International Journal of Linguistics, Literature and Translation
ISSN: 2617-0299 (Online); ISSN: 2708-0099 (Print)
DOI: 10.32996/ijllt
Journal Homepage: www.al-kindipublisher.com/index.php/ijllt

IJLLT
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Underproduction and Overproduction: Comparison of the Use of English Relative Clauses in Chinese EFL Learners' and Native Speakers' Academic Writing

Qi Zhao[1]✉, and Shuning Ma[2]
[1]*Inner Mongolia University of Technology, School of Foreign Languages, Hohhot, China*
[2]*Inner Mongolia University of Technology, School of Foreign Languages, Hohhot, China*
Corresponding Author: Qi Zhao **E-mail**: qizhao0123@outlook.com

| ABSTRACT
This study investigates Chinese ESL learners' production of English relative clauses in academic writing with corpus-based methodologies. With the help of a series of retrieval codes for Antconc software designed for the investigation, different types of relative clauses in both L1-Chinese students' and L1-English students' essays are identified for statistical analysis and textual analysis. The results validate a phenomenon reported in previous studies that Chinese ESL learners generally underproduce relative clauses. Meanwhile, it is found that they especially underproduce finite restrictive relative clauses and non-finite present participial relative clauses. On the other hand, we unexpectedly find they overproduce non-restrictive relative clauses compared with native speakers. The analysis of some representative examples reveals that their behaviours could be attributed to the following factors: 1) the insufficient mastery of the grammatical structure of relative clauses; 2) the way they organize information; 3） the transfer from Chinese "run-on sentences. These findings could provide insights for ESL teachers to understand L1-English students' problems with producing relative clauses and better teach them how and when to use certain types of relative clauses.

| KEYWORDS
English relative clauses, Second language acquisition, Chinese ESL learners, Underproduction, Overproduction,

ARTICLE INFORMATION
ACCEPTED: 01 March 2025 **PUBLISHED:** 24 March 2025 **DOI:** 10.32996/ijllt.2025.8.3.31

1. Introduction

The English relative clause (RC) is a subordinate clause modifying the preceding noun phrase (also called the antecedent). It is typically introduced and marked by a relativizer (a relative pronoun or adverbial), which takes the position of the relativized NP (noun phrase) in the clause and anaphorically refers to the antecedent. The RC is considered one of the most challenging grammatical structures to acquire for ESL (English as a Second Language) learners whose first languages (L1) have typologically different noun-modifying clauses (Yip & Matthews, 1991; Gao, 2014; Cho & Lee, 2016).Chinese ESL learners' acquisition of RCs receives particular attention as Chinese RC-like noun-modifying clauses are different from English RCs in both the word order and the way of marking. They are neither post-nominal (instead, they are before the nouns) nor marked by relativizers (instead, they are marked by an attributive marker "*de*")(see example 1) (L. Li, 1998).

(1) *The book that is on the desk belongs to me.*
Shuzhuo shang de shu shi wode.
desk on attributive marker book be mine

The phenomenon that Chinese ESL learners produce substantially fewer RCs in writing was first reported in Schachter's study (1974). Schachter speculated that Chinese learners tried to "avoid" using RCs to make fewer errors as they found it difficult to be transformed from their L1 counterparts (Schachter, 1974). Chinese ESL leaners' production of RCs has been received continuous

attention, for understanding how second language learners acquire and produce such complicated grammatical structures under potential negative interference of L1 (Yip & Matthews, 1991; J. Li, 1996; Chan, 2004, 2010; Jin & Qiao, 2010; Yan, 2011; Parkinson & Musgrave, 2014; Gao, 2014; Chandavengerwa & Matende, 2020).

However, only a few studies provided quantitative data from sufficiently large samples to validate that Chinese ESL learners' underproduction of RCs is a universal but not merely individual phenomenon. Besides, non-typical RCs (such as non-restrictive RCs and non-finite/reduced RCs), which have different pragmatic functions or ways of relativizing, have not been considered in most studies. Although various factors and hypotheses are proposed to explain this underproduction phenomenon, they have not been comprehensively examined with a qualitative analysis of learners' texts. Given these concerns, the present study adopts a corpus-based methodology to quantitatively and qualitatively analyze different types of RCs produced in Chinese ESL learners' and native speakers' writing to investigate and understand this phenomenon.

2. Literature review

2.1 The psychological and linguistic factors to explain underproduction of RCs

In previous studies, L2 learners' less frequent use of certain linguistic structures has been interpreted as conscious avoidance, ignorance of the linguistic knowledge, and subconscious underproduction caused by unawareness (Schachter, 1974; Kleinmann, 1977; Zhao, 1989; Li, 1996).

