

Article

Relative clauses in written Hong Kong English and British English: A corpus-based comparison

Chi Wui Ng

Academic Unit of Social Contexts and Policies of Education, The University of Hong Kong, Hong Kong, China; ngchiwui@connect.hku.hk

CITATION

Ng CW. Relative clauses in written Hong Kong English and British English: A corpus-based comparison. *Language Exploration*. 2025; 1(1): 3081.
<https://doi.org/10.59400/le3081>

ARTICLE INFO

Received: 20 January 2025
Accepted: 28 February 2025
Available online: 17 March 2025

COPYRIGHT



Copyright © 2025 by author(s).
Language Exploration is published by Academic Research Services LLC (ARSL). This work is licensed under the Creative Commons Attribution (CC BY) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract: Knowledge of relative clauses in Hong Kong English as a variety of the English language is limited. This study aims at comparing frequencies of occurrence of relative clauses in written Hong Kong English and their counterparts in British English. *wh* relatives and *that* relatives in the Hong Kong and Great Britain components of the International Corpus of English were identified and compared with each other quantitatively. Results indicate that relative clauses appear much more frequently in written British English than in written Hong Kong English, and frequencies of occurrence of most types of relative clauses in written British English are either higher than or equal to those of their counterparts in written Hong Kong English. Given the low frequencies of occurrence of some types of relative clauses in both Hong Kong English and British English, it may not be necessary to require learners to learn those types of relative clauses in formal instruction.

Keywords: English relative clauses; Hong Kong English; British English; corpus linguistics

1. Introduction

English relative clauses (RCs) are commonly taught to ESL and EFL learners across the globe. In Hong Kong, upper primary students are expected to “use the relative pronouns ‘which, who, that, whose, where’ to link ideas or add information to a noun or noun phrase” in the English language subject [1].

Research on English RCs is rich yet predominantly focuses on standard in lieu of regional varieties of the English language. For instance, Biber et al. [2] conducted a large-scale corpus analysis of written and spoken English and discovered remarkably higher frequencies of occurrence of English RCs across written registers, ranging from informal ones like fiction to formal ones like academic prose, than in conversational English. In Hong Kong, Gisborne [3] and Newbrook [4] attempted to conduct corpus-based studies on RCs in Hong Kong English (HKE), yet these studies were qualitative in nature. They identified the morphosyntactic attributes of RCs in HKE without any quantitative components illuminating the HKE users’ overall usage patterns of English RCs or comparing the frequencies of occurrence between RCs in HKE and those in other English varieties.

The present study quantitatively compares RCs in HKE with those in British English (BrE), which is one of the “standard” English varieties. Knowledge on HKE users’ production of RCs in their written language informs second language grammar instruction by aligning English language education in the classroom with primary language data in language users’ natural language production.

2. Literature review

This section commences with a descriptive grammatical description of English RCs. Following a description of the grammar of English RCs is a review of antecedent studies on RCs in HKE as well as an account of the research gap.

2.1. Grammatical description of English RCs

The following grammatical description of English RCs, which follows the paradigm of descriptive grammar, is provided to elucidate grammatical terminologies applicable throughout the article.

Being an overt or covert subordinate clause anaphorically linked to an antecedent, a RC is a modifier of an antecedent and comprises within the structure a relativized element whose interpretation is determined by the antecedent [5]. Being postnominal clauses, English RCs follow the modified heads. An instance of an English RC is presented in (1), the bracketed part of which is an RC. Introducing the RC, the relativized element *which* is a subordinator marking the RC as a subordinate clause, which serves as a postmodifier of the underlined noun *breach*. Interpretation of such a relativized element is contingent upon the antecedent *beach* in the main clause. English RCs can be categorized by their form into *wh* relatives, *that* relatives, and bare relatives. *wh* relatives possess overt relativized elements realized occupying either the subject position or pre-nuclear position, videlicet before the subject and predicate of the entirety of the RC (ibid.). Alternatively, RCs can also be classified by the syntactic role of the relativized element in the RC or the syntactic role of the antecedent in the main clause.

- 1) We went to a beach ; [which Ali had recommended ___ ; to us].
Hewings [6]

RCs can also be classified into seven categories by the syntactic role played by the relativized element in the RC as presented in **Table 1** or five categories by the syntactic role played by the antecedent in the main clause as presented in **Table 2** with bracketed parts being RCs [5,7]. Specific restrictions are imposed on the types of RC that are permissible with distinct relativized elements. More specifically, *wh* relatives are permissible with all relativized elements while *that* relatives are impermissible in RCs with relativized genitive determiners and complements of auxiliary verbs (ibid.). Only bare relatives are permitted in RCs with relativized objects, subject complements, prepositional complements, and adjuncts (ibid.).

