Investigating Linguistic Features of Narration Among Non-Chinese Speaking Students in Hong Kong*

Cindy Man-fong Lam[1], Kevin Kin-wing Chan[2], Sheung Ping Wong[1]
Hong Kong Metropolitan University[1], Hong Kong Shue Yan University[2]

Abstract
In existing research on Chinese as a second language (CSL) acquisition in Hong Kong, the Cantonese oral proficiency of non-Chinese speaking (NCS) students receives far less attention than the challenge of developing their literacy skills in standard written Chinese (SWC). This study focuses on the second language (L2) Cantonese oral language abilities of NCS students with advanced speaking proficiency, describing several salient linguistic features in their Cantonese speech based on elicited oral narrative speech data. The test group comprises 16 senior secondary students of Pakistani heritage recruited from a single school in Hong Kong (mean age = 16.44 years), with an average duration in formal Cantonese learning experience of 12.50 years. Students' oral performance was assessed through a storytelling task using the wordless illustrated storybook *Frog, Where Are You?* (Mayer 1969). Prominent linguistic features observed from audio recording transcriptions include the extensive blending of SWC elements across various lexical and morphological categories, English code-mixing retaining source-language morphosyntax, and a limited inventory of classifiers. Comparison with native Cantonese speaker controls reveals systematic errors (Corder 1967) involving such features, suggesting the emergence of a new interlanguage variety of spoken Cantonese among South Asian NCS students in Hong Kong. The findings of this study offer valuable insights regarding how advanced adolescent CSL learners in Hong Kong simultaneously acquire multiple Chinese varieties in their formal education. This research carries significant implications for pedagogy and curriculum design that aims to improve Cantonese oral proficiency among this group of learners.

Key words
Hong Kong Cantonese, interlanguage, non-Chinese speaking (NCS) students, oral narratives, standard written Chinese (SWC)

* The work described in this paper was fully supported by Hong Kong Metropolitan University Research Grant No. RD/2022/2.2 and the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS16/H15/19).

1. Introduction

Within the Hong Kong school system, non-Chinese speaking (NCS) students of minority ethnic backgrounds face the challenge of simultaneously developing different language abilities in two distinct varieties of Chinese, namely vernacular Cantonese for listening and speaking, in addition to standard written Chinese (SWC) for reading and writing (Chan 2019). Existing research recognizes that a discrepancy exists between the oral and literacy abilities of NCS students learning Chinese as a second language (CSL), with such students generally achieving greater proficiency in listening and speaking, given the prevalence of vernacular Cantonese in society (Census and Statistics Department 2022, Ku et al. 2005). Nevertheless, CSL acquisition remains an under-researched topic in the Hong Kong context, with more publications placing greater emphasis on issues including learner identities, learning motivations, and language policy (Gao et al. 2019, Wang and Tsung 2022). Even among acquisition-oriented studies, less attention has been paid to the Cantonese oral skills, particularly those of adolescent NCS learners who may have achieved advanced speaking proficiency through social immersion and formal education.

Following the approaches developed through the body of "frog story" research pioneered by Berman and Slobin (1994), this study explores the second language (L2) Cantonese speaking abilities of NCS students in Hong Kong through an analysis of oral narratives based on the story of *Frog, Where Are You?* (Mayer 1969). Employing a case study approach, it reports preliminary findings from an experiment conducted on adolescent L2 learners undergoing secondary education. Specifically, this study seeks to address the following research questions:

1. What are the salient linguistic features present in the spoken Cantonese of NCS students with advanced or near-native speaking proficiency?
2. Which of the above features are unique to NCS students?
3. Which of the above features are not found among NCS students (or other L2 Cantonese learners) with elementary proficiency?

The findings show that respondents exhibited a series of distinct errors in their spoken Cantonese, with these patterns suggesting the emergence of a new interlanguage variant among certain South Asian users of Cantonese in Hong Kong. The use of *errors* here refers to Corder's (1967:167) "systematic errors of the learner from which we are able to reconstruct his knowledge of the language to date, i.e., his transitional competence," as opposed to simply *mistakes* in performance.

This study is organized as follows. The following section reviews existing research on CSL acquisition in Hong Kong, as well as previous studies on the oral narrative abilities of L2 Cantonese learners and the linguistic features of their spoken Cantonese. Section 3 describes the methodology of this study, while Section 4 presents the results of the current analysis. Section 5 closes with a discussion of the findings and offers concluding remarks.

2. Previous research

2.1 CSL acquisition in Hong Kong

For NCS students, the specialized learning of distinct Chinese varieties based on different usage domains—Cantonese for oral communication, SWC for written communication—in an environment where Cantonese is the majority language means a higher chance of achieving better performance in the former. Previous surveys support this tendency, having consistently found that among the predominant South Asian ethnic minority groups in Hong Kong (Pakistanis, Indians, Nepalese, and Filipinos), Chinese language abilities in reading and writing are generally poorer than in speaking (Census and Statistics Department 2022, Ku et al. 2005, Li and Chuk 2015). However, these surveys make this claim based on qualitative data—particularly self-reported

questionnaire responses by NCS learners—rather than empirical performance measures for evaluating their language skills. In addition, discussions of the linguistic abilities of NCS students are often framed in terms of benchmarking their performance against some abstract proficiency levels, rather than scrutinizing the actual linguistic outputs they produce.

Furthermore, as mentioned above, research on CSL acquisition in Hong Kong has largely centered on the development of literacy skills among NCS students to close the gap with their relatively advanced oral skills, with the primary emphasis being frameworks for facilitating effective Chinese reading and writing performance among NCS students (Tang 2018, Wang and Tsung 2022, Wong 2010). Conversely, few studies have probed the oral language performance of these students further, much less investigated whether they have successfully mastered vernacular Cantonese, or continue to carry certain errors in their speech. This study aims to substantiate this body of research by presenting an error analysis of the Cantonese speech of more advanced NCS learners.

2.2 Oral narrative abilities of L2 Cantonese learners

From as early as the preschool stage, narrative storytelling represents an important learning activity in language education. Oral narration is thus a familiar and accessible discourse genre for learners of all ages, and represents a valuable medium for eliciting extended, naturalistic speech data for assessing the oral abilities of L2 learners (Berman and Slobin 1994, Gagarina et al. 2016, Klamer and Moro 2020). Moreover, the particular discursive purpose of conveying information about a sequence of events in a narrative task allows for the study of L2 learners' ability to implement various pragmatic language skills, including linguistic devices of evaluation and coherence (Fernández 2011).