The notion of "avoidance" was proposed by Schachter (1974) and enjoyed a classic status in the field of SLA. She analysed compositions written by Persian, Arab, Chinese, and Japanese students, and found that Chinese and Japanese students made fewer errors in the RCs but also produced fewer numbers of RCs in total. As Chinese and Japanese merely have prenominal noun-modifying clauses in their L1 while Persian and Arab have postnominal RCs as English, Schachter suggested that Chinese and Japanese students could find it more challenging and thus tried to avoid English RC to reduce the chance of making errors.

The avoidance behaviour was re-interpreted in Kleinmann's study (1977. They argued that the precondition of "avoidance" is that learners must have appropriate knowledge of the linguistic feature they choose not to use; otherwise, the non-use or the less frequent use would just indicate ignorance. In his study, he found that the non-use and less frequent use of the target structures among his subjects were not correlated to their level of knowledge but to their tendencies to avoid failure. Therefore, he suggested it was reasonable to consider their behaviour as "avoidance".

Li (1996) introduced a broader concept that covers "conscious avoidance" and "subconscious underproduction". Li (1996) argued that conscious avoidance presupposes L2 learners are "aware of the need" to use the target structure. To find out if Chinese ESL learners "consciously" avoid English relative clauses, Li (1996) asked his subjects to complete sentence-making tasks to test their ability to produce RCs. After the tasks, he asked the subjects if they had ever thought about using RCs in the tasks in which they did not do so. None of the subjects claimed they were avoiding RCs and worried about making errors. Instead, they just felt that other devices were more suitable than RCs in that circumstance (Li 1996). Then he asked the subjects to re-do similar tasks with an explicit requirement that "each sentence should contain a relative clause". Interestingly, this time, all participants correctly produced RCs for each sentence. Therefore, Li suggested the RCs-underproduction observed in the first test was subconscious and due to lack of awareness but not inability. Although this is a plausible explanation, the causes of this lack of awareness are yet to be explored.

2.2 The accessibility of different types of RCs and underproduction

Keenan and Comrie (1977) proposed an easy to difficult accessibility of RCs which depends on the position of the relativized NP in the subordinate clause called Noun Phrase Accessibility Hierarchy (NPAH): SU (subject) > DO (direct object) > IO (indirect object) > PREP (object of a preposition) > GEN (genitive case) (Keenan & Comrie, 1977). Yip and Matthews applied this hypothesis to explain a Chinese student's performance on RCs. They spotted some anaphorically related sentences where RCs are expected but also found that this student was able to produce RCs in other contexts (example 2). They suggested that this was the exact case indicted by NPAH, as the RCs that the student avoided (GEN and PREP) were the more difficult types in the hierarchy, while the RCs he produced were the easiest ones (Yip & Matthews, 1991).

(2) *a. Rome is under the military rule of Caesar. <u>And Caesar's</u> ambition is more and more obvious...*
(The underlined part could be relativized by "whose" to form a GEN RC)
b. This can be shown by looking at Caesar's words. <u>In his words,</u> he shows his contempt...
(The underlined part could be relativized by "in which" to form a PREP RC)

c. … against Caesar, the one who will probably become the king of Rome. (SU)
d. Except Brutus who actually rebels against Caesar for the common good… (SU)

Jin and Qiao (2010) particularly examined the use of non-restrictive relative clause (NRCs) in Chinese students' compositions with a hypothesis that NRCs would be even more problematic than restrictive relative clauses (RRCs) as the former does not exist in Chinese "at all". Their analysis of NRCs' frequencies suggested elementary Chinese ESL learners tended to underproduce NRCs while the intermediate and advanced groups did not. Therefore, they suggested that "the effect of L1 transfer is diminished with the increase of the proficiency level" (Jin & Qiao 2010, p.120). Interestingly, Chandavengerwa and Matende Matende (2020) proposed an opposite view that the structure of NRCs is similar to the structure of Chinese "run-on clauses connected by commas".

The non-finite/reduced RCs are the other type of RC considered to be absent in Chinese (Li, 1996). Chan's study (2010) on the errors made in RCs by Chinese ESL learners revealed they were incompetent in constructing non-finite RCs. He observed some erroneous RCs that seemed to be the mixture of finite and non-finite RCs (example 4). Chan suggested that a possible reason why the students construct such sentences was that they are confused with the rules that a finite RC requires a relativizer and a finite verb, whereas a non-finite RC removes the relativizer and includes a non-finite verb (2010, p.206).

(4) *a. You are the first come to Hong Kong.* (Finite RC: You are the first who come to Hong Kong.)
b. I have a large family which including grandmother. (Non-finite RC: I have a large family including grandmother.)

