This paper attempts to examine whether Chinese English learners adopt the underproduction or overproduction behavior when using English relative clauses in academic writing. To achieve the objective of the study, the researchers used a corpus-based method to analyze and compare the production of the restrictive relative clause among Chinese students and native English students. The aims were to find out if the Chinese students underproduce the English relative clause and also examine how frequently they use that, which, and zero relative pronouns as reported in the sub-corpus of the Written English Corpus of Chinese Learner (WECCL) and the British Academic Written English corpus (BAWE). Based on the findings, conclusions are drawn as follows; compared to native English students, Chinese EFL learners generally tend to underproduce the restrictive relative clause in academic writing; and on the other hand, overuse the relative pronoun that. The observations from this study concur with the findings from previous studies.

**KEYWORDS**

English relative clause; underproduction; overproduction; academic writing

**Introduction**

The relative clause (RC) is an important feature in language grammar and its importance for the language learner can hardly be overemphasized. It is a universal feature that has attracted considerably widespread attention amongst researchers due to its distinctive syntactic behaviors and unique forms, thereby, making relative clause acquisition one of the most interesting areas of study for linguistic, psychological or pedagogical purposes (Izumi, 2003). However, although universal in many languages of the world, the RC tends to vary from language to language in terms of the position of the relative clause with respect to the head noun, the ways RCs have to be marked, and the position of prepositions in the RC, among other things (Marefat & Rahmany, 2009). It is this variability that has been the main source of confusion and errors in the process of RC acquisition among second language learners.

In foreign language learning the ability to achieve a certain level of autonomy when it comes to the overall mastery of the language is of paramount importance, hence a growing number of studies have placed much emphasis towards the acquisition and production of this complicated and problematic feature, the relative clause. Thus, with English increasingly becoming the lingua franca it has changed countless foreign students’ learning experiences in that they must have a good command of convectional skills in the academic environment (Hyland, 2002). Accordingly, more and more attention has been placed on the understanding of how foreign language
learners acquire such complicated grammatical features (such as the relative clause) in an attempt to improve both theoretical and pedagogical practices.

However, due to the differences between English and Chinese, and the complex grammatical attributes such as restrictive or non-restrictive clause, among others, the relative clause has been considered to be one of the most difficult areas of English to master for Chinese as a foreign language (EFL) learners (Cho & Lee, 2016). Existing studies have reported that due to these difficulties Chinese (EFL) learners eventually end up producing significantly fewer relative clauses compared to other second language learners from other linguistic backgrounds (Persian, Arabic etc.) (Li, 1996; Liao & Fukuya, 2004; Schachter, 1974).

To this end, the current study adopts a corpus-based methodology in an attempt to provide useful insights towards the understanding of Chinese EFL learners’ use and production of the restrictive relative clause in academic writing and compare them native English students. As well as, investigate further the claim that Chinese EFL students do not only underproduce restrictive relative clauses (RRCs) in academic writing, but also tend overuse some relative pronouns.

Literature Review

Differences between English and Chinese RCs

RCs are subordinate clauses modifying a nominal element. They modify the antecedent restricting the range of possible referents or provide additional information about the referents (Chang 2004). This additional information is usually marked by certain signal words that are called relative pronouns. In English relative pronoun variations include who, whom, which, that and whose while, in Mandarin Chinese the equivalent of the RC is sorely the connecting particle “de” (的). The RC can be further classified based on their semantic functions into restrictive and non-restrictive RCs. While RRCs are used as modifiers of the nominal heads and usually introduced by that or wh-relative markers, non-restrictive markers are used to modify the nominal heads which they are by themselves sufficiently identifiable and are introduced by which preceded with a comma. Moreover, the RRC provides crucial information to clarify the noun phrase (NP) it modifies, whilst the NRC provides interesting information, yet not necessary (Lock, 1996).