Table 1. Categorization of relative clause by syntactic role of relativized element.

Syntactic role of relativized element	Example
Subject (S)	The <u>man</u> <u>;</u> [who ___ <u>;</u> came to dinner] turned out to be from my home town.
Object (O)	This is the <u>letter</u> <u>;</u> [that she received ___ <u>;</u>] from the Governor.
Subject complement (SC)	Her book displays the fine sceptical intelligence of the <u>scholar</u> <u>;</u> [she is ___ <u>;</u>].
Prepositional complement (PC)	The <u>penknife</u> <u>;</u> [that he was trying to cut it with ___ <u>;</u>] was blunt.
Adjunct (A)	Do you remember the <u>day</u> <u>;</u> [we met Kim at the races ___ <u>;</u>]?
Genitive determiner (GEN)	One cannot tailor a suit for a <u>client</u> <u>;</u> [whose <u>;</u> measurements remain unknown].
Complement of auxiliary verb (AUX)	He told me to <u>design it myself</u> <u>;</u> [which I simply can't ___ <u>;</u>].

Note: Reprinted from Huddleston et al. [5].

Table 2. Categorization of relative clause by syntactic role of antecedent.

Syntactic role of antecedent	Example
Subject (S)	The <u>girl</u> _i [who ___ _i speaks Basque] is my cousin.
Direct object (dO)	I know the <u>girl</u> _i [who ___ _i speaks Basque].
Indirect object (iO)	We gave the <u>boy</u> _i [who ___ _i] broke the window a warning.
Prepositional complement (PC)	I talked with the <u>girl</u> _i [who ___ _i speaks Basque].
Subject complement (SC)	Mr. Thomas is a <u>teacher</u> [who ___ _i prepares his lessons].

Note: Reprinted from Larsen-Freeman et al. [7].

Not only can English RCs be categorized by formal attributes, but they can also be classified into integrated and supplementary RCs in terms of grammatical functions performed. Integrated RCs, also known as restrictive RCs, are integrated into the main clause and usually function as modifiers of noun phrases restricting the denotation of the modified head noun [5]. The RC in (1) *which Ali had recommended to us* is an example of integrated RC in that it restricts the denotation of the antecedent *beach*. In contrast, supplementary RCs, also known as non-restrictive RCs, are not fully embedded into the main clause and offer additional information of the antecedent in lieu of delimiting the modified antecedent (*ibid.*). In other words, an integrated RC is an integral component of a main clause, omission of which utterly changes the proposition of the entirety of the sentence, whilst a supplementary relative is a separate unit of information, which can be omitted without any drastic changes in the meaning of remaining segments of the sentence.

2.2. RCs in HKE

There has been a growing body of literature on HKE, yet it is a vexed issue whether HKE is regarded as an independent variety of the English language or simply an interlanguage [8–12]. The English language was massively spread in Hong Kong as a contact language between the Chinese and the British during its colonial period between 1842 and 1997 and has become a value-added language in the territory after 1997 albeit its use amongst Hong Kong Chinese for intra-ethnic communication is limited [11–13]. At present, there is a widespread agreement that HKE is a nativized or emergent variety by reason of one-of-a-kind and systematic linguistic attributes making it distinguishable from other varieties of English, especially in the phonological and lexical aspects [14]. The public's attitude towards the acceptability of HKE is however ambivalent, as observed in people's conception of Hong Kong as a monolingual Chinese society, conception of external norms as standards, and conception of distinctive linguistic attributes as grammatical errors reflecting learners' falling standards [8,15–18]. Such ambivalence makes it difficult for HKE to reach the phase of endonormative stabilization in Schneider's [14] dynamic model of the evolution of postcolonial Englishes, which is characterized by achievement of an autonomous status in society, albeit Hansen Edward [19] observed attributes suggesting the movement of HKE towards endonormative stabilization and even the final phase of differentiation. Despite the debate over its status, widely employed in academia to denote the variety produced by Cantonese learners of English in Hong Kong [20–26], the notion of HKE is operationalized in the present study.