Considering the above, the present study will not only investigate the overall production of RCs but also the production of RCs of different relativized positions, embedded positions, and the way of relativizing.

3. Research questions and Methodology

3.1 The research questions

The present study intends to re-examine the use of RCs in the compositions of Chinese learners to address the following questions:
1) Do Chinese ESL learners investigated in this study underproduce RCs in general compared with native speakers?
2) Do Chinese ESL learners tend to underproduce certain types of RCs, considering RCs of different relativized positions (NPAH) and the way of relativizing (RRCs, NRCs, and non-finite/reduced RCs)?
3) If they do underproduce RCs or certain types of RCs, what could be the plausible explanations (considering conscious avoidance, ignorance of the linguistic knowledge, and subconscious underproduction)?

3.2 Methodology

To investigate the underproduction phenomenon, analysing the real language data of learners or conducting tests on learners are the two main methods adopted in the previous studies. Yip and Matthew's study (1991) and Chan's studies (2004, 2010) analysed compositions written by Chinese students and took certain anaphorical and redundant sentences or ungrammatical RC-like sentences as evidence of their inability. However, their studies merely focused on individual cases and did not involve a sufficiently large sample. Li's study (1996) and Gao's study (2014) both used sentence-making tests to examine their subjects' competence to produce RCs and quantitatively analyzed their scores. A problem with their designs is that they merely contrast the subjects' performances with the standard answers without any comparison with native speakers.

Jin and Qiao (2010), Yan (2011), and Chandavengerwa and Matende (2020) all conducted a corpus-based methodology and involved a relatively large collection of writings, respectively consisting of 37123 words, 120000 words, and 20000 words. They compared the frequencies of RCs in Chinese ESL learners' compositions with those in the native speakers'. While Jin and Qiao did not reveal their method of identifying and counting RCs, Yan and Chandavengerwa and Matende provided an efficient method to deal with such a mass of data: take relative pronouns as the prompts of RCs and retrieve them in the AntConc software. However, this method has a significant flaw: the pronouns used as relativizers (which, that, who, etc.) do not necessarily indicate RCs; they could be used in interrogatives or substantive clauses. Considering this flaw, Yan sorted out pronouns that did not indicate RCs were sorted out after the retrieval, while Chandavengerwa and Matende appeared to be not. We compared their statistics with Parkinson and Musgrave's study (2014), in which all the RCs were manually identified and counted. Table 1 shows that the RC frequencies calculated in other studies were all around 10, whereas those in Chandavengerwa and Matende's study (2020) were much higher (16 and 55), which implies that this flaw could cause the overcounting of RCs.

Table 1. The comparison of statistics in previous studies using corpus

Statistic resource	Subject Group	Frequency(per 1000 words)
Parkinson and Musgrave, 2014	IELTS 6.5- group	6.9
	IELTS 6.5+ group	10.56
Jin and Qiao, 2010	Chinese students	6.82
	native speakers	12.42
Yan, 2011	Chinese HK Group	7.65
	British Group	8.85
Chandavengerwa and Matende, 2020	College students in China	16
	College students in the UK	55

Therefore, the present study revised the method and designed a series of codes composed of symbols in AntConc and CLAWS5 POS-tags (Table 2) as the prompts for the retrieval. The design of the codes is based on the key grammatical structure of different types of RC (Table 3). For RCs relativized by Wh-pronouns or Wh-adverbs, the key structure is "NP+ Wh-relativizer". For the RCs whose relativizers are omitted (Zero-RCs), the key structure is "NP + the subordinate clause (subject + verb)". For the non-finite RCs, it is "NP+ participial". The code is the linear combination of the corresponding tags of these grammatical components. Before the data processing, each code has been tested and proven to be feasible.

Table 2. The symbols and CLAWS5 POS-tags utilized in the retrieval codes

Symbols/Tags	Grammatical component	Tags	Grammatical component
+	Zero or one character	PUR	right bracket
@	zero or one word	AT0	article
N++	covers all forms of nouns	PNP	personal pronoun
DTQ	Wh-determiner	V++	covers all forms of verbs
PNQ	Wh-pronoun	AV+	covers all types of adverbs
AVQ	Wh-adverb	VVG	ing-particles of lexical verbs
CJT	that	VBG	ing-particles of 'be'
PRP	preposition	VVN	past particles of lexical verbs
PUN	general punctuation		

