

《婦人良方大全》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

2008/3/26 陳建安 製作

一、《婦人良方大全》一書(婦人良方大全.xml)使用舊版造字 44 個，字頻 128 次，詳如附件一。這 44 個造字中，40 個可轉成 Windows XP 能顯示的字，字頻 122 次；另外 4 個字必須轉成構字式，字頻 6 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(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、異體字問題：新版漢籍考量到使用者檢索及使用時的便利性，將用字原則改為除專詞等特殊情形之外，一律改用標準字呈現。如編號 1403 的「穴△兔」係「冤」字之異體，故以「冤」字取代。又編號 1641 的「胆」字，係「膽」字之簡體字，亦以正體的「膽」字取代。

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 4145 的「羈」字，Unicode 字型呈現為「羈」，實際上仍為同一字。

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 2260 的「癩」字，Unicode 編碼為 3FD7，由於 Windows XP 並不支援，仍須使用構字式「疒△癩」。

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、《婦人良方大全》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
556	商	1	1	FDB6	5547	商		
783	𠂇	1	1	FEFC			𠂇	
1278	碍	1	1	9155	788D	碍		
1403		1	1	91F4			穴厶兔	冤，異體字問題。
1641	胆	6	6	93A8	80C6	胆		膽，異體字問題。
1781	藎	1	1	9475	854B	藎		
1885	蠱	3	3	9540	882D	蠱		
2260	癩	2	2	977D	3FD7		疒厶頰	
3064	鰓	4	4	9CB2	9C02	鰓		
3066	𠄎	1	1	9CB4	519D	𠄎		宜，異體字問題。
3145	擗	1	1	9D44	64C0	擗		
3778	併	9	9	8149	5002	併		併，異體字問題。
3792	凉	6	6	8157	51C9	凉		涼，異體字問題。
3796	决	2	2	815B	51B3	决		決，異體字問題。
3797	凭	1	1	815C	51ED	凭		憑，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3814	却	20	20	816D	5374	却		卻，異體字問題。
3858	冤	3	3	81BB	5BC3	冤		冤，異體字問題。
3931	椀	4	4	8245	6900	椀		
4036	窰	4	4	82D0	7AB0	窰		
4038	竈	7	7	82D2	7AC8	竈		
4063	藥	1	1	82EB	7CF5	藥		
4068	綫	1	1	82F0	7DAB	綫		線，異體字問題。
4112	蔴	1	1	835D	8534	蔴		
4114	藥	4	4	835F	8616	藥		
4131	𪔐	4	4	8370	8842	𪔐		
4145	羈	1	1	837E	898A	羈		Unicode 呈現差異。
4184	迹	1	1	83C7	8FF9	迹		
4186	遡	1	1	83C9	9061	遡		
4195	鉢	1	1	83D2	9262	鉢		
4218	韭	4	4	83E9	97EE	韭		
4223	顙	1	1	83EE	984B	顙		
4232	隲	2	2	83F7	96B2	隲		鷺，異體字問題。
4267	鼈	8	8	845B	9F08	鼈		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4307	罵	1	1	84A5	99E1	罵		罵，異體字問題。
4423	臀	1	1	855A	81CB	臀		臀，異體字問題。
4501	獮	1	1	85CA	7356	獮		
4515	呪	1	1	85D8	546A	呪		
4767	懷	1	1	8778	61B9	懷		
4803	腭	1	1	87BE	816D	腭		
5220	毌	1	1	8A66	6BCC	毌		母，異體字問題。
5299	戟	2	2	8AD7	39B8		卓𠂇戈	戟，異體字問題。
5408	紕	1	1	8BA7	7D4D	紕		
5549	藁	5	5	8C75	85F3	藁		
5666	污	5	5	8D4D	6C5A	污		污，異體字問題。