

《清實錄》（滿洲實錄）舊版造字轉碼說明

中研院資訊所文獻處理實驗室

2007/6/21 趙苑曲

一、《清實錄》滿洲實錄部分(滿洲實錄.xml)使用舊版造字 13 個，字頻 59 次，詳如附件一。這 13 個造字中，11 個可轉成 Windows XP 可顯示的字，字頻 57 次；另外 2 個字必須轉成構字式，字頻 2 次。

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：造字在文件中的出現次數

丁、Big5：造字的 Big5 碼

戊、Unicode：造字所對應的 Unicode 碼

己、WinXP：造字在 Windows XP 的對應字形

庚、構字式：Windows XP 若無對應字形，則改採用構字式。

辛、備註凡例：

1、錯字：與原書字形不符，用正確的字取代原造字，如編號 2099 的「錢」字，原字應為「鍍」。

2、Unicode 字型呈現差異：Unicode 字型與舊漢籍造字有些微差異，但只是字體風格差異，實際上仍為同一個字，因此仍取 Unicode 字型。如編號 4081 的「翺」字，Unicode 字型呈現為「翺」，實際上仍為同一字。

3、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 5533 的「𪛗米」。

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 2099 的「錢」字，Unicode 編碼為 4939，由於 Windows XP 並不支援，仍須改用構字式「金𪛗」。

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》「滿洲實錄.xml」造字分析表

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
2099	鏹	1	9679	4939		金𠂔𠂔	𠂔金厂炎𠂔，錯字。
3796	決	6	815B	51B3	決		
3817	厠	1	8170	53A0	厠		
3992	嘗	2	82A4	751E	嘗		
4024	礮	1	82C4	792E	礮		
4033	稽	2	82CD	7A2D	稽		
4081	翺	1	82FD	7FFA	翺		Unicode 字型呈現差異
4085	耻	1	8342	803B	耻		
4176	輒	1	83BF	8F19	輒		
4643	徃	40	86BB	5F83	徃		
5164	肇	1	89ED	8088	肇		
5533		1	8C65			𪛗𠂔米	
5697	煑	1	8D6C	7151	煑		Unicode 字型呈現差異