Table 3. The retrieval codes of different types of RCs

Types	The key grammatical structure	Codes
Restrictive Wh-RCs	NP which/whose …	N++@DTQ
	NP preposition which/whose …	N++@PRP@DTQ
	NP who/ whom …	N++@PNQ
	NP preposition whom …	N++@PRP@PNQ
	NP when/where/why …	N++@AVQ
Non-restrictive Wh-RCs	NP, which; Np (which/whose …	N++@PUN/PUR@DTQ
	NP, preposition which/whose; Np (preposition which/whose …	N++@PUN/PUR@PRP@DTQ
	NP, who/whom …; NP (who/whom …	N++@PUN/PUR@PNQ
	NP, preposition whom …	N++@PUN/PUR@PRP@PNQ
	NP, when/where/why …	N++@PUN/PUR@ AVQ
Restrictive That-RCs	NP that…	N++@CJT
Restrictive Zero-RCs	NP (zero-relativizer) subject verb…	N++ @N++@V++
		N++@AT0@N++@V++
		N++@PNP@N++@V++
	NP (zero-relativizer) subject adverb verb…	N++ @N++@AV+@V++
		N++@AT0@N++@AV+@V++
		N++@AT0@PNP @AV+@V++

Non-finite RCs	NP V-ing …	N++@VVG
		N++@VBG
	NP V-ed …	N++@VVN

The written texts investigated in the present study are selected from the British Academic Written English corpus (BAWE). The BAWE corpus consists of 2,897 academic assignments written by college students studying in three British universities (Warwick, Reading, and Oxford Brookes) with various L1s. We selected 30 assignments of L1-Chinese students (86908 words) and 30 of L1-English students (82524 words) from the BAWE to construct two sub-corpora (Table 4), with the size that is larger than samples in any other previous studies. The original texts are firstly annotated by a free web tagger with CLAWS 5 POS tags, then imported into AntConc (the license of AntConc 3.5 was issued by AntLab Solutions and the permission of non-Commercial was granted) for the retrieval. The returned concordances are finally imported into Excel worksheets for manual checking and categorizing.

Table 4. The composition of L1-Chinese and L1-English sub-corpus

	L1-Chinese Sub-corpus			**L1-English Sub-corpus**		
Discipline	Biology	Engineering	Sociology	Biology	Engineering	Sociology
Number of authors	6	10	4	10	10	8
Number of texts	10	10	10	10	10	10
Number of words	18,565	30,560	37,783	18,609	30,009	33,906
Average length	1,687	2,778	3,434	1,691	2728	3,082
Total number of words	86908			82524		

4. Results and discussion

4.1 The underproduction of RCs in general

To answer the first question, we count RCs in each assignment and calculate their frequencies as clauses per 1,000 words. The mean RC frequency of L1-Chinese sub-corpus is 9.61 while that of L1-English sub-corpus is 12.94 (Table 5). Levene's test shows that RC frequencies of the 30 assignments in L1-Chinese sub-corpus and those of the 30 in L1-English sub-corpus are qualified for t-test. The result of the t-test shows the frequency of RCs in the L1-Chinese sub-corpus is significantly lower than that in the L1-English corpus, which indicates that L1-Chinese students investigated in this study do generally underproduce RCs.

Table 5. The frequencies of overall RCs in the L1 Chinese and the L1 English corpus

Group	**Mean**	**SD**	**Levene's test for equality of variances**	**T-test for significance**
L1-Chinese	9.61	3.25	Equal variances assumption is accepted (p= 0.33)	The difference is significant (p= 0.003)
L1-English	12.94	5.06		

4.2 Comparison of the frequencies of different types of RCs in the two sub-corpora

We then calculated the frequencies of different types of RC according to their ways of relativizing for comparison. As presented in table 6, though, the mean frequencies of both finite and non-finite RCs in the L1-Chinese sub-corpus are lower than that in the L1-English sub-corpus (5.97 versus 7.36; 3.73 versus 5.60). Though the t-test shows the difference of the overall finite RCs is not significant (p=0.086), a detailed analysis of the sub-types reveals significant differences. These differences suggest that L1-Chinese students underproduce finite restrictive RCs, especially "That-RCs" (p=0.006) and "Wh-RCs" (p=0.014); unexpectedly, they overproduce non-restrictive "Wh-RCs" (p=0.024). For non-finite RCs, the mean frequencies of both past participle RCs and present participle RCs in the L1-Chinese corpus are lower than that in the L1-English sub-corpus (2.78 versus 3.58; 0.74 versus 1.56), while only the difference of present participle RCs is significant (p=0.009).

