

香港個別字字形調整
與
香港字符集橫向擴展(Horizontal Extension)的建議

建議人：陸勤和陳輝恒

背景和問題

在整理香港字形的工作中，第 57 次會議¹所討論的「溫/𤇗」相關（含〔𤇗/𤇗〕組部件）等字，確認香港的慣常寫法是「𤇗」。由於以〔𤇗/𤇗〕作為部件的多個漢字在 ISO/IEC 10646 中，沒有認同，且分開編碼，會議的決議是：「含“𤇗”和“𤇗”的異體字如分別編碼，應保留“𤇗”和“𤇗”兩種寫法；如只有一個編碼，楷體和宋體都用“𤇗”」。

會後我們額外發現有兩組香港使用漢字（含〔兌/兌〕及〔告/告〕兩組部件）其字形有與〔𤇗/𤇗〕類同之處，在 ISO/IEC 10646 中，同樣沒有認同，且分開編碼。舉例來說，大五碼中的「兌」(Big5 0xA749)，香港的慣常寫法是「兌」。但是，「兌」「兌」等字在 ISO/IEC10646²被編上獨立內碼。如果對「兌」對應的大五碼（和相應的 U 碼：U+514C）的字形進行修正，更改後的香港字形會和 ISO/IEC10646 的另一個漢字字形相同（U+5151）。

由於大五碼對 U 碼的映射關係不可以改變³，如果更改大五碼所對應的字形，就會產生兩個問題：第一，我們的修正造成了同一個字形有兩個 U 碼。這會造成對 ISO/IEC10646 編碼原則的破壞，造成不同時期生成的文件所對應的字形出現不一致性。第二個問題是，在使用符合香港標準字形的中文平台上將只能顯示「告」、「兌」，而無法使用常見的「告」字、傳統的「兌」字。

針對這些問題，我們亦對香港的電腦使用進行了調查，發現現存大部分的家用商用電腦作業系統均基於 U 碼；無論使用微軟視窗或是蘋果電腦，無論在中文（中國）、中文（香港）或是中文（台灣）輸入環境下，

¹ 中文電腦用字工作小組文件編號 2014/04《制訂香港中文異體字的編碼及字形規範》頁 15 問題 2.3

² 為方便我們將 ISO/IEC 10646 的編碼簡稱“U 碼”

³ ISO/IEC 10646 規定所有編碼字符集向 U 碼的映射(mapping)不可以更改。否則會產生不同系統雙向轉換的不一致性(inconsistency in round trip conversion)。

兩種字形均可分別打出，例如輸入〔金口竹山〕，用戶分別可選〔1 兌 2 兌〕。而當使用者選 2 輸出「兌」時，所輸出的內碼也不是大五碼所對應的 U 碼 U+514C，而是 U+5151。基於這一調查，我們認為最好的解決方案是當異體字在 ISO/IEC 10646 中分開編碼，保持與大五碼對應的 U 碼使用「告」、「兌」字形。

另外調查發現有不少系統都依賴使用者電腦預設字體區分兩種異體字，比如香港特別行政區政府教育局課程發展處中國語文教育組所製的《小學學習字詞表》網上版中，「告」字一條之「《常用字字形表》註」內容為「或作告，今作告」。若果兩字不再區分，或讓現存數位資料失去原義，對現存系統構成重大衝擊，影響深遠。

ISO/IEC 10646 對應的中日韓表意文字達 8 萬餘字，不少為生闢正字或異體字，若電腦輸入法或字體廠商需要全數支援，不合乎成本效益；大五碼與香港增補字符集成為必須支援字元的最大參照。但現存的大五碼與香港增補字符集並無包含一些香港常用、在 ISO/IEC 10646 分開編碼的異體字。如果輸入法或字體僅僅支援大五碼與香港增補字符集在 ISO/IEC 10646 對應的字元，無法在現有電腦平台上打出或顯示這些香港慣用的字形。因此有必要考慮將 ISO/IEC 10646 中香港慣常使用的漢字納入一個適應香港本地化⁴使用的香港字集。

