

## 蔚為大觀的粵拼網上資源

張群顯

香港理工大學

香港語言學學會的粵語拼音方案（以下簡稱“粵拼”）面世 20 年於茲，與其他拼式相比，它在使用上的最大優勢在於網上。

“粵拼”所指稱的並不一定是香港語言學學會的粵語拼音方案，因此，在網上搜尋“粵拼”不能說明甚麼問題。“粵拼”的粵拼是 *jyutping*，是個相當獨特的字母串，基本可代表本文的論述對象“粵拼”。《維基百科》的英文版 *Wikipedia* 及《維基百科》所衍生的跨語詞典 *Wiktionary* 都有“*Jyutping*”條，所指稱的正是本文論述對象“粵拼”。在 Google 搜尋“*jyutping*”，一般有 125 萬條材料左右。由於“Yale”的語境比“*jyutping*”闊得多，因此兩者無從比較。為了營造相似語境，我們分別用不加引號的“*jyutping Cantonese*”和“*Yale Cantonese*”來搜尋，結果前者有 441 萬條，後者有 92.7 萬條。<sup>1</sup> 粵拼與耶魯大學粵音拼式在網上的相對熱度可見一斑。

語言學學會的網頁上有一份材料 *Websites related to Jyutping*（粵拼網頁一覽），僅一頁篇幅；重在羅列，而介紹的文字不多，希望本文能補上述網頁材料之不足。網上材料的介紹，有一定的時間性，正好配合本刊的“通訊”本色。

首先要介紹的，是語言學學會本身的網頁內有關粵拼的部分，包含下列子網頁（括號內為對該部分的簡單介紹）：

- *Installation of Jyutping Input Method* / 安裝粵拼輸入法（指導如何使用微軟 Office 2007 內置的粵拼輸入法。）
- *The Jyutping Scheme* / 粵拼方案（顧名思義）
- *Websites related to Jyutping* / 粵拼網頁一覽（見上文）
- *Work Group on Jyutping* / 粵拼工作組（對該工作組的介紹）
- *FAQ on Jyutping* / 粵拼答問（對常見疑問的回應和說明）

由於這是語言學學會的“官方”網址，事事都力求客觀和審慎，應用性不強。要真正領略粵拼在網絡世界的威力，還待其他相關的、友好的，甚至是“發燒友”式的網頁。

---

<sup>1</sup> 2013 年 11 月 2 日香港時間上午 11 時 40 分。

上面說過，粵拼“用場的最大優勢在於網上”，原因何在？根本的原因在於粵拼的創設具有信息技術的前瞻性，所用字母限於在網絡上的傳輸能確保無往而不利的“alphanumerals / 文數字”字符集，具體來說就是限於 1-6 和 a-z。這個取向，加上語言學學會的學術權威和代表性，使得香港特區政府的中文界面諮詢委員會從 1999 年成立開始便一直以粵拼為其工作上唯一粵音拼式。較傳統的語文工作者或以為“中文界面”主要涉及電腦，少涉及語文，或以為對語文來說，“中文界面”不成氣候；境外的語文工作者或以為區區香港一隅，其“中文界面”工作的影響力過不了深圳河和伶仃洋；因而都不以為意。

2001 年，中文界面諮詢委員會的《香港增補字符集》出台，粵語字的編碼和電腦顯示問題逐步迎刃而解，而這些粵語字在香港特區政府官方材料上的標音是唯一用的粵拼。2004 年，香港特區政府把《電腦用漢字粵語拼音表》掛在網上。<sup>2</sup> 本表收漢字 29,145 個，是迄今為止收字最多的漢字粵音標音材料。除了收字多外，本表還有另一個特色，就是對一些認受性較高的系統性語音變異予以承認。這些變異，包括凡 n- 可接受 l-、凡 ng- 可接受零 -、凡零 -（虛語素除外）可接受 ng-、凡在 [ɔ:] 前的 gw / kw 可接受 g / k、凡 ng 音節可接受 m 音節。此外，從拼法（而非音系）的角度看，凡在 [u:] 前的 g / k 也可接受拼作 gw / kw。上面的音，一方面接受異讀 / 異拼，另一方面以較多人主張的讀音 / 拼法作為首選音（稱為“無標記音”）。這麼一來，在態度上承認變異，在操作上突出標準，兩者取得平衡。