Table 6. The frequencies of different types of RCs according to relativizers/ways of relativizing

RC-types	L1 Chinese (mean)	L1 English (mean)	T-test for significance
Total finite RCs	6.05	7.36	No (p=0.086)
Restrictive Wh-RCs*	2.09	3.17	Yes(p=0.014)
Non-restrictive Wh-RCs*	2.27	1.36	Yes(p=0.024)
That-RCs*	1.23	2.29	Yes (p=0.006)
Zero-RCs	0.38	0.54	No (p=0.328)
Total non-finite RCs*	3.53	5.33	Yes (p=0.022)
Past participle RCs	2.78	3.75	No (p=0.155)
Present participle RCs*	0.74	1.59	Yes (p=0.009)

For the finite RCs, the number of RCs of different relativized positions is also counted. We did not do the t-test since the frequencies of the RCs at some positions in a single text are quite low. Table 7 presents the average frequency of RCs of different positions in the two sub-corpora. As expected by NPAH (subject > direct object > indirect object > object of a preposition > genitive case), both L1-Chinese and L1-English students produce more RCs of easier positions in the hierarchy. And except for the genitive position, the frequency of each type in the L1-Chinese corpus is lower than that in the L1-English corpus. The exception seemingly suggests that L1-Chinese students overproduce RCs in the genitive position.

Table 7. The frequencies of different types of RCs according to relativized positions

RC-types/the relativized position	L1 Chinese (mean)	L1 English (mean)
Subject	4.52	4.77
Object of a verb	0.77	0.87
Object of a preposition	0.67	0.76
Genitive	0.13	0.08

4.3 The underproduction of finite RRCs and the overproduction of RCs in the genitive position

With respect to the results in 4.2, we analyse the essays written by L1-Chinese students in which finite RRCs are rarely used or not used at all (in which the frequency is lower than 1) to find possible explanations for their underproduction. According to Biber et al. 's study based on *Longman Spoken and Written English Corpus* (1999), elaborated post-modification is most commonly used for the first mentions of the reoccurring noun phrases in academic prose. Li (1996) also assumed that RCs are very likely to be used in definitions of objects. Therefore, we sorted out the sentences in which the reoccurring terms first occur in the essays for analysis. It is found that in L1-English students' essays, they tend to define the terms by indicating the "broad category" or the "nature" of the terms and then confining them with post-modifying RRCs (example 5).

(5) *a. The stock turnover ratio* ***is an efficiency ratio which*** *measures how much stock is left as part of the company's assets. (0228e)*
b. Carabidae ***are arthropods that*** *occur commonly around the world… They* ***are ground beetles that*** *belong to the family Coleoptera… (6011g)*

By contrast, the L1-Chinese students who rarely use RRCs went straight to the description of the feature of the term without indicating its category or nature, which a simple sentence is competent with (example 6). If taken the "broad category + post-modification" format, 6a could be "*DNA and RNA* ***are two molecules that*** *have very similar structures*", 6 b could be "*hypothesis testing* ***is a method that is used*** *to compare…*", which would also involve RRCs. The comparison of these two sets of examples seemingly implies that the production of RRCs is relevant to the way the writer introduces or explains the academic terms.

(6) *a. DNA and RNA have very similar structures. They both consists of a ribose sugar, phosphate group and…(0100b)*
b. Hypothesis testing is used to compare observed and expected results, or to compare two or more sample means. (0036a)

Though we cannot simply attribute these two L1-Chinese students' (student 0100 and student 0036) choices of this way of expression to avoidance, ignorance or subconscious underproduction of RCs without personal interviews and other tests, we find

other clues in their essays. For student 0100, we find the student is fully capable of producing NRCs in the other text (example 7). The NRC has a similar grammatical structure to RRCs in example 5 but a different textual function. Unlike the defining function of RRCs in example 5, the NRC in example 7 simply adds unessential information to the term it modifies. Although the sentence is also where the term "*first order reaction*" occurs, the student seemingly does not feel the necessity to define it formally. Therefore, we suggest the non-use of RCs in student 0100's case is not conscious avoidance or ignorance, but subconscious awareness. But the necessity to use defining RRCs in such cases is another issue yet to be discussed.

(7) *This reaction* ***is a first order reaction, where*** *the rate is proportional to the concentration of the substrate. (0100a)*

For student 0036, we find they actually produce some RCs with grammatical errors in the same text (example 8). Example 8a and 8b were not returned by the Antconc when we did the RC-retrieving, as they do not conform to the grammatical structure of a "That-RC" and "Wh-RCs" nor a "Zero-RC". The two sentences constructed by the student are relativized on the position of the subject with no relativizer, whereas a "Zero-RC" only allows the relativizer to be omitted when the relativized position is the object. Though this case contributes to the statistical result of the low RC frequency, it does not imply any use or conscious avoidance of RCs. Instead, it suggests that student 0036 may not be fully competent to produce RCs but attempted to do so.