In the 21st century, several studies have employed spoken narratives to investigate the oral language proficiency of NCS students in Hong Kong. For instance, Cheung (2019) uses data from a story-retelling task to explore whether the age of arrival in Hong Kong among South Asian NCS students correlates with narrative performance. Next, drawing on the traditions of "frog story" research (Berman and Slobin 1994), Zhu (2019) examines both the macrostructural and microstructural narrative performance of Primary 6 NCS students, making comparisons among low-performing NCS, high-performing NCS, and native Chinese-speaking (CS) students. Likewise, Chan et al. (2023) analyze both levels of narrative performance among a group of Urdu-Cantonese bilingual primary school children, reporting on the associations between both levels in both languages. However, no studies thus far have investigated the oral narrative abilities of more advanced NCS learners, specifically secondary school students, some of whom may have achieved a speaking proficiency approaching the level of native Cantonese speakers. This study addresses this research gap by reporting on the characteristics of NCS secondary students' spoken language, as demonstrated through their performance in oral narration.

2.3 Linguistic features of NCS students' L2 spoken Cantonese

Lastly, another area that remains under-studied in research on NCS students is the specific features at various levels of linguistic structure that are present in their L2 spoken Cantonese. A recent study by Yu et al. (2021) reports that South Asian NCS students have a reduced tone system smaller than the standard six-tone inventory of Hong Kong Cantonese, with tone mergers occurring in both perception and production. Oral narrative studies investigating NCS students' Cantonese (Chan et al. 2023, Cheung 2019, Zhu 2019) have also remarked on the use of elements such as classifiers, connectives, manner modifiers, and relative clauses as part of the analyses

concerning microstructural performance overall. Nevertheless, these studies lack more comprehensive descriptions of the lexical items and grammatical structures produced by NCS students. This study thus seeks to contribute a substantial descriptive analysis of NCS students' spoken Cantonese through reporting prominent linguistic features encountered throughout their oral narrations.

3. Methodology

3.1 Participants

NCS test group participants were recruited from a mainstream government-sub-vented secondary school in Hong Kong in July 2023, with 24 students participating. The school uses Cantonese as the medium of instruction supplemented with English; in addition, senior secondary students receive instruction in Putonghua from Form 4 (10^{th} grade) onwards. Among the participants, 19 Pakistani students, representing the predominant ethnicity among the participants (79.17%) were selected for analysis; three were excluded due to incomplete narrations, leaving 16 (including one mixed-ethnicity Pakistani-Filipino student) who completed the experimental task. The final test group was roughly balanced in terms of grade level (nine Form 4 students, seven Form 5 students), but not in terms of gender (11 males, 5 females). Participants' ages ranged between 15 and 19 years, with the average age being 16.44 years ($\sigma = 1.09$).

Regarding the language environments of test group participants, 13 reported speaking Urdu at home as a family language. Five were born in Hong Kong and thus exposed to Cantonese as the majority language since birth, while six reported speaking Cantonese at home as an additional language alongside Urdu or English. Participants' formal Cantonese learning experience through schooling ranged between 6 and 16 years, with an average duration of 12.50 years ($\sigma = 2.76$). In particular, 13 participants (81.25%) reported learning Cantonese for at least 10 years.

To enable comparison with CS students, a control group was recruited from another private secondary school in September 2023, with 11 students participating; all completed the experimental task and are included in the current analysis. Due to offering both the local Hong Kong Diploma of Secondary Education (HKDSE) curriculum and the International Baccalaureate Diploma Programme (IBDP) curriculum, the school adopts both Cantonese and English as mediums of instruction; Putonghua is offered as a second language alternative to Cantonese in the Mandarin *ab initio* course for students pursuing IBDP qualifications. The control group was roughly balanced in terms of gender (6 males, 5 females), but not in terms of grade level (two Form 1 students, two Form 2 students, seven Form 3 students). Participants' ages ranged between 11 and 15 years, with the average age being 13.36 years ($\sigma = 1.12$). All spoke Cantonese at home, while their formal Cantonese learning experience ranged between 9 and 14 years, with an average duration of 12.50 years ($\sigma = 1.96$).

To report findings without revealing personal details, participants were each assigned a code including their school (SA for the test group, SC for the control group), ethnicity (Ch = Chinese, Pa = Pakistani, Ph = Filipino), and gender (M/F). Table 1 provides more detailed information about the demographic characteristics of both participant groups.

Table 1. Demographic characteristics of sample participants.

	Test group (NCS, *n*=16)	Control group (CS, *n*=11)
Age		
Range, years	15–19	11–15
Mean	16.44 (σ = 1.09)	13.36 (σ = 1.12)
Grade distribution	Form 4: 9 (56.25%) Form 5: 7 (43.75%)	Form 1: 2 (18.18%) Form 2: 2 (18.18%) Form 3: 7 (63.64%)
Gender		
Male	11 (68.75%)	6 (54.55%)
Female	5 (31.25%)	5 (45.45%)
Language environment		
Born in Hong Kong	5 (31.25%)	9 (81.81%)
Cantonese spoken at home	6 (37.50%)	11 (100.00%)
Urdu spoken at home	13 (81.25%)	0 (0.00%)
Cantonese learning experience		
Range, years	6–16	9–14
Mean, years	12.50 (σ = 2.76)	12.50 (σ = 1.96)
≥10 yrs.	13 (81.25%)	8 (72.73%)

3.2 Materials and procedures

Following the approach of Berman and Slobin (1994), examiners interviewing the participants instructed them individually to tell the story depicted in the wordless illustrated storybook *Frog, Where Are You?* (Mayer 1969) in Cantonese. In order to reduce the burden of memorization and elicit more spontaneous responses, participants first looked through all the images once to familiarize themselves with the story, before narrating the story while going through the images again one by one. During narration, examiners kept verbal feedback to a minimum to avoid influencing participants' individual expressions of the story.

The third author then transcribed audio recordings of the narrations for further analysis. In the process of reviewing the transcripts, various lexical, morphological, syntactic, and pragmatic features observed as salient across the narratives of multiple participants were tallied. In particular, occurrences of the following features were counted for each narrative: (a) SWC lexical items, (b) English lexical items, (c) classifiers, and (d) connectives.