《電腦用漢字粵語拼音表》的發布，是粵拼發展的一個轉捩點。即使不是網上作業，這個實質上的漢字粵語標音總彙也是十分重要的參考資料。網上作業的話，其重要性更不言而喻。網上 Unicode 漢字字集和編碼查詢的終極參考網 Unihan<sup>3</sup> 就是用粵拼標注粵音的。Unicode 是舉世字符集的標準，在一切涉 Unicode 的官方論述上，標注粵音都是用的粵拼，這使得粵拼在電腦字符的領域已成了實質上的國際標準。

粵拼是語言學學會的產物，但該會並沒有開發基於粵拼的輸入法。《電腦用漢字粵語拼音表》的發布，卻催生了好幾個不同組織或個人所開發的粵拼輸入系統。香港特區政府本身就提供了“網上中文輸入工具”，<sup>4</sup> 可用粵拼輸入上述全部 29,145 個漢字，在網上作業，不用下載到機上。同樣是網上作業的，還有 Online CPIME 的 Cantonese Pinyin Input Method (Online version)。<sup>5</sup> CPIME 還同時提供其他三種粵音拼

<sup>2</sup> 網址：[http://www.ogcio.gov.hk/tc/business/tech\\_promotion/ccli/cliac/canton\\_pronun\\_list.htm](http://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/cliac/canton_pronun_list.htm) 或 <http://www.iso10646hk.net/jp/index.jsp>

<sup>3</sup> 網址：<http://www.unicode.org/charts/unihan.html>

<sup>4</sup> 網址：<http://www.ccli.gov.hk/webime.html>

<sup>5</sup> 網址：<http://cpime.hk/online/?type=2>

式的輸入。由於其他拼式不像粵拼有 29,145 個漢字大字符集對應資料，凡是不限於粵拼而是兼容其他拼式的廣義粵語拼音輸入系統，就不可避免地只能輸入較小範圍的漢字，例如大五碼字符集。隨機的、能離線作業（相對於網上作業）的粵拼輸入系統中，微軟 Office 2007 所內置的粵拼輸入系統是個里程碑。它只限粵拼，不能用其他拼式。這就保證了它可輸入上述全部 29,145 個漢字。微軟的輸入系統還反映了上文提到的容納變異性的語言學處理。比方說，在輸入單字的時候，鍵入傳統所不容的 l- 拼法也一樣能找到以 n- 拼法為“正”的字，例如“你、男、女”等，只不過在字列中會排在較後的位置就是。

接著要介紹一個也許最不需要我們來介紹的網頁：《粵語審音配詞字庫》。<sup>6</sup> 之所以這麼說，是因為它本身的知名度本身就很高。從 2003 年迄今，瀏覽次數已超過一千五百萬。對粵語字音有興趣的人，鮮有不認識這個網頁的。它是在語言學學會的《粵語拼音字表》基礎上加工而成的，其加工包括添加何文匯、朱國藩（1999）的音。由於它使用了一些由《粵語拼音字表》編者所提供的未發表材料，在《粵語拼音字表》中僅附星號“\*”和井號“#”的音，本網頁卻一一列具個別編者姓名，以配合其文責自負之旨。本網頁固然非常有用，但限於宗旨和體例，其字彙範圍僅限於大五碼字符集。

粵語言語群體不限於香港，澳門也用粵語，而內地的粵語人比香港更多。由於種種原因，澳門和內地的語文學者和政府組織普遍對粵拼不感興趣。正因為如此，一群內地粵語發燒友組成“粵語協會”，建造大型網頁，<sup>7</sup> 大力推廣粵拼，就真是一件教香港人感到欣慰的事。說他們“大力推廣粵拼”，真是一點不假。早在微軟 Office 2007 內置粵拼輸入系統之前，他們已推出基於粵拼的“輕鬆粵拼輸入法”，供免費下載。這套輸入系統，詞彙量比微軟那套的大，而且可選擇輸入繁體字還是簡體字，是筆者個人首選的輸入系統。除了輸入系統外，該網頁還介紹和教授粵拼。可以說，語言學學會該做而未做的，他們做了。

“粵語協會”的網頁規模不小，但說到漢字與粵拼對照材料之多，還不及另一個內地網頁“粵語學習網”。<sup>8</sup> 本網頁並沒有開宗明義以粵拼為準，也沒有特別介紹粵拼，但在材料方面絕大部分實際使用粵拼。究其原因，我相信是由於它在材料的預備上得益於網上一些由漢字轉換為粵拼的免費資源（詳下文）。一小部分不用粵拼的內容，應是直接取自內地一些拼音材料，因而並沒有使用網上漢字轉換粵拼資源。網址內“fyan8”所呼應的是該網頁的另一稱謂“方言吧”。網頁雖有其他方言的材料，但遠遠不及粵語材料之多。

