

《外臺秘要》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
2008/6/16 黃榮順 製作

一、《外臺秘要》一書(外臺秘要.xml)使用舊版造字 116 個，字頻 266 次，詳如附件一。這 116 個造字中，72 個可轉成 Windows XP 可顯示的字，字頻 155 次；另外 44 個字必須轉成構字式，字頻 111 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 2091 的「广△雍」字，原字應為「癰」。

2、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 4878，舊漢籍造字為「宀」，比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「宀」者，在 xml 檔中部分已人工修改為「宀」，故配合修改，將編號 4878 的「宀」字，手動取代為「宀」。

- 3、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「𦉑」，部件外字為「𦉑」，一一比對原書，確認是「𦉑」者，再以手動取代。
- 4、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 2408 的「𦉑」係「𦉑」字之異體，故以「𦉑」字取代。又如編號 1641 的「胆」係「膽」字之簡體，故以「膽」字取代。
- 5、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。
- 6、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 4919 的「𦉑 𦉑鹿」。

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。所以如造字編號 888 的「𦉑」字，Unicode 編碼為 3B88，由於 Windows XP 並不支援，仍須改用構字式「木 𦉑奈」。

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：舊漢籍造字。
- 丙、取代：使用字形與造字欄不同時，於此欄列出。
- 丁、檔案原文摘錄：摘錄檔案(外臺秘要.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。
- 戊、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。

