigate and study the distinctive features of Chinese English learners’ use of the definite article “the”, which are different from those of native English speakers, so as to find out the reasons for the distinctive features. Finally, based on these differences and reasons, some enlightenment is provided for the English teaching of the definite article “the”.

The thesis consists of five chapters. Chapter One introduces the background and purpose of this study and the structure of the thesis. Chapter Two reviews achievements of definite article at home and abroad. Chapter Three shows the theoretical foundation. Chapter Four involves the research design, which clearly introduces the research content, research questions and research tools used in the study as well as research procedures. Chapter Five describes the main results and findings of this study. In addition, it analyzes and discusses the results of the present study on the basis of the data and instances. And the reasons responsible for the results and findings are presented in this chapter. Chapter Six shows the conclusions drawn from the research.

2. Literature Review

This chapter consists of three sections. Section 2.1 summarizes the previous related studies at abroad. Section 2.2 reviews the previous related studies at home.

2.1. Previous Researches Abroad

In English-speaking countries, the study of the definite article “the” covers many fields, such as linguistics, cross-language research, cognitive linguistics, second language acquisition research and so on. Based on the insights of sociolinguistic and pragmatic research, Acton made an in-depth study of the social phenomenon of Americans using English definite article, presenting that semantic meaning is part of that same web of interrelations. He thinks the exact interpretation of a particular the-plural on a given occasion of use depends importantly upon its indexical character, the beliefs of the speech participants, and myriad other contextual factors, but is nonetheless constrained in a principle way.

From the viewpoint of cognitive linguistics [2]. Kim explains the usage of the English definite article and is further to demonstrate the excellence of its explanatory adequacy based on viewing and cognitive models, which is hoped to be helpful for English courses [3].

At the same time, there are also many studies on non-native English speakers. For example, Beavitt and Popova put forward the common problems of non-native English learners. That's while existing pedagogical approaches are successfully used to teach non-native English learners near-native competency in the use of English articles, the final stage of native-equivalent competency continues to evade. Meanwhile, their research also pointed out that existing conceptual approaches to explicating the function of the English article system have limited utility and a comprehensive system has yet to be developed [4].

In addition, some foreign scholars have investigated the use of English articles by non-native English speakers, including Chinese subjects, and put forward several cases of their excessive use of definite articles [5].

Except for the above studies, some scholars also provide some strategies to help master definite articles through research. Through comparative experiment, it revealed that corrective feedback provided had a significant effect on the correct use of articles and the students who received corrective feedback outperformed those who did not receive CF and the students significantly improved their ability in using the targeted functions of the English article system accurately [6].

2.2. Previous Researches at Home

Judging from the current status of domestic research, Chinese researchers have continued to conduct in-depth research on the English definite article "the", paying attention to both theoretical linguistic analysis and practical teaching applications. Early Chinese scholars' research on the definite article mainly focused on supplementary research on a certain aspect of the existing usage or collocation rules of "the", and the research level was relatively single.

However, Wang Jin's study of the definite article "the" at the discourse level can expand the scope of his research and turn the perspectives of many researchers to the discourse fields involving pragmatics, semantics and cognition. In particular, Wu Jun discussed the referential meaning of the English definite article in discourse from the relationship between language and context, including context and actual situation, which provides important results for promoting people to better understand the discourse [7].

The study of linguistics is inseparable from real language materials. These language materials enable researchers to find more problems worthy of exploration and broaden their research perspectives. For example, Wang Lin, Jin Lingzi and Yu Jinyan used in-class tests and statistics to investigate the non-generic usage of definite article and the acquisition of zero article by English majors, and pointed out several misuses of definite article [8].

Besides, corpus can also be one of the important tools to provide authentic materials for Chinese English learners. Through the corpus, the researchers have completed many studies that can deepen the understanding of the definite article, which cover a wide range of fields, such as summarizing the types and characteristics of the definite article the used by Chinese English learners, analyzing its composition characteristics and high-frequency collocations

[9].

The rich data of the corpus provide the possibility for various types of comparative studies, such as the comparison between high school and college English learners, the comparison between second language learners and native speakers, or the comparison between indefinite articles and definite articles[10]. Furthermore, Yan Lili used AntConc search tool and CLEC corpus to find that Chinese students' acquisition of English articles has certain regularity, and learners at all levels have their own characteristics. Finally, she summarized the characteristics and reasons of these acquisition situations to provide some enlightenment for future English teaching[11].

However, after reading the literature, I found that there are many limitations in the field of definite article in domestic research and so do some other researchers[1,12]. Compared with the international research in this field, most of the domestic research is about the misuse and usage characteristics of definite article. The corpus of current Chinese English learners is mainly written materials, so most of the research is also based on written language, while the amount of research on oral English is very small. At the same time, they also pointed out that in terms of the factors affecting the use of articles, more studies only focus on a single factor.

3. Theoretical Foundation

The application of corpus methods in linguistic research is primarily concerned with analysing and understanding language use and change through large-scale textual data. This approach relies on large quantities of actual linguistic material, usually stored in computerised databases, which are called corpora.