---

<sup>6</sup> 網址：<http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/>

<sup>7</sup> 網址：<http://www.cantonese.asia/>

<sup>8</sup> 網址：<http://www.fyan8.com/>。為方便網上準確搜尋，凡是出以簡化字的網上資源名稱，本文保持其簡化字身分；下同。

以上兩個內地網頁，內容龐雜。接著要介紹的，同樣是內地網頁，但內容相對專注，主要是提供粵語流行曲的歌詞。網頁的名稱是“越K越愛”。<sup>9</sup>這跟粵拼有甚麼關係呢？它為每一首歌的歌詞提供了粵拼對照。不光粵拼，還提供該音節讀音的音頻。多少首？不是數十、不是數百，而是數千首。本來該網頁一直提供歌曲數目，自大約兩年前開始卻不再提供數目。印象中，它最後一次提到的數字是四千多。至於歌手數目，現時提供的數字是 148。歌手許冠傑個人有 83 首，卻竟然找不到站在粵拼的立場最想看到的《有酒今朝醉》。這首歌其中一句是“點知 zoet2 聲跟咗老許”。凡是不區分長而較低的 [œ:] 韻腹和短而較高的 [e] 韻腹的拼式（包括把兩者歸併為不標長短的“œ”的坊間所謂國際音標標音），都不能把 oet 韻和 eot 韻區別開來。<sup>10</sup>《有酒今朝醉》發表於 1976 年。香港語言學者 1993 年發布的粵拼，沒有昧於香港粵語音系的新發展，把兩個韻腹清楚地區別開來。

可以設想，這個網頁背後有漢字與粵拼對照以及每個音節的音頻兩種資料庫。該網頁就利用這兩個資料庫來為歌詞提供粵拼和每字讀音音頻。既然如此，由漢字到粵拼和音頻的過程就不需人手操作。這由漢字連繫粵拼再連繫音頻過程的自動化，衍生出兩種應用系統：一、字典式的逐字查音（包括粵拼和音頻），二、整段文本（包括歌詞）的逐字粵拼注音和提供音頻。這第二項應用，使得該網頁為歌詞注粵拼和逐字提供音頻的工序可以自動化。這種操作方式，該網頁並不以為諱，還特別為訪客提供了“自助粵語歌詞注音工具”，並說明提供此工具的緣由：“因站长工作繁忙，没有时间更新加入歌曲。所以製作了这个自助注音工具。”這個“由漢字轉換為粵拼和音頻”的工具，適用於任何文本，而並不限於歌詞。

這樣為一段文本注粵拼的工具，對使用粵拼的人來說十分有用。網絡上，提供同類功能的免費工具還有好幾個，有的還提供音頻。下表把它們與《越K越愛》放在一起，並把各工具轉換“行路去銀行”成粵拼的結果列出，以資比較。

網頁	音頻	“行路去銀行”轉換成粵拼結果
粵拼易 <sup>11</sup>	✘	[haang4 hang4 hang6 hong4] lou6 [heoi2 heoi3] ngan4 [haang4 hang4 hang6 hong4] 行 路 去 銀 行
Chinese Word Parser <sup>12</sup>	✘	haang4 lou6 {heoi3 heoi2} ngan4 hong4

<sup>9</sup> 網址：<http://ykyl.net/>

<sup>10</sup> 在網上搜尋漢字和拼音對照的《有酒今朝醉》，有兩樁有趣的發現。先是找到一處看似提供對照的，卻原來所注的音是普通話漢語拼音，真是啼笑皆非。另一處真的提供了對照，卻是吧 zoet2 的漢字寫作“著”，然後據此漢字注音作 zoek（不標調）。

<sup>11</sup> 網址：<http://www.jyutpingeasy.net/scgi-bin/toJyutPing.cgi>

<sup>12</sup> 網址：[http://www.cantonese.sheik.co.uk/scripts/parse\\_chinese.php?action=parse](http://www.cantonese.sheik.co.uk/scripts/parse_chinese.php?action=parse)

电子导盲犬粤语词典 <sup>13</sup>	✓	行 haang4 路 lou6 去 heoi3 銀 ngan4 行 hong4
Pinyinsoft <sup>14</sup>	✗	行 路 去 銀 行 haang4 lou6 heoi2 ngan2 haang4
粵語發音詞典 <sup>15</sup>	✓	同上
越K越愛	✓	行 hang4 路 lou6 去 heoi3 銀 ngan4 行 hang4

