

# 粵語研究新資源：《香港二十世紀中期粵語語料庫》

錢志安

香港教育學院

## 1. 摘要

本文介紹一個新建構的粵語語料庫，語料來源為上世紀五六十年代香港粵語電影對白。構建語料庫的主要目的是為粵語共時和歷時研究提供新資源和新角度。其中十四套已處理的電影對白語料（約十九萬字，共八千六百多個詞條，總詞頻十四多萬）可在網上（<http://hkcc.livac.org/>）搜索。

## 2. 粵語歷時研究的一些挑戰

粵語研究主要以口語材料為對象，要有系統和嚴謹地收集和處理口語語料一直是粵方言研究的一大挑戰。上世紀七十年代以來，不少粵語研究成果陸續面世（如 Yue-Hashimoto 1972，張洪年 1972/2007，高華年 1980，Matthews and Yip 1994/2010），部分更涉及大規模的田野調查（如詹伯慧和張日昇 1988, 1994, 1998）。這些研究主要是對不同地區的粵語作出共時（synchronic）描述，語料主要來自個別發音人或編者的語感。

除了共時研究，不少學者開始進行粵語歷時（diachronic）研究，有關研究主要通過十九世紀的方言材料構擬早期粵語的面貌。這些研究指出過去兩百年，粵語在詞匯、語音、語法等各方面經歷了不少重要演變：完成體標記“嚟”由“咗”取代（Cheung 1997，郭必之和片岡新 2006，片岡新 2010）、中性問句的 VP-Neg 句型發展為 V-Neg-VP 句型（Yue 2004，Cheung 2001）和間接賓語標記“過”轉為“畀”和工具標記“畀”的消失（Chin 2009, 2011）等，<sup>1</sup> 以上部分特徵現在大多已經消失或只散見於某些粵方言。<sup>2</sup> 假如沒有這些早期材料，我們不可能知道這些散見於個別方言的語言特徵原來是早期粵語的痕跡。由此可見，這些方言材料在粵語歷時研究上有很重要的價值（Yue 2004，錢志安 2011）。

---

<sup>1</sup> 詳細的文獻資料可參閱錢志安（2011）。

<sup>2</sup> 如從化粵語中間接賓語標記仍然用“過”（Chin 2011）；開平方言的中性問句仍多用 VP-Neg 句型（余靄芹 1992）。

不過，我們注意到大部分粵語歷時研究所採用的材料在十九世紀和二十世紀初出版，最早的是馬禮遜（Robert Morrison）1828年的 *A Vocabulary of the Canton Dialect*（Morrison 1828），此書於2001年由倫敦的 Ganesha 出版社重印（Morrison 2001），而最晚的是趙元任1947年出版的 *Cantonese Primer*（Chao 1947，中譯為《粵語入門》），1950年之後的材料較少使用。其中一個原因可能是從二十世紀中期到現在只有六十多年的光景，那時候出生的人到現在才六十多歲，所以如果要了解二十世紀中期的粵語狀況，我們可以跟這些六十多歲的長者進行調查，但是這種“非實時”（*apparent time*）研究背後有一個重要假設，就是發音人現在所用的語言跟他們出生和長大時期的語言沒有很大分別（錢志安 2011），然而我們都很清楚語言絕對不是一成不變，而是不停地演變，這在一些幾代同堂的家庭裏尤為明顯，年青一輩跟年長家庭成員的語言不斷互相影響，所以我們單憑長者的語料並不能完全準確地構擬語言的早期狀況。因此，只有以實時的語料為基礎的歷時研究，才能充分了解和掌握語言的具體發展（Claridge 2008）。

上文提及一些粵語早期特徵的消失時間大約是二十世紀中期，換句話說，新舊特徵很有可能在那段時期的語言中共現。假如我們能夠收集那段時期的實際語料，必定能夠增加我們對粵語演變的了解。但是，有關研究有至少三項挑戰：

