

《清實錄》(宣統政紀) 舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中研院史語所漢籍電子文獻工作小組
2007/5/24 黃榮順 製作

一、《清實錄》宣統政紀實錄部分(宣統政紀.xml)使用舊版造字 39 個，字頻 460 次，詳如附件一。這 39 個造字中，15 個可轉成 Windows XP 可顯示的字，字頻 228 次；另外 24 個字必須轉成構字式，字頻 232 次。而這 39 個造字，全部使用了重碼的構字部件，屬於不需轉碼的狀況。

二、附件一的造字分析表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字

丙、字頻(txt)：造字在「.txt」文件的出現次數

丁、字頻(xml)：造字在「.xml」文件的出現次數

戊、Big5：造字的 Big5 碼

己、Unicode：造字所對應的 Unicode 碼

庚、WinXP：造字在 Windows XP 的對應字形

辛、構字式：Windows XP 若無對應字形，則改採用構字式

壬、備註凡例：

1、重碼部件，不轉碼：同時使用舊漢籍造字及構字式，導致一個字碼代表兩種字形的重碼問題，必須一一比對原書字形，以判斷應取何者之字形，取構字式部件者不轉碼。「宣統政紀」的缺字皆屬此類情形，如編號 5334，舊漢籍造字字形為「暉」，構字式部件為「扌」，比對原書發現書中皆取構字式部件字形「扌」，故不需轉碼。

2、待造字：舊版漢籍造字字形為電腦缺字，且漢字構形資料庫尚未收錄此字形，等待補造字中，所以「造字」欄空白無法看到字形。如編號 5001 的「召𠂔己」。

三、Unicode 目前收錄的漢字總數為 70194，分屬於三個不同區段，詳如表一。目前 Windows XP 只支援 CJK 認同表意文字區的 20902 個字，內碼為 4E00-9FFF。所以造字編號 5004 的「𧇗」字，Unicode 編碼為 22927，由於 Windows XP 並不支援，仍須改用構字式「𧇗𠂔心」。

表一、Unicode 的字數及編碼區段

Unicode	字集子集合	新增字數	新增編碼區段	總字數	WinXP
1.1 版	CJK 認同表意文字區	20902	4E00-9FFF	20902	支援
3.0 版	CJK 認同表意文字擴充 A 區	6582	3400-4DFF	27484	不支援
3.1 版	CJK 認同表意文字擴充 B 區	42710	20000-2A6D6	70194	不支援

附件一、《清實錄》「宣統政紀.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4176	輒	1	1	83BF	8F19	輒		灾，重碼部件，不轉碼。
4254	鷗	10	10	844E	9D7E	鷗		愛，重碼部件，不轉碼。
4350	侶	4	4	84D0	5058	侶		虫，重碼部件，不轉碼。
5001		1	1	88E7			召 厶 己	爰，重碼部件，不轉碼。 待造字。
5004	顰	2	2	88EA	22927		頰 厶 心	半，重碼部件，不轉碼。
5012	媵	2	2	88F2	36D0		女 厶 更	茲，重碼部件，不轉碼。
5178	禿	1	1	89FB	462A		禿 厶 充	龜，重碼部件，不轉碼。
5233	覽	2	2	8A73	3F49		臨 厶 瓦	月，重碼部件，不轉碼。
5329	鬻	6	3	8AF5	26269		形 鬻 缶 回 口	亻，重碼部件，不轉碼。
5330	趨	44	21	8AF6	47C6		走 厶 畢	辵，重碼部件，不轉碼。
5332	歐	28	4	8AF8	22FDB		區 厶 攴	宀，重碼部件，不轉碼。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5333	撻	7	7	8AF9	22BC6		扌 𠂔 𠂔	攵，重碼部件，不轉碼。
5334	暉	83	8	8AFA	6692	暉		扌，重碼部件，不轉碼。
5338	暎	27	27	8AFE	668E	暎		艹，重碼部件，不轉碼。
5342	曆	10	10	8B43			形 厂 林 日 口	卩，重碼部件，不轉碼。
5344	磬	7	7	8B45	55F8	磬		攵，重碼部件，不轉碼。
5347	憾	2	2	8B48	617D	憾		四，重碼部件，不轉碼。
5351	肱	285	2	8B4C	80AC	肱		𠂔，重碼部件，不轉碼。
5352		3	3	8B4D			形 日 〇 丰 丩 口	丩，重碼部件，不轉碼。 待造字。
5356	繇	2	2	8B51	221CD		繇	忄，重碼部件，不轉碼。
5358	麓	2	2	8B53	23BDB		形 牙 攵 厂 毛 口	𠂔，重碼部件，不轉碼。
5365	隄	7	7	8B5A	9681	隄		卩，重碼部件，不轉碼。
5368	轟	331	2	8B5D	66E7	轟		人，重碼部件，不轉碼。
5378	巒	2	2	8B67	5DD9	巒		衣，重碼部件，不轉碼。
5382		1	1	8B6B			馬 厶 卩	足，重碼部件，不轉碼。 待造字。
5387	憇	3	2	8B70			形 亻 夂 心 口	忄，重碼部件，不轉碼。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5391	擡	1	1	8B74	648E	擡		ㄩ，重碼部件，不轉碼。
5416	頰	8	8	8BAF	2947A		冉厶頁	𠂇，重碼部件，不轉碼。
5468	逕	1	1	8BE3	28553		形之四于	然，重碼部件，不轉碼。
5479	颯	56	1	8BEE	4D77		圭厶黽	𠂇，重碼部件，不轉碼。
5493	之	1	1	8BFC			之	彡，重碼部件，不轉碼。
5643	涎	2	2	8CF5	6D8F	涎		𠂇，重碼部件，不轉碼。
5689	𠂇	2	2	8D64	20087		𠂇	爻，重碼部件，不轉碼。
5796	𠂇	1760	73	8DF1	26248		𠂇厶缶	厶，重碼部件，不轉碼。
5797	灑	1048	149	8DF2	7054	灑		厶，重碼部件，不轉碼。
5798	虧	114	27	8DF3	48DC		虛厶卩	厶，重碼部件，不轉碼。
5800	邃	118	4	8DF5	908C	邃		〇〇，重碼部件，不轉碼。
5807	睥	456	28	8DFC	4079		目厶辟	形，重碼部件，不轉碼。
5808	賀	456	28	8DFD	27D4D		𠂇厶貝	𠂇，重碼部件，不轉碼。