以上六個網頁可分成五組，每組所給的音各不相同，可知是五起分頭開發的。它們的分歧在於處理一字多音時的不同取向。《粵拼易》的處理最審慎、負責，但加工量也最小。它把每個字所有的音都依字母序羅列出來，讓用者自選；看來這個資料庫的“行”沒有 hong2 音，“銀”沒有 ngan2 音，所據的應是《粵語拼音字表》而不是《電腦用漢字粵語拼音表》。*Chinese Word Parser* 自己建立了粵拼對照的詞彙（詳下），因此可按詞標音。然而，單音節詞“去”還是有兩個可能讀音。在這樣的情況下，該系統也審慎地列出兩個音供使用者自選。“去”的兩音，《粵拼易》與 *Chinese Word Parser* 次序不同。看來，*Chinese Word Parser* 並非字母序，而是以常用的排前。《电子导盲犬粤语词典》把每個字都準確拼出，可見它背後也有粵拼對照的詞庫作為依據。相對於 *Chinese Word Parsar* “去”這個詞兩音並存，《电子导盲犬粤语词典》不見得一定更可靠。試用單字詞“行”來檢驗，*Chinese Word Parsar* hang4, haang4, hong4, hang6 四音並舉固然太寬，但《粵語詞典》只給 hang4 音卻又太窄了。*Pinyinsoft* 和《粵語發音詞典》的處理相同，是最機械、最沒道理的，硬著頭皮取字母序排頭的音；結果，“去”和“銀行”都錯了，“行路”雖然對了，也不過是巧合。《越K越愛》也是只取其中一音。它的取音原則跟 *Pinyinsoft* 不同，取的是該字最常用的音；卻也是不管上下文，硬著頭皮就取該音，效果也好不了多少：“去”是對了，“行路”和“銀行”卻是錯的。

背後有粵拼對照詞彙作依據的兩個網頁，顯然有較佳表現。現在就讓我們介紹它們所從屬的母網頁以及其粵音詞彙。

首先是 *Chinese Word Parser* 的母網頁。<sup>16</sup> 老實說，筆者搞不清楚這個網頁的名稱是甚麼，就姑且參考網址稱它為 *Cantonese.Sheik* 好了。這是個大型網頁，版權擁有人是 Adam Sheik（石亞當）。就性質來說，這是個以英語為媒介的教授粵語和有關漢字的網頁。我們這裡要重點介紹的，是它的網上粵英詞典 *CantoDict*。它的自我介紹

<sup>13</sup> 網址：[http://www.eguidedog.net/cantonese\\_cn.php](http://www.eguidedog.net/cantonese_cn.php)

<sup>14</sup> 網址：<http://pinyinsoft.com/hp/cantonese-to-jyutping/>

<sup>15</sup> 網址：<http://www.yueyv.cn/>

<sup>16</sup> 網址：<http://www.cantonese.sheik.co.uk/>

說：“This dictionary is a collaborative project started in November 2003. Entries are added and mistakes corrected by a team of volunteers from around the world.”意思是：“本詞典是合作式項目，始自 2003 年 11 月。詞條的增加和更正都是靠的一班世界各地的志願者。”在 2003 年 11 月 4 日，該詞典收詞 58350 條，涵蓋 5190 個不同漢字。這是收詞量僅次於 Cowles (1965) 的粵音詞典。

《电子导盲犬粤语词典》的母網頁就是《电子导盲犬》。這是個非常獨特的網頁。地域屬中國內地，卻是全面地書面中英雙語、音檔普粵英三語。網頁自我介紹，“電子導盲犬 (eGuideDog) 項目旨在編寫適合盲人使用的自由軟體。”建立粵拼對照粵語詞彙，是其中一個子項目。“電子導盲犬項目”的負責人是黃冠能，他本人是區分高平與高降調的粵語人。該網頁有一段說明：

我們的粵語字典是基於粵拼系統的。粵拼系統的版權屬於香港語言學學會。然而，為了便於廣州人使用，我們把這個系統擴展到 7 個音調。我們把音調 (contour) 53 從 55 中分離出來，標記為 7 調。例如，“衫、煙 (酒)”是 55 調而“三，(濃)煙”是 53 調。香港人可能沒有察覺這裡面的差別。這個系統並不被語言學所承認，只是我個人為了改善 Ekho TTS 而做的修改。

CantoDict 的特點是詞彙量大，除書面語詞彙外，還收粵語口語詞；但由於是志願者合作式項目，沒有權威性可言。《电子导盲犬粤语词典》的詞彙量不詳，估計也在數萬之譜，也一定包含口語詞。與此兩個詞彙剛好相反，香港特別行政區政府 (2007) 的粵拼注音詞彙收詞較少，僅 9706 條，而且僅限於書面語；但是在香港的小學教育圈中具有權威性。香港特別行政區政府 (2007) 以及它的中英詞典版香港特別行政區政府 (2008) 都有網上版本。<sup>17</sup>