- (1) 上文已經指出上世紀中期出版的粵語材料寥寥可數。本文作者的雙賓句研究（Chin 2009）參考了44份早期粵語材料，當中只有四分之一在1940年後出版。
- (2) 大部分早期材料主要是用來學習粵語，因此材料中的句子大都很簡短，結構也不複雜。本文作者收集的雙賓句中，幾乎所有賓語（直接賓語和間接賓語）都是由數詞、量詞和名詞組合而成，甚少含有複雜的成分，如關係子句。這些材料未必能夠完全反映當時語言實際使用狀況。此外，七十年代後出版的粵語研究專著都以語言學框架作為分析和描寫基礎，著作的語料主要基於編者的語感，而不是完全由大量和客觀的語料歸納得來。
- (3) 第三個挑戰是語料的組織和處理。現時粵語歷時研究主要依靠學者自己收集的材料，有時會互相借用材料，以增加語料數量。但是當中最迫切的問題是這些語料沒有以嚴謹的語料庫語言學技術處理和組織。所謂處理和組織並不只是把材料掃描成電子文檔，而是先把語料電子化，然後進行分詞和不同類別的標注（如語音和詞性等），再配以檢索系統，方便查詢和統計。這些技術不但省卻每次因為新研究課題而重新翻查所有材料的時間，更可以提高效率和研究結果的可靠性和準確性。

### 3. 現有的粵語語料庫

除了通過跟個別發音人進行方言調查和收集方言語料之外，很多學者開始收集自然（natural）和真實（authentic）語料，並構建語料庫。現時已公開的現代粵語語料庫主要有以下四個：

- (1) 香港兒童粵語語料庫（Hong Kong Cantonese Child Language Corpus - CANCELP）（Lee and Wong 1998）
- (2) 香港雙語兒童語料庫（Yip and Matthews 2007）
- (3) 香港粵語語料庫（Hong Kong University Cantonese Corpus）（Wong 2006）
- (4) The Hong Kong Cantonese Adult Corpus（Leung and Law 2001）

首兩個語料庫的名稱已經明確顯示它們是為語言習得研究而構建，第三和第四個語料庫的資料主要取自二十世紀 90 年代大學生對談和電台節目訪談。這些語料庫部分語料可在網上查閱。

早期粵語方面，有張洪年開發的語料庫，該語料庫原計劃錄入十七份早期粵語材料，但是構建過程中涉及不少技術問題，如粵語俗字的表示和注音方式，最後只處理了七部材料（楊榮祥 2008），語料庫現已在網上公開（[http://ccl.ust.hk/useful\\_resources/useful\\_resources.html](http://ccl.ust.hk/useful_resources/useful_resources.html)）。

### 4. 另一種適合粵語歷時研究的語料

要選擇合適的口語材料構建語料庫並不容易，尤其是建構一個以過去語言為基礎的語料庫，除非語料事先已記錄下來，否則要追溯有關的時期的語言使用狀況，幾乎是不可能的。根據我們的觀察，有一種二十世紀中期香港粵語口語材料很值得粵語研究者重視和關注，那就是當時在香港十分流行的粵語電影，一般被稱為“粵語長片”。過去已有學者通過多媒體（如電視劇、電台廣播劇和電影）收集和分析粵語口語材料（如 Chan 1996，歐陽偉豪 2006，梁仲森 2005，片岡新 2010，Chor 2010，Leung 2010 等）。當中李雄溪和許子濱（2005）、劉鎮發和蕭佩珊（2010）專門分析和比較粵語長片和現代香港粵語的語言特徵，如語音和句末助詞。

香港電影業有超過一百多年的歷史（鍾寶賢 2004），是香港娛樂事業中的重要一環，尤其是二十世紀中期。根據 Jarvie（摘自鍾寶賢 2004: 177），在短短的二十年間（五十至七十年代），粵語電影的累積產量達到約三千部，年產量平均約 150 部（即差不多每兩天生產一部電影），有關資料見表一。

年份	粵語電影數量	國語電影數量
1952 - 1955	627	222
1956 - 1960	963	314
1961 - 1965	928	206
1966 - 1970	361	286
<b>總數</b>	<b>2879</b>	<b>1028</b>

表一 二十世紀五六十年代香港電影生產量

這些統計數據顯示，粵語長片中的對白可以轉化成一個十分有用的語料庫，作為研究香港半世紀前粵語的重要語料。但是，過往的學者只是分析幾套電影的對白，並沒有把相關語料系統化，製成語料庫。此外，香港以往的電影研究大多把重點放在社會文化方面，如電影故事如何反映社會狀況或市民大眾的心理質素等（如林年同 1978，石琪 1988 等），對電影對白進行語言學研究則較少見。這是由於電影或電視劇中使用的語言過去都被認為不夠自然和真實，因為這些對白都是編劇、導演和演員預先編寫和安排。（有關這些早期見解，參見 Quaglio 2008, 2009，Bednarek 2010，Richardson 2010，Alvarez-Pereyre 2011 的回顧和分析）。不過，近幾年出現了不少以電視劇對白為對象的語言研究，Alvarez-Pereyre (2011) 指出，雖然電視劇對白是預先寫好，但是它們都是集體創作的成果，當中選用的詞匯和句式大多是社會上的最常用和流通的，換句話說，這些對白都可以反映當時語言的使用狀況。