Approaching English RCs produced by learners in Hong Kong from a sociolinguistics perspective, both Newbrook [4] and Gisborne [3] identified morphosyntactic attributes of RCs in HKE. Newbrook's [4] study was not exclusive to RCs in HKE, but it analyzed how the syntactic formations of RCs in different English varieties differed from one another and provided plausible explanations for such attributes. One attribute identified was the omission of subject relative pronouns, as in “*This is the who student did it.” [4], even though Hong Kong students have been taught that such a structure is non-standard or unacceptable. Another attribute identified was participial clauses with relative markers, as in “*This is the student who admitted last year.” [4], which was likely to result from learner's confusion between the finite past tense and non-finite past participle forms of verbs. Gisborne [3] built upon Newbrook's [4] study of English RCs across varieties and conducted the most corpus-based research to date on morphosyntactic attributes of RCs in HKE in a systematic fashion. In an attempt to provide evidence proving the status of HKE as an independent English variety, the study utilized an incomplete Hong Kong component of the International Corpus of English (ICE) to delineate the morphosyntactic attributes of RCs in HKE. The morphosyntactic features of RCs in HKE identified in Gisborne [3] were zero-subject relatives, participial clauses with relative markers, use of the relative adverb ‘where’ in RCs with abstract nouns as antecedents, missing prepositions, presence of resumptive pronouns, and an unclear distinction between integrated and supplementary RCs. Some of these attributes were also reported in Chan [27], who reviewed two untutored and unaided free-writing tasks completed by Cantonese-English ESL learners in Hong Kong and attested to some of the lexicogrammatical attributes of RCs in HKE identified by Newbrook [4] and Gisborne [3]. For instance, she attributed learners' production of participial clauses with relative markers to their inadequate mastery of the finite nature of RCs [27].

Gisborne [3] has been the only corpus-based study to date on RCs in HKE exclusively. In spite of its provision of a detailed description of morphosyntactic features of RCs in HKE, the study possessed two limitations. First, the Hong Kong database of the ICE had not been completed when the study was conducted. Second, a quantitative component was absent in the study. Without information on frequency counts of the morphosyntactic attributes of RCs in HKE identified, it was doubtful how prominent those attributes were. Also, without a quantitative component, the study provided no information on the usage patterns or frequencies of use of RCs amongst users of HKE. These limitations have motivated the present study.

2.3. Research problem and question

The present study is a segment of a larger study on RCs in written HKE. Analysis from the perspective of second language development was presented in Ng [25] whereas this paper presents the sociolinguistics component of the study. It fills research gaps in research on RCs in Hong Kong by following up Gisborne [3] with primary language data from a complete Hong Kong database for the International Corpus of English and quantitative analysis, which was absent in the original study, to identify quantitative attributes of RCs in HKE in comparison with their counterparts

in BrE, one of the standard varieties of the English language, with the following research question:

How do different types of RCs in written HKE quantitatively differ from their counterparts in written BrE?

Attributed to higher frequencies of occurrence of English RCs across written registers [2], the present study will pinpoint written HKE for more focused and meticulous analysis.

3. Research methodology

3.1. Research design

Corpus linguistics, which is a research methodology investigating language using corpus data and methods, was capitalized upon in the present study [28]. A corpus denotes “a finite collection of machine-readable text, sampled to be maximally representative of a language or variety” [29].

3.2. Sample

Primary language data were collected from the Hong Kong and Great Britain components of the ICE. The entirety of the ICE project, which involves construction of electronic corpora of distinct regional varieties of the English language in regions where English is spoken either as a first language or official additional language, commenced in 1988 with an aim of collecting for synchronic studies of world Englishes [30]. Each ICE component corpus comprises 500 texts of approximately 2000 words each produced after 1989, which make up a total of approximately one million words, with a common corpus design with different genres of text [30]. Authors and speakers of the texts are adults educated in English-speaking settings, who were either born in the region where the respective corpus was constructed or migrated there in their early years (*ibid.*). A host of social variables, *videlicet* sex, age, education, and occupation, were taken into consideration in the course of text selection to maximize representation of disparities in such social variables [31].

Being the British component of the ICE, the International Corpus of English—Great Britain (ICE-GB) served as a pilot project for other components of the ICE and was constructed at the Survey of English Usage, University College London between 1990 and 1998 with texts dated between 1990 and 1993 inclusive [32]. Authors and speakers of the texts are British people born in England, Scotland, or Wales who may not necessarily use “standard” or “educated” BrE (*ibid.*).

Being the Hong Kong component of the ICE, the International Corpus of English—Hong Kong (ICE-HK) was constructed under the auspices of a grant from the Research Grants Council with texts dated from the 1990s [33]. It has been the most comprehensive database of HKE ever compiled.

Only written texts in the ICE-HK and ICE-GB were analyzed in the current study. There are a total of 200 written texts with 499898 words in the ICE-HK and a total of 200 written texts with 431171 words in the ICE-GB.

3.3. Data collection procedures

Research instruments and procedures were piloted prior to the launch of the main study. Data in the pilot study comprised 10 untimed student essays in the ICE-HK (23,593 words) and their counterparts in the ICE-GB (21,959 words); the data constituted a segment of the written texts in the two corpora and accounted for about 5% of the data in the main study. Details and findings of the pilot study were reported in Ng [24].

