

## IICore 香港特用字音義試釋

張群顯

香港中文界面資訊委員會/香港理工大學  
ctkhcheu@polyu.edu.hk

### 1. 香港特用字過去的際遇

中文電腦使用初期，主要使用兩種編碼字符集，見表[1]。

[1]

年分	字集名	地區	字數	字彙性質	漢字體系
1981	國(家)標(準)	中國內地	6763	現代標準漢語用字	中國內地體系
1984	Big5/大五(碼)	台灣	13053	同上	傳統體系

香港和澳門沒有自己的漢字編碼字符集，是可以理解的。用中文電腦，首先要對字符集作出選擇。香港和澳門都使用傳統體系的漢字，很自然就選擇了大五碼字集。大五碼字集雖然在體系上符合港澳的要求，但是它畢竟在字符集的建立過程中完全沒有考慮港澳的需要。澳門且不說，香港的人名用字“葛”、地名用字“埗”、屋邨用字“邨”、音譯 lift(升降機)的“輦”，大五碼字集都沒有。在那些日子，香港、澳門特用字在信息技術中缺乏支援，可以說，港、澳特用字是漢字世界的孤兒。這個情景在張群顯(1997)早已詳細論及。

香港人是怎樣解決香港特用字的問題的呢？回答是不同時代有不同的做法。作者在自己的教學工作中弄了個有關發展的時序表，見表[2]。

[2]

年份	方式	字碼環境	主要問題
1992 前	隨機隨“機”造“字” (電腦字符)	大五	重複勞動、每機獨立
1992	ISO10646 內 59 字	Unicode	字數少、字碼(Unicode)當時不是常規平台
1995	政府通用字庫	大五	認受性不高
1999	香港增補字符集 /ISO10646	大五及 Unicode	(對字符輸入支援不足)

1992年前，大家隨機地隨“機”在大五碼字集的用戶造字區添加所需字符。由於基本上個人在獨立運作的電腦上造字，大量重複勞動不用說，也甭想對相同字符的碼位編配有任何協調。

1992年，在那含二萬許字符的CJK字符集定案前夕，香港語言學學會有機會向國際標準組織(ISO)的表意文字小組(IRG)游說要在字集中照顧香港的用字需要，並獲得接納。表意文字小組依據香港語言學學會所提交的字符清單把以下59個香港特用字納入含20902字的CJK字符集，見[3]。這59個字，當然起到一定的作用，可它有兩個缺點。第一，字太少了，像“葛”、“埗”、“邨”、“𪗇”等至關重要的字，都沒有包括在內。第二，它只能在Unicode的字碼環境使用，而Unicode在相當長的一段時間並非常規的字碼環境。

[3] 𪗇 𪗈 𪗉 𪗊 𪗋 𪗌 𪗍 𪗎 𪗏 𪗐 𪗑 𪗒 𪗓 𪗔 𪗕 𪗖 𪗗 𪗘 𪗙 𪗚 𪗛 𪗜 𪗝 𪗞 𪗟 𪗠 𪗡 𪗢 𪗣 𪗤 𪗥 𪗦 𪗧 𪗨 𪗩 𪗪 𪗫 𪗬 𪗭 𪗮 𪗯 𪗰 𪗱 𪗲 𪗳 𪗴 𪗵 𪗶 𪗷 𪗸 𪗹 𪗺 𪗻 𪗼 𪗽 𪗾 𪗿 𪗸 𪗹 𪗺 𪗻 𪗼 𪗽 𪗾 𪗿

1995年，配合互聯網在香港起步，港英政府把內部使用的大五碼外字集公開，讓各界可以下載安裝這批字，使字符與字碼的對應開始有個共識，名為“政府通用字庫”。這套字庫有兩方面的問題：第一，字庫的建立沒有計劃、沒有原則；第二，一些在外字區早已造了一定數量字符的用家，寧可使用和推廣自己那套。總的來說，“政府通用字庫”的認受性有待提高。

1999年，回歸後的香港政府成立了中文界面諮詢委員會，在政府通用字庫的基礎上發展出“香港增補字符集”。字符集依據一定的原則去決定字符能否入集，多年來行大五碼和Unicode雙軌制，為2008年起為新入集的字只編配Unicode字碼的做法提供過渡。經過多年的努力，常規字集以外香港所需要的字，今天絕大部分都已包括在內了。因此，這個字符集今天有相當高的認受性。

## 2. IICore 與香港特用字

CJK字符集今天的總數約9萬，有點尾大不掉，因而幾年前已有International Ideograph Core (IICore)的倡議。對香港人來說，IICore的研製是個很好的機會讓香港的特用字能進入一個長遠來說比國標0集6763字和大五碼13053字的國際認受性更高的小字集。若嫌9萬字太累贅，內地可退到國標6763字，台灣可退到大五碼13053字，但香港不能。在香港司空慣見的“𪗇”字，Unicode是個五位數字，即在字符集延伸區(Extension) B的範圍，足見其邊緣性。