Since existing literature has demonstrated that RC constructions between English and Chinese differ in so many ways (Li, 1996; Liao & Fukuya, 2004), there is need to look at the variations between Chinese and English RCs in order to fully understand the dynamics at play. An example below extracted from Yip & Matthews (1991) which illustrates further on the variations between English and Chinese RCs is worth noting:

(1a) Xihuan wo de ren (subject relative clause)
    Like me REL person
    ‘The person/ people who I like.’
(1b) wo xihuan de ren (Object relative clause)
    I like REL person
    ‘The person / people who I like.’ (p. 112).

From the illustration above three major differences are apparent. Firstly, the branching direction between the two languages is different. English is a right branching direction (RBD) language, whereas Mandarin Chinese relies on the left branching direction (LBD) (Cho & Li 2016). Also apparent is the fact that, unlike English, Chinese, being a head final language, is head-final for restrictive RCs. Also, the Chinese restrictive CR is left adjoined to the noun that it is modifying and is introduced by complementizer “de” (的), while English RCs are introduced by a variety of relative pronouns and relative adverbs. Last, pronoun retention in Chinese RCs is allowed, while on the other hand, in English RCs it is not allowed (Chang, 2004). Therefore, given the completely opposite ways in which the two languages conceptualize the RC, transfer issues are bound to occur.
Thus, the above-mentioned differences in RC construction between Chinese and English RCs serve to highlight possible areas of difficulty that Chinese EFL learners may encounter in the process of producing English RCs. Moreover, the above mentioned differences further justify the rationale for choosing the restrictive RC as the target structure for the present study which lies in the fact that English restrictive RCs are difficult constructions for learners to produce, comprehend, and even imitate (Sadighi, 1994), and therefore, provide a suitable ground for further analysis.

**Background studies on the use of the English RCs among Chinese students**

Existing studies concerning the RC have reported that the problems with which English learners are generally confronted with concern the first language (L1) influence and avoidance (Li, 1996; Schachter, 1974). The groundbreaking study by Schachter (1974) on the avoidance phenomenon was first to outline the challenges that non-native speakers of English face during the process of relative clause production. In her study on the use of the restrictive RC, Schachter reported that, while Arabic and Persian speaking students produced as many RCs as native English speakers in writing, Chinese and Japanese students produced less RCs. Whilst, on the other hand, Persian and Arabic students whose mother tongues had the same post-nominal RCs as English produced about the same number of RCs in their writing as the native students. Another important finding was that Chinese and Japanese students simply opted to paraphrase or simplify the sentences in fear of making errors. But, when they used the RCs, they produced them rather more accurately. Schachter (1974) then concluded that the structural differences between the mother tongue and the target language were the cause of this avoidance behavior. However, since the study did not control for proficiency level or the frequency of the RC in the texts and lacked proof that the learners had the ability to use RCs, it was heavily criticized (Liao & Fukuya 2004).

Conversely, Odlin (1989) postulated that the underproduction of certain linguistic features did not necessarily suggest avoidance and, therefore, proposed an alternative term ‘underproduction’ (Li 1996). In an attempt to explore further on the underproduction phenomenon, Bley-Vroman & Hough (1988) in Li (1996) compared the frequency of RCs in Chinese and English using the first five chapters of an American literary work *The Great Gatsby* and its Chinese translation. They reported on the underproduction of RCs since only a third of the original English RCs were translated into Chinese. Bley-Vroman & Hough (1988) then concluded that the underproduction of English RCs by Chinese EFL students was not necessarily a product of avoidance caused by learning difficulty, but rather a product of transfer of the distribution, frequency, and function patterns from first language.

Li (1996) later took up both Schachter and Bley-Vroman & Hough’s lines of argument and postulated that there exist two kinds of underproduction, namely conscious avoidance and subconscious underproduction. According to Li (1996:p.173) subconscious underproduction referred to the situation which L2 learners underproduce certain structures in the target language, therefore, since Chinese learners only transferred those Chinese RCs that totally correspond to ‘pure’ English RCs, namely those full RCs which do not have other functions besides being noun modifiers, into their English writing, and produced fewer RCs. Hence, suggested that the term subconscious underproduction seemed to be the most appropriate term. Li (1996) further concluded that pragmatic differences rather than structural differences between English and Chinese RCs were the cause for underproduction.