## 5. 香港二十世紀中期粵語語料庫

鑑於粵語長片可以加強粵語歷時方面的研究，2011-2012 年間，我們構建了一個香港二十世紀中期粵語語料庫。語料庫將收錄了 1950 年至 1970 年之間二十部粵語電影的對白，對白以漢字轉寫後，進行分詞，<sup>3</sup> 並配以粵語拼音（主要用作語料查詢）。其中十四套已處理的電影對白（約 19 萬字）可通過互聯網搜索引擎查詢（網址：<http://hkcc.livac.org/>）。除了字詞和讀音之外，用戶也可以運用非語言參數，如演員名字、性別、電影名稱、生產年份等查詢語料。除了粵語研究之外，我們認為語料庫也可以應用到其他語言學範疇的研究，如話語分析、語言和性別、語言和文化互動關係等。以下我們介紹語料庫的一些主要功能。

<sup>3</sup> 以香港教育學院語言資訊科學研究中心過往處理漢語語料分詞原則為基礎。



圖一 香港二十世紀中期粵語語料庫主頁 (<http://hkcc.livac.org/>)

查詢語料需要先登入系統。新用戶可以先開設帳戶（圖二）。

歡迎使用「香港二十世紀中期粵語語料庫」，要檢索語料庫，您需要有一個帳戶，請按這裡註冊。

Welcome to *The Linguistics Corpus of Mid-20th Century Hong Kong Cantonese*. To access the corpus, you need an user account. To sign up, please click [HERE](#).

**登入帳戶/Login**

用戶名 / 登記電郵  
Username/Registered email

密碼  
Password

如引用本語料庫，可參考以下條目：  
香港教育學院語言資訊科學研究中心《香港二十世紀中期粵語語料庫》，檢自於日期，<http://hkcc.livac.org>

圖二：帳戶開設和系統登入介面

語料庫提供三種方式給用戶搜索語料。

- (1) 語料庫資料一覽 (**Global Information**) – 用戶可以按照頻率（由高至低或由低至高排列）瀏覽“總字表”和“總詞表”（見圖三）。

**Word List 總詞表**

Total 8620 word types and 144128 occurrences. The list is sorted by descending frequency.  
共有8620個詞種出現了144128次。以高至低頻率排序。按這裡以低至高排序。

ID 序號	Word 詞語
1	你
2	我
3	呀
4	唔
5	係
6	呢
7	噏
8	佢
9	嗰
10	亮

圖三：語料庫總詞表（首十個高頻詞）

在詞表和字表中，用戶可以點擊詞語，瀏覽詞語訊息。圖四列出含有“畀”（按頻率由高至低排序，位於總詞表第 27 位）的電影數目（14 套）、對話數目（834 個）和總詞頻（901 次）。另外，系統提供詞語的“粵拼”<sup>4</sup>和統一碼（Unicode）。用戶然後可以“查看例句”（圖五）。

**Word 詞語: 畀**

The word 畀 has appeared 901 time(s) in 834 dialog(s) and 14 movie(s).

詞語「畀」於14套電影，共834個對話，出現了901次。

ID 序號	Character 字元	Jyutping 粵拼	Unicode 統一碼
1	畀	bei2,bei3	U+7540

[Check examples 查看例句](#)

[Home 首頁](#) | [Intro 簡介](#) | [Search 搜尋](#)

