

《肘後備急方》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

2008/2/25 陳建安 製作

一、《肘後備急方》一書(肘後備急方.xml)使用舊版造字 49 個，字頻 71 次，詳如附件一。這 49 個造字中，40 個可轉成 Windows XP 能顯示的字，字頻 60 次；另外 9 個字必須轉成構字式，字頻 11 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、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 5516 的「○○束」字，原字為「棘」。
- 2、校對問題，新版漢籍錯字，人工取代：漢籍在舊轉新的過程中，以人工修改方式取代舊漢籍造字，但是卻取代為錯誤的字形或加以刪除，因此出現舊版漢籍沒有，卻在新版漢籍出現的錯誤情形。處理方式為把被修改為錯字的部分，以及沒被修改到的舊漢籍造字，一併以對應的 Unicode 或構字式取代。如編號 4848 的「輩」字，在 xml 檔中被刪除了一次，故頻次不同，處理方式為在被刪除處補上「輩」字，並將其他未經修改的部分，直接手動取代為「輩」。
- 3、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「糸 厂 墨 

4、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 5705 的「廐」字，Unicode 字型呈現為「廐」，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 2452 的「𪛗」字，Unicode 編碼為 45F6，由於 Windows XP 並不支援，仍須使用構字式「虫𪛗」。

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

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(肘後備急方.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。
- 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。

附件一、《肘後備急方》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
975	𪛗	1	1	8F60	4C1B		𪛗𪛗	
1067	𪛗	1	1	8FDE	68CA	𪛗		
1079	總	3	3	8FEA	7DEB	總		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1187	盖	2	2	90B9	76D6	盖		蓋，異體字問題。
1664	腸	1	1	93BF	8193	腸		
1688	魷	1	1	93D7	9B6B	魷		
2021	趲	2	2	95EA	8D82	趲		
2325	睭	1	1	97E0	77A4	睭		
2452	蠮	1	1	98C2	45F6		虫𧈧葛	
3048	叅	1	1	9CA2	53C5	叅		
3770	个	1	1	8141	4E2A	个		個，異體字問題。
3793	減	2	2	8158	51CF	減		
3796	決	2	2	815B	51B3	決		
3840	坂	1	1	81A9	5742	坂		
3850	姪	2	2	81B3	59D9	姪		
3884	忽	1	1	81D5	6031	忽		
3909	擡	1	1	81EE	64E1	擡		
3944	氷	3	3	8252	6C37	氷		
3951	湏	1	1	8259	6E4F	湏		
3974	牀	2	2	8270	7240	牀		
3996	疎	1	1	82A8	758E	疎		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4063	藁	1	1	82EB	7CF5	藁		
4093	舐	1	1	834A	445B		舌𠂔氏	舐，異體字問題。
4115	藁	3	3	8360	85C1	藁		
4131	𠂔	2	2	8370	8842	𠂔		
4176	輒	2	2	83BF	8F19	輒		
4218	菲	1	1	83E9	97EE	菲		
4254	鷗	1	1	844E	9D7E	鷗		
4307	罵	1	1	84A5	99E1	罵		
4432		1	1	8563			𠂔 厂墨	纏，異體字問題。
4447	捏	1	1	8572	63D1	捏		
4515	呪	3	3	85D8	546A	呪		
4546	鷗	2	2	85F7	9D76	鷗		
4767	懷	1	1	8778	61B9	懷		
4794	藁	2	2	87B5	8448	藁		
4848	輩	2	1	87EB	8EF0	輩		輩，校對問題，新版漢籍錯字，人工取代。
4904	尅	1	1	8864	5C05	尅		
5074	尅	1	1	8971	5C2B	尅		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5082	訖	1	1	8979	279BD		言厶尔	
5099	惣	1	1	89AC	60E3	惣		
5108	惛	3	3	89B5	60FD	惛		
5176		2	2	89F9			氵厶冗	沉，異體字問題。
5219	妃	1	1	8A65			女厶已	妃，校對問題，舊版漢籍錯字，可用程式全部取代。
5478	飮	1	1	8BED	4D77	飮		
5516		2	2	8C54			〇〇束	棘，校對問題，舊版漢籍錯字，可用程式全部取代。
5558	麓	1	1	8C7E	9E84	麓		
5621	汨	1	1	8CDF	6CEA	汨		
5705	廐	1	1	8D74	5ED0	廐		Unicode 呈現差異。
5783	敦	1	1	8DE4	3A8D		孛厶攵	

附件二、《肘後備急方》手動取代表

編號	檔案原文摘錄	手動取代結果
4848	蝦魚厶且食此多致	蝦魚厶且輩食此多致
	先𠂔所用藥	先輩所用藥