

《千金翼方》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
2008/5/29 黃榮順 製作

一、《千金翼方》一書(kinshin.xml)使用舊版造字 112 個，字頻 274 次，詳如附件一。這 112 個造字中，80 個可轉成 Windows XP 可顯示的字，字頻 199 次；另外 32 個字必須轉成構字式，字頻 75 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、舊版漢籍錯字，必須人工取代：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，雖然頻次相同，其為原書所無之誤增字，仍須手動刪除。如編號 2465 的「金~~𠄎~~敷」，原書所無，為誤增字，手動刪除。
- 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「𦉳」，部件外字為「𦉳」，一一比對原書，確認是「𦉳」者，手動取代。

- 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 819 的「𠄎 𠄎 侯」係「候」字之異體，故以「候」字取代。又如編號 5034 的「弃」係「棄」字之簡體，故以標準字「棄」取代。
- 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 5466 的「倏」字，Unicode 字型呈現為「倏」，實際上仍為同一字。
- 5、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 4005 的「鬼 𠄎 天」。

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「齶」字，Unicode 編碼為 26848，由於 Windows XP 並不支援，仍須改用構字式「雍 𠄎 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，或頻次雖同，須以手動方式處理，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、檔案原文摘錄：摘錄檔案(千金翼方.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(手動刪除)者，表示原書無字，必須手動刪除。
- 丁、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(手動刪除)者，表示原書無字，必須手動刪除。

