《清實錄》（太祖）舊版造字轉碼說明

中研院資訊所文獻處理實驗室
2006/12/28 趙苑曲

一、 《清實錄》太祖實錄部分(太祖.xml)使用舊版造字 20 個，字頻 28 次，詳如附件一。這 20 個造字中，18 個可轉成 Windows XP 可顯示的字，字頻 26 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。

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 1616 的「⿺辶里」字，Unicode 編碼為 284E6，由於 Windows XP 並不支援，仍須改用構字式「⿺辶里」。

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》「太祖.xml」造字分析表

編號	造字	字頻	Big5	Unicode	WinXP	構字式
1616	𨑨	1	936D	284E6		辶△里
1885	𧊒	1	9540	882D	𧊒	
3791	况	5	8156	51B5	况	
3801	剗	1	8160	5257	剗	
3816	厓	1	816F	5393	厓	
3846	壻	2	81AF	58FB	壻	
3858	寃	1	81BB	5BC3	寃	
3900	拕	1	81E5	62D5	拕	
3968	煅	1	826A	7145	煅	
4010	皐	2	82B6	7690	皐	
4028	禀	1	82C8	7980	禀	
4167	賫	1	83B6	8CEB	賫	
4176	輙	1	83BF	8F19	輙	
4309	刼	1	84A7	523C	刼	
5240	擡	1	8A7A	6481	擡	
5302	昬	1	8ADA	662C	昬	
5541	蠹	2	8C6D	8827	蠹	
5564	厭	1	8CA6			厭△木
5666	汚	1	8D4D	6C5A	汚	
5693	鹻	2	8D68	9E7B	鹻	