圖四：詞語“畀”在語料庫中的訊息

ID 序號	Movie 電影	Actor 演員	Text 內文
1	標準丈夫	女	你快啲話畀佢聽同佢娶老婆咪得囉！
2	標準丈夫	張英才	媽媽！我咪出過公司幫佢手嘍囉！點知佢話我唔在行！又話我阻手阻腳，唔畀我做咩嘍！係咪咩？
3	標準丈夫	張英才	嗱！大把……我成個相簿貼滿晒！我攞出嚟畀你哋揀下呀！
4	標準丈夫	男	不過我聲明先呀！如果係啲啲噉噉舞女呀，交際花之類呢，你就無謂畀我睇喇！免至噉心機，揮眼腳！
5	標準丈夫	男	噉喇！有邊啲至係正當當，身家清白嘅女仔，你自己揀出嚟畀我睇喇！
6	標準丈夫	女	鬼咁眼腳！不如噉喇！你聽朝揀到畀我哋喇！
7	標準丈夫	女	好簡單啫！首先我唔想咁快嫁住，就算我係嫁，都唔會嫁畀你哋噉噉花花公子！
8	標準丈夫	李香琴	哦！噉咪話我認叻呀！如果你真係畀我忠靠過我管教呢，包你不出三個月暗，我就可以將佢訓練成一個一標準丈夫！
9	標準丈夫	男	嗱！你叫佢出嚟等我話畀佢知！
10	標準丈夫	李香琴	不過你一定要將管教威全權交晒畀我至得噉囉！
11	標準丈夫	張英才	哦～媽媽講晒畀我聽噉喇！哎！倩姑我哋呢道，都有咩面立足噉喇！噉過去轉換下環境，都好噉

圖五：詞語“畀”在語料庫中的例句

圖五列出含有“畀”字的例句，除了例句外，系統會顯示電影名稱和演繹該對白的演員名字（如不清楚演員名字，系統只顯示演員的性別）。例句已做了初步分詞，並以超鏈結表示，用戶可點擊任何一個詞語進行另一組查詢。用戶也可以點擊演員名

<sup>4</sup> 粵拼全名為“香港語言學學會粵語拼音方案”（<http://www.lshk.org/>），於一九九三年發佈，是一套普及、簡單、合理、易學、易用的粵語拼音轉寫方案。語料庫所顯示的粵拼讀音來自 Unihan Database（<http://www.unicode.org/charts/unihanrsindex.html>），部分標示的粵語讀音可能較為生僻或少用（如“畀”讀成去聲 [bei3]）。要注意的是，語料庫標註讀音並不是用來反映電影中的實際讀音，而是讓用戶多一個條件搜索語料。我們認為即使語料庫現時保留生僻的讀音，對搜索結果應該不會有太大的影響，不過我們以後會對所有讀音進行校對和篩選。



字，找出該演員有份演出的電影，見圖六（以李香琴為例），然後點擊“參與對白”來檢視該演員在語料庫中的所有對白。

**李香琴**

**Gender 性別: F 女**

Acted in 參演電影:

ID 序號	Movie 電影	Year 年分	Director 導演	Role 角色
1	標準丈夫	1965	盧雨岐	七姨三四
2	春色滿園宮	1966	陳焯生	陸蘭香
3	化身情人	1965	楚原	謝美蓮
4	我愛雲蘿蘭	1966	楚原	孕婦陳玉英
5	小偷捉賊記	1962	楊工良	施太、梅姑娘

[Check dialogs 參與對白](#)

[Home 首頁](#) | [Intro 簡介](#) | [Search 搜尋](#)

圖六：李香琴參與的電影

- (2) **基本搜尋 (Basic Search)** – 用戶可以字或詞（根據系統分詞標準）或粵語讀音作為搜索條件，語音系統以粵拼為本，輔以國際音標，方便不諳粵拼的用戶。圖七示範如何查詢語料庫中含有 [p<sup>h</sup>a:i<sup>33</sup>]（粵拼為 paai3）的詞語。圖八顯示檢索結果。

- Search for character/word 輸入字/詞
- Search by Jyutping 粵拼
 

p (/p<sup>h</sup>/) 怕

aai (/ɑ:i/) 挨

3 (陰去, 中入: 33,3) 訓,發

圖七：以粵語讀音 [p<sup>h</sup>a:i<sup>33</sup>] 作為檢索條件

ID 次序	Word 詞語	Occurrences 出現次數
1	派	14 ( <a href="#">check examples 查看例句</a> )
2	特派員	5 ( <a href="#">check examples 查看例句</a> )
3	寫實派	2 ( <a href="#">check examples 查看例句</a> )
4	印象派	1 ( <a href="#">check examples 查看例句</a> )
5	特派專員	1 ( <a href="#">check examples 查看例句</a> )

圖八：以粵語讀音 [p<sup>h</sup>a:i<sup>33</sup>] 作為條件的檢索結果

搜索結果（圖八）顯示含有 [p<sup>h</sup>a:i<sup>33</sup>] 的詞語和頻率，用戶可以點擊“查看例句”檢視例句。

- (3) **進階搜尋 (Advanced Search)** – 除了字詞和粵語讀音外，用戶可以加入非語言參數來檢索語料，如電影名稱（可多選）、演員名字或性別（見圖九）。