A corpus is a specially collected, structured, representative collection of linguistic material of a certain size that can be retrieved by a computer programme[13]. As a linguistic tool, a corpus adopts a bottom-up approach to research. This approach emphasises the large-scale collection, storage and analysis of linguistic data from actual linguistic materials, assisted by computers. This approach makes corpus linguistics both scientific and humanistic, capable of describing aspects of the language system based on real evidence and explaining language patterns on this basis. Divided the research methods into three categories and they are corpus-informed approach, corpus-based approach and corpus-driven approach according to the degree of dependence on corpus[14].

The corpus-informed approach, advocated by applied linguist Michael McCarthy, is mainly used for the development of textbooks. Corpus-based research starts more from existing theoretical frameworks and uses the corpus as a tool to validate these theories. This approach emphasises on testing and supporting existing theories. On the other hand, the corpus-driven approach does not see the corpus as a methodology, but rather sees the corpus as having research value in itself, which can be analysed to discover new linguistic phenomena or laws [15]. This approach is more data-driven.

This study relies on Zhang Zhenbang's theoretical perspective on the functional classification of the definite article to guide the direction of the research, takes real cases from the BNC and CLEC corpora as the source of the research data, searches for the definite article by using the indexing tools of AntConc 3.5.9 and Sketch Engine, and analyses and

processes the data on the basis of the indexing, so that this study adopted is a corpus-based approach[16].

The beginning of any corpus study is the creation of the corpus itself. The decisions that are taken about what is to be in the corpus, and how the selection is to be organized, control almost everything that happens subsequently[13]. The use of corpus is a key step in the contrastive study of English and Chinese. Here are some steps.

The first consideration is the aim of the activity of corpus creation. This will help to select the appropriate corpus and analysis methods. To select the appropriate English-Chinese parallel corpus according to the research needs is next step. And in terms of the selection of the most suitable corpus for English-Chinese contrastive study, it needs to comprehensively consider the amount of data, quality, representativeness, diversity and whether it meets the needs of specific research methods and topics. After selecting the corpus, data preprocessing is required, including text cleaning, standardization and annotation to ensure data quality and comparability. This may involve the removal of irrelevant characters, unified part-of-speech tagging, word segmentation and other operations. In the contrastive study of English and Chinese, effective data analysis and statistics are also very important. In the step, we can choose the right data analysis tool such as Antconc and Sketch Engine. Besides, researchers can also apply the statistical analysis methods to go on data analysis. It means according to the evidence, the index comparative analysis method is one of the most commonly used methods in statistical analysis.

This method reveals the quantitative differences and changes between things by comparing different indicators. In the contrastive study of English and Chinese, this method can be used to analyze the differences between the two languages in specific indicators (such as frequency of use, grammatical structure, etc.). Finally, researchers explain the results and organize the research process, methods, results and conclusions into a written report, so as to obtain scientific and empirical support for the research.

4. Research design

This chapter refers to the research methodology of this study, which can be divided into three sections. Section 4.1 presents the research questions of the study. Section 4.2 introduces the research instruments used in the study. And the last section addresses the whole procedure of this study in detail.

4.1. Research Content and Research Questions

In this study, a corpus-based approach was used to investigate the use of definite articles in written language by Chinese English learners and English native speakers in order to find out the similarities and differences between them. At the same time, the distinctive features of Chinese English learners' use of the definite article "the" which are different from those of native English speakers will be investigated to find out the reasons for the distinctive features. Finally, based on these differences and reasons, some enlightenment is provided for the English teaching of the definite article "the".

According to the literature review and data, the research questions of this paper are as follows :

(1)What are the similarities and differences in the use of definite articles between Chinese EFL learners and native English speakers ?

(2)What are the reasons for the differences in the use of

definite articles between Chinese English learners and native English speakers ?

(3)Through the analysis of the above two questions, what enlightenment can be obtained in teaching ?

4.2. Research Instruments

In the following section, the research design will be illustrated, including the selection of corpus and processing tools used in this study.

4.2.1. Corpus

The purpose of this study is to explore the unique distribution characteristics of definite article in Chinese English learners' writing, so the author needs to compare Chinese learners with native English speakers.

Therefore, this study adopts Chinese Learners English Corpus and British National Corpus, which can provide a large number of authentic language materials written by Chinese learners and native English speakers. The general description of these two corpora is as follows.

1) Native Corpus

The British National Corpus (BNC) is one of the most representative contemporary English corpora in the world with 100 million words representing written and oral language samples of British English in the late 20th century. However, the main purpose of this study is to compare the differences in the use of definite articles between Chinese English learners and native English speakers in written language, so oral language samples are excluded. However, it is worth mentioning that the written part of BNC accounts for 90 %, including national and regional newspapers, various journals, academic books and popular novels, published and unpublished letters and university papers, which can provide rich English texts for this study [17]. In addition, each text contains complete classification, context and bibliographic information, which also provides feasibility for selecting samples and classification.

2) Learner Corpus

Chinese Learner English Corpus (CLEC) is the largest and most authoritative language database in China. The database collects more than one million words of five kinds of students' corpus, including middle school (St2), college English level 4 (St3) and level 6 (St4), junior English majors (St5) and senior English majors (St6), and scientifically labels and classifies language errors[18]. Since the purpose of this study is to provide suggestions for the teaching of definite articles for middle school students, this study selects the real corpus of middle school students as one of the research objects of this study, that is St2.