五、「千金翼方.xml」發現之缺字外錯字共 2 處，修正情形詳列附件三，附件三「千金翼方.xml」缺字外錯字修改表體例說明如下：

- 甲、檔案原文摘錄：摘錄檔案(千金翼方.xml)中包含該錯字的文句段落，其中紅字底線的部分，為需要手動取代的原文。
- 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、「千金翼方.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
681	𠂇	1	1	FE74			𠂇	
783	亼	1	1	FEFC			亼	
784	亻	1	1	FEFD			亻	
787	𩚑	1	1	8E41	26848		雍𠂇肉	
819	候	2	2	8E61	202EB		亻 𠂇 候	候，異體字問題。
847	𩚑	1	1	8E7D	7F3B	𩚑		
904	𩚑	1	1	8ED8	2437D		𩚑 𠂇 火	熱，異體字問題。
915	遲	1	1	8EE3	905F	遲		
972	犇	2	2	8F5D	7287	犇		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
975	甌	1	1	8F60	4C1B		鬲厶瓦	
981	豺	3	3	8F66	72B2	豺		
1054	邠	1	1	8FD1	2867B		丹厶卩	
1067	碁	5	5	8FDE	68CA	碁		碁，異體字問題。
1146	蕘	7	7	906E	7506	蕘		
1310	蟣	7	7	9175	45EB		虫厶麻	
1390	糶	1	1	91E7	7A64	糶		
1418	劬	1	1	9244	4FF2	劬		
1465	炒	1	1	9273	7C86	炒		
1530	紉	3	3	92D6	7CFA	紉		糾，異體字問題。
1890	葱	5	5	9545	44D7		艹厶念	蔥，異體字問題。
1951	味	1	1	95A4	20C0C		口厶末	
2021	趲	1	1	95EA	8D82	趲		
2147	蟪	1	1	96CB	87EE	蟪		
2374	冢	1	1	9852	51A1	冢		冢，異體字問題。
2415	鬪	1	1	987B	95D8	鬪		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2449		1	1	98BF			名△隹	雒，異體字問題。
2465	鍤	1	1	98CF	495E		金△敦	舊版漢籍錯字，手動刪除。
2496	鰈	1	1	98EE	4C5C		魚△昔	
2853	式	6	6	9B5A	5F0D	式		
2893	属	1	1	9BA4	5C5E	属		屬，異體字問題。
2977	却	6	6	9BF8			去△卩	卻，異體字問題。
2996	双	21	21	9C4C	53CC	双		雙，異體字問題。
3199	駸	1	1	9D7A	9ABD	駸		
3770	个	1	1	8141	4E2A	个		
3826	叶	1	1	8179	53F6	叶		
3836	喻	2	2	81A5	564F	喻		
3859	尔	1	1	81BC	5C14	尔		
3864	峯	1	1	81C1	5CEF	峯		峰，異體字問題。
3869	币	1	1	81C6	5E00	币		
3876	廻	12	12	81CD	5EFB	廻		迴，異體字問題。
3903	箒	1	1	81E8	25BE4		从△韋	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4005		1	1	82B1			鬼△天	待造字。
4012	鼓	2	2	82B8	76B7	鼓		鼓，異體字問題。
4013	盃	2	2	82B9	76CC	盃		
4027	裨	1	1	82C7	7986	裨		裨，異體字問題。
4036	窰	2	2	82D0	7AB0	窰		
4113	簡	1	1	835E			艹△問	
4148	魂	2	2	83A3	4C1F		云△鬼	魂，異體字問題。
4172	蹶	6	6	83BB	8E77	蹶		蹶，異體字問題。
4198	鑛	2	2	83D5	945B	鑛		
4203	隣	1	1	83DA	96A3	隣		鄰，異體字問題。
4218	菲	1	1	83E9	97EE	菲		
4223	顛	1	1	83EE	984B	顛		
4226	飡	1	1	83F1	98E1	飡		
4236	瞞	2	2	83FB	77D2	瞞		
4265	鼃	2	2	8459	9F03	鼃		
4283		2	2	846B	622F	戲		戲，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4309	刼	4	4	84A7	523C	刼		劫，異體字問題。
4315		1	1	84AD			形 𣎵 冫 自 寸 𠂇	形 林 缶 冫 𠂇 𠂇，異體字問題。
4400	兇	6	6	8543	5160	兇		
4455	劒	6	6	857A			僉 刀	劍，異體字問題。
4496	段	3	3	85C5	53DA	段		段，異體字問題。
4512	咤	1	1	85D5	549C	咤		
4531	攢	1	1	85E8	6505	攢		攢，異體字問題。
4582	榛	1	1	865C	3B8F		木 柰	
4643	徃	4	4	86BB	5F83	徃		往，異體字問題。
4668	噐	1	1	86D4	5650	噐		器，異體字問題。
4704	犝	1	1	86F8	7271	犝		
4716	埜	2	2	8745	57FF	埜		
4735	壠	5	5	8758	58E0	壠		
4746		2	2	8763	28932		形 牙 女 厘 𠂇	釐，異體字問題。
4764	摠	1	1	8775	64A7	摠		舊版漢籍錯字，手動刪除。
4854		1	1	87F1			彳 疑	凝，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4878	宀	2	2	884A	5B82	宀		宀，異體字問題。
4885	糝	3	3	8851	7CDA	糝		
4904	尅	2	2	8864	5C05	尅		尅，異體字問題。
4907	橐	1	1	8867	69D6	橐		橐，異體字問題。
4909	弥	1	1	8869	5F25	弥		彌，異體字問題。
4980		1	1	88D2			藝 _火	藝，異體字問題。
5027	牕	1	1	8942	724E	牕		
5034	弃	4	4	8949	5F03	弃		棄，異體字問題。
5071	忝	1	1	896E	5FE9	忝		
5108	惛	2	2	89B5	60FD	惛		
5119	博	3	3	89C0	613D	博		
5120	虛	1	1	89C1	8657	虛		
5147	恠	3	3	89DC	6060	恠		
5149	胥	7	7	89DE	2676B		要 _月	腰，異體字問題。
5302	昏	6	6	8ADA	662C	昏		昏，異體字問題。
5306		2	2	8ADE			𣎵 _間	簡，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5315	沿	1	1	8AE7	3CC2		氵厶公	沿，異體字問題。
5318	蔭	1	1	8AEA	4516		艹厶陵	
5351	肫	3	5	8B4C	80AC	肫		重碼問題，部分採用舊漢籍造字「肫」，部分採用部件外字「𠂇」。
5408	絳	2	2	8BA7	7D4D	絳		
5433	爰	1	1	8BC0	590B	爰		
5439	栢	2	2	8BC6	6822	栢		
5447	躡	1	1	8BCE	8E83	躡		
5450	滄	1	1	8BD1	6E4C	滄		
5466	倏	1	1	8BE1	5010	倏		Unicode 字型差異。
5533		13	13	8C65			穀厶米	穀，異體字問題。
5549	藁	1	1	8C75	85F3	藁		
5551	賈	1	1	8C77	8CF7	賈		
5554	撥	1	1	8C7A	6A43	撥		
5556	隸	1	1	8C7C	96B7	隸		隸，異體字問題。
5558	龕	3	3	8C7E	9E84	龕		
5652	蝨	1	1	8CFE	27312		或厶虫	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5666	汚	7	7	8D4D	6C5A	汚		汚，異體字問題。
5678	𡗗	1	1	8D59	757A	𡗗		
5720	埇	3	3	8DA5	578D	埇		
5723	𦍋	2	2	8DA8	4367		牛 𦍋 羊	
5742	澁	1	1	8DBB	6F81	澁		澁，異體字問題。
5752	摠	4	4	8DC5	63D4	摠		
5850	厶	2	2	C6C9		厶		

附件二、「千金翼方.xml」手動取代表

編號	造字	檔案原文摘錄	手動取代結果
2465	金 𦍋 𦍋	如 <u> </u> 釵股長一尺五寸	如釵股長一尺五寸 (手動刪除)
4764	摠	淨潔 <u> </u> 養	淨潔養 (手動刪除)
5351	𦍋	黑子去 <u> </u> 息肉疽	黑子去 <u> </u> 息肉疽
		一名蝕 <u> </u>	一名蝕 <u> </u>
		灰療黑子 <u> </u> 贅	灰療黑子 <u> </u> 贅

附件三、「千金翼方.xml」缺字外錯字修改表

檔案原文摘錄	手動取代結果
形 ⁺⁺ △(自△(△米))△皿□	形 ⁺⁺ 自□米皿□
養。洗刷飲飼	養 _之 。洗刷飲飼