To this end, studies on the use of the English relative clause among Chinese EFL learners have emphasized the need to examine not only the L2 forms that were actually produced by the learners of a foreign language, but also the L2 forms they seem to avoid using consistently (Liao & Fukuya, 2004). Since, the difficulty of the RC for Chinese and Japanese students manifested itself not in the number of errors made by these two groups of learners, but rather in the number of RCs produced, which was much smaller than that produced by other second language learners (Sadighi, 1994).
Corpus-based studies on relative clause production among EFL learners

Although previous research on the use of RCs by EFL learners has revealed interesting findings, comparative corpus-based studies are still scant. Several reports show that, despite the variability within the non-native learners’ interlanguage when producing English RCs, the results of corpus-based studies have produced a similar pattern in that Chinese EFL learners generally under produce restrictive RCs (Biber et al., 1999; Cho, 2009; Cho and Lee, 2016). A study by Biber et al. (1999) compared the use of relative pronouns from different writing genres such as newspapers, academic papers and fictional stories. The results of the study indicated that among the relative pronouns which was the most frequently used relative pronoun and that the use of which had a frequency of 4,950 cases per one million, that had 2,400 cases, who had 1,300 cases, whose had 160 cases and whom had 130 cases. More specifically, in academic prose the restrictive relative pronoun which had a frequency of 3810 cases, indicating that which is more frequently used than that. Contrastively, fiction and news genres show a more frequent use of that than which. Biber et al. (1999) then concluded that this trend illustrated that the relative pronoun which is more formal and conservative than that. In another study Cho and Lee (2016) investigated the RCs in science and engineering journal papers under two categories: restrictive RCs and non-restrictive RCs. Findings from this study reported very interesting results. Thus, although Chinese EFL learner’s underproduced the restrictive RC, an extreme use of that over which in the all the science and engineering journal papers selected in the study was reported. To this end, Cho and Lee (2016) then concluded that the preferable use of simple sentences in science and engineering papers may have been the reason behind the less frequent use of restrictive RCs. From the results of these corpus-based studies, it can be concluded that despite the general tendency to underproduce restrictive RCs, Chinese EFL learners tend to overuse some relative pronouns in their writing which they deem more familiar. Moreover, the availability of inconsistencies within the existing literature serve as evidence that the investigation of how Chinese EFL learners use the restrictive RC is an area that deserves the outmost attention from second language acquisition researchers, for it is ripe with evidence.

However, despite the availability of several studies based on English RCs, corpus-based comparative investigation of RCs is still an area that has not been actively conducted. Therefore, in line with the inquiry of the aforementioned studies, the present study investigated the frequency of restrictive RC production using data from two corpora and made a comparison between native and non-native learners’ productions.

Methodology

This section focuses on the research questions and research instruments used to answer these questions.

Research Questions

The present study seeks to investigate the Chinese students’ production of restrictive and non-restrictive RCs in their academic writing and make a comparison between the sub-corpus of the Written English Corpus of Chinese Learner (WECCL) and the sub-corpus of the British Academic Written English (BAWE). The research questions are as follows:

1) What is the frequency in the use of restrictive relative clauses in academic writing of native and non-native English learners in the WECCL and the BAWE corpus?

2) Do Chinese EFL learners underuse or overuse relative pronouns in their academic writing? If so, what are the possible reasons for it?