4.2.2. Data processing Tools

The research tools used in this paper include AntConc 3.5.9 and Sketch Engine. The followings are their specific introduction.

1) AntConc 3.5.9

AntConc 3.5.9 is a free corpus search tool that has many functions such as indexing, lexicon generation, collocation and cluster extraction. By searching for words, phrases or other terms in the corpus, it retrieves and displays concurrent contexts, word frequency tables and keyword lists.

This study primarily uses the word search of AntConc. Retrieving all terms of a word or phrase from a huge corpus is an essential feature of AntConc. Its searches are divided into basic and advanced searches. In this paper, we use the basic search function for our study. The basic search function enables the retrieval of a single word or phrase. The search

results can be sorted according to the sorting function and saved as a text file. Its search function allows frequency counting of linguistic phenomena in native and learner texts, which helps researchers identify similarities and differences between native and learner texts, and understand the developmental characteristics and rules of foreign language learners. In this study, AntConc was used to retrieve frequencies and index lines of target words in CLEC.

2) Sketch Engine

Sketch Engine is a corpus query system. Currently, there are corpora in Chinese, English, German, and other languages in the system, with a total word count of hundreds of millions of words. Through online searching, it can provide users with diversified analysis results.

The advantage of Sketch Engine is that not only the word frequency of the target words can be retrieved, but also the collocations of the target words and the corresponding information. In addition, the index lines of the target words can be randomly sampled. The Sketch Engine system contains the British National Corpus. Therefore, Sketch Engine is a very useful tool for the study of target words in BNC. It can be of great help for collocational use and class conjunction use in this study. In this study, by using the system to retrieve target words in BNC, word frequency, collocations and indexing lines for analysis can be obtained.

4.3. Data Analysis Procedures

By comparing and analyzing the data retrieved of “the” from BNC and CLEC, the entire analytical process consists of the following steps.

Step one: The first step is to ensure the comparability of the data. Since the text of St2 in CLEC comes from the written writing of secondary school students, the authors will omit the spoken part of BNC and limit the written data of native speakers to “essay-school”, which means that the data of BNC comes from the writings of school students (except college students), so that the data sources of both the Chinese learners and the native speakers are under the two ranges of “students” and “written language” to ensure the reliability of the data comparison in terms of text level and type. Finally, as for the sampling of BNC, I use Sketch Engine to search out all index lines of “the” in the BNC, then use Sketch Engine’s random sampling function to extract a part of index lines to save and analyze. As for the sampling of CLEC, the authors use AntConc to search index lines of node words in the CLEC, and utilise a random number generator for sampling. In this way, the authors succeeded in randomly selecting 100 pieces of data from each of the two corpora, BNC and CLEC, to be analysed.

Step two: Data collation is the theme of this part, the most important of which is to determine the basis of classification. As for the research basis of the definite article mentioned in this study, the author adopts Zhang Zhenbang’s point of view. He thinks it has ideographic function including generic reference and specific reference [19].

Generic reference refers to the category, which refers to a class of people or things, and definite articles, indefinite articles and zero articles all have this function.

Specific reference refers to a specific object in a class of people or things. It is divided into definite specific reference and indefinite specific reference, and only “the” belongs to definite specificity. Definite specific reference can be divided into anaphoric specific reference, cataphoric specific reference, situational reference. The specific explanation is as

follows :

(1)Anaphoric specific reference: Use the definite article to represent the person or thing mentioned above. The first mention usually uses indefinite articles. The second mention, if it is inconvenient to use pronouns, use definite articles.

(2)Cataphoric specific reference: The anaphor is not in the above, but in the

following, or more precisely, the designated role of the definite article depends on the post-modifier of the noun head.

(3)Situational reference:The reference relationship based not on the context, but on the knowledge shared by both sides of the conversation. After that, the classified text is saved in word documents named “native language classification” and “non-native language classification” respectively.

Step three: In this step, data were collected and compared. On the one hand, the difference in the frequency of the definite article “the” between native speakers of English and Chinese learners of English in each ideographic function is studied. On the other hand, it is a collocation study. In detail, it is to observe and analyse what kinds of collocations are used by native speakers of English and Chinese learners of English to achieve the corresponding ideograms, as well as the differences between them in terms of the types of collocations and the frequency of their use.

Step four: At the end of the comparison step, the results will be discussed based on the data analyzed to answer the research questions of this paper.

The findings and conclusions will be described and possible reasons for the differences and similarities will be deduced.

5. Results and Discussion

In this part, we will present and discuss the data results of the ideographic function of the definite article. However, the results of the study show that there is a special case in the selected corpus, that is, part of the corpus belongs neither to the scope of the specific reference nor to the scope of the generic reference. So this study decides to use “others” to refer to this special case, and discusses and describes this kind of case alongside the specific reference and the generic reference. This type of situation is discussed and described side by side with special and analogous references.

5.1. Similarities for Use of “The” Between Chinese Learners and Native Speakers

In this part, we will discuss the similarities between Chinese learners and native speakers in the areas of specific reference, generic reference and “others”.