(8) *a. Bia is another type of error Ø may take place in observational studies. (0036a)*
b. There were a few confounders Ø have been considered in this study. (0036a)

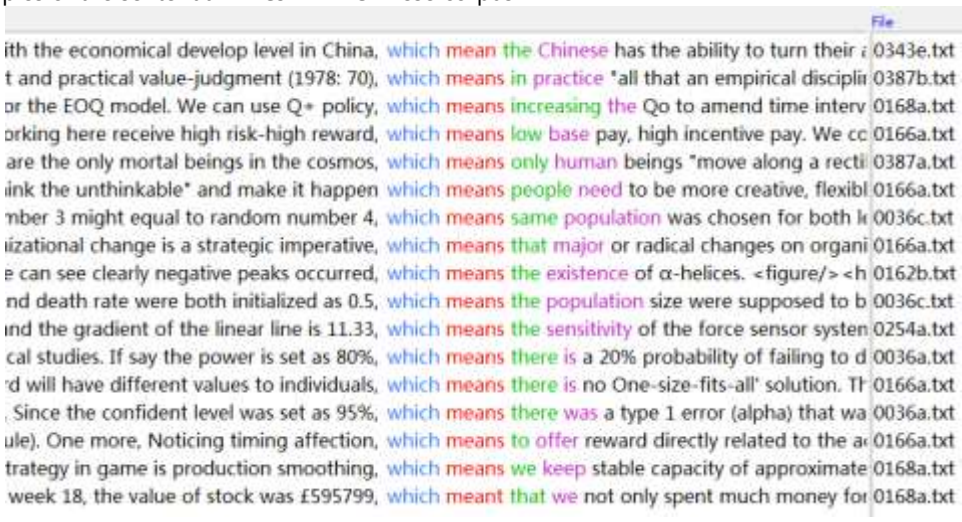
Concerning the exception that the frequency of RCs in the genitive position in the L1-Chinese corpus is slightly higher than that in the L1-English corpus, we re-examine all the "Whose-RCs" in both corpora. Interestingly, we find that 5 out of 11 "Whose-RCs" produced by L1-Chinese students modify inanimate objects (example 9), while all 5 "Whose-RCs" produced by L1-English modify animate objects (example 10). This broadening application could explain L1-Chinese students' "overproduction" of "Whose-RCs".

(9) *a. …**commodity exchange, whose** boundaries…(0319a)*
*b. …**immunoglobulins, whose** duty is…(0162c)*
*c. …new chain of **happenings whose** eventual outcomes…(0387a)*
*d. …**philosophical thinking, whose** object is…(0387a)*
*e. …**modern science, whose** spirit is…(0387a)*
(10) *a. …**women whose** case was reported to …(0214b)*
*b. …**responsible individuals, whose** health status is…(0252g)*
*c. …**moth whose** larvae mine through the leaf…(6176c)*
*d. …**Maffesoli, whose** object is to find enjoyment…(0320j)*
*e. …**Leah Betts whose** picture, …, was front page news…(0320j)*

4.4 The overproduction of finite NRCs

In contrast to the underproduction of finite RRCs, the statistical results in 4.2 indicate that L1-Chinese students produce more NRCs than L1-English students. We examined L1-Chinese students' essays with high frequencies of NRCs and L1-English students' essays with few NRCs. The textual analysis reveals two factors that contribute to the L1-Chinese students' overproduction of NRCs.The first one is the extensive use of sentential NRCs. A typical example is the highly frequent sentential NRC beginning with "*which means*" accruing in their essays (Figure 1). The sentential "*which means*" NRCs could be translated to Chinese as two independent sentences, and the subordinate clause would be "*zhe (**This**) yowzah (means)…*". Though sentential NRCs and their "Chinese equivalents" have this structural difference, it seems not to cause difficulties for L1-Chinese students to produce them.

Figure 1. The examples of the sentential NRCs in L1-Chinese corpus



Chandavengerwa and Mantende propose a view that non-restrictive RCs are similar to a kind of Chinese structure with "run-on clauses" (2020, p.128). We do observe some run-on clauses that might be directly transmitted from their Chinese forms, which are two sentences with different subjects connected by the comma or the conjunction "and" (example 11). The second subject "*this*" actually refers to the first whole sentence. The sentential NRC and this "run-on" structure have a similar textual function that it relates the two sentences and a similar construction that a pronoun at the beginning of the second sentence refers to the first. L1-Chinese students who tend to use "run-on clauses" to connect information in Chinese may find the NRCs a useful device to make the information coherent while keeping the sentence grammatical, therefore producing them frequently.