香港既有“香港增補字符集”，是否就乾脆要求把該字符集2004版的4941個字悉數放進IICore呢？那樣做雖然簡單省事，卻是不智的，主要原因是：“香港增補字符集”礙於要把“政府通用字庫”的3,049字悉數接收下來，無論此後

漢字的入集考量如何嚴謹，也改變不了那 3,049 字中包含大量按新尺度本不該入集的字符。換句話說，即使是認受性比“政府通用字庫”高得多的“香港增補字符集”，也沒能集中反映香港最需要哪些字符。

在這樣的背景下，香港政府中文界面諮詢委員會轄下的中文電腦用字工作組花了相當長的時間和相當多的心力，開了多次會議，整理出一份香港人所需要的字符清單，獲委員會通過，提交給表意文字小組考慮放入 IICore。香港人的期望是，過去因不在國標 0 集和大五碼字集內以致長期欠缺支援的香港特用字，可以進入 IICore，不用再當孤兒。

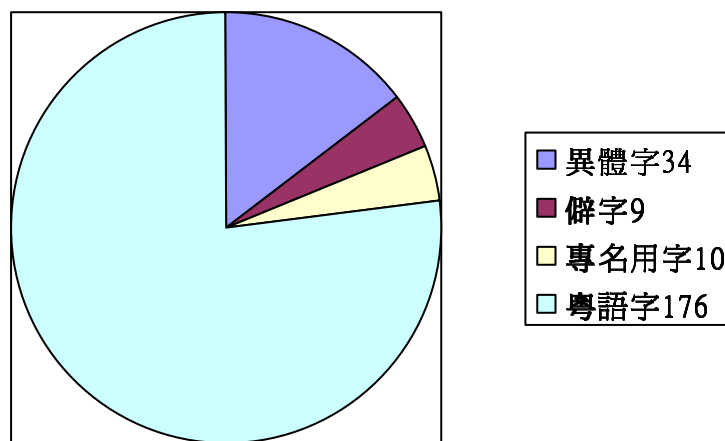
在 2004 年制定的 IICore，收 9810 字，其中有 833 字同時見於“香港增補字符集”。我們相信這 833 字已包含了香港的特用字，不過這 833 字也同時包含了別的地方也使用甚至更需要的異體字和簡化字。因此，這 833 字也未足以集中反映香港的特用字。

獨特地為香港人所需要的是哪些字呢？IICore 內，有 229 個字（也在 833 字之內）是唯一由香港為 IICore 的研製而提交的。香港中文界面諮詢委員會把這批字定性為核心香港用字。考慮到這批字跟香港關係的特殊性，我們認為“香港特用字”的提法最能反映其本質。

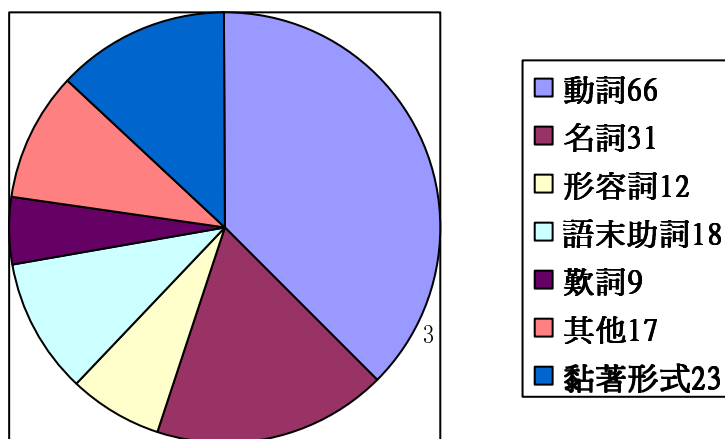
### 3. 香港特用字分類

這 229 個字可首先分為四大類，如圖[4]所示，而那 176 個粵語字本身又可據詞類或語法功能分為圖[5]所示各類。

[4]



[5]



#### 4. 各字身份的確定

一個漢字的屬性傳統分形音義三方面。對編碼字符集內的字符來說，字的基本形體是現成的。要確定某字符的身份，就是要確定那字符的音和義。香港在提交字符給表意文字小組考慮納入 IICore 的時候，已為每個字符提供了粵語讀音。因此，就這 229 個香港特用字來說，讀音也是現成的。形音義三方面，唯一欠缺的是字義。