Instruments

Guided by the above-mentioned research questions, the current study adopted the corpus-based methodology because of the several advantages associated with it. For instance, the corpus approach is a valuable tool which assists the researcher in dealing with language uses in real contexts. More so, the availability of large compilations of authentic language samples stored and accessed electronically allows researchers to extract authentic examples of uses of certain lexical items from a vast amount of data in a short while (Flowerdew, 2009). Also due
to variability among the results reported by previous studies concerning the RC modern studies have turned to the corpus-based methodology as the most favorable way to elicit more natural learner language samples. Consequently, over the past decades, the corpus-based approach has been continuously gaining interest in the field of linguistics as the most favorable method to use when investigating features of language, resulting in the establishment of several corpora from native and non-native English speaking countries such as China, Malaysia, Britain, America and many more.

The current research adopts corpus linguistics as its methodology(using the British Academic Written English and the Written English Corpus of Chinese Learner), whose value to linguistic investigations seem to have been convincingly demonstrated in corpus studies in recent years (Biber et al., 1999; Hyland, 2002).

**British Academic Written English (BAWE)**
Employment of the BAWE corpus is one of the two ways that the researcher uses in order to provide an analysis of recurring language patterns, which are based on empirical data rather than introspection or through elicitation techniques (Flowerdew, 2009). BAWE corpus is a collection from the project, ‘An Investigation of Genres of Assessed Writing in British Higher Education’. The project was funded by the Economic and Social Research Council from the year 2004-2007. This corpus contains assignments that were collected from various universities in the UK including Warwick University, Reading University, Oxford Brookes University, and City. These assignments were judged by the subject tutors and as standard productions in their areas. Therefore, they are accredited representative of written assignments of students across UK universities. It contains just under 3000 good-standard written student assignments with length range from 500 to 5000 words. With a total of 6,506,995 words, it is a medium-large size corpus. For this reason, this corpus will provide a more detailed picture when it is compared to other corpora.

**Written English Corpus of Chinese Learner (WECL)**
This corpus was developed by Nanjing University and Foreign Language Teaching and Research Press from the year 2003-2005. The WECL has one-million-word collection of written samples, from first year to fourth year English majors from nine universities in China. The writings in this corpus have two different types, in-class (limited) and after-class (unlimited) argumentative, narrative and expositive writings, each containing 200-800 words. The WECL has 3578 articles with 1,186, 215words and it is a tagged corpus.

**Data analysis and Procedure**
The study followed a four-step procedure in the process of collecting and qualitatively and quantitatively analyzing the data. First, in order to identify the relative clauses in both corpora the frequencies of use of the relative pronouns *that* and *which* were ran in the AntiCoc 3.5.0 software and recorded. Second, since the size of the corpora is not exactly the same, with sub-corpus BAWE having a corpus size of 255 essays with a total of 16,900 words, and, the sub-corpus WECL having a corpus size of 151 essays with a total of 42,013 words, a general comparison would not be reliable enough to give a conclusive picture on the use of the RC. The researchers, therefore, selected a much smaller and equal number of words for the sample to investigate. In this case 1000 words were selected as the sample size for both corpora, AntiCoc (Edit pad pro) was used to import the samples and count them. Third, the sample was then manually analyzed so as to identify cases in when *which* was used as a restrictive employed as a relative clause and identify zero relative pronouns. The percentage of use for each type of RRC in each sub-corpus was then calculated. Lastly, comparisons between the two corpora were made in terms of frequencies and patterns of use.

**Results and Discussion**
This section presents the findings and discussion of the investigation. The presentation of results and discussion follows the order set by the research questions.
Results

The results of the restrictive relative pronoun from the sub-corpus WECCL and BAWE were analyzed as follows,

Table 1. Frequency of restrictive RCs by native and non-native English language students per 1000 words

<table>
<thead>
<tr>
<th>Restrictive Relative Clause</th>
<th>Subject R</th>
<th>Object R</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAWE</td>
<td>49 (89%)</td>
<td>6 (10.9%)</td>
<td>55</td>
</tr>
<tr>
<td>WECCL</td>
<td>11 (68.7%)</td>
<td>5 (31.2%)</td>
<td>16</td>
</tr>
</tbody>
</table>