附件一、「外臺秘要.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
606	彡	1	1	FDE8			彡	
763	卩	1	1	FEE8			卩	
888	捺	1	1	8EC8	3B88		木厶奈	
899	馳	1	1	8ED3	99DE	馳		
1000		1	1	8F79			王厶巳	玨，異體字問題。
1002	鳩	1	1	8F7B	9D02	鳩		
1146	甕	13	12	906E	7506	甕		甕，校對問題，新舊版缺字並存。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1238	戾	1	1	90EC	77E6	戾		
1338		2	2	91B3			火△目	待造字。
1512		2	2	92C4			彳△具	待造字。
1530	紕	1	1	92D6	7CFA	紕		
1641	胆	2	2	93A8	80C6	胆		膽，異體字問題。
1867	菽	1	1	94ED			艹△禾△丸△	
1890	葱	1	1	9545	44D7		艹△念	蔥，異體字問題。
1893	袞	1	1	9548	886E	袞		
2021	趁	2	2	95EA	8D82	趁		趁，異體字問題。
2091	癰	1	1	9671	222CA		广△雍	癰，校對問題，舊版漢籍錯字，可用程式全部取代。
2147	蟞	4	4	96CB	87EE	蟞		
2159	𪗇	2	2	96D7	4D35		麥△并	
2295		1	1	97C2			艹△彙	稿，校對問題，舊版漢籍錯字，可用程式全部取代。
2304	鉶	4	4	97CB	28965		金△卩	
2316	滴	1	1	97D7	6EF3	滴		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2322	鏘	2	2	97DD	4948		金 𠂔 莽	
2408	鬧	1	1	9874	9599	鬧		鬧，異體字問題。
2461	葱	3	2	98CB	44E4		艹 𠂔 忽	蔥，校對問題，新舊版缺字並存。
2516	餈	1	1	9943	9919	餈		
2564	刺	2	2	9973	34E8		夾 𠂔 丿	
2572	銜	1	1	997B	8858	銜		銜，異體字問題。
2937		2	2	9BD0			𠂔 𠂔 立	豎，異體字問題。
2996	双	2	2	9C4C	53CC	双		雙，異體字問題。
3064	鯛	1	1	9CB2	9C02	鯛		
3148	棗	1	1	9D47	2346F		木 𠂔 兆	桃，異體字問題。
3153	筓	2	2	9D4C	7BFC	筓		
3191		1	1	9D72			𠂔 𠂔 𠂔 𠂔 𠂔 戊	箴，異體字問題。
3257	藕	1	1	9DD6	85CA	藕		
3776		6	6	8147			𠂔 𠂔 棧	待造字。
3813	邛	1	1	816C	536D	邛		
3822	叁	1	1	8175	53C1	叁		參，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3860	𪗇	1	1	81BD	5C20	𪗇		
3866	崑	1	1	81C3	5D10	崑		崑，異體字問題。
3874		2	2	81CB			忄 厶 稟	慄，異體字問題。
3903	箐	2	2	81E8	25BE4		竹 厶 韋	
3924	菴	5	5	81FD	671E	菴		
3940	薊	3	3	824E	6B1D	薊		
3942	毡	1	1	8250	6BE1	毡		
3973		1	1	826F			衤 厶 扁	褊，異體字問題。
3977	狗	1	1	8273	72E5	狗		
4017		1	1	82BD	880F	蟹		蟹，異體字問題。
4023	碓	1	1	82C3	78D3	碓		
4027	裨	1	1	82C7	7986	裨		裨，異體字問題。
4031	杭	1	1	82CB	79D4	杭		
4033	稽	1	1	82CD	7A2D	稽		
4036	窰	1	1	82D0	7AB0	窰		
4047		3	3	82DB			形 竹 口 乂 比 口	篋，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4060	穀	3	3	82E8	7CD3	穀		
4078		1	1	82FA			巳厶升	异，異體字問題。
4085	耻	1	1	8342	803B	耻		恥，異體字問題。
4093	舐	3	3	834A	445B		舌厶氏	舐，異體字問題。
4094	館	1	1	834B	8218	館		館，異體字問題。
4095	舩	2	2	834C	8229	舩		
4113	簡	1	1	835E			艹厶間	
4118	藜	1	1	8363	8602	藜		
4145	羈	1	1	837E	898A	羈		Unicode 字型差異。
4172		4	4	83BB	8E77	蹶		蹶，異體字問題。
4174	𨾏	1	1	83BD	8EB6	𨾏		
4186	遡	1	1	83C9	9061	遡		
4195	鉢	17	17	83D2	9262	鉢		
4198	鑛	1	1	83D5	945B	鑛		
4203	隣	1	1	83DA	96A3	隣		鄰，異體字問題。
4206	榆	1	1	83DD	3BD3		木厶翕	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4218	萑	1	1	83E9	97EE	萑		
4222	頰	3	3	83ED	9847	頰		
4247	鮎	1	1	8447	9B8E	鮎		
4251	鰕	3	3	844B	9C15	鰕		
4274		35	35	8462			扌 Δ 友	拔，異體字問題。
4299	炳	1	1	847B	242F2		火 Δ 苜	炳，異體字問題。
4353	菟	2	2	84D3	83AC	菟		菟，異體字問題。
4373	鼯	1	1	84E7			鼠 Δ 畠	
4423	醫	2	2	855A	81CB	醫		
4437	廩	1	1	8568	5EEA	廩		廩，異體字問題。
4493	特	1	1	85C2	3E40		牛 Δ 寺	特，異體字問題。
4502	俎	1	1	85CB	723C	俎		俎，異體字問題。
4513	眇	1	1	85D6	7717	眇		
4519	咲	1	1	85DC	54B2	咲		
4525	哏	3	3	85E2	54EF	哏		
4531	攢	1	1	85E8	6505	攢		攢，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4546	鷓	1	1	85F7	9D76	鷓		
4560	協	2	2	8646	604A	協		協，異體字問題。
4628	垲	1	1	86AC	2129F		土厶目	
4671	鷓	2	2	86D7	4D0F		燕厶鳥	
4716	埜	1	1	8745	57FF	埜		
4762	旃	1	1	8773	3ACB		舛厶冉	
4803	腭	2	2	87BE	816D	腭		
4812	羗	1	1	87C7	26351		羗厶、	羗，異體字問題。
4878	宄	3	2	884A	5B82	宄		宄，校對問題，新舊版缺字 並存。
4919		5	5	8873			亻厶鹿	待造字。
4950	嶮	1	1	88B4	5D18	嶮		
5008	幪	1	1	88EE	5E5E	幪		
5026		2	2	8941			形广回示	廩，異體字問題。
5074	尫	1	1	8971	5C2B	尫		尫，異體字問題。
5076	爇	2	2	8973	2435C		火厶參	
5102	曷	1	1	89AF	3B02		日厶禺	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5108	惛	16	16	89B5	60FD	惛		
5333	撻	1	1	8AF9	22BC6		扌 㒼 爻	
5351	肱	2	9	8B4C	80AC	肱		重碼問題，部分採用舊漢籍造字「肱」，部分採用部件外字「𠂇」。
5434		1	1	8BC1			形 灬 白 工 冫	涅，異體字問題。
5441		1	1	8BC8			艹 厶 檠	待造字。
5447	躡	1	1	8BCE	8E83	躡		
5478	飴	2	2	8BED	98F0	飴		
5541	蠹	2	2	8C6D	8827	蠹		蠹，異體字問題。
5544	硎	3	3	8C70	2543B		石 厶 冈	硎，異體字問題。
5554	撥	1	1	8C7A	6A43	撥		
5558	麓	1	1	8C7E	9E84	麓		
5677	𦉳	4	4	8D58	4674		𦉳	
5753	潜	2	2	8DC6	6F5C	潜		潛，異體字問題。
5768	潤	1	1	8DD5	6FF6	潤		潤，異體字問題。

附件二、「外臺秘要.xml」手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果
1146	蕘		__ (全部取代)	<u>蕘</u> (全部取代)
2461	𠄎 忽	蔥	忌生__生菜	忌生 <u>蔥</u> 生菜
			勿著__豉醋	勿著 <u>蔥</u> 豉醋
4878	宀	穴	是以不取灸 <u>宀</u> 者耳	是以不取灸 <u>穴</u> 者耳
			灸禁 <u>宀</u> 及老少加減法	灸禁 <u>穴</u> 及老少加減法
5351	肫		易虛則生 <u>𠄎</u> 小者	易虛則生 <u>肫</u> 小者
			<u>𠄎</u> ，疑當作疣	<u>肫</u> ，疑當作疣