4. Results

4.1 Narrative performance

In terms of general narrative performance measures, the performances of both the test and control groups were rather similar. Looking at duration, the values for the NCS students ranged between 2:44 and 5:17, with the average duration being 4:04. Similarly, the values for the CS students ranged between 2:21 and 5:20, with the average duration being 4:24. The averages for both groups are also comparable to the overall sample average (4:21) for the cohort of Primary 6 students discussed in Zhu (2019). Turning to unfiltered token count—which includes elements such as "filled pauses, repetitions, reformulations, and interjections" (Zhu 2019:71)—the values for the NCS students ranged between 322 and 911 tokens, with an average of 552 tokens. Similarly, the

token counts for the CS students ranged between 319 and 962 tokens, with a slightly greater average of 653 tokens.

4.2 SWC code-mixing

The most notable discovery of this study is that NCS students incorporated SWC elements in vernacular Cantonese speech, a tendency rarely discussed among this group. Within the current sample, Twelve NCS participants (75.00%) featured SWC elements in their narrations, compared to nine CS participants (81.81%). However, the nature and extent of the blending between SWC and Cantonese is different between both groups. Table 2 shows that among the 11 distinct categories of SWC elements that emerged through the process of transcript annotation, three were observed exclusively among the test group, while another three were observed exclusively among the control group. This section presents a selection of examples that best illustrate the distinctive uses of SWC elements by NCS students.

Table 2. SWC categories observed in Cantonese narrations.[1]

Participants	SWC categories
Both groups	Nouns, verbs and adjectives, adverbs; diminutive affixes
NCS only	Directional verbal particle -下 (*-haa6*), postverbal particle -得 (*-dak1*), linking particle 的 (*dik1*), adverbializing suffix -地 (*-dei6*)
CS only	Locatives, prepositions, aspectual markers

With respect to lexical categories, SWC lexical items produced by NCS students were attested across the major categories of nouns, verbs, (predicative) adjectives, and adverbs. For nouns, the lexical set of SWC items across nine participants (56.25%) included time nouns used as temporal adverbials, such as 晚上 (*maan5soeng6*) for *night* in (1), along with family and kinship terms, such as *little brother* to refer to the boy, as in (2). Among predicative elements, the lexical set for four participants (25.00%) primarily consisted of verbs, with the exception of one adjective 害怕 (*hoi6paa3*) for *scared* as in (3).

(1) ?呢個 我哋 睇到 係 晚上。
?ni1-go3 ngo5-dei6 tai2-dou2 hai6 **maan5soeng6.**
this-CL 1PL see-can be **night**
'This we can see is nighttime.' (SAPa2M, l. 3)

(2) ?弟弟 起身 嗰陣 時 呢，
?dai6dai6 hei2san1 go2-zan6 si4 ne1,
little.brother wake.up that-CL time TOP
'When little brother woke up,' (SAPa18F, l. 15)

[1] Cantonese examples are transcribed using the Jyutping romanization scheme developed by the Linguistic Society of Hong Kong (LSHK), available online at: jyutping.org/en/jyutping/. Abbreviations that deviate from Leipzig glossing rules: ADVZ=adverbializer; CL=classifier; DIM=diminutive; DIREC=directional verbal particle; INT=interjection; LP=linking particle, PERS=perspective verbal particle; SFP=sentence final particle.

(3) ?但係 個 deer 都 好 害怕 佢 啦，
?daan6hai6 go3 deer dou1 hou2 **hoi6paa3** keoi5 laa1,
but CL deer also very **scared** 3SG SFP
'But the deer was also very scared of him,' (SAPa13M, l. 40)

For *little brother* in particular, the participant (SAPa18F) shifts between the standard pronunciation *dai6dai6* as in (2), where both syllables have a low-level tone, and the vernacular pronunciation *dai4dai2* as in (4), where a diminutive tone change applies to the second syllable and it is realized with a high-rising tone. In (4), the participant manages to self-correct from the standard pronunciation to the vernacular one.

(4) 嗷 嗰個 弟弟 都 會 好… 弟弟
gam2 go2-go3 **dai6dai6** dou1 wui5 hou2… **dai4dai2**
then that-CL **little.brother** also would very **little.brother\DIM**
好 嬲，
hou2 nau1,
very angry
'Then that little brother would also [get] very… (restart) little brother [got] very angry,' (SAPa18F, l. 19)

Turning to morphological phenomena, salient SWC features observed among NCS students include the diminutive affix 小- (*siu2*-), the adverbializing suffix -地 (*-dei6*), the linking particle 的 (*dik1*), and the directional verbal particle -下 (*-haa6*), the latter three only occurring among the test group. Regarding diminutives, whereas certain participants in both the NCS and CS groups (three each) used *siu2*- with nouns referring to the primary characters (the boy, the dog, and the frog), only those in the test group applied *siu2*- to the near-total exclusion of the vernacular diminutive affix (DIM) -仔 (*-zai2*), with the latter entirely absent from the narration of one NCS student (SAPa13M). For this set of NCS participants, the standard diminutive forms for certain nouns (e.g., 小狗 *siu2-gau2*, *little dog*) have supplanted the corresponding vernacular forms (e.g., 狗仔 *gau2-zai2*, *doggie*) as the default, if not the sole option for referring to certain characters with diminutive affect, as in (5) through (7).

(5) 個 小鳥 都 好 嬲
go3 **siu2-niu5** dou1 hou2 nau1
CL **DIM-bird** FOC very angry
'The little bird was also very angry' (SAPa13M, l. 33)

(6) 從前 就 有 個 小男孩 叫 小明 啦，
cung4cin4 zau6 jau5 go3 **siu2-naam4-haai4** giu3 siu2-ming4 laa1,
Formerly then have CL **DIM-male-child** call Siu Ming SFP
'Once upon a time there was a little boy called Siu Ming,' (SAPaPh1F, l. 2)

(7) 跟住 佢 同 個 小狗 就 一齊 瞓覺。
gan1zyu6 keoi5 tung4 go3 **siu2-gau2** zau6 jat1cai4 fan3gaau3.
then 3SG with CL **DIM-dog** then together sleep
'Then he slept together with the little dog.' (SAPaPh1F, l. 4)

Similarly, for four NCS participants (25.00%), the standard adverbializing suffix (ADVZ) *-dei6* represented the default derivational strategy for adverbs derived from adjectives, as opposed to the vernacular suffix 嗰 (*-gam2*),[2] as in (8) and (9).[3] The realization of *slowly* as *maan6maan2-dei6* in (8) is particularly remarkable for juxtaposing the standard suffix with the vernacular diminutive tone change on the adjectival root, thus representing a standard/vernacular blend occurring *below* the word level.