Key: Subject R = Subject Relatives, Object R = Object Relatives

Table 1 displays the frequency of the restrictive relative clause in the two papers analyzed. The average number of sentences in sub-corpus BAWE was 20.3 words per sentence and sub-corpus WECCL had an average of 21.6 words. By comparison, it is striking to see that, Chinese students, as reported in the data from WECCL, tend to underproduce the restrictive RC in their academic writing. Thus, a frequency of 77.5% for native students in the BAWE corpora compared to a frequency of 22.5% for non-native speakers in the WECCL corpora was recorded. T-tests performed to measure the statistical differences between native English students and Chinese students’ production of the restrictive relative clauses in the two corpora presented results which show that there was a significant difference between native and non-native speakers in sub-corpus BAWE and WECCL at the level of 0.02. Therefore, with regard to the first question, it was concluded that Chinese learners underproduced the relative clause in academic writing.

However, concerning the production of both subject and object restrictive RCs, the frequency distribution of both types of RCs in the corpora presents an interesting picture. With the frequency use of the subject RC being 89% and 68.7 % respectively in the BAWE and WECCL, object RCs being 10.9% and 31.2 % respectively. Chinese learners tend to prefer object RCs possibly due to the influence of Chinese language which prefers the object head NP which is modified by the object RC.

Table 2 below presents a frequency on the use of the restrictive relative pronoun which, that, and zero pronoun. The overall results report that the RC_THAT is the most preferred in both corpora.

Table 2. Frequency of use on the restrictive relative pronoun which, that and zero pronoun in each sub-corpus

<table>
<thead>
<tr>
<th></th>
<th>BAWE Freq.</th>
<th>WECCL Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC_WHICH</td>
<td>12 (21.8%)</td>
<td>4 (25%)</td>
</tr>
<tr>
<td>RC_THAT</td>
<td>42 (76.3%)</td>
<td>12 (75%)</td>
</tr>
<tr>
<td>Zero relative pronoun</td>
<td>1 (1.8%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Key: Freq. = Raw Frequency per 1000 words

In order to present a much more detailed picture on the frequency of use of each particular relative pronoun Table 2 presents the frequency distribution of restrictive RC led by which, that and the zero-relative pronoun in the sub-corpus BAWE and WECCL. In terms of the use of the relative pronouns Chinese learners tend to use that more frequently among the restrictive relative pronouns. That shows the highest frequency with 75%, followed by which with a frequency of 25%, and zero relative pronoun with 0%. Thus, these results are contrary to the statistical results by Biber et al. (1999) who postulated that among restrictive relative clauses which was more commonly used in academic texts and constituted about 70% of the RCs that they investigated. Instead the findings from these study concur with the findings presented by Cho and Lee (2016).
Discussion
The findings of the study seem to suggest two trends with regard to the use of restrictive RCs in the WECCCL and BAWE data. Thus, the general tendency of Chinese EFL learners to underproduce the restrictive RCs compared to native speakers and the tendency to overuse *that* among other restrictive RCs in academic writing. This section discusses the results of the study as well as look into the possible reasons for the trends.

The overall tendency to underproduce restrictive RCs
The findings presented in the results section reveal that the use of the English RC among Chinese students is significantly lower than that of native English students. Therefore, implied in the results is the fact that Chinese EFL learners tend to underproduce the restrictive RCs in writing. The possible reason for this trend being the presence of syntactical differences between the branching directions of the two languages. As previously mentioned, since English restrictive RCs are head-initial whereas those of Chinese are head final, it is most likely that Chinese EFL learners would find this particular construction in the target language too complex and difficult to comprehend, and thereby eventually end up underproducing the restrictive RC.