In the main study, *wh* relatives and *that* relatives were identified in the entire datasets of ICE-HK and ICE-GB through a concordance search for all relative words and the subordinator ‘that’ with the concordance program *WordSmith* 5.0 [34]. The *wh* relatives and *that* relatives identified in the datasets constituted the data for the study.

3.4. Data analysis procedures

Once located, the RCs identified in ICE-HK and ICE-GB were syntactically classified into specific types (based on the syntactic roles of relativized elements and antecedents) prior to calculation of frequency counts. The analysis measured both the absolute numbers and percentage distributions of each RC type across the two varieties. Finally, a quantitative comparison was made between the frequencies of distinct RC types in the two components of the ICE using descriptive statistics to identify quantitative differences between them.

4. Results

The research question of the study is how the quantitative distributions of different types of RCs in written HKE differ from those in written BrE. Therefore, frequency counts of distinct types of RCs in written HKE and BrE were compared to each other.

The total frequency distribution of *wh* and *that* relatives in the ICE-HK are displayed in **Table 3**. A total of 2421 RCs are identified in the written texts, with *wh* relatives and *that* relatives accounting for 71.79% and 28.21% of the instances respectively. RCs with relativized subjects (76.13%) dominate the data while those with relativized subject complements and relativized complements of auxiliary verbs are notably scarce and even totally absent. RCs with prepositional complements (39.45%), subjects (20.57%), and direct objects (19.25%) as antecedents are the most frequent with frightfully rare occurrence of RCs with indirect objects as antecedents amongst the 2421 instances.

Table 3. Frequencies of occurrence of different types of RCs in the ICE-HK.

Syntactic role of antecedent	Syntactic role of relativized element							Total
	S	O	PC	GEN	SC	A	AUX	
S	407 (16.81)	30 (1.24)	23 (0.95)	18 (0.74)	0 (0.00)	20 (0.83)	0 (0.00)	498 (20.57)
dO	371 (15.32)	37 (1.53)	32 (1.32)	16 (0.66)	0 (0.00)	9 (0.37)	1 (0.04)	466 (19.25)

Table 3. (Continued).

Syntactic role of antecedent	Syntactic role of relativized element							Total
	S	O	PC	GEN	SC	A	AUX	
iO	1 (0.04)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.04)
PC	697 (28.79)	40 (1.65)	67 (2.77)	42 (1.73)	0 (0.00)	108 (4.46)	1 (0.04)	955 (39.45)
SC	235 (9.71)	17 (0.70)	24 (0.99)	6 (0.25)	0 (0.00)	43 (1.78)	0 (0.00)	325 (13.42)
*	132 (5.45)	10 (0.41)	16 (0.66)	5 (0.21)	0 (0.00)	13 (0.54)	0 (0.00)	176 (7.27)
Total	1843 (76.13)	134 (5.54)	162 (6.69)	87 (3.59)	0 (0.00)	193 (7.97)	2 (0.08)	2421 (100)

Note: Percentages of occurrence of different types of English RCs are in parentheses. * indicates RCs modifying antecedents with unidentified syntactic roles or serving as adjuncts in sentences.

Total frequency counts of distinct types of *wh* and *that* relatives in the ICE-GB are presented in **Table 4**. 3304 RCs are identified in the written texts in the ICE-GB. 76.51% and 23.49% are *wh* relatives and *that* relatives respectively. RCs with relativized subjects (67.52%) dominate the data and constitute the vast majority of all RCs whereas those with relativized subject complements and relativized complements of auxiliary verbs occur the least frequently. RCs with prepositional complements (46.79%), direct objects (18.37%), and subjects (17.92%) as antecedents appear the most frequently in the data with frightfully rare occurrence of RCs with indirect objects as antecedents amongst the 3304 instances.

Table 4. Frequencies of occurrence of different types of RCs in the ICE-GB.

Syntactic role of antecedent	Syntactic role of relativized element							Total
	S	O	PC	GEN	SC	A	AUX	
S	441 (13.35)	39 (1.18)	62 (1.88)	15 (0.45)	1 (0.03)	34 (1.03)	0 (0.00)	592 (17.92)
dO	402 (12.17)	71 (2.15)	96 (2.91)	18 (0.54)	1 (0.03)	19 (0.58)	0 (0.00)	607 (18.37)
iO	3 (0.09)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	3 (0.09)
PC	997 (30.18)	99 (3.00)	179 (5.42)	60 (1.82)	1 (0.03)	209 (6.33)	1 (0.03)	1546 (46.79)
SC	226 (6.84)	21 (0.64)	37 (1.12)	10 (0.30)	0 (0.00)	46 (1.39)	0 (0.00)	340 (10.29)
*	162 (4.90)	11 (0.33)	17 (0.51)	4 (0.12)	0 (0.00)	20 (0.61)	2 (0.06)	216 (6.54)
Total	2231 (67.52)	241 (7.29)	391 (11.83)	107 (3.24)	3 (0.09)	328 (9.93)	3 (0.09)	3304 (100)

Note: Percentages of occurrence of different types of English RCs are in parentheses. * indicates RCs modifying antecedents with unidentified syntactic roles or serving as adjuncts in sentences.