(8) ?跟住 佢哋 細聲啲， 慢慢地 出去。
?gan1zyu6 keoi5-dei6 sai3-seng1-di1, **maan6~maan2-dei6** ceot1-heoi3.
then 3PL small-voice-bit **slow~slow\DIM-ADVZ** out-go
'Then they slightly more quietly, slowly went out.' (SAPa2M, l. 36)

(9) ?好 開心地 返 屋企
?hou2 **hoi1sam1-dei6** faan1 uk1kei2
very **happy-ADVZ** return home
'Very happily returned home' (SAPa3M, l. 39)

Next, one participant (SAPa2M) featured a single instance of the standard linking particle (LP) *dik1*, which appears between a noun and its preceding modifier. In (10), the particle appears between the noun 位 (*wai2, place*) and the preceding relative clause (RC) containing the verb 瞓覺 (*fan3gaau3, sleep*) to mark the former. However, the participant also has the vernacular linking particle 嘅 (*ge3*) in his lexicon, given that it occurs immediately before the relative clause linking it to the possessor determiner phrase (DP) 佢哋 (*keoi5dei6*, *they*). Indeed, immediately after first uttering the relativized noun phrase (NP), the participant is able to self-correct and restate the construction with the vernacular linking particle.

(10) 跟住 佢哋 睇 佢哋 嘅 瞓覺 的 位…
gan1zyu6 keoi5-dei6 tai2 keoi5-dei6 **ge3** fan3gaau3 **dik1** wai2…
Then 3PL look 3PL **LP** sleep **LP** place
瞓覺 嘅 位 搵 青蛙。
fan3gaau3 **ge3** wai2 wan2 cing1waa1.
sleep **LP** place find frog
'Then they looked at their place where they sleep… place where they sleep to find the frog.' (SAPa2M, l. 6)

[2] However, other homonymic forms of *gam2* represented by the same character were observed in the narrations of these participants, including one form serving as a connective. See Section 4.4 for a discussion on this category.

[3] Adverbializing *-dei6* was observed for one CS student in the construction 好自動地 (*hou2 zi6dung6-dei6*, "very automatically;" SCCh11F, l. 53). As this was an isolated occurrence, the use of *-dei6* is not considered to be systematic among the control group.

Lastly, two NCS participants (12.50%) expressed downward motion in certain instances using only the standard directional verbal particle (DIREC) *-haa6*, as in (11) and (12). While permissible in writing for some contexts, in vernacular Cantonese the corresponding directional particle 落 (*-lok6*) must be followed by a perspective particle (PERS), either 去 (*heoi3, "go"*) or 嚟 (*lai4, "come"*) to indicate the speaker's perception of the motion event. Of the two participants, only one (SAPa2M) managed to self-correct and restate the verb with both particles, as in (12).

(11) *之後 個 男仔 都 快啲 跳下，
*zi1hau6 go naam4-zai2 dou1 faai3-di1 **tiu3-haa6**,
then CL boy-DIM also fast-bit **jump-DIREC**
Intended: 'Then the boy also quickly jumped down,' (SAPa11F, l. 16)

(12) *跟住 嗰個 bee nest 跌… 跌咗下…
*gan1zyu6 go2-go3 bee nest dit3… dit3-zo2-**haa6**…
then that-CL bee nest fall fall-PFV-**DIREC**
跌咗落嚟 啊！
dit3-zo2-**lok6**-lai4 aa3!
fall-PFV-**DIREC**-PERS SFP
Intended: 'Then that bee nest fall… fell down… fell down!' (SAPa2M, l. 20)

4.3 English code-mixing

Anticipating that some NCS students may not recall or even know certain Cantonese lexical items pertinent to the story, such as 鹿 (*luk2*) for *deer* or 蜂巢 (*fung1caau4*) for *beehive*, participants had permission to use English lexical items in those specific instances before resuming their narration in Cantonese. Ultimately, 11 NCS participants (68.75%) featured at least one substantive English lexical item[4] in their narratives, including three who uttered over 20 English tokens throughout their narrations. The number of English lexical items appearing in a narration ranged between 0 and 26 tokens, with the average being 9.00 tokens ($\sigma = 9.42$). These English tokens accounted for between 0.00% and 7.12% of the total tokens in a narrative, with the average ratio being 1.74% ($\sigma = 2.17\%$). In total, 39 English word types were observed across the NCS narratives, of which 26 (66.67%) were nouns and 10 (25.64%) were verbs. Notably, 5 participants (31.25%) used the English word *frog* rather than the Cantonese item 青蛙 (*cing1waa1*) in their narrations, despite the examiners priming the use of the latter during the initial instructions[5] given to participants. Other word types attested across multiple NCS participants include *(rein)deer, climb(ing), hole(s), bee(s), (bee)hive, cliff, jar, tree, forest, pet, rock, and squirrel.*

Interestingly, whereas CS participants only incorporated the root forms of English lexical items, NCS participants also produced forms employing additional morphology. For instance, both singular nouns, such as *tree* and *hole*, as well as plural nouns such as *holes*, were observed across the narratives of four NCS participants, as in (13) and (14). Regarding verbs, both bare infinitive and inflected forms were borrowed and then combined with Cantonese elements to encode

[4] Excluding items such as *BB*, a colloquial abbreviation of "baby," along with the filler words *OK* and *what*.

[5] 「呢度有一本書。佢講緊一個故事。故事入便有一個男仔[指封面上的男孩]，一隻狗[指狗]，同埋一隻青蛙[指青蛙]。宜家你先睇一次，留意每一幅圖。睇完之後，再講個故事畀我聽。」("Here is a book. This book tells a story about a boy [point to picture on cover], a dog [point], and a frog [point]. First, you will look at all the pictures. Pay attention to each picture. After that you will tell the story.") (Zhu 2019:66)

aspectual meanings. In (15), the present participle form of *climb* follows the copula 係 (*hai6*) to express the present progressive action "is climbing." In the following examples, the participant (SAPa17M) first utters the past participle form *grabbed* as in (16); he then repeats the statement with the bare infinitive form *grab* followed by the Cantonese aspectual marker 咗 (*zo2*) as in (17), thus mapping the perfective aspect to the latter instead of the English suffix *-ed*.