5.1.1. Similarities in Specific Reference

According to Table 1, Chinese learners of English use “the” to refer to something in 66% of the 100 items of the selected corpus. Obviously, the frequency of “the” used to refer to specific reference is already more than half of the selected corpus, and the same situation also occurs in native speakers, who use “the” to realise the function of specific reference is 85% in frequency. It indicates that “the” is more widely used by Chinese English learners and native speakers, far more than the other two cases, to express specific reference.

According to the previous research design, specific reference can be subdivided into anaphoric specific reference, situational reference and cataphoric specific reference. Next, we will discuss the similarities between Chinese English learners and native speakers in these three aspects.

In terms of frequency, from Table 1, we can also see that the frequency of using “the” for post-tabular reference is 12 percent. the frequency of Chinese English learners in using “the” to realize the anaphoric specific reference is 32 %, the frequency of using “the” to realize the situational reference is 22 %, and the frequency of using “the” to realize the cataphoric specific reference is 12 %. These data shows that the frequency of the definite article used by Chinese English learners decreases in the order of anaphoric specific reference, situational reference and cataphoric specific reference. The same situation also occurs in native speakers. The frequency of native speakers in using “the” to achieve anaphoric specific reference is 36 %, the frequency of using “the” to achieve situational reference is 30 %, and the frequency of using “the” to achieve cataphoric specific reference is 19 %.

Table 1. The frequency of four types of ideographic function in BNC and CLEC

| Ideographic function | Corpus | Frequency |
|-------------------------------|--------|-----------|
| Specific reference | BNC | 85 |
| | CLEC | 66 |
| Anaphoric specific reference | BNC | 36 |
| | CLEC | 32 |
| Aituational reference | BNC | 30 |
| | CLEC | 22 |
| Cataphoric specific reference | BNC | 19 |
| | CLEC | 12 |

In terms of collocation, this study will combine examples to discuss the similarities between Chinese English learners and native speakers.

In the case of anaphoric specific reference, both Chinese English learners and native speakers use two kinds of collocation structures to refer to the specific reference. One is “the+n+prep”, the other is “the+n+postpositional determinative”.

(1) Examples of “the+n+prep” in CLEC and BNC

1) CLEC

① The night life in Guang Dong Commercial College is more colourful than that in our institute.

② The only way to qualify the people as a skillful worker is to practice for thousands of time.

③ I went to primary school at the age of 7, and left it at 13 after successfully passing the entrance examination to junior Middle School.

2) BNC

① The measure was necessary partly to halt the fall in output due to exhausted and outdated equipment.

② The National Assembly approved the establishment of this commission on May 14.

③ When Paris claims he loves Helen above thought, he is providing piercing insight into the twisted motives behind the adultery that has infected Troy.

(2) Example of “the+n+postpositional determinative” in CLEC and BNC

1) CLEC

① Both of the ways mentioned above enable us to gain insight into the world.

② They would get the biggest steaks they could find.

③ This is the reason why haste makes waste.

2) BNC

① Clearly, the observations made by Darwin during his voyage on the Beagle, for example, would have been inconsequential for science had they remained Darwin’s private experiences.

② The little sounds I make against your skin don’t mean anything.

③ The Design Change (DC) identifier has been CANCELLED (state 5) and no further work can be done on the modules listed.

It can be seen from Table 2 that when “the” is used by Chinese English learners to realize the anaphoric reference function, the frequency of using the former collocation is 87.5 %, and the frequency of using the latter collocation is 12.5 %. When “the” is used by native speakers to realize the anaphoric reference function, the frequency of using the former collocation is 75 %, and the frequency of using the latter collocation is 25 %.

The data prove that the ability of Chinese English learners to use the “the + n + prep” to refer to specific reference is almost the same as that of native speakers. Besides, these two collocations can be used as a kind of sign to help students understand and use “the” as a special reference.

Table 2. The frequency of two types of collocations in BNC and CLEC

| Collocations in anaphoric specific reference | Corpus | Frequency |
|--|--------|-----------|
| the+n+prep | BNC | 27 |
| | CLEC | 28 |
| the+n+postpositional determinative | BNC | 9 |
| | CLEC | 4 |

In the case of situational reference, it is one of the most frequently used ideographic functions among both Chinese learners and native speakers of English, second only to the anaphoric specific reference. In situational reference, the prevailing structure is “the + n”, and there is no special marking structure as in the case of situational reference. The definition of this function also implies that one of the most important bases for judging whether “the” denotes a contextual referent is the knowledge shared by both parties to the conversation, in other words whether they know what the noun modified by the definite article refers to.

(1) Examples of in CLEC and BNC

1) CLEC

① On the contrary, the life expectancy had been improved from 40 years old in 1960 to 60 years old in 1990.

② The infant mortality was 200 deaths per 1,000 births, but it was dropped to 100 deaths per 1,000 births in 1990.

③ He got to the place but his friend didn’t.

2) BNC

① One night Len’s wife, a small woman, had repeatedly asked a trio of Teds from the next village to cut down their noise and stop carving the backs of the seats.

② But I knocked you to the floor, and the blow was so hard that it made my hand throb for hours afterwards.

③ In future there will be no access through these gardens but it is only two swims and the rest of the canal can be

accessed via other routes.