Overall speaking, RCs are more frequent in the ICE-GB than in the ICE-HK, with the latter exhibiting approximately 73% of the total RC frequency observed in the former. Most RC types in the ICE-GB either exceed or match their corresponding frequencies in the ICE-HK. Even though some types of RCs in the ICE-GB, such as

SGEN RCs and SCS RCs, occur less frequently than their counterparts in the ICE-HK do, the difference in frequency counts remains minimal.

On the basis of the above comparison, three findings are worth deliberation. First of all, RCs with relativized prepositional complements occur less frequently than those with relativized adjuncts in the ICE-HK but more frequently than their counterparts in the ICE-GB; frequent use of RCs with relativized adjuncts in the ICE-HK can be analyzed at length with reference to specific instances of primary language data from the corpus. Another finding is that RCs with subjects as antecedents occur more frequently than those with direct objects as antecedents do in the ICE-HK but less frequently than their counterparts in the ICE-GB do; a higher frequency count of RCs with subjects as antecedents in the ICE-HK is predominantly contributed by *wh* relatives in lieu of *that*-relatives. Last but surely not the least, RCs with relativized subject complements are totally absent in the ICE-HK. That said, only does a corpus collect a limited sample of primary language data, and only positive evidence in lieu of any negative evidence is available. Therefore, by no means does the absence of a language feature in a corpus necessarily imply a lack of knowledge of such a language structure in the language.

5. Discussion

Frequencies of different types of RCs in written HKE and BrE were compared. The discussion below focuses on a quantitative comparison between all types of RCs in lieu of merely *wh* relatives or *that* relatives. Key findings as well as possible explanations of the quantitative differences will be deliberated at length.

First and foremost, RCs are markedly more frequent in the ICE-GB than in the ICE-HK. This appears to show that BrE users use RCs more extensively than HKE users do. First, users of HKE may possess a less robust command of English RCs than BrE users do. Learners' consistent production of a certain language structure entails their possession of knowledge on the structure, so learners' higher level of development of a structure will effectuate its higher frequency of occurrence in naturalistic production data [35]. Most HKE and BrE users are foreign language learners and native speakers of English respectively. This distinction likely explicates why BrE users, who probably have more and earlier exposure to the English language, possess better mastery of English RCs than HKE users do and thereby produce the structure more extensively in their language output. Second, HKE users may exhibit avoidance behavior in their usage of RCs. Being unfamiliar with complex sentence structures in English, one instance of which is RCs, users of HKE may shun their use of such structures lest grammatical errors be committed [36]; this effectuates a lower frequency of occurrence of RCs in the written production of users of HKE, which fails to truly reflect their linguistic knowledge of English RCs.

Notwithstanding a lower frequency of occurrence of RCs in written HKE when compared to written BrE, it is unnecessary to require Cantonese learners of English in Hong Kong to profoundly increase their frequency of production of RCs as long as ideas can be expressed accurately and fluently inasmuch as it is language users' own choice to make their own decisions on language structures applied for conveyance of meanings. More importantly, it is unnecessary to require learners to develop or

produce language structures that never occur in primary language data, such as RCs with relativized subject complements, to align language learning with authentic language use. That said, should learners refuse to produce target structures simply owing to avoidance, it is suggested to expose learners to RCs more frequently in a bid to enhance their familiarity with the target structures and detract from their anxiety.

