

《繡像金瓶梅詞話》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
2008/6/30 陳維君 製作

一、《繡像金瓶梅詞話》一書(繡像金瓶梅詞話.xml)使用舊版造字 87 個,字頻 472 次,詳如附件一。這 87 個造字中,62 個可轉成 Windows XP 能顯示的字,字頻 432 次;另外 25 個字必須轉成構字式,字頻 40 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,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 2151,舊漢籍造字為「𠂔」,比對 txt 檔及 xml 檔後,發現原先在 txt 檔使用舊漢籍造字「𠂔」的構字式「𠂔久缶」,在 xml 檔中部分已人工修改為「隆」,故配合修改,將使用編號 2151 造字「𠂔」的構字式「𠂔久缶」,手動取代為「隆」。

2、重碼問題,採用部件外字,不轉碼: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」。然而漢籍在舊轉新的過程中,若已將原舊漢籍造字的部分全數改為

Unicode 或構字式，則在 xml 出現的造字頻次皆採用部件外字，此時便不需轉碼。如編號 5797，字碼為 8DF2，對應的舊漢籍造字為「灩」，部件外字為「𠂔」，比對原書發現書中皆取部件外字「𠂔」，故不需轉碼。

- 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 4171 的「𠂔疋厂封◻」係「躡」字之異體，故以「躡」字取代。又編號 2206 的「袂」字，係「襖」字之簡體字，亦以標準字的「襖」字取代。
- 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 5440 的「昧」字，Unicode 字型呈現為「昧」，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 1268 的「𠂔𠂔鑾」。

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 14 的「胤」字，Unicode 編碼為 38E7，由於 Windows XP 並不支援，仍須使用構字式「𠂔彳么月廾◻」。

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、造字：舊版造字字形。
- 丙、檔案原文摘錄：摘錄檔案(繡像金瓶梅詞話.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。

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。

戊、備註：其他需註明事項。

附件一、《繡像金瓶梅詞話》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
14	胤	1	1	FA4D	38E7		𠂇彡幺月廾	
426	亲	1	1	FCD1			亲	
761	久	1	1	FEE6			久	
781	丿	1	1	FEFA			丿	
814	墓	2	2	8E5C	87C7	墓		𡗗，異體字問題。
852	貉	1	1	8EA4	72E2	貉		
893	峩	1	1	8ECD	5CE9	峩		峨，異體字問題。
900	嵯	1	1	8ED4	5D73	嵯		嗟，異體字問題。
912	鷓	1	1	8EE0	9DC0	鷓		
1014	玠	1	1	8FA9	73CE	玠		
1268		1	1	914B			彡 厶 鑾	待造字。
1310	蠲	3	3	9175	45EB		虫 厶 麻	
1467	纂	5	5	9275	25C67		竹 厶 秦	
1616	逕	4	4	936D	284E6		辶 厶 里	
1646	脰	2	2	93AD	8117	脰		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
1905	視	6	6	9554	88E9	視		
1994	達	1	1	95CF	6FBE	達		
2021	趁	1	1	95EA	8D82	趁		
2042	疏	1	1	9640	47FD		足厶尢	
2072		1	1	965E			艹厶糜	靡，異體字問題。
2151	𠃉	2	1	96CF	961D	𠃉		校對問題，新舊版缺字並存，人工取代，參見附件二。
2206	袂	5	5	9747	8884	袂		襖，異體字問題。
2452	蠍	1	1	98C2	45F6		虫厶葛	
2459	騷	1	1	98C9	4BAB		馬厶婁	
2505	翻	1	1	98F7	98DC	翻		
2770	絳	1	1	9AC6	7D98	絳		
2905	挀	1	1	9BB0	6455	挀		
2937		1	1	9BD0			𠃉厶立	待造字。
3029	鋤	5	5	9C6D	941D	鋤		
3096	歲	1	1	9CD2	5D57	歲		歲，異體字問題。
3113	𡗗	1	1	9CE3	6245	𡗗		
3160	墳	33	33	9D53	575F	墳		墳，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3164		1	1	9D57			土𠂔冊	土𠂔冊，異體字問題。
3177	𧈧	3	3	9D64	7F4E	𧈧		
3250	扌	1	1	9DCF	6304	扌		
3770	个	34	34	8141	4E2A	个		個，異體字問題。
3807	驗	1	1	8166	9A10	驗		
3815	𧈧	2	2	816E	537A	𧈧		𧈧，異體字問題。
3824	叙	9	9	8177	53D9	叙		敘，異體字問題。
3829	咏	3	3	817C	548F	咏		
3834	嗶	2	2	81A3	5637	嗶		
3836	喻	1	1	81A5	564F	喻		
3850	姪	2	2	81B3	59D9	姪		
3852	嫻	1	1	81B5	5A63	嫻		
3877	迺	1	1	81CE	5EFC	迺		迺，異體字問題。
3927	柏	1	1	8241	6855	柏		
3976	犁	1	1	8272	7282	犁		
4001	瘧	1	1	82AD	7602	瘧		
4028	稟	1	1	82C8	7980	稟		稟，異體字問題。
4036	窰	4	3	82D0	7AB0	窰		校對問題，新舊版缺字並存，人工取代，