(13) ?跟住 嗰個 仔 係 去 好大個 **tree** 入便
?gan1zyu6 go2-go3 zai2 hai6 heoi3 hou2-daai6-go3 **tree** jap6-bin6
Then that-CL kid be go very-big-CL **tree** in-side
睇 一個 **hole.**
tai2 jat1-go3 **hole.**
look one-CL **hole**
'Then that kid is going inside a very big tree to look at a hole.' (SAPa2M, l. 22)

(14) 入咗 嗰啲 **holes** 度
jap6-zo2 go2-di1 **holes**-dou6
enter-PFV those-CL **holes**-LOC
'Went into those holes' (SAPa4M, l. 22)

(15) 跟住 佢 係 **climbing** 一個 佢 覺得 係 一個 木。
gan1zyu6 keoi5 hai6 **climbing** jat1-go3 keoi5 gok3dak1 hai6 jat1-go3 muk6.
then 3SG be **climbing** one-CL 3SG think be one-CL wood
'Then he is climbing what he thinks is a wood [=tree].' (SAPa2M, ll. 27–28)

(16) 個 動物 就 **grabbed** 個 細路仔
go3 dung6mat6 zau6 **grabbed** go3 sai3lou6-zai2
CL animal then **grabbed** CL kid-DIM
'The animal [=deer] then grabbed the little boy' (SAPa17M, ll. 25–26)

(17) 嗰個 動物 就 **grab 咗** 嗰個 細路仔
go2-go3 dung6mat6 zau6 **grab-zo2** go2-go3 sai3lou6-zai2
that-CL animal then **grab-PFV** that-CL kid-DIM
'That animal [=deer] then grabbed that little boy' (SAPa17M, ll. 27)

The particular example of English code-mixing presented in (18), where the intended utterance is "a beehive," is especially remarkable for flouting the grammatical rules of both English and Cantonese. First, the English compound noun *beehive* is broken up by the Cantonese linking particle *ge3*. Second, English plural morphology is realized on *bees* to mark a collective referent, which would be unnecessary for the corresponding Cantonese item 蜜蜂 (*mat6fung1*) provided that it is quantified by a collective classifier; however, the classifier immediately preceding *bees* is the general sortal classifier 個 (*go3*) that is actually quantifying over *hive*, resulting in a difficult-to-parse expression equivalent to "a hive of bees." The occurrence of such an expression, an example of shift-induced interference (Thomason and Kaufman 1988), indicates that the participant remains at an interlanguage stage in developing Cantonese as a target language, continuing to apply English syntagmatic relations to Cantonese lexical items.

(18) 跟住 佢 嘅 小狗 就 發現 一個 **bees** 嘅 **hive.**
gan1zyu6 keoi5 ge3 siu2-gau2 zau6 faat3jin6 jat1-go3 **bees** **ge3** **hive.**
then 3SG LP DIM-dog then discover one-CL **bees** **LP** **hive**
'Then his dog discovered a bees' hive.' (SAPa13M, l. 22)

4.4 Classifiers

Given the multifarious complexity of the Cantonese classifier system (Erbaugh 2002, Matthews and Yip 2011:109, Tse et al. 2007), the acquisition of classifiers poses a formidable challenge to achieving native-like proficiency for NCS students. Examining the context of oral narratives in particular, Zhu (2019:177-8) finds among Primary 6 NCS students classifier repertoires smaller than that of their CS counterparts; the number of classifier types observed among the lower- and higher-performing NCS groups were four and six respectively, compared to 14 for the CS group. In this study, the classifier inventory across all test group participants included 10 types, roughly half of the 19 found for the control group. While the larger figure among the NCS participants here compared to Zhu's (2019) groups indicates further acquisition overall beyond the primary school level, individual repertoires remain significantly limited; the average size for test group participants was 2.63 types, much lower than the average of 6.91 types for the control group. In particular, two NCS participants (SAPa2M, SAPa16M) used only *go3* throughout their narratives, while another five used only *go3* and the general collective classifier 啲 (*di1*, *some*). Notably, only three participants (SAPa7M, SAPa17M, SAPaPh1F) had in their repertoires the sortal classifier 隻 (*zek3*) for animate referents such as animals; even then, this specific classifier competes with *go3* among noun phrases individuating the animal characters in the story.

In agreement with previous studies investigating classifier acquisition among L1 and L2 child learners alike (Tse et al. 2007, Wei and Lee 2001, Zhu 2019), the current study recognizes the overgeneralization of *go3* among NCS students, to the extent that it serves as the only classifier for two participants. Moreover, this study observes certain instances of redundant classifiers, previously mentioned in Tse et al. (2007), occurring with certain nouns among three participants (SAPa5M, SAPa9M, SAPa11F), where *go3* appears superfluously before the classifiers specifically associated with those nouns, as in (19) and (20). Regarding the sole case found in their data, as reproduced in (21), Tse et al. (2007:509) analyze it as an instance of self-correction, with the participant beginning the utterance with *go3* and reiterating until eventually arriving the target classifier *gaan1* for "shop." However, the NCS participants in this study consistently add *go3* only before the target classifier-noun (CL-N) pairs shown in (19) and (20), i.e., "a tree" is always realized as **go3 po1-syu6*,[6] without doing so elsewhere. A possible explanation is that for these participants, the CL-N pair has been incorrectly acquired as an entire, inseparable lexical unit for representing the noun referents, which they then enumerate using *go3*.

(19) *嗰 呢一個幅 圖 入便
*gam2 ni1-jat1-**go3-fuk1** **tou4** jap6-bin6,
then this-one-**CL-CL** **picture** in-side
Intended: 'So inside this picture' (SAPa9M, l. 1)

[6] The character representing the sortal classifier 棵 *po1* for trees may also be realized using the pronunciation *fo2*, as is adopted in Tse et al. (2007). Zhu (2019) represents the former pronunciation using the alternate character 夥.