Another key finding is that RCs with relativized prepositional complements occur less frequently than those with relativized adjuncts in the ICE-HK yet more frequently than their counterparts in the ICE-GB. In practice, RCs with relativized prepositional complements and those with relativized adjuncts are interchangeable at times. For instance, in (1), the relative adverb ‘where’ can be superseded by the complex relative phrase ‘in which’ without modifying the meaning of the entirety of the sentence. The choice between RCs with relativized prepositional complements and those with relativized adjuncts is thereby sometimes freely made by language users. The more frequent occurrence of RCs with relativized adjuncts in HKE plausibly signifies the preference of users of HKE for RCs with relativized adjuncts over those with relativized prepositional complements. Being notoriously difficult to learn, prepositions can hardly be mastered well even by advanced second language learners of English who have already learned the language for a considerable amount of time, so users of HKE, who are foreign language learners of English, may prefer RCs with relativized adjuncts to those with relativized prepositional complements should they be given a choice to shun the use of prepositions [7]. The other reason why RCs with relativized adjuncts are frequent in HKE is that some RCs are acquired by second language learners as phrasal vocabulary in lieu of syntactic structures. One concrete instance is the expression ‘the reason why’ as in (2). The bracketed structure can be analyzed as an RC with ‘the reason’ and ‘why’ as the antecedent and relative word, respectively. Meanwhile, ‘the reason why’ can be conceived to be a formulaic sequence by virtue of its widespread use in language, use for a specific purpose of introducing the reason, and use as a lexical item in lieu of syntactic structure, enhancing the fluency in production [37]. Such expressions substantially enhance the frequency of occurrence of RCs with relativized adjuncts in written HKE.

(1) Moreover there are some elaboration rites such as the grand parade and scrambling of the bun towers in which Cheung Chau people feel excited just same as the clan in Australia [where the members of the clan feel excited through the totem rites such as dancing and singing].

<ICE-HK: W1A-007 #98.1>

(2) The reason [why Durkheim did not use supernatural or God to define religion] is that some religions have no God or supernatural but they only have some orthodox principles.

<ICE-HK: W1A-007 #20.1>

Regarding the lower frequency of occurrence of RCs with relativized prepositional complements than those with relativized adjuncts in written HKE, rarely does language users’ arbitrary choice between RCs with relativized prepositional complements and those with relativized adjuncts distort the communication of meanings. That said, solidification of learners’ knowledge of prepositions and complex relative phrases may boost their confidence in the use of RCs with prepositional complements and thereby provide them with more choices in their

language production. On one hand, English prepositions being notoriously difficult for second language learners to master, any language structures involving may pose a challenge to learners [7]. It is recommended to draw upon concepts of cognitive linguistics like landmark and trajector to enhance learners' understanding of English prepositions by using schematic diagrams to present meanings of prepositions with respect to the aforementioned cognitive linguistic concepts [38]; this lays a strong basis for them to develop RCs with relativized prepositional complements. On the other hand, the relationship between relative markers and the rest of the RC is implicitly specified in East Asian languages at times with covert relative clauses. Learners may thereby find complex relative phrases particularly difficult to master. However, current instructional materials of English RCs fail to introduce complex relative phrases to learners at length [39]. Syntactic structures of complex relative phrases ought to be explicated to learners systematically to facilitate their development of such target structures.

Added to the above, RCs with subjects as antecedents appear more frequently than those with direct objects as antecedents in the ICE-HK yet less frequently than their counterparts in the ICE-GB. The reason why such a finding is noteworthy is that it deviates from the Perceptual Difficulty Hypothesis (PDH) [40]. RCs with subjects as antecedents and those with direct objects as antecedents involve center-embedding and right-embedding respectively. In accordance with the PDH, the latter ought to occur more frequently than the former does; all the same, empirical corpus data reflects the opposite. One plausible explanation is that there are barely any distinct types of RC taught in Hong Kong on the basis of their natural acquisition or instructional orders. The instance provided in the intended document, as shown in (3), is an RC with a subject as the antecedent. Planning their implemented curriculum on the basis of the intended curriculum, English teachers in Hong Kong may take the instance provided in the curriculum document as an exemplar for teaching and focus on teaching RCs with subjects as antecedents in lessons. Exposed to RCs with subjects as antecedents more frequently in English language lessons, students may acquire such RCs better and produce them more frequently in their language use.

(3) [The boy who usually waters the plants] is called John.
Curriculum Development Council [1]

Last but not least, RCs with relativized subject complements are totally absent in the ICE-HK. In reality, RCs with relativized subject complements are frightfully rare in written BrE with only 3 tokens amongst the 3304 RCs identified; this means that such structures are uncommonly adopted in written language production. For such a reason, it is reasonable for the frequency of occurrence of RCs with relativized subject complements to be low in written HKE, but the utter absence of such structures may plausibly imply HKE users' lack of linguistic knowledge of such a language structure. It is implausible to verify such a speculation on the basis of corpus data, which can solely provide positive evidence in lieu of negative evidence of language users' language production. In other words, only does the corpus tell what language users are capable of producing in their naturalistic production, but it also fails to tell language structures that users fail to produce.