建議一

我們認為，如果對這些大五碼內的漢字直接更改字形，不僅會破壞 ISO/IEC 10646 的編碼原則，還會造成用戶使用的不方便，因此，我們建議，按照第 57 次會議所討論的溫溫等字（含〔𠂇/𠂈〕組部件）所作決定（議程 2.2 序號 2.3）——「含“𠂇”和“𠂈”的異體字如分別編碼，應保留“𠂇”和“𠂈”兩種寫法；如只有一個編碼，楷體和宋體都用“𠂇”，提議凡含〔兌/兌〕及〔告/告〕兩組部件的異體字如分別編碼，應保留〔兌/兌〕及〔告/告〕兩組部件兩種寫法；如只有一個編碼，楷體和宋體都用〔兌〕及〔告〕」。

建議二

建議將含〔𠂇/𠂈〕、〔兌/兌〕及〔告/告〕三組部件、在 ISO/IEC 10646 分開編碼的這批字直接進行香港字符集對 ISO/IEC 10646 的橫向擴展。也就是說，將 ISO/IEC 10646 原本不是香港提出編碼的漢字作為香港字符集的一個部分。通過橫向擴展可以確認香港的慣用字形，同時也不破

⁴ 與本地化（localization）相應的技術稱之為：locale。電腦平臺以香港 locale 來組織和支持相應的字形、輸入法、以及其他與本地化有關的信息和關聯技術。

壞大五碼和 ISO/IEC 10646 的對應關係，一舉兩得。此建議中所涉及的需要橫向擴展的 22 個字符列於附表中。

建議三

鑒於方便廠商對香港漢字及相應的本地化技術的支援，我們建議在 ISO/IEC10646 編碼集內建立一個完整的稱為《香港字符集》的命名字符集⁵。字符包括目前香港使用的所有大五碼的漢字（和其他符號）、香港增補字符集的漢字（和其他符號），以及香港確認的橫向擴展漢字（或符號）。除本文件附錄所包括的 22 個漢字，應更改現行的香港增補字符集增收原則，確定接受對 ISO/IEC10646 的縱向⁶和橫向擴展的原則。

我們認為，這三點建議能夠幫助香港電腦系統擺脫一直以來沿用大五碼所造成的限制，使香港能夠更加全面和完善的完成從大五碼向 ISO/IEC 10646 國際編碼平臺的轉換。

⁵ 在 ISO/IEC10646 為支援本地化，容許定義命名字符集（named character set）。

⁶ 現在《香港增補字符集》的增收原則實際上是對 ISO/IEC 10646 進行的縱向（vertical）擴展。

附件：建議進行橫向擴展的字表

序號	大五碼	U 碼	大五碼字形	香港字形	橫向擴展的 U 碼
1	0xA769	U+544A	告	告	U+543F
2	0xA749	U+514C	兌	兑	U+5151
3	0xAEAE	U+6085	悅	悦	U+60A6
4	0xD1BE	U+6329	掙	掙	U+635D
5	0xD5A7	U+6553	斂	敛	U+655A
6	0xD5BF	U+68B2	稅	税	U+68C1
7	0xD258	U+6D97	浼	浼	U+6D9A
8	0xB57C	U+7A05	稅	税	U+7A0E
9	0xB2E6	U+812B	脫	脱	U+8131
10	0xB8C0	U+86FB	蛻	蜕	U+8715
11	0xBBA1	U+8AAA	說	說	U+8AAC
12	0xBE55	U+92B3	銳	銳	U+92ED
13	0xBE5C	U+95B1	閱	閱	U+95B2
14	0xB6FE	U+5ABC	媼	媼	U+5AAA
15	0xB759	U+614D	慍	愠	U+6120
16	0xBA72	U+6C33	氫	氢	U+6C32
17	0xE2BE	U+7185	熅	焮	U+7174
18	0xEAD5	U+7E15	縕	縕	U+7DFC
19	0xE3A6	U+8183	膾	膾	U+817D
20	0xC4AD	U+860A	蘊	蕴	U+85F4
21	0xEEC1	U+8F40	輻	輻	U+8F3C
22	0xC1DF	U+919E	醞	醞	U+9196