Search for character/word 輸入字/詞   
 Search by Jyutping 粵拼  
 All Initials 所有聲母  All Finals 所有韻母  Any Tone 所有聲調   
 Movie 電影  Select Movie(s) 選擇電影 \*  
 All artists 所有演員  Male 男  Female 女  Artist 演員  Select Artist(s) 選擇演員 \*

圖九：進階搜尋介面

## 6. 結語

本文介紹《香港二十世紀中期粵語語料庫》的建構目的和主要檢索功能。我們認為有關語料可以為粵方言研究開闢多一個途徑，填補粵語歷時研究材料在質和量方面的空缺。不過，這個語料庫只代表一個起步點，我們希望它能吸引更多志趣相投的專家學者加入粵語研究行列，進一步加深我們對粵語發展的認識和了解。

### 鳴謝

本語料庫有賴以下機構、學者和同事支持和協助，才能成功開發，作者深表謝意：

- (1) 香港教育學院校內研究基金（項目名稱：*Spoken Corpus Construction and Linguistic Analysis of Mid-20th Century Cantonese*, RG41/2010-2011）；
- (2) 顧問：鄒嘉彥教授、梁仲森先生；
- (3) 語料庫開發團隊成員：黃得森先生、黃國燊先生、周嘉寶先生、莊煒霖先生、孫芷盈小姐、吳詠賢小姐。

本文部分內容曾在第十六屆國際粵方言研討會（2011年12月，香港理工大學）和 *Workshop on Innovations in Cantonese Linguistics*（2012年3月，美國俄亥俄州立大學）發表，並獲得不少與會學者指正。

### 參考文獻

- 高華年。1980。《廣州方言研究》。香港：商務印書館。
- 郭必之、片岡新。2006。早期廣州話完成體標記“曉”的來源和演變。《中國文化研究所學報》第46期，頁91-116。
- 李雄溪、許子濱。2005。五、六十年代香港粵語電影語言研究——以語氣詞“嘍”、“啫”為例。發表於第十屆國際粵方言研討會，香港中文大學。
- 梁仲森。2005。《當代香港粵語語助詞的研究》（*A Study of the Utterance Particles in Cantonese as Spoken in Hong Kong*）。香港：香港城市大學語言資訊科學研究中心。
- 林年同。1978。五十年代粵語電影研究中的幾個問題。收錄於《五十年代粵語電影回顧展》。香港：香港市政局，頁11-16。



- 劉鎮發、蕭佩珊。2010。香港語言變化的探討：透過六十年代粵語電影比較今昔粵語語音。發表於第十五屆國際粵方言研討會，澳門理工學院。
- 歐陽偉豪。2006。從周星馳對白的舊粵語到動詞後置成份的句法構築。發表於香港語言學學會 - 粵語討論會 (WOC-5)：從社會語言學看粵語近代的轉變，香港中文大學。
- 片岡新。2010。粵語體貌詞尾“緊”的演變和發展。香港中文大學哲學博士論文。
- 錢志安。2011。粵語語法的多角度研究。《中國語文研究》第1、2期，頁33-44。
- 石琪。1988。香港片的社會心態。收錄於《第十二屆香港國際電影節——香港電影與社會變遷》。香港：香港市政局，頁10-14。
- 楊榮祥。2008。粵語歷史文獻數據庫的製作與應用。《中國語文研究》第1期，頁1-8。
- 余靄芹。1992。廣東開平方言的中性問句。《中國語文》第4期，頁279-286。
- 詹伯慧、張日昇。1988。《珠江三角洲方言詞匯對照》。廣州：廣東人民出版社。
- 詹伯慧、張日昇。1994。《粵北十縣市粵方言調查報告》。廣州：暨南大學出版社。
- 詹伯慧、張日昇。1998。《粵西十縣市粵方言調查報告》。廣州：暨南大學出版社。
- 張洪年。1972/2007。《香港粵語語法研究》。香港：香港中文大學出版社。
- 鍾寶賢。2004。《香港影視業百年》。香港：三聯書店。
- Alvarez-Pereyre, Michael. 2011. Using film as linguistic specimen. In *Telecinematic Discourse: Approaches to the Language of Films and Television Series*, ed. Roberta Piazza, Monika Bednarek and Fabio Rossi, 47-67. Amsterdam: John Benjamins.
- Bednarek, Monika. 2010. *The Language of Fictional Television: Drama and Identity*. London: Continuum International Publishing Co.
- Chan, Marjorie. 2006. Gender-marked speech in Cantonese: The case of sentence-final particles *je* and *jek*. *Studies in the Linguistic Sciences* 26(1/2): 1-38.
- Chao, Yuen-Ren. 1947. *Cantonese Primer*. Cambridge: Harvard University Press.
- Cheung, Samuel. 1997. Completing the completive: (Re)constructing early Cantonese grammar. In *Studies on the History of Chinese Syntax* (Journal of Chinese Linguistics Monograph, Series No.10), ed. Sun Chaofen, 133-165. Berkeley: Project on Linguistic Analysis.
- Cheung, Samuel. 2001. The interrogative construction: (Re)constructing early Cantonese grammar. In *Sinitic Grammar: Synchronic and Diachronic Perspectives*, ed. Hilary Chappell, 191-231. Oxford: Oxford University Press.
- Chin, Andy Chi-on. 2009. The verb GIVE and the double-object construction in Cantonese in synchronic, diachronic and typological perspectives. Doctoral dissertation, University of Washington.
- Chin, Andy Chi-on. 2010. The GO-type and the GIVE-type indirect object markers in the Conghua dialect. In *Studies in Honor of Jerry Norman* (《羅杰瑞先生七秩晉三壽慶論文集》), ed. W. South Coblin and Anne O. Yue, 245-276. Hong Kong: The Ng Tor-tai Chinese Language Research Centre, Institute of Chinese Studies, The Chinese University of Hong Kong.
- Chin, Andy Chi-on. 2011. Grammaticalization of the Cantonese double object verb [pei<sup>35</sup>] 畀 in typological and areal perspectives. *Language and Linguistics* 12(3): 529-563.
- Chor, Winnie. 2010. Inclusion of the outsider – grammaticalization of the verbal particle *maai* in Cantonese. *Taiwan Journal of Linguistics* 8(2): 33-62.
- Claridge, Claudia. 2008. Historical corpora. In *Corpus Linguistics: An International Handbook*, Volume 1, ed. Anke Ludeling and Merja Kyto, 242-258. Berlin: Mouton de Gruyter.