The corpus of CLEC is taken from Chinese secondary school students' writings, and these writings are centred on specific topics. For example, the first example of CLEC presented here is a writing on the topic of life expectancy, and the third example of CLEC presented here, "He got to the place but his friend didn't.", is from a story writing around the background of an event. The third example sentence "He got to the place but his friend didn't." comes from story writing that revolves around the background of an event, and the reader can easily understand that the "place" here refers to the place where the protagonist and his friend agreed to go after understanding the background of the story. But whether they are topic writing or story writing, they both give learners a specific context. Similarly, the three example sentences of BNC shown here are all taken from a book, and the readers, by understanding the content of the book, put themselves in the corresponding contexts and thus understand what a noun refers to. For example, "But I knocked you to the floor, and the blow was so hard that it made my hand throb for hours afterwards." The reader, after learning the events of the incident, can easily know that the word "floor" here refers to the floor where the two characters clashed.

In the case of cataphoric specific reference, it is the least frequently used type of specific reference among both Chinese learners of English and native speakers. As with situational referents, the universal structure of this function is "the + n", with no special marking structure as in the case of anaphoric specific reference. The definition of this function also implies that an important criterion for determining whether "the" denotes a post-contextual feature is whether or not a noun has already been mentioned in the preceding text. However, in the course of the study, the authors found that there is no absolute boundary between situational reference and cataphoric specific reference. This finding is true for both Chinese learners and native speakers.

Examples of in CLEC and BNC

1) CLEC

① On the contrary, the life expectancy had been improved from 40 years old in 1960 to 60 years old in 1990.

② The infant mortality was 200 deaths per 1,000 births, but it was dropped to 100 deaths per 1,000 births in 1990.

③ He got to the place but his friend didn't.

BNC

① One night Len's wife, a small woman, had repeatedly asked a trio of Teds from the next village to cut down their noise and stop carving the backs of the seats

② But I knocked you to the floor, and the blow was so hard that it made my hand throb for hours afterwards.

③ In future there will be no access through these gardens but it is only two swims and the rest of the canal can be accessed via other routes.

In this study, we will take the same examples of situational reference as the one discussed in the previous section. The third example of CLEC, "He got to the place but his friend didn't." comes from a story written in the context of a certain event, where the word "place" has already been mentioned in the context of the story, so "the" modifying "place" is used to refer to the place. Similarly, the tense of "But I knocked you to the floor, and the blow was so hard that it made my hand throb for hours afterwards." means the scene described in the statement has already happened, and the scene is mentioned in the preceding text, so "floor" has already been mentioned

in the preceding text when describing the conflict between the characters. "The" in "the floor" is used to achieve cataphoric specific reference after it has been mentioned in the preceding text.

5.1.2. Similarities in Generic Reference

According to Table 3, in the selected 100 corpora, Chinese English learners use "the" to realize the ideographic function with a frequency of 20 %, which is the second most frequently used ideographic function by Chinese English learners after specific reference, and it is also the most frequently used ideographic function by native speakers. The frequency of native speakers using "the" to realize the generic reference is 14 %.

Table 3. The frequency of three types of ideographic function in BNC and CLEC

| Ideographic function | Corpus | Frequency |
|----------------------|--------|-----------|
| Specific reference | BNC | 85 |
| | CLEC | 66 |
| Generic reference | BNC | 14 |
| | CLEC | 20 |
| Others | BNC | 1 |
| | CLEC | 14 |

In addition, they generally use the "the + n" to achieve this function, and these nouns contain both singular and plural nouns, as well as uncountable nouns.

(2) Examples of in CLEC and BNC

1) CLEC

① When we walk in the street, we can see all kinds of advertisements on both sides of the road.

② The bicycle is the most popular means of transportation in China.

③ So the people in Europe began to drink the tea.

2) BNC

① The miners' strike will give us an opportunity to discuss the militarization of the police, one of the most noticeable as well as one of the most controversial trends of recent years.

② It will be in everyone's interest that the precise arrangements are specified, eg whether the partnership has a tenancy or contractual licence to occupy and in consideration of what payments, or whether the property is to be enjoyed rent free, or whether the firm is given only a non-exclusive personal licence which will determine on the death or retirement of the partner who grants the same.

5.1.3. Similarities in "others"

In the process of classification, it is found that there are samples that are neither specific nor generic, which will be collectively referred to as "others" in the paper.

5.2. Differences for Use Of "The" between Chinese Learners and Native Speakers

This part will mainly discuss the differences between Chinese English learners and native speakers in specific reference and "others".

5.2.1. Differences in Specific Reference

According to Table 4, in terms of the total number of specific reference, native speakers use "the" to indicate specific reference much more frequently than Chinese EFL

learners, and in the three cases of anaphoric specific reference, cataphoric specific reference and situational reference, the frequency of native speakers is also higher than that of Chinese English learners.

Table 4. The frequency of four types of ideographic function in BNC and CLEC

| Ideographic function | Corpus | Frequency |
|-------------------------------|--------|-----------|
| Specific reference | BNC | 85 |
| | CLEC | 66 |
| Anaphoric specific reference | BNC | 36 |
| | CLEC | 32 |
| Situational reference | BNC | 30 |
| | CLEC | 22 |
| Cataphoric specific reference | BNC | 19 |
| | CLEC | 12 |

On the whole, Chinese learners and native speakers almost always use the simple “the + n”. However, since most of the Chinese learners’ corpus comes from their English writing, which is formed around a specific topic, most of the nouns used are topic words or common nouns related to topic. In contrast, native speakers select a wide range of sources in the corpus, and the corresponding written language involves a much richer variety of content types, even to the extent that words like “Beagle”, “Design Change (DC)”, “the 1977 Act”.

