

《備急千金要方》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

2008/3/24 陳建安 製作

一、《備急千金要方》一書(備急千金要方.xml)使用舊版造字 134 個，字頻 379 次，詳如附件一。這 134 個造字中，86 個可轉成 Windows XP 能顯示的字，字頻 289 次；另外 48 個字必須轉成構字式，字頻 90 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、校對問題，舊版漢籍錯字，可用程式全部取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中；若該造字的所有頻次，皆屬於錯誤使用的錯字情形，可以用程式全部取代為正確字形。如編號 4728 的「衣衾」字，原字為「卒」。

2、重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代：舊漢籍造字與構字式部件外字，皆是建立在 Big5 的使用者造字區中，兩者所使用的碼區是重疊的，在同時使用構字式和舊漢籍造字的新版漢籍中，這些重疊的字碼可能被用以表達兩個不同字形，我們稱之為「重碼問題」。在使用構字式的 xml 檔中，舊漢籍造字也會呈現為重碼的部件外字，因而必須將採用舊漢籍造字部分以對應的 Unicode 或構字式取代，才能呈現正確字形；至於採用部件外字者則不需轉碼。如編號 5351，字碼 8B4C，對應的舊漢籍造字為「肱」，部件外字為「𠂇」，一一比對原書，確認是「肱」者，再以手動取代。

- 3、校對問題，新舊版缺字並存：漢籍在舊轉新的過程中，以人工修改方式取代舊漢籍造字，雖然修改的結果是正確的，但是只修改到部分造字，沒有全部修改到，導致舊漢籍造字在新版漢籍檔案中，一部分已被取代成 Unicode 或構字式，另一部分仍為舊漢籍造字的並存情形。處理方式為配合漢籍工作室的修改，將沒被修改到的舊漢籍造字，以對應的 Unicode 或構字式取代。如編號 3961，舊漢籍造字為「灾」，比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「灾」者，在 xml 檔中部分已人工修改為 Unicode 的「灾」字，故配合修改，將編號 3961「灾」字，手動取代為 Unicode 的「灾」字。又如編號 773 的「厶」字，txt 檔中共出現 2 次，有 1 次當作構字式部件使用，而 xml 檔已將此構字式修改為 Unicode 的字形，因而造成頻次不同，處理方式為將剩下的 1 次修改為「厶」。
- 4、標記問題：標記問題主要是漢籍電子文獻在舊轉新的過程中，採用了新的標記語言，而在修改標記時，對於某些標題句做了增刪的動作，因為這些標題句的內容包含了缺字，所以這些更改標記的動作造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號 4789 的「蝮」字，xml 檔在 449-1 頁新增了一個標題，因而造成頻次不同，處理方式為將所有頻次皆修改為「蝮」。
- 5、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 4432 的「𦉰 𦉱 𦉲」係「纏」字之異體，故以「纏」字取代。又編號 2996 的「双」字，係「雙」字之簡體字，亦以正體的「雙」字取代。
- 6、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 1925 的「𪔐」字，Unicode 字型呈現為「𪔐」，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 787 的「𪔐」字，Unicode 編碼為 26848，由於 Windows XP 並不支援，仍須使用構字式「雍 𪔐 肉」。

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

四、 txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，詳細取代內容可參考附件二的手動取代表。附件二的手動取代表的欄位說明如下：

- 甲、 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- 乙、 檔案原文摘錄：摘錄檔案(備急千金要方.xml)中包含該造字的文句段落，其中紅字底線的部分，為需要手動取代的原文。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。
- 丙、 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。

五、 對於在缺字之外發現的錯字，詳細取代內容可參考附件三的校出錯字表。附件三校出錯字表的欄位說明說明如下：

- 甲、 檔案原文摘錄：摘錄檔案(備急千金要方.xml)中包含錯字的文句段落，其中紅字底線的部分，為需要手動取代的字形。
- 乙、 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。