(20) *好 大力 郁緊 嗰個棵 樹
*hou2 daai6-lik6 juk1-gan2 go2-**go3-po1** **syu6**
very big-strength move-PROG that-**CL-CL** **tree**
Intended: 'Very vigorously moving that tree' (SAPa11F, ll. 23–24)

(21) *有 個… 有 個間 買 糖果 嘅 店 喺度。
*jau5 go3… jau5 **go3-gaan1** maai5 tong4gwo2 ge3 **dim3** hai2dou6.
have CL have **CL-CL** buy candy LP **shop** LOC
Intended: 'There is, there is a shop [for] buying candy there.' (Tse et al. 2007:509)

4.5 Connectives

Given the importance of cohesion as a skill reflecting a speaker's ability to organize content at the macrostructural level of the entire narrative discourse (Chan et al. 2023, Fernández 2011, Zhu 2019), this study examines NCS students' production of connectives (sentential conjunctions) to establish logical relations between clauses and form links between events in the story. Referring to the functional classification system for connectives developed in Halliday and Matthiessen (2014), the current analysis groups the connective types found throughout the narrations into five categories: additive, adversative, causal-conditional, clarifying, and spatio-temporal. On average, NCS students produced slightly more connective tokens than CS students (27.69 vs. 23.54). This is not entirely surprising given that Cantonese does not always require connectives to link clauses together, with the logical relations between them instead inferred from context. Consequently, connective use may actually be expected to decrease as proficiency increases.

Table 3 presents the relative frequencies of connective categories averaged across the narratives for participants in each group. For both the test and control groups, most connectives were spatio-temporal types such as 跟住 (*gan1zyu6, next*), 之後 (*zi1hau6, afterwards*), and 嗰 (*gam2*, *then*). The predominance of spatio-temporal connectives may be attributed to the nature of the narrative task. Given that participants were instructed to narrate the story one page at a time, they may feel obliged to explicitly indicate the transitions between events depicted on separate pages using connectives as literary devices for expressing cohesion.

Table 3. Relative frequencies of connective categories in Cantonese narrations.

Connective category (Halliday and Matthiessen 2014)	Test group (NCS) (%)	Control group (CS) (%)
Additive	12.18	6.12
Adversative	15.67	9.70
Causal-conditional	7.86	7.84
Clarifying	1.85	3.84
Spatio-temporal	62.44	72.50

5. Discussion

5.1 Proficiency level of NCS students

Based on both general narrative performance measures and salient linguistic features observed across narrations, this study determines that the current group of NCS students comprises advanced-level speakers. Using the performance descriptors in the American Council on the

Teaching of Foreign Languages (ACTFL) proficiency guidelines for speaking as a reference, they are able to complete an oral narration over a sustained time period of several minutes, during which they demonstrate some control over time frames using aspect and deployment of pragmatic language devices to produce connected discourse throughout the narrative. Nevertheless, the occurrence of extended mazes in some participants—including hesitations and self-corrections as they seek to retrieve appropriate vocabulary and grammatical structures—signals lower fluency compared to the most advanced speakers. In addition, both the use of English code-mixing to express unknown or unrecalled vocabulary items, and the limited repertoire of classifiers, demonstrate incomplete mastery of the target language lexicon. Nevertheless, their performance achievements align them with at least the Advanced Low level for speaking in the ACTFL guidelines (2024:18).

5.2 Prevalence of SWC code-mixing

Of the various linguistic features reported in this study, the manifestation of numerous types of SWC phenomena among a majority of NCS students is remarkable as an indicator of an interlanguage stage in their L2 Cantonese learning. The blending of SWC elements in their speech indirectly demonstrates that their CSL learning has progressed beyond the elementary stage, given that beginner Cantonese learners in the process of acquiring basic speaking competencies would not yet be receiving significant exposure to standard and literary language from written texts. Interference from speaking knowledge of Putonghua (Mandarin) is largely ruled out, given that except for one participant (SAPaPh1F) who successfully completed the narrative task in Putonghua, all NCS students in the test group failed to formulate even basic sentences in Putonghua, leaving the written language as the primary source of SWC elements.

Of course, the incorporation of SWC in Cantonese speech also occurs among the CS students in the control group, and in neither of the groups do SWC elements make up the entirety of their narrations. However, unlike native Cantonese speakers in Hong Kong who consciously differentiate between vernacular Cantonese and SWC, these NCS students do not distinguish between the two varieties, perceiving both to be part of a unified "Chinese" (中文, *zung1man2*)[7] that they are learning, and thus creating unconscious standard/vernacular blends in their speech. Furthermore, even as some of the NCS students in this study displayed some awareness of alternations between vernacular Cantonese and SWC through self-correction (examples (4), (10), and (12)), the chances of committing such blending errors will only increase as they encounter more SWC through learning the written language. Still, the fact that a minority of NCS students could narrate the story without invoking any SWC elements shows that attaining such native-speaker intuitions remains possible despite the learning challenges.

5.3 Emergence of new interlanguage varieties of spoken Cantonese

The systematicity of linguistic features such as SWC elements and a limited classifier inventory among the Pakistani students of the test group suggest the emergence of new interlanguage varieties of spoken Cantonese among South Asian NCS students more broadly. The blurring of boundaries between vernacular Cantonese and SWC in these new interlanguage varieties necessitates discussion of how the involuntary language mixing of NCS students in Hong Kong differs from the discretionary movements between varieties by native Cantonese speakers.

---

[7] In fact, the authors note that when asked about the languages they spoke at home, NCS students typically responded first with "Chinese," resulting in the need for clarifying whether they were indicating Cantonese or Putonghua.

Figure 1 illustrates the relations between the different established varieties of Chinese in Hong Kong. The linguistic environment of native Cantonese speakers in Hong Kong with limited proficiency of Putonghua is unique in the world (except for neighboring Macau) in that it includes both the local vernacular (III) alongside a regional spoken realization of the national standard written language (II), which takes the form of SWC spoken using Cantonese pronunciations (IV). Such a rich linguistic repertoire is not available to CSL learners elsewhere in the world, where the target varieties of spoken Putonghua and SWC all originate from the same standardized stratum, without direct lexical and grammatical input from other regional Chinese varieties such as Cantonese (Yue) or Fujianese (Min). In the figure, the separate expressions for the third-person plural pronoun *them*, in both their spoken and written forms, reveal the intricacies of Cantonese and SWC coexisting as parallel strata within the geographical variety of Hong Kong Chinese (港式中文, *gong2sik1 zung1man2*; Shi et al. 2006).