5.2.2. Differences in “others”

As shown in Table 5, the number of “the” used by Chinese learners in “others” is 14, while the number of “the” used by native speakers is only 1. And by observing this kind of sample, the author summarizes its form as follows :

(1) Used for “The + comparative level, the + comparative level”

1) Samples in CLEC:

① In other words, the more one study, the more he will get.

② It means that the more you practise, the better you will get.

2) Samples in CLEC: No samples

(2) Used to modify the highest level of adjectives

1) Samples in CLEC:

① But the most important point is that we have to notice speaking and excising every day and must be every day not one or two days.

② The most important one is through mass media.

2) Samples in CLEC: No samples

However, it is worth noting that “the+superlative adjective” does not appear alone, but often appears in the form “the + superlative adjective + noun”. However, under this rule, the function of “the” is mainly reflected in the modification of adjectives, so it cannot be classified into “the + noun” to indicate specific reference or generic reference.

(2) Used as a fixed collocation to express specific meaning

1) Samples in CLEC:

① Modern society is complex and developing all the time.

② On the contrary, the life expectancy had been improved from 40 years old in 1960 to 60 years old in 1990.

③ If health and wealth can’t be obtained at the same time, which one do you choose?

2) Samples in CLEC:

On the other hand, more efficient production was a significant benefit to those who had been on courses.

From the above examples, it can be found that Zhang Zhenbang’s classification of the ideographic function of definite articles does not cover all examples. Chinese learners of English do not limit their use of “the” to its ideographic function, but also show frequent use of fixed phrases containing “the”.

Table 5. The frequency of three types of ideographic function in BNC and CLEC

| Ideographic function | Corpus | Frequency |
|----------------------|--------|-----------|
| Specific reference | BNC | 85 |
| | CLEC | 66 |
| Generic reference | BNC | 14 |
| | CLEC | 20 |
| Others | BNC | 1 |
| | CLEC | 14 |

5.3. Reasons for Differences

Therefore, in view of these differences, the author investigates the possible causes of this situation from different angles. Based on the previous research results on article system and personal learning experience, the following section introduces the possible reasons for this difference between Chinese learners and native English speakers.

The first reason will be talked about is cultural differences. The cultural differences between Chinese and English expressions not only affect Chinese EFL learners’ understanding and application of the language form and structure of definite articles, but also involve deeper differences in cultural writing. For example, Westerners are more inclined to explicitly point out the object of discussion in communication, even in informal situations, which prompts them to frequently use “the” in sentences to refer to specific objects. In contrast, in Chinese, the concept of articles is not as clear and necessary as in English. Chinese expressions often rely on context to determine the specific or general reference of nouns, rather than distinguishing by articles. And it also confirmed by the data of this study, showing that native speakers only use “the” to indicate specific reference much more frequently than Chinese EFL learners, and in the three cases of anaphoric specific reference, cataphoric specific reference and situational reference, the frequency of native speakers is also higher than that of Chinese EFL learners.

The second reason will be mentioned is mother tongue transfer. Chinese learners of English do not limit their use of “the” to its ideographic function, but also show frequent use of fixed phrases containing “the”. This may be influenced by native language transfer. Since the Chinese language does not have an article system, Chinese learners of English need to rely on fixed phrases to achieve communicative purposes other than referring to something in particular or in a similar way. For example, expressions such as “one...the other” and “at the same time” in Chinese do not have a clear grammatical structure in the English system of articles, and they can only be expressed by using fixed phrases containing “the”. We can only use the fixed phrases containing “the” to realize meaning transfer with Chinese, and we have no way of knowing what

the definite articles in these fixed phrases do. This difference in linguistic structure makes Chinese learners rely more on their mother tongue when using the definite article, which leads to frequent use of fixed phrases containing the definite article that cannot be interpreted in terms of ideographic function.

Another factor that always appears with mother tongue transfer is grammar teaching. Chinese EFL learners usually choose the semantics they want to express in Chinese and then search for the corresponding English expressions stored in their memory. In addition, from the perspective of personal education experience, middle school English teachers mainly focus on the grammatical teaching of definite articles from the perspective of distinguishing the two ideographic functions of specific reference and generic reference and the need to use the grammatical rules of “the”. Chinese learners of English acquire these usages mainly from their English teachers, so the learners’ level of acquisition is related to the teaching methods of the English teachers. Grammar books are relatively comprehensive in describing the usage of the definite article, but most learners accept the usage of the definite article listed in the grammar books as emphasized by their English teachers in class, together with a lot of written practice, so that these usages are deeply memorized by the students and as long as they will relate these expressions to the corresponding Chinese meanings. However, this also reflects a negative effect, which can lead to the solidification of expressions among Chinese EFL learners. In fact, in order to avoid repetitive use of a word or phrase, native speakers tend to replace it with its synonym, pronoun, or more concise language in order to make their language more varied and validated.