The following examples extracted from the papers with the least frequent use of relative pronouns show possible situations in which the restrictive RC could have been produced:

*As we all know, rubbish is a treasure in the wrong palace. We are supposed to make full use of them to save the limited resource and meanwhile protect our environment. (Sample 1)*

*Thirdly it is very lonely for old people to live without their children. They can accompany with dogs or cats. They can go out for a walk with their pets. It is healthy for their body and heart. (Sample 2)*

From these examples it is evident that two types of behavior dominate here and throughout the rest of the samples. First, multiple sentences seem to be preferred rather than using the relative clause; secondly, instead of using the RC, learners eventually end up using long sentences which are separated by several commas and in some instances shorter but multiple sentences. Since Chinese learners tend to prefer this construction, they eventually end up underproducing the restrictive RC and over produce the non-restrictive RC where this style is more acceptable. Also, another possible explanation for the underproduction of RCs lies in the fact that learners may underproduce a certain structure in the TL, not because they feel difficulty in producing the structure, but because the difference between the L1 and L2 is too subtle to be noticed (Li, 1996). Thus, according to this notion the presence of these subtle pragmatic differences will influence the Chinese EFL learners to underproduce this feature without realizing they are doing so.

Similar trends of underproduction of have been reported in previous studies (Bley-Vroman & Hough, 1988; Biber et al., 1999; Cho, 2009; Cho and Lee, 2016; Liao & Fukuya 2004). Thus, the results of this study support the claim that Chinese learners underproduce the English restrictive relative clause.

The possible tendency to overuse *that*
The results from Table 2 indicate that in WECCCL there is a tendency to overuse relative pronoun *that* among Chinese EFL learners. Thus, based on the results presented, *that* constituted more than half of the total of the RRCs clauses used with a frequency of 75% in use. The possible implication for this tendency is that when it comes to restrictive RCs the use of *that* takes precedence because Chinese EFL learners tend to overuse *which* as a non-restrictive pronoun, since the structure of non-restrictive RCs is quite similar to the structure of Chinese sentences which feature run-on clauses connected by commas. Hence, when it comes to restrictive RC the use of *which* is limited.
Possible instances of over production are found in the following example:

As an English major student, pronunciation is so important that nobody can ignore it...and an e-dictionary contains words that we need in CET-4, CET-6, TOEFL, GRE, etc. so that we can focus on words that are needed for those kind of examinations. (Sample 3)

Thus, from this example the subject could have possibly used other relative pronouns to make the essay more varied instead of using the same word over and over again.

The findings from the current study concur with the results presented by Cho and Lee (2016) who reported that in CELL and JACS corpora that accounted for 97.3% and 84.3%, respectively. Although contrary to the assumption by Biber et al., (1999) who postulated that in academic writing which is the highly preferred relative pronoun than that. To this end Cho and Lee (2016) conclude that the possible reason for this current trend is that the relative pronoun that seems to have become the norm for restrictive clauses since which is no longer recommended for restrictive relative clause in the discourse community.

In summary, the outcomes of this study follow similar patterns to the results presented in the literature review, which point out that despite the general tendency to underproduce, Chinese EFL learners tend to overuse some relative pronouns more frequently than others. However, taking the entire range of restrictive RCs into account, in the current study concludes that there exists the tendency to overproduce the relative pronoun that.

Conclusion
The current investigation of restrictive RCs was particularly aimed at validating two claims. First, that Chinese learners underproduce the restrictive RC and; secondly, that Chinese learners overproduce some RCs. Based on the results and discussion presented above several insights can be drawn. Firstly, with regard to the Chinese EFL learners’ frequency of use on the English relative clauses in academic writing, there is a general tendency to underproduce restrictive relative clauses in English. As for the question whether they Chinese learners underproduce or overproduce RCs, the current study attests to the fact that as far as RC production in academic writing is concerned, Chinese EFL learners tend to resort to underproduce the restrictive RC. Secondly, compared to native English students, Chinese EFL learners tend to overuse certain restrictive RCs out of the need to avoid an unfamiliar feature. With this in mind, a clear distinction of the restrictive relative clauses with which Chinese learners are most likely to overuse is of paramount importance so as to assist second language learners and instructors on how to better understand and master this difficult feature. However, the current study is by no means complete, therefore, further discussion on the issues raised in this study are needed.

References