This study complements Gisborne [3] and Newbrook [4] to provide a more comprehensive picture of the usage of RCs in HKE. While Gisborne [3] and

Newbrook [4] identified some unique attributes that make RCs in HKE different from RCs in standard English varieties morphosyntactically, the present study focuses on the quantitative comparison of the RCs in HKE and those in BrE. It shows that the RCs produced by users of HKE do not only deviate from RCs in standard English varieties by form, but they also differ in terms of frequencies of occurrence. This study shows that RC is a less commonly used syntactic structure in HKE than in BrE, and the lower frequency of occurrence is shown in most types of RCs.

6. Conclusion

The study fills the research gap in research on RCs in Hong Kong English by quantitatively comparing morphosyntactic features of RCs in HKE and those in BrE. The key finding is that most types of RCs are more frequently observed in the ICE-GB than in the ICE-HK. Knowledge of RCs in written HKE is expected to advance understanding of HKE as an English variety. One implication of the study is that some types of RCs seldom occur in ICE-GB and even never appear in ICE_HK., Given the low frequency counts of those types of RCs, such as RCs with relativized subject complements, in naturally occurring language data, it is doubtful whether it is necessary to require Cantonese learners of English in Hong Kong to profoundly increase their frequency of production of RCs through formal instruction. In other words, the curriculum may have to be revised to ensure that the language taught in the classroom is needed or used in real-life communication.

The study possesses two limitations. The first limitation is the omission of bare relatives. In contrast with *wh* relatives and *that* relatives, which possess relativized elements realized as relative phrases and the subordinator *that* respectively, bare relatives possess relativized elements realized as a gap [5]. Being difficult to identify by means of concordance search, bare relatives were omitted in the course of data analysis; this detracts from the internal validity of the study, for frequencies of occurrence of RCs computed in the study fail to truly reflect frequencies of occurrence of all types of RCs but solely *wh* relatives and *that* relatives. Corpora with syntactic tagging may address such a pitfall. Another limitation of the study is the presence of solely one rater in the course of coding. Qualitative studies often embody reliability checks through peer checking, in which coding is performed not only by the researcher but also by an additional rater to compare correspondence between the two sets of outcomes, to enhance both the measurement validity of the data and the reliability of the entirety of the study (ibid.). Such a step was, however, skipped in the current study out of difficulty in gauging a second rater without any funding.

In response to the aforementioned limitations, it is suggested that bare relatives be incorporated into the data analysis to enhance the internal validity of the study, and it is recommended to perform reliability checks in future research to further enhance the measurement validity of the data and reliability of the study.

Acknowledgments: I would like to express my sincere gratitude to Carmen Lee and Honorary Gerald Nelson from the Department of English, The Chinese University of Hong Kong for their supervision of this research project.

Conflict of interest: The author declares no conflict of interest.