Therefore, the frequent use of “the” by Chinese EFL learners in their inability to recognize such samples precisely reflects the fact that English teaching in the lecture-plus-practice mode tends to make students inert and dependent, and too lazy to look for other expressions that can convey the same meaning, which will greatly affect their use of the definite article in speech, as well as the characteristics and developmental patterns of the English language of Chinese learners.

5.4. Pedagogical Implications

Based on the results and findings, this study is expected to help teachers and students understand the use of definite article better and to provide a reference for updating the teaching and learning process.

5.4.1. Emphasis on Using Examples

English teachers in middle schools should clearly distinguish the specific reference and generic reference. Although the study also compares the three types of specific reference: anaphoric specific reference, situational reference, and cataphoric specific reference. However, as mentioned earlier, there is no clear categorisation of situational reference and cataphoric specific reference, and secondary school students have not yet formed a level of thinking that can understand these three types of specific reference, so it may be more difficult for them to acquire the definite article if they are taught in detail about these three types of specific reference. Therefore, when teaching the definite article in English to secondary school students, it is important that the English teacher teaches the students to make a clear distinction between the generic reference and the specific reference.

Firstly, when teaching articles, it needs to be clearly explained to learners that the definite article “the” is used to refer to a specific person or thing, whereas the indefinite articles “a” or “an” are used to refer to a generic situation, i.e., they do not refer to a specific person or thing. Understanding can be deepened by matching proportional sentences, such as using the indefinite article when something is mentioned for the first time in a text and the definite article when it is mentioned for the second time. Specific examples can also be used to illustrate scenarios where the definite article is used. For example, the definite article should be used when referring to a person or object that has already been mentioned above, or that is known to both parties. In addition, examples of category usage can be used to help learners understand, for example, the use of the article in conjunction with singular countable nouns to indicate category.

5.4.2. Emphasis on Context

Besides, as for teachers in middle schools, in the process of English teaching, teachers should emphasize context. In the past, researchers have investigated the use of definite articles by Chinese learners in tests to reveal the important role of “the” in affecting people’s interpretation and understanding of discourse, especially its referential specificity and context dependence[20,21]. In the purpose, content and method of grammar teaching in the basic stage of English, Zhang Zhenbang also mentioned that one of the principles followed by New English Grammar is to deal with grammatical problems in connection with pragmatics and discourse, and emphasized that sentences do not exist in a vacuum, but always appear in a specific context. Therefore, when describing syntax, we should not only describe the grammatical structure of sentences, but also contact the use of occasions, so that learners can know how to use grammar appropriately in specific communicative situations[19]. In addition, translation exercises can be used to improve learners’ comprehension and application skills. English sentences containing the definite article can be selected for learners to try to translate into Chinese, and they can be instructed on how to translate accurately using the definite article.

5.4.3. Emphasis on Designing Exercise

What’s more, Teachers should pay attention to grammatical rules and exercises, so that students can practice the sentences and contexts of the definite article “the” repeatedly. But teachers must change the teaching idea of simply imparting grammatical knowledge, and replace it with more practice while speaking. Otherwise, simply instilling grammatical clauses and simply performing grammatical analysis will also make grammar teaching boring and become a burden for learners. Some fill-in questions, multiple-choice questions or correction questions can be designed so that learners can master the correct use of definite articles in practice. In addition to these written exercises, teachers can often add translation training when explaining sentences or commenting on daily homework to improve learners’ application ability through translation practice. Some English sentences containing definite articles can be selected, so that learners can try to translate them into Chinese and guide them how to use definite articles accurately for translation. Besides, teachers should focus on the differences in students’ acquisition of articles and correct common errors in the learning process, so as to improve their awareness and ability to use English articles correctly. In addition, teachers can make use of the great power of corpus and learn how to integrate the information and functions from the corpus.

Because corpus can provide typical and real language materials for definite article used by native English speakers, making use of corpus can not only help improve teaching method based on these written materials, but help learners understand the real usage and collocation of more words and improve their expression ability. It can also help learners better understand the meaning and usage of definite articles in specific contexts, rather than memorizing them in isolation. Students can also observe the use of various grammatical rules of definite articles in language materials, which helps Chinese learners better understand their complex usages, actively discover the typical usages of definite articles, stimulate their desire for knowledge in English learning, and cultivate students' subjective initiative to find problems.

6. Conclusion

This section briefly summarizes the main results of this study and presents the limitations of this study.

First of all, in terms of similarities, both Chinese EFL learners and native English speakers frequently use the definite article "the" to express specific reference, especially when expressing anaphoric specific reference, followed by situational reference and cataphoric specific reference, indicating that both of them are proficient in the use of "the" to achieve the function of specific reference. Therefore, Chinese EFL learners' mastery of the ideographic function of "the" is gradually moving closer to that of native English speakers. From the point of differences, there are some distribution features of definite article the in the writings by Chinese EFL learners in comparison with native English speakers. There are significant differences in the use of "the" between Chinese EFL learners and native English speakers in the two situations of specific reference and indistinct reference.