- Lee, Thomas, and Colleen Wong. 1998. CANCELP: The Hong Kong Cantonese child language corpus. *Cahiers de Linguistique - Asie Orientale* 17(2): 211-228.
- Leung, Man Tak, and Sam Po Law. 2001. HKCAC: The Hong Kong Cantonese Adult Language Corpus. *International Journal of Corpus Linguistics* 6(2): 305-325.
- Leung, Wai Mun. 2010. *On the Synchrony and Diachrony of Sentence Final Particles*. Germany: Lambert Academic Publishing Co.
- Matthews, Stephen, and Virginia Yip. 1994/2010. *Cantonese: A Comprehensive Grammar*. London: Routledge.
- Morrison, Robert. 1828. *A Vocabulary of the Canton Dialect*. Macau: The Honorable East India Company's Press.
- Morrison, Robert. 2001. *A Vocabulary of the Canton Dialect* (reprint of the 1828 edition). London: Ganesha.
- Quaglio, Paulo. 2008. Television dialogue and natural conversation: Linguistic similarities and functional differences. In *Corpora and Discourse*, ed. Annelie Adel and Randi Reppen, 189-210. Amsterdam: John Benjamins.
- Quaglio, Paulo. 2009. *Television Dialogue: The Sitcom Friends vs. Natural Conversation*. Amsterdam: John Benjamins.
- Richardson, Kay. 2010. *Television Dramatic Dialogue: A Sociolinguistic Study*. New York: Oxford University Press.
- Wong, Ping-Wai. 2006. The specification of POS tagging of the Hong Kong University Cantonese Corpus. *International Journal of Technology and Human Interaction* 2(1): 21-38.
- Yip, Virginia, and Stephen Matthews. 2007. *The Bilingual Child – Early Development and Language Contact*. Cambridge: Cambridge University Press.
- Yue-Hashimoto, Anne. 1972. *Studies in Yue Dialects 1: Phonology of Cantonese*. Cambridge: Cambridge University Press.
- Yue, Anne. 2004. Materials for the diachronic study of the Yue dialects. In *The Joy of Research: A Festschrift in Honor of Professor William S-Y. Wang on His Seventieth Birthday*, ed. Shi Feng and Shen Zhongwei, 246-271. Nankai: Nankai University.

通訊地址：香港 新界 大埔 香港教育學院 語言學及現代語言系

電郵地址：andychin@ied.edu.hk

收稿日期：2012年9月8日

接受日期：2012年11月18日