為這 229 字分類以及交代它們的意義，兩者相關而互動。以上的分類，是交代它們意義的第一步。不同類別的字符，交代字義的具體做法會不一樣。

先從字數最少的兩類說起，即僻字和專名用字。

表[6]的九個字，跟香港和粵語都沒有內在關係。譬如“𦘔”，因“𦘔𦘔”多用於簡化字環境，所以“繁體”的“𦘔”主要用於香港。其他的字都因各不相同的理由而於香港用得較多。這一組字的普通話讀音都是現成的。有個別字，多個以交代大字符集字符屬性為務的網站都只提供粵音，而沒有提供普通話讀音，“狽”是其中一例。其實，“狽”字《漢語大字典》明明標了 gēng 音。

[6]

字	粵音	漢語拼音	釋義	用例	用例釋義
𦘔	mun5	mǎn	一種微生物		
鉅	bat1	bù	一种金属元素，符号 Pu。		
𦘔	sil	sī		~擦音	Sibilant(語音學術語)
𦘔	gul	gū		~𦘔	車輪
疱	paau3	pào		~疹	一種疾病
獁	Maa5	mǎ		猛~	←mammoth; 古哺乳动物，长毛象
芪	kei4	qí		北~	一種中藥材
狽	gang1	gēng	沙皮狗(terrier)		
𦘔	bang1	bēng	擬聲詞；←bang		

表[7]的十個字，都是專名用字。只要有用例，就已交代了那字的身分，不用強加釋義。人名所涉兩人，都是知名人士。

[7]





字	粵音	詞類	附註
嘅	ge3	結構助詞	
嘅	hai2	介詞	
响	hoeng2	介詞	與“響”通，作為介詞多作“响”。
嗰	naa1	介詞	
嚟	gan2	動詞補語	
咗	zo2	動詞補語	
埗	bung6	量詞	
嚟	gau6	量詞	
啲	go2	指示詞	
嘍	baang4	擬聲詞	
嚇	but1	擬聲詞	
啲	cyut2	擬聲詞	
噍	dut1	擬聲詞	
噍	gwut2	擬聲詞	
噍	syut1	擬聲詞	
啲	di1	數量詞	
𠬞	ham6	數量詞	

“黏著形式”要添加其他成份成為詞才能作詞類劃分。表[11]為每字添上其他成份，然後據所構成的詞作詞類劃分。

[11]

字	粵音	構成詞	附註	詞類
嘍	baang6	𠵼 <sup>~</sup> laang6		數量詞
糍	ci4	糯米 <sup>~</sup>		名詞
躑	dan1	踏 <sup>~</sup>		動詞
哋	dei2	肥肥 <sup>~</sup>	具能產性	形容詞
囡	doe4	圓 dam4 <sup>~</sup>		形容詞
慙	fit1	揸 <sup>~</sup>		動詞
喺	gaal	而 <sup>~</sup>	也作“家”	時間詞
鐳	gaan2	殺手 <sup>~</sup>	gaan3 變調而成	名詞
𠵼	gaat6	<sup>~</sup> zaat2	“甲由”又作“由甲”	名詞
蠔	hyun2	黃 <sup>~</sup>	不限粵語，但粵語常用。	名詞
咁	jaa1	哎[ai1] <sup>~</sup>		嘆詞
𠵼	kaa4/kaa2	kaa4kaa2(聲)		擬聲詞
𧈧	kam4	<sup>~</sup> 𧈧		名詞
𠵼	laang6	𠵼 baa6 <sup>~</sup>		數量詞
𠵼	lam1	<sup>~</sup> leot3	見 Cheung & Bauer 2002:273	動詞
𧈧	maang1	泥 <sup>~</sup>		名詞
𠵼	mol	<sup>~</sup> 囉[lo1]		名詞
𠵼	pik1	<sup>~</sup> lik1paak1laak1		擬聲詞
𠵼	san2/seon2	𠵼 <sup>~</sup>		名詞
𠵼	zaa6	𠵼[ngaa6] <sup>~</sup>		形容詞
由	zaat2	甲 <sup>~</sup>	“甲由”又作“由甲”	名詞
蠅	zil	蚊 <sup>~</sup>		名詞
𠵼	zin2	牛 <sup>~</sup>		名詞

以上對 229 個香港特用字的身分，按其類屬作出了初步的勾勒，相信對 IICore 字符屬性資料庫的建立不無幫助。

### 參考書目

- Cheung, Kwan-hin & Robert Bauer. 2002. *The representation of Cantonese with Chinese characters*, Journal of Chinese Linguistics Monograph Series #18.
- 漢語大字典編輯委員會. 1992. 《漢語大字典(縮印本)》，湖北辭書及四川辭書。
- 李學銘主編. 2000. 《常用字字形表(二零零零年修訂本)》，香港:香港教育學院。
- 台灣“中華民國”教育部. 2004. 《異體字字典》，<http://140.111.1.40/main.htm>.
- 張群顯. 1997. 《漢語信息技術與香港語言生活的配合問題》，載《語文建設》，109 期，1997 年 7 月，頁 20-24.