粵拼是須要學習的。一套好的拼音，小孩子學起來一點也不困難。倒是先入為主用慣另一套拼式的成年人要他們改用粵拼或有心理障礙。要學粵拼的話，老師輩不一定幫得上忙。好在網上自學粵拼的資源多的是。香港理工大學編有中英雙語的《網上粵拼教室》。<sup>18</sup>有一位李銳華先生，在網上提供了共 80 頁的 ppt 檔案《粵語拼音教程》，<sup>19</sup>供免費下載使用。上面提到的“粵語協會”，也為粵拼提供了一些說明和教學。

除了上述雙語的或純粹中文的教程外，網上還有好些純英文粵拼教程。下面是其中一種，由三個部分合成一套，是 YouTube 視頻：

<sup>17</sup> 網址：<http://www.edbchinese.hk/lexlist/>

<sup>18</sup> 網址：<http://www.iso10646hk.net/jp/learning/index.jsp>

<sup>19</sup> 網址：<http://www.senseasy.net/leeyuiwah/CHS/Jyutping-tutorial.latest.ppt>

- Learning Cantonese Chinese - Jyutping Pronunciation Part 1<sup>20</sup>
- Learning Cantonese Chinese - Jyutping Pronunciation Part 2<sup>21</sup>
- Learning Cantonese - Jyutping Pronunciation Part 3<sup>22</sup>

對於不諳粵語的人來說，學習粵拼與學習粵語語音兩件事是一件事，也是粵語（不限語音）學習的一部分。上面提到的 *Cantonese.Sheik* 就是以英語為媒介的粵語第二語學習網頁。香港中文大學也有三個粵語第二語學習網頁：*Online Cantonese*<sup>23</sup> 是以英語為媒介的；《粵語拼音速遞》<sup>24</sup> 和《港式粵語速遞》<sup>25</sup> 則是以中文為媒介的。

雪球在滾，網上的粵拼資源正不斷增加。以上憑個人有限的經驗聊作記錄。過不了多久，本文就要過時了。但願如此。

#### 參考文獻

- 何文匯、朱國藩。1999。《粵音正讀字彙》。香港：香港教育圖書公司。
- 香港特別行政區政府教育局課程發展處中國語文教育組。2007。《香港小學學習字詞表》。香港：香港特別行政區政府教育局。
- 香港特別行政區政府教育局課程發展處中國語文教育組。2008。《中英對照香港學校中文學習基礎字詞》。香港：香港特別行政區政府教育局。
- 香港語言學學會。1997 / 2002。《粵語拼音字表》。香港：香港語言學學會。
- Cowles, Roy Thomas. 1965. *The Cantonese Speaker's Dictionary*. Hong Kong: Hong Kong University Press.

通訊地址：香港 九龍 紅磡 香港理工大學 中文及雙語學系

電郵地址：kwan.hin.cheung@polyu.edu.hk

收稿日期：2013年11月8日

接受日期：2013年11月11日

---

<sup>20</sup> 網址：<http://www.youtube.com/watch?v=8XLQTPqL2hU&list=PL0si1gzCQMBIbiw22QVi2Qj-G3VT SFOVd&index=1>

<sup>21</sup> 網址：[http://www.youtube.com/watch?v=\\_FA4JqwYJIU&list=PL0si1gzCQMBIbiw22QVi2Qj-G3VT SFOVd&index=2](http://www.youtube.com/watch?v=_FA4JqwYJIU&list=PL0si1gzCQMBIbiw22QVi2Qj-G3VT SFOVd&index=2)

<sup>22</sup> 網址：<http://www.youtube.com/watch?v=Pm-oIq8OyAE&list=PL0si1gzCQMBIbiw22QVi2Qj-G3VT SFOVd&index=3>

<sup>23</sup> 網址：[http://www.ilc.cuhk.edu.hk/chinese/can\\_eng/layout.html](http://www.ilc.cuhk.edu.hk/chinese/can_eng/layout.html)

<sup>24</sup> 網址：[http://www.ilc.cuhk.edu.hk/chinese/cantontones/\\_home/home.html](http://www.ilc.cuhk.edu.hk/chinese/cantontones/_home/home.html)

<sup>25</sup> 網址：[http://www.ilc.cuhk.edu.hk/chinese/canton\\_express/others/frontpage.html](http://www.ilc.cuhk.edu.hk/chinese/canton_express/others/frontpage.html)