附件一、《備急千金要方》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
348	与	1	1	FC61			与	
396	急	2	2	FCB3			急	
572	囟	2	2	FDC6			囟	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
656	音	5	5	FE5B			音	
660	𠂇	2	2	FE5F			𠂇	
698	𠂇	1	1	FEA7			𠂇	
711	𠂇	1	1	FEB4			𠂇	
718	𠂇	3	3	FEBB			𠂇	𠂇，異體字問題。
746	𠂇	2	2	FED7			𠂇	
749	𠂇	1	1	FEDA			𠂇	
773	厶	2	1	FEF2			厶	厶，校對問題，新舊版缺字並存。
777	𠂇	2	2	FEF6			𠂇	
784	𠂇	2	1	FEFD			𠂇	𠂇，校對問題，新舊版缺字並存。
787	𠂇	1	1	8E41	26848		雍𠂇肉	
792	𠂇	1	1	8E46	5E2C	𠂇		
844	𠂇	2	2	8E7A	9ACE	𠂇		
1059	𠂇	2	2	8FD6	51A3	𠂇		最，異體字問題。
1067	𠂇	4	4	8FDE	68CA	𠂇		棋，異體字問題。
1146	𠂇	17	17	906E	7506	𠂇		
1238	𠂇	1	1	90EC	77E6	𠂇		侯，異體字問題。
1310	𠂇	4	4	9175	45EB		虫𠂇麻	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1374	紆	1	1	91D7	25F8D		纟厶干	
1530	紕	3	3	92D6	7CFA	紕		
1544	紕	2	2	92E4	7D59	紕		
1638	洩	1	1	93A5	3CDB		氵厶與	
1664	腸	2	2	93BF	8193	腸		
1818		1	1	94BC			莉厶余	藜，異體字問題。
1905	棍	2	2	9554	88E9	棍		
1925	羈	1	1	9568	8989	羈		Unicode 呈現差異。
1999	贊	2	2	95D4	8CDB	贊		贊，異體字問題。
2091	癰	1	1	9671	222CA		广厶雍	癰，校對問題，舊版漢籍錯字，可用程式全部取代。
2144	胚	1	1	96C8	80A7	胚		
2159	𪔐	1	1	96D7	4D35		麥厶并	
2195	攢	2	2	96FB	3A4C		扌厶費	
2275	鰓	2	2	97AE	4C49		魚厶且	
2374	冢	1	1	9852	51A1	冢		冢，異體字問題。
2452	蠍	2	2	98C2	45F6		虫厶葛	
2461	葱	1	1	98CB	44E4		艹厶忽	葱，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2501	颺	1	1	98F3	98C3	颺		
2505	飜	1	1	98F7	98DC	飜		
2564	剌	1	1	9973	34E8		夾𠂔 𠂔	
2572	街	4	4	997B	8858	街		
2628	莢	2	2	99D5	845C	莢		
2709		1	1	9A67			𠂔𠂔十𠂔見	葱，異體字問題。
2915	悤	1	1	9BBA	6964	悤		
2938	簪	1	1	9BD1	7C2E	簪		簪，異體字問題。
2996	双	1	1	9C4C	53CC	双		雙，異體字問題。
3040	畝	2	2	9C78	7552	畝		畝，異體字問題。
3257	藕	4	4	9DD6	85CA	藕		
3770	个	5	5	8141	4E2A	个		個，異體字問題。
3776		1	1	8147			𠂔𠂔𠂔棧	木𠂔𠂔箋，異體字問題。
3829	咏	1	1	817C	548F	咏		
3830	咤	1	1	817D	54A4	咤		
3836	喻	3	3	81A5	564F	喻		
3858	寃	2	2	81BB	5BC3	寃		冤，異體字問題。
3860	𡵓	4	4	81BD	5C20	𡵓		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3931	椀	5	5	8245	6900	椀		
3932	椶	1	1	8246	6936	椶		
3940	欝	2	2	824E	6B1D	欝		
3944	氷	1	1	8252	6C37	氷		
3948		1	1	8256	6CAD	沝		沐，校對問題，舊版漢籍錯字，可用程式全部取代。
3961	灾	2	1	8263	707E	灾		灾，校對問題，新舊版缺字並存。
3966	烟	6	6	8268	70DF	烟		
3993	畀	1	1	82A5	2233F		畀	
4002	瘡	1	1	82AE	7607	瘡		
4016	着	17	17	82BC	7740	着		著，異體字問題。
4017	蟬	1	1	82BD	880F	蟬		
4023	碓	1	1	82C3	78D3	碓		
4060	穀	1	1	82E8	7CD3	穀		穀，校對問題，舊版漢籍錯字，可用程式全部取代。
4063	藜	16	16	82EB	7CF5	藜		
4093	舐	2	2	834A	445B		舌𠂔氏	舐，異體字問題。