The reason for the above differences may be the influence of cultural differences, mother tongue and grammar teaching. In fact, in order to avoid the repeated use of a word or phrase, native speakers tend to use its synonyms, pronouns or more concise language to replace it. The explanation of definite article usage in Chinese middle school English textbooks is concise and clear, which is easy for beginners to understand and master, but at the same time, it is also incomplete and ambiguous [22]. Therefore, in the actual teaching process, most teachers usually appropriately extend the focus of English articles and some fixed phrases according to the content of the textbook, which may lead to students' limited acceptance of relevant grammatical knowledge and easy to produce fixed thinking, lazy to use other flexible expressions, and also affect Chinese students' current inability to use authentic expressions completely like English native speakers.

There are also some shortcomings in this study inevitably. In this part, the author clearly points out the limitations of this study as follows :

(1) In order to align with the writing level of native speakers, the author chooses Chinese English learners at the university level as the subjects of the survey, so the language materials at the senior high school level are excluded. Therefore, the particularity of language materials in senior high school is not within the scope of this study. The author really hopes that the further study of definite articles can look at it more comprehensively.

(2) There are few samples collected in this study, so there are other special cases that exist but are not taken into account. It is hoped that this study can be continued in the future.

References

- [1] Hu, H. L.: "A Corpus-Based Study on the Use of Definite Articles by English Major Students" (MA, Northwest A&F University, China 2012), p.1-58.
- [2] Acton, E. K.: "Pragmatics and the Social Life of the English Definite Article," *Language*, Vol. 95 (2019) No.1, p.37-65.
- [3] Kim, H. D.: "Cognitive Linguistic Explanatory Adequacy on the Definite Article of English," *The Journal of Linguistics Science*, Vol. 29 (2004), p.107-124.
- [4] Beavitt, T. A. & N. G. Popova: "The Role of the English Article System in Developing Dialogical Context: A View from Russian Science," *Journal of Language and Education*, Vol. 6 (2020) No.3, p.52-68.
- [5] Cho, H.: "The Use of English Articles by Non-Native Speaker of English," *Asia-Pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, Vol. 8 (2018) No.5, p.127-139.
- [6] Soori, A. & A. A. Samad: "The Efficacy of Immediate and Delayed Corrective Feedback in the Correct Use of English Definite and Indefinite Articles," *The Journal of American Science*, Vol. 42 (1992) No.2, p.349-354.
- [7] Wu, J.: "On the Discourse Referential Meaning of English Definite Article THE and Its Application in Translation," *Journal of Changchun University of Science and Technology*, Vol. 9 (2011), p.60-62.
- [8] Wang, L. et al.: "A Study on the Non-Generic Use of English Definite Article and the Acquisition of Zero Article," *Journal of Shaoxing University (Education Edition)*, Vol. 12 (2023), p.85-87.
- [9] Sun, H. Y. & C. F.: "Based on Corpus Study of Learner Definite Article THE Formal Characteristics," *Foreign Languages and Translation*, Vol. 2 (2023), p.58-63.
- [10] Tang, B. X.: "A Preliminary Study on the Regularities of Chinese English Learners' Acquisition of English Articles—A Study Based on the Chinese Learner English Corpus," *Journal of Hainan Radio and Television University*, Vol. 2 (2004), p.71-94.
- [11] Yan, L. L.: "A Preliminary Study on Chinese Students' Acquisition of English Articles—A Study Based on the Chinese Learner English Corpus," *Foreign Language Teaching and Research*, Vol. 3 (2003), p.210-214.
- [12] Guo, H. J. & Zhou, Q. Q.: "The Acquisition of Indefinite Articles by Chinese English Learners: A Study of Variation," *Journal of PLA University of Foreign Languages*, Vol. 2012 (2012) No.1, p.54-58.
- [13] Sinclair, J.: *Corpus, Concordance, Collocation* (Oxford University Press, United Kingdom 1991).
- [14] Liang, M. C. et al.: *Corpus Application Tutorial* (Foreign Language Teaching and Research Press, Beijing, China 2013).
- [15] Heine, B. & H. Narrog: *The Oxford Handbook of Linguistic Analysis* (Oxford University Press, United Kingdom 2015).
- [16] Zhang, Z. B.: "The Purpose, Content, and Method of English Grammar Teaching in the Basic Stage," *Foreign Language World*, Vol. 1 (1992), p.1-5.
- [17] Mao, Y. X.: "Based on BNC and CLEC Corpus Analysis of English Verb Control Collocation Characteristics," *Overseas English*, Vol. 12 (2023), p.85-87.
- [18] Zhang, J. & Z. B. R.: "Analysis of Chinese English Learners' Misuse of Articles Based on CLEC Corpus," *Journal of Huaihu University*, Vol. 9 (2008), p.103-104.
- [19] Zhang, Z. B.: *New English Grammar Tutorial* (Shanghai Education Press, China 2017).

- [20] Geng, J.: "The Semantic Analysis of the Definite Article 'Misuse' by Chinese Learners of English," *Asian Social Science*, Vol. 6 (2010) No.7, p.180-184.
- [21] Wang, J.: "The Cohesion and Coherence of English Definite Article in Discourse," *Journal of Sichuan International Studies University*, Vol. 6 (2001), p.68-71.
- [22] Zhu, X. Y.: "Cognitive Analysis of Misuse of English Definite Article by Chinese Second Language Learners," *Modern English*, Vol. 3 (2020), p.103-105.