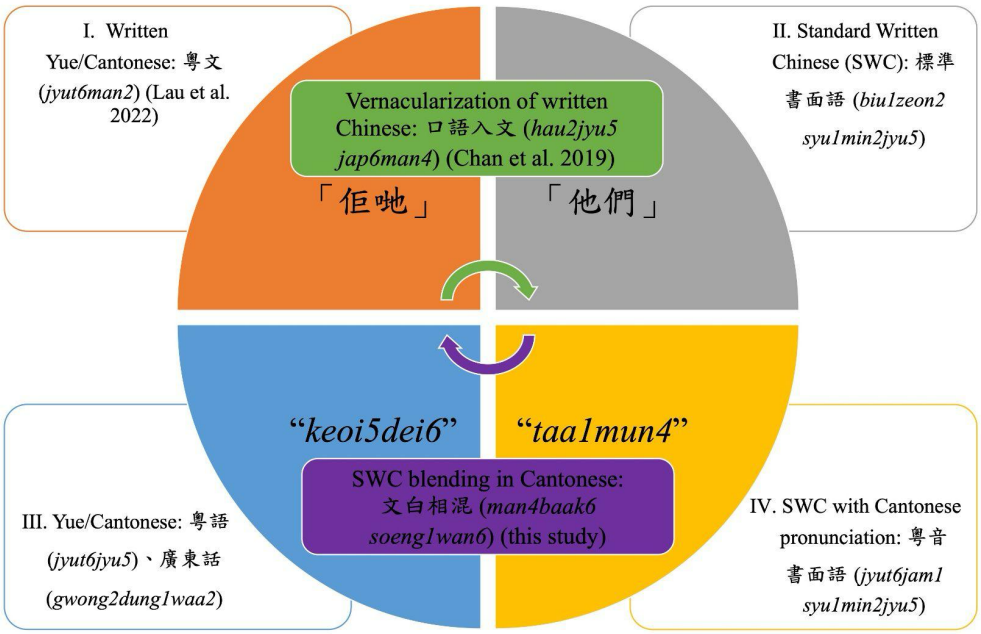

Figure 1. Established and hybrid varieties of Chinese in Hong Kong.

Additionally, the two arrows represent distinct movements across pairs of varieties in the same usage domain, resulting in admixtures in the in-between spaces. The green arrow represents the vernacularization of written Chinese, where written forms of colloquial Cantonese lexical items and grammatical structures are incorporated into SWC (Chan 2019, Chan et al. 2019). On the other hand, the purple arrow represents movement in the reverse direction among the spoken varieties as observed for the NCS students in this study, tentatively labeled as "SWC blending in Cantonese," where the spoken forms of SWC elements are infused into colloquial spoken Cantonese. The difference lies simply in which domain these blends are realized. For instance, the phrase "their things" is 佢哋嘅嘢 (*keoi5dei6 ge3 je5*) in Cantonese and 他們的東西 (*taa1mun4 dik1 dung1sai1*) in SWC, both having the structure [3PL-LP-thing]; however, the blends ?佢哋嘅東西 (?*keoi5dei6 ge3 dung1sai1*) and ?他們的嘢 (?*taa1mun4 dik1 je5*) are both possible in writing and speech alike.

In the past, movements between varieties in a usage domain by native speakers have not been discussed seriously because of the assumption that they know to separate the varieties, only performing code-mixing purposefully to serve particular pragmatic functions. Such movements are even more under-studied among NCS students, who perform them unintentionally, with only a few projects and publications (Chan 2019, Chan et al. 2019) exploring the interlanguage space these students create through the vernacularization process of written Chinese. The occurrence of unconscious SWC blending in Cantonese among NCS students even with advanced proficiency, then, means that it is now imperative to examine the mutual interactions between the two oral varieties in their speech, and the resulting interlanguage spectrum that is currently developing among NCS students of different ethnic and educational backgrounds across schools in Hong Kong.

6. Conclusion

This study described salient linguistic features in the spoken Cantonese among a group of South Asian NCS secondary students with advanced oral language proficiency, observing certain usage patterns in their oral narrative speech data that distinguishes this group from both elementary L2 Cantonese learners and native Cantonese speakers. Focusing on a feature analysis of oral language output, the current analysis encompasses a broad scope, reporting on various phenomena ranging from the lexical and morphological to the syntactic and pragmatic levels. A further advantage of the research design is the inclusion of a control group for comparison, enabling the determination of certain features and patterns as being exclusive to NCS students in the test group. Further studies ought to investigate whether such features are also abundant in their spontaneous speech outside of the oral narration context.

The discovery of extensive SWC blending in the Cantonese speech of NCS students revealed various idiosyncratic, systematic errors that point towards the formation of at least one new interlanguage spoken variety of Chinese among South Asian NCS students in Hong Kong, with the vernacular Cantonese substrate being influenced by the importation of SWC lexical items and morphosyntactic strategies. Such interlanguage varieties emerge because NCS students lack the awareness to distinguish consciously between the Cantonese and SWC strata in their target language education even as they are simultaneously acquiring both, leading to them regarding both as being the same "Chinese." This obscurity is further reinforced by the fact that under current educational practices, neither stratum is acquired completely across all usage domains. Most mainstream public schools in Hong Kong do not organize a separate "Cantonese language" course to instruct NCS students explicitly in what counts as vernacular spoken Cantonese, even as Cantonese is often the primary medium of instruction (MOI), and some schools may identify Putonghua as a distinct spoken variety of Chinese through offering Putonghua as an additional language course. Consequently, because of specialized learning based in different domains, NCS students acquire neither pure Cantonese nor pure SWC, but a hybrid.

The occurrence of SWC blending in NCS students' spoken Cantonese poses important implications for both language pedagogy and curriculum development in Hong Kong. On the pedagogical side, schools may need to review the MOI they select for their NCS students to learn Chinese in. However, regardless of whether Cantonese or Putonghua is ultimately chosen, the school should incorporate in their CSL education explicit instruction regarding the differences between vernacular Cantonese and SWC. Such efforts may be supported by new curriculum materials, especially targeting NCS students with at least intermediate Cantonese speaking proficiency who are beginning the simultaneous acquisition of SWC in writing.

References

American Council on the Teaching of Foreign Languages [ACTFL]. (2024) *ACTFL Proficiency Guidelines 2024.* 4th edn. Alexandria, VA: American Council on the Teaching of Foreign Languages.

Berman, Ruth A. and Dan I. Slobin. (1994) *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chan, Angel, Sarah Chen, Saboor Hamdani, Bernard Tse, and Kelly Cheng. (2023) Storytelling in bilingual Urdu−Cantonese ethnic minority children: Macrostructure and its relation to microstructural linguistic skills. *Frontiers in Psychology*, 14. 924056. www.doi.org/10.3389/fpsyg.2023.924056.

Chan, Shui Duen. (2019) From oracy to literacy: The role of Cantonese in the learning of Chinese reading and writing. In John C. Wakefield (ed.), *Cantonese as a Second Language: Issues, Experiences and Suggestions for Teaching and Learning*. 209-231. Oxford and New York: Routledge.