4095	舩	1	1	834C	8229	舩		
4113	蔺	1	1	835E			艹会間	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4148	𩇑	1	1	83A3	4C1F		云𩇑鬼	魂，異體字問題。
4172	𩇒	3	3	83BB	8E77	𩇒		蹶，異體字問題。
4173	𩇓	1	1	83BC	8EAD	𩇓		
4176	𩇔	20	20	83BF	8F19	𩇔		
4216	𩇕	1	1	83E7	97C8	𩇕		
4218	𩇖	4	4	83E9	97EE	𩇖		
4226	𩇗	5	5	83F1	98E1	𩇗		
4260	𩇘	11	11	8454	9E9E	𩇘		
4282	𩇙	1	1	846A	2A4A0		黑𩇙敢	
4325	𩇚	7	7	84B7	8CD6	𩇚		𩇚，異體字問題。
4395	𩇛	1	1	84FD	9721	𩇛		
4423	𩇜	3	3	855A	81CB	𩇜		𩇜，異體字問題。
4432		1	1	8563			𩇝𩇞 厂𩇟𩇠	纏，異體字問題。
4490	𩇡	1	1	85BF	4E9D	𩇡		齋，異體字問題。
4495		7	7	85C4			麥𩇢少	麥𩇢少，異體字問題。
4520	𩇣	1	1	85DD	285FF		𩇤𩇥𩇦	邈，異體字問題。
4525	𩇧	6	6	85E2	54EF	𩇧		
4546	𩇨	1	1	85F7	9D76	𩇨		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4560	協	1	1	8646	604A	協		
4561		1	1	8647			𠂔 ㄣ 厂 曲 比 𠂔	麤，異體字問題。
4643	徃	2	2	86BB	5F83	徃		
4668	噐	4	4	86D4	5650	噐		噐，異體字問題。
4716	塡	1	1	8745	57FF	塡		
4728	卒	1	1	8751	461A		衣 𠂔 十	卒，校對問題，舊版漢籍錯字，可用程式全部取代。
4789	虺	9	10	87B0	878B	虺		虺，標記問題。
4794	蓁	5	5	87B5	8448	蓁		
4795	礪	5	5	87B6	7847	礪		
4797	堡	2	2	87B8	57A1	堡		
4812	羴	3	3	87C7	26351		羴 𠂔、	羴，異體字問題。
4814	健	7	7	87C9	5FA4	健		
4848	輩	1	1	87EB	8EF0	輩		輩，異體字問題。
4850	薛	1	1	87ED	26D6E		薛	
4980	藝	1	1	88D2			藝 𠂔 火	藝，異體字問題。
4981	皛	1	1	88D3	9CEC	皛		
4999	巘	1	1	88E5	5DD7	巘		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5013	彳	1	1	88F3			彳	
5026		1	1	8941			𠂇广回示	廩，異體字問題。
5034	弃	1	1	8949	5F03	弃		棄，異體字問題。
5074	𠂇	3	3	8971	5C2B	𠂇		
5075	彳	11	11	8972			彳	
5095	来	4	4	89A8	6765	来		來，異體字問題。
5133	憊	2	2	89CE	6199	憊		
5180	拘	1	1	89FD	6285	拘		
5192	隙	2	2	8A4A	969F	隙		
5292	𠂇	1	1	8AD0	65C9	𠂇		
5306		1	1	8ADE			𠂇 𠂇 間	簡，異體字問題。
5351	𠂇	2	5	8B4C	80AC	𠂇		與 𠂇 重碼，重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代。
5447	𠂇	2	2	8BCE	8E83	𠂇		
5478	𠂇	2	2	8BED	4D77	𠂇		
5482	𠂇	1	1	8BF1	8812	𠂇		
5533		5	5	8C65			𠂇 𠂇 米	穀，異體字問題。
5541	𠂇	1	1	8C6D	8827	𠂇		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5558	龕	20	20	8C7E	9E84	龕		
5588	恡	1	1	8CBE	6061	恡		
5694		2	2	8D69			𠂇𠂇𠂇𠂇𠂇	概，異體字問題。
5702	厨	2	2	8D71	53A8	厨		廚，異體字問題。
5704	𧈧	5	5	8D73	6C0A	𧈧		
5709	𧈧	2	2	8D78	872F	𧈧		
5713	𧈧	1	1	8D7C	481E		𧈧𧈧𧈧	
5720	𧈧	1	1	8DA5	578D	𧈧		
5803	𧈧	2	2	8DF8	4ABC		出𧈧頁	

附件二、《備急千金要方》手動取代表

編號	檔案原文摘錄	手動取代結果
773	字形。	𠂇字形。
784	人解𧈧亦脊	人解𧈧𧈧亦脊
3961	皆得攘鮮滅惡	皆得攘災滅惡
4789	𧈧（全部取代）	𧈧（全部取代）
5351	寒則生𧈧	寒則生𧈧
	𧈧則陰病	𧈧則陰病

附件三、《備急千金要方》校出錯字表

檔案原文摘錄	手動取代結果
諸本篇作𧈧𧈧？围	諸本篇作算
治產後往𧈧來寒熱	治產後往來寒熱