

《清實錄》(世祖) 舊版造字轉碼說明

中研院資訊所文獻處理實驗室

2007/1/3 趙苑曲

一、《清實錄》世祖實錄部分(sezhu.xml)使用舊版造字 63 個，字頻 160 次，詳如附件一。這 63 個造字中，39 個可轉成 Windows XP 可顯示的字，字頻 111 次；另外 24 個字必須轉成構字式，字頻 49 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 1493 的「鐸」字，原字應為「鐸」。
- 2、異體字：與原書字形不符，但不是錯字，只是異體字，改用原書字形取代。如編號 1240 的「𠄎革卅𠄎伐」字是「鞭」的異體字，「𠄎革卅𠄎戍」也是「鞭」的異體字，因書中採用的是「𠄎革卅𠄎戍」字，故以「𠄎革卅𠄎戍」字取代「𠄎革卅𠄎伐」。
- 3、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 4144 的「霸」字，Unicode 字型呈現為「霸」，實際上仍為同一字。
- 4、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 1073 的「算么」。

部分「造字」欄空白者，因已使用其他字形取代，故不另外補造字。如編號 1240 的「𠄎革卅𠄎伐」字，已由「𠄎革卅𠄎戍」取代，故不另造「𠄎革卅𠄎伐」。

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 1136 的「歛」字，Unicode 編碼為 28A22，由於 Windows XP 並不支援，仍須改用構字式「金厶炊」。

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》「世祖.xml」造字分析表

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
383	灰	1	FCA6			灰	
675	食	1	FE6E			食	
841	礲	1	8E77	7932	礲		
1073		1	8FE4			算厶么	待造字。
1088	暘	1	8FF3	757C	暘		
1136	歛	2	9064	28A22		金厶炊	
1189	璽	2	90BB	24A71		殿厶玉	
1218	焯	2	90D8	24423		火厶鼎	
1240		1	90EE			𠄎革卅𠄎伐𠄎	𠄎革卅𠄎𠄎，異體字。

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
1493	鐸	1	92B1			金厶革	「鐸」，錯字，取代。
1551	紘	1	92EB	7D8B	紘		
1633		1	937E			彡厶涸	待造字。
1885	逢	1	9540	882D	逢		
1923		1	9566			弼厶木	待造字。
1927	榱	2	956A	3BC6		木厶產	
1932	鏘	2	956F	28A74		金厶穿	
1935	鑠	1	9572	9443	鑠		
2389	鏗	1	9861	28BF5		金厶興	
2394	磬	1	9866	40D5		𠂔厶石	
2648	鴻	1	99E9	2A19F		氵厶勅厶鳥	
3287	榿	1	9DF4	234BB		𠂔木气土	
3793	減	1	8158	51CF	減		
3802	劓	1	8161	529A	劓		
3806	劓	1	8165	52E0	劓		
3816	厓	1	816F	5393	厓		

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
3843	堦	11	81AC	5826	堦		
3875		5	81CC			豈厶支	待造字。
3916	敷	4	81F5	656D	敷		
3975	牕	3	8271	7255	牕		
3981	珉	5	8277	73C9	珉		
3990	甄	2	82A2	750E	甄		
4027	裨	11	82C7	7986	裨		
4035	穽	1	82CF	7A7D	穽		
4036	窰	5	82D0	7AB0	窰		
4144	霸	1	837D	8987	霸		Unicode 字型呈現差異。
4180	皐	1	83C3	8FA0	皐		
4198	鑛	1	83D5	945B	鑛		
4216	鞞	4	83E7	97C8	鞞		
4234	羸	12	83F9	9A58	羸		
4254	鷓	9	844E	9D7E	鷓		
4274		3	8462			扌厶友	待造字。

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
4283	戲	11	846B	622F	戲		
4314	洽	1	84AC	6D5B	洽		形: 讠 令 口 口, 錯字。
4374	漼	1	84E8	3D36		讠 厶 隹	
4409	回	12	854C	518B	回		
4668	噐	1	86D4	5650	噐		
4730	璽	1	8753	58D0	璽		
4746	齧	1	8763	28932		形: 牙 夂 厘 口	
4778	鞞	1	87A5	97B4	鞞		
4933	崗	1	88A3	5CDD	崗		
4955	龜	1	88B9	9F9C		龜	
4961	盼	1	88BF	80B9	盼		
5223		5	8A69			形: 扌 穴 人 夕 口	待造字。
5437	并	1	8BC4	5E77	并		
5463	岢	1	8BDE	65F9	岢		
5491	榱	1	8BFA	6967	榱		
5493	之	1	8BFC			之	

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
5556	隸	1	8C7C	96B7	隸		
5588	恠	3	8CBE	6061	恠		
5730	阨	1	8DAF	9638	阨		
5753	潜	3	8DC6	6F5C	潜		
5764	籓	2	8DD1	7C5E	籓		
5782	瀨	1	8DE3	701E	瀨		