Chan, Shui Duen (陳瑞端), Yingnan Li (李英男), and Xin Liang (梁鑫). (2019) Brief introduction of the Interlanguage Corpus of Non-Chinese Speaking (NCS) Learners of Chinese in Hong Kong primary schools (「香港非華語小學生語料庫搜索系統」簡介). *Research and Development Project Dissemination Conference of SCOLAR*. PS08. Hong Kong: Standing Committee on Language Education and Research (SCOLAR). scolarhk.edb.hkedcity.net/sites/default/files/uploads/RnD/Project/Dissemination/2019/PS_PPTs/PS08.pdf.

Cheung, Hin Tat. (2019) The effect of age of arrival to the development of Cantonese narrative in ethnic minority children of Hong Kong. In John C. Wakefield (ed.), *Cantonese as a Second Language: Issues, Experiences and Suggestions for Teaching and Learning*. 232-250. Oxford and New York: Routledge.

Census and Statistics Department. (2022) *2021 Population Census – Thematic Report: Ethnic Minorities.* www.census2021.gov.hk/doc/pub/21c-ethnic-minorities.pdf. Hong Kong: Information Services Department, Government of the Hong Kong Special Administrative Region.

Corder, Stephen P. (1967) The significance of learners' errors. *International Review of Applied Linguistics* 5, 161-169.

Erbaugh, Mary S. (2002) Classifiers are for specification: Complementary functions for sortal and general classifiers in Cantonese and Mandarin. *Cahiers de linguistique - Asie orientale* 31.1, 33–69. www.doi.org/10.3406/clao.2002.1602.

Fernández, Camila. (2011) Mindful storytellers: Emerging pragmatics and theory of mind development. *First Language* 33.1. 20-46. www.doi.org/10.1177/0142723711422633.

Gagarina, Natalia, Daleen Klop, Ianthi M. Tsimpli, and Joel Walters. (2016) Introduction: Narrative abilities in bilingual children. *Applied Psycholinguistics* 37.1, 11-17. *Special Issue: Narrative Abilities in Bilingual Children.* www.doi.org/10.1017/S0142716415000399.

Gao, Fang, Chun Lai, and Christine Halse. (2019) Belonging beyond the deficit label: The experiences of 'non-Chinese speaking' minority students in Hong Kong. *Journal of*

Multilingual and Multicultural Development 40.3, 186-197. www.doi.org/10.1080/01434632.2018.1497042.

Halliday, Michael A. K. and Christian M. I. M. Matthiessen. (2014) *Halliday's Introduction to Functional Grammar*. 4th edn. London and New York: Routledge.

Klamer, Marian and Francesca R. Moro. (2020) What is "natural" speech? Comparing free narratives and Frog stories in Indonesia. *Language Documentation and Conservation* 14, 238–313. hdl.handle.net/10125/24921.

Ku, Hok Bun, Kam Wah Chan, and Karamjit Kaur Sandhu. (2005) *A Research Report on the Education of South Asian Ethnic Minority Groups in Hong Kong*. The Hong Kong Polytechnic University.

Lau, Chaak Ming, Grace Wing-yan Chan, Raymond Ka-wai Tse, and Lillian Suet-ying Chan. (2022) *Words.hk*: A comprehensive Cantonese dictionary dataset with definitions, translations and transliterated examples. *Proceedings of the 1st Workshop on Dataset Creation for Lower-Resourced Languages (DCLRL)*. 53-62. Marseille, France: European Language Resources Association.

Li, David Chor-shing and Joanne Yim-ping Chuk. (2015) South Asian students' needs for Cantonese and written Chinese in Hong Kong: A linguistic study. *International Journal of Multilingualism* 12, 210-224.

Matthews, Stephen and Virginia Yip. (2011) *Cantonese: A Comprehensive Grammar*. 2nd edn. London and New York: Routledge.

Mayer, Mercer. (1969) *Frog, Where Are You*? New York: Dial Press.

Shi, Dingxu (石定栩), Jingmin Shao (邵敬敏), and Zhiyu Zhu (朱志瑜). (2006) *Comparison Between Hong Kong Chinese and Standard Chinese* (港式中文與標準中文的比較). 2nd edn. Hong Kong: Hong Kong Educational Publishing Company (香港教育圖書出版社).

Tang, Tsz Wan (鄧芷雲). (2018) *The Relationship Between Chinese Orthographic Knowledge, Transcription Skills and Writing Performance on Ethnic Minority Primary Students in Hong Kong* (香港少數族裔小學生的中文字形知識、轉寫能力和寫作表現的關係). Doctoral dissertation, The Chinese University of Hong Kong.

Tang, Zhixiang (湯志祥). (2000) *Xianggang Yueyu ciyu de kouyuhua ji qi chengyin* (香港粵語詞語的口語化及其成因) [Colloquialization of Hong Kong Cantonese Words and Its Causes]. In Chow-Yiu Sin (單周堯) and Kang Kwong Luke (陸鏡光) (eds.), *Proceedings of the 7th International Conference on Yue Dialects* (《第七屆國際粵方言研討會論文集》). 295-306. Hong Kong: Commercial Press.

Tse, Shek Kam, Hui Li, and Shing On Leung. (2007) The acquisition of Cantonese classifiers by preschool children in Hong Kong. *Journal of Child Language* 34, 495-517.

Thomason, Sarah Grey and Terrence Kaufman (1988). *Language Contact, Creolization and Genetic Linguistics.* Berkeley and Los Angeles: University of California Press.

Wang, Danping and Linda Tsung. (2022) Teaching Chinese to ethnic minority students in Hong Kong: A systematic review. *Chinese Journal of Applied Linguistics* 45.3, 375-393.

Wei, Li and Sherman Lee. (2001) L1 development in an L2 environment: The use of Cantonese classifiers and quantifiers by young British-born Chinese in Tyneside. *International Journal of Bilingual Education and Bilingualism* 4.6, 359-382. www.doi.org/10.1080/13670050108667738.

Wong, Yu Ka. (2010) *Acquisition of Chinese Literacy by Ethnic Minority Children in Hong Kong Primary Schools*. Doctoral dissertation, The Chinese University of Hong Kong.

Zhu, Yue. (2019) *Chinese Oral Language Abilities of Primary Six Non-native Chinese Learners in Hong Kong*. Doctoral dissertation, The University of Hong Kong.

cmflam@hkmu.edu.hk