References

1. Curriculum Development Council. English Language Education Key Learning Area Curriculum Guide (Primary 1 - Secondary 6). Hong Kong: Hong Kong Government Printer; 2017.
2. Biber D, Johansson S, Leech G, et al. Longman grammar of spoken and written English. Harlow, Essex, England: Pearson Education Ltd.; 1999.
3. Gisborne N. Relative clauses in Hong Kong English. *World Englishes*. 2000; 19(3): 357-371. doi: 10.1111/1467-971x.00184
4. Newbrook M. Which way? That way? Variation and ongoing changes in the English relative clause. *World Englishes*. 1998; 17(1): 43-59. doi: 10.1111/1467-971x.00081
5. Huddleston R, Pullum GK, Peterson P. Relative constructions and unbounded dependencies. In: *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press; 2002.
6. Hewings M. *Advanced grammar in use*, 3rd ed. Cambridge: Cambridge University Press; 2013.
7. Larsen-Freeman D, Celce-Murcia M, Frodesen J, et al. *The grammar book: form, meaning, and use for English language teachers*, 3rd ed. Boston, MA: National Geographic Learning, Heinle Cengage Learning; 2016.
8. Bolton K. The sociolinguistics of Hong Kong and the space for Hong Kong English. In: *Hong Kong English: Autonomy and creativity*. Hong Kong: Hong Kong University Press; 2002.
9. Bolton K, Bacon-Shone J, Luke KK. Hong Kong English. In: *The Handbook of Asian Englishes*. Hoboken, NJ: John Wiley and Son; 2020.
10. Fuchs R. Hong Kong English: Structural features and future prospects. In: *Multilingual Global Cities: Singapore, Hong Kong, Dubai*. London: Routledge; 2020.
11. Kirkpatrick A. *World Englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press; 2007.
12. Luke KK, Richards JC. English in Hong Kong. *English World-Wide*. 1982; 3(1): 47-64. doi: 10.1075/eww.3.1.04kan
13. Li DCS. The Functions and Status of English in Hong Kong. *English World-Wide*. 1999; 20(1): 67-110. doi: 10.1075/eww.20.1.03li
14. Schneider EW. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press; 2007.
15. Chan KLR. The future of Hong Kong English: Codification and standardization?. In: Tang W (editor). *Hong Kong: Past, present and future*. Nova Science; 2020.
16. Chan KLR. The Construct of English Native Speaker in Hong Kong. *Journal of English and Applied Linguistics*. 2024; 3(1). doi: 10.59588/2961-3094.1074
17. Hung TTN. Chapter 7. Hong Kong English. In: *English in Southeast Asia: Features, policy and language in use*. Amsterdam: John Benjamins; 2012.
18. Sung CCM. Hong Kong English: Linguistic and Sociolinguistic Perspectives. *Language and Linguistics Compass*. 2015; 9(6): 256-270. doi: 10.1111/lnc3.12142
19. Hansen Edwards JG. Hong Kong English: attitudes, identity, and use. *Asian Englishes*. 2015; 17(3): 184-208. doi: 10.1080/13488678.2015.1049840
20. Gonzales WDW. From tweets to trends: analyzing sociolinguistic variation and change using the Twitter Corpus of English in Hong Kong (TCOEHK). *Asian Englishes*. 2023; 27(1): 1-24. doi: 10.1080/13488678.2023.2251771
21. Gonzales WDW, Zhang Y. Reshaping language learners' languaging habitus: A world-Englishes-informed critical pedagogy. *RELC Journal*; 2025.
22. Ladegaard HJ, Chan KLR. Teachers' attitudes towards varieties of Hong Kong English. *English World-Wide*. 2022; 44(2): 251-275. doi: 10.1075/eww.21060.lad
23. Ng CW. Subject-Operator Disagreement in English Existential Constructions: A Corpus-Based Analysis of American English and Hong Kong English. *Applied Linguistics Research Journal*. 2020; 4(3), 1-12.
24. Ng CW. Relative clauses in written Hong Kong English: a corpus-based study of untimed student essays. *Asian Englishes*. 2023; 26(2): 372-390. doi: 10.1080/13488678.2023.2251732
25. Ng CW. A revisit of three hypotheses about second language development of English relative clauses. *Pedagogical Linguistics*. 2024; 6(1): 78-108. doi: 10.1075/pl.23008.ng
26. Zhang Y, Gonzales WDW. World English pedagogy: Constructing learner identity. *ELT Journal*; 2024.

27. Chan AYW. Toward a Taxonomy of Written Errors: Investigation Into the Written Errors of Hong Kong Cantonese ESL Learners. *TESOL Quarterly*. 2010; 44(2): 295-319. doi: 10.5054/tq.2010.219941
28. McEnery T, Hardie A. *Corpus Linguistics: Method, theory and practice*. Cambridge: Cambridge University Press; 2011.
29. McEnery T, Wilson A. *Corpus linguistics: An introduction*, 2nd ed. Edinburgh, Edinburgh University Press; 2001.
30. Greenbaum S. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press; 1996.
31. Greenbaum S, Nelson G. The International Corpus of English (ICE) Project. *World Englishes*. 1996; 15(1): 3-15. doi: 10.1111/j.1467-971x.1996.tb00088.x
32. Nelson G, Wallis S, Aarts B. *Exploring Natural Language: Working with the British component of the International Corpus of English*. Philadelphia, John Benjamins Publishing Company; 2002.
33. Bolton K, Nelson G, Hung J. A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*. 2002; 7(2): 165-182. doi: 10.1075/ijcl.7.2.02bol
34. Scott M. *WordSmith Tools version 5.0*. Liverpool: Lexical Analysis Software; 2010.
35. Gass SM, Mackey A. *Data Elicitation for Second and Foreign Language Research*. Routledge; 2007.
36. Chan AYW. Syntactic Transfer: Evidence from the Interlanguage of Hong Kong Chinese ESL Learners. *The Modern Language Journal*. 2004; 88(1): 56-74. doi: 10.1111/j.0026-7902.2004.00218.x
37. Schmitt N, Carter R. Formulaic sequences in action: An introduction. In: Schmitt N (editor). *Formulaic sequences*. Amsterdam: John Benjamins; 2004.
38. Wong MHI, Prange J. A Bayesian approach to (re)examining learning effects of cognitive linguistics–inspired instruction. *Studies in Second Language Acquisition*. 2024; 46(5): 1493-1513. doi: 10.1017/s0272263124000603
39. Harris C. *Developing Skills: Grammar & Usage for Junior Secondary Learners 2 (Set B)*. Hong Kong: Aristo Educational Press Ltd.; 2012.
40. Kuno S. The position of relative clauses and conjunctions. *Linguistic Inquiry*. 1974; 5(1): 117-136.