(11) a. *Internet advertising spent was £ 1,366.4 million in 2005**, this is** a 7.8% share of all advertising spend*. (0353a)
b. *…the two soil samples within the same quadrat seem to have substantially different salt levels **and this** lead to …* (6251e)

The second factor is the production of NRCs without the comma, like example 14 by some L1 English students. In example 14, as '*EMS*' is fully identifiable alone, the following RC is not identified to comment on the behaviour of "*Centrica's businesses*". Though this RC is not separated by a comma, it is actually an NRC. Such "unseparated NRCs" are found in 5 of 6 L1 English students' assignments where the standard form of NRCs is absent. This results in our miscount of the actual number of NRCs produced by L1-English students and thus contributes to L1-Chinese students' relative overproduction.

(12) *Centrica's businesses not adopting ISO 14001 might fail with their EMS **which** may lead to some breach of government consents…(0222e)*

4.5 The underproduction of present participial RCs

For the non-finite/reduced RCs, the results shown in Table 6 in 4.2 indicate that L1-Chinese students produce significantly fewer present participial RCs than L1-English students. The textual analysis reveals two situations that may explain their underproduction of present participial RCs. The first situation is that the use of the present participial form of verbs as post-modifiers is related to certain themes. For example, "*coding*" and its synonyms were used to describe the function of a gene several times in the L1 English sub-corpus (example 12). The heavy use of these structures in L1-English students' essays (9 times) contributes to the overall high frequency of present participle RCs, whereas none of these verbs occur in L1-Chinese students' essays. But, it is noted that "gene," which is the theme that triggers these RCs, is not involved in any of L1-Chinese students' essays either, even though we do include essays for biology courses.

(12) *a. Where thr = **gene coding** for threonine, arg = **gene coding** for arginine…*
*b. …as the mini-white **gene encoding** red is lost… (0035a)*
*c. **Genes expressing** neck and tail sheath protein… (0141h)*

The second situation is that the verbs that tend to be constructed as present participle RCs by L1-English students are used in the full finite RCs or substituted by prepositions with the same functions in L1- Chinese students' essays. We find that "*allow*", "*consist*

of," and "*affect*" are constructed in present participle RCs in L1-English corpus (example 13), while the same verbs are constructed in finite RCs in L1-Chinese students' essays where the present participle RCs are absent (example 14).

(13) *a. … have also built a good reputation* ***allowing*** *them to sell goods… (0228e)*
b. … a manometer is a device ***consisting of*** *columns of liquid in a U-tube…(0023d)*
c. …the economic environment is dominant role ***affecting*** *the impetus…(0343e)*
(14) *a. … by carrying out a longitudinal study* ***which allowed*** *him to follow… (0350d)*
b. The apparatus was called Analogue Experimental Transducer, ***which consisted of*** *a… (0254a)*
c. …identify significant factors ***which affect*** *forecasts… (0168a)*

The present participle RCs started with "surrounding" to is also found several times in the L1-English corpus (example 15). By contrast, we find some prepositions, such as "around", are used to perform the same role in a similar context in L1-Chinese students' essays (example 16). Interestingly, although present participle RCs are absent in the students' essays mentioned above (examples 14 and 16), we find these students (student 0350, 0254, 0168, 0320, 0316) at least 8 past participle RCs in their essays. This also accords with the results in 4.2 that the difference between the frequency of past participle RCs in L1-Chinese and L1-English corpus is not significant (p=0.155). This finding implies that past participle RCs may be easier than present participle RCs for L1-Chinese students to acquire.

(15) *a. … by knowledges and* ***discourses surrounding*** *sexuality, such as… (0401a)*
b. …the boundaries of medicalisation and the moral ***discourse surrounding*** *health (0252g)*
c. The ***debates surrounding*** *the topic of delinquency are complex… (0320i)*
(16) *a. A* ***discourse upon*** *the origin …(0319b)*
b. …a critical ***discussion of*** *how the media in general… (0316c)*
c. …whole ***discussion around*** *"media imperialism" …(0316c)*

5. Conclusion

The current study investigates the use of RCs in academic assignments written by Chinese ESL learners and native English speakers to explore if Chinese ESL learners underproduce English RCs and why. The study adopts a corpus-based methodology and examines a sample of students' writings with the largest size among previous studies and designs a series of retrieval codes for Antconc software to identify and count different types of RCs more precisely. Based on the statistical analysis, we suggest Chinese ESL learners in our study generally underproduce RCs and specifically underproduce finite RRCs and non-finite present participle RCs. In contrast to the overall tendency of underproduction, Chinese ESL learners tend to overproduce NRCs, and some of them tend to overuse "*Whose-RCs*".

Further textual analysis reveals some factors that may explain the underproduction and the overproduction. First, some Chinese ESL learners may not fully master the grammatical structures of certain types of RCs, thus producing erroneous RCs, which were not identified and counted. Secondly, Chinese ESL learners generally tend to utilize NRCs to provide additional information for the antecedents instead of RRCs to define or constrain the terms. They may have different thoughts from native English speakers on if certain information is defining or complementary to the antecedent even under similar contexts. Third, the overproduction of sentential NRCs may be transferred from the use of the "run-on" sentences in their L1. From the analysis of some representative examples, we do not consider that the less use of RCs by Chinese ESL learners is due to conscious avoidance for making fewer errors but is more likely caused by inability and their choices on ways to organize the information. These findings could be helpful for ESL teachers to understand Chinese students' problems with producing RCs and better instruct them. But it is worth noting that the Chinese students investigated are at least upper-intermediate L2 speakers of English as they were admitted and studied in UK universities. Chinese students with lower English levels may have different performances and other issues.

The current study also raises a question that needs further exploration: does the underproduction phenomenon found among Chinese ESL learners actually affect the quality of their writing? As least for non-finite present participle RCs, we find the students who do not use them at all apply other "compensatory" which also convey their meanings. In other words, under what situations is a certain type of RC necessary or the better choice for writers? For such a complex and challenging grammatical structure ,ESL learners, would master them better if they knew not only how to produce but also when to use them.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Biber, D., Johansson, S., Leech, G., Conrad, S. and Edward Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow. Longman.
[2] Chan, A. Y. W. (2004). Syntactic Transfer: Evidence from the Interlanguage of Hong Kong Chinese ESL Learners. *The Modern Language Journal*, *88*(1), 56–74. https://doi.org/10.1111/j.0026-7902.2004.00218.x
[3] Chan, A. Y. W. (2010). Toward a Taxonomy of Written Errors: Investigation Into the Written Errors of Hong Kong Cantonese ESL Learners. *TESOL Quarterly, 44*(2), 295–319. https://doi.org/10.5054/tq.2010.219941
[4] Chandavengerwa, N. M., & Matende, T. (2020). Underproduction or Overproduction? A Corpus-based Study of Chinese EFL Learners' Use of English Relative Clauses. *International Journal of Linguistics, Literature and Translation, 3*(3), 122–130. https://doi.org/10.32996/ijllt.2020.3.3.13
[5] Cho, D. W., & Lee, K. (2016). English relative clauses in science and engineering journal papers: A comparative corpus-based study for pedagogical purposes. *Ampersand*, 3, 61–70. https://doi.org/10.1016/j.amper.2016.03.002
[6] Gao, Q.-Q. (2014). Chinese EFL Learners' Acquisition of English Relative Clauses. *International Journal of English Linguistics, 4*(3), p82. https://doi.org/10.5539/ijel.v4n3p82
[7] Jin, X., & Qiao, M. (2010). Interlingual Factors in Chinese College Students' Acquisition of English Relative Clauses. *Chinese Journal of Applied Linguistics*, 33, 110-126+128.
[8] Keenan, E. L., & Comrie, B. (1977). Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry, 8*(1), 63–99.
[9] Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning, 27*(1), 93–107. https://doi.org/10.1111/j.1467-1770.1977.tb00294.x
[10] Li, J. (1996). Underproduction Does Not Necessarily Mean Avoidance: Investigation of Underproduction Using Chinese ESL Learners. In L. F. Bouton (Ed.), *Pragmatics and language learning* (Vol. 7, pp. 171–187). https://eric.ed.gov/?id=ED400710
[11] Li, L. (1998). A comparison of word order in English and Chinese. *Papers and Studies in Contrastive Linguistics*, 34, 153–161.
[12] Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14, 48–59. https://doi.org/10.1016/j.jeap.2013.12.001
[13] Schachter, J. (1974). An error in error analysis. *Language Learning, 24*(2), 205–214. https://doi.org/10.1111/j.1467-1770.1974.tb00502.x
[14] Yan, W. (2011). *Relative Clauses in Hong Kong English: A Corpus-based Perspective. Chinese Journal of Applied Linguistics, 34*(4). https://doi.org/10.1515/cjal.2011.031
[15] Yip, V., & Matthews, S. (1991). Relative Complexity: Beyond Avoidance. *CUHK Papers in Linguistics*. https://eric.ed.gov/?id=ED363103
[16] Zhao, R. (1989). A Discourse Analysis of Relative Clauses in Chinese and English: An Error in "An Error in Error Analysis." *IDEAL*, 4, 105–117.