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
								參見附件二。
4091	臯	3	3	8348	81EF	臯		
4129	虬	1	1	836E	866C	虬		
4171		1	1	83BA			𠃉足厂封	躡，異體字問題。
4232	隲	6	6	83F7	96B2	隲		隲，異體字問題。
4247	鮎	1	1	8447	9B8E	鮎		
4252	鰓	4	4	844C			魚𩺰眾	
4262	麩	4	4	8456	9EAA	麩		
4263	麩	1	1	8457	9EAF	麩		
4460	勅	4	4	85A1	52D1	勅		
4463	擢	1	1	85A4	3A72		扌𩺰擢	
4520	邈	1	1	85DD	285FF		辶𩺰狼	
4560	協	2	2	8646	604A	協		
4572	嘍	1	1	8652	20DFA		口𩺰勃	
4715	嘍	1	1	8744	5688	嘍		
4759	襴	2	2	8770	8947	襴		
4784	鞞	1	1	87AB	9793	鞞		
4903	吋	8	8	8863	5C00	吋		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4966	癩	1	1	88C4	764E	癩		
4981	鳧	2	2	88D3	9CEC	鳧		
5032	焜	1	1	8947	3B0E		焜	
5075	彡	1	1	8972			彡	
5094	翬	2	2	89A7	26407		翬	
5123	悞	1	1	89C4	6130	悞		
5147	恠	1	1	89DC	6060	恠		
5173	漱	2	2	89F6	6F44	漱		漱，異體字問題。
5245		1	1	8AA1			掉𠃉戈	擻，異體字問題。
5299	戟	4	3	8AD7	39B8		卓𠃉戈	校對問題，新舊版缺字並存，人工取代，參見附件二。
5439	栢	7	7	8BC6	6822	栢		
5440	昧	1	1	8BC7	771C	昧		Unicode 呈現差異。
5447	躡	1	1	8BCE	8E83	躡		躡，異體字問題。
5543	鯨	1	1	8C6F	9B9D	鯨		
5555	迥	1	1	8C7B	9008	迥		迥，異體字問題。
5558	龕	1	1	8C7E	9E84	龕		
5694		1	1	8D69			𠃉自无木	概，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5704	𧯛	2	2	8D73	6C0A	𧯛		
5753	潛	3	3	8DC6	6F5C	潛		潛，異體字問題。
5797	灩	1	241	8DF2	7054	灩		與𪗇重碼，重碼部件，不轉碼。

附件二、《繡像金瓶梅詞話》手動取代表

編號	造字	檔案原文摘錄	手動取代結果	備註
2151	𧯛	以𧯛 _𧯛 久缶 _𧯛 聖 治事	以隆聖治事	「𧯛 _𧯛 久缶 _𧯛 」為「隆」之異體字，以「隆」取代「𧯛 _𧯛 久缶 _𧯛 」。
4036	窰	多居住 _約 中	多居住 _窰 中	
		說是 _約 上人	說是 _窰 上人	
		官 _約 雙箍鄧漿盆	官 _窰 雙箍鄧漿盆	
5299	卓 _𪗇 戈	_𪗇 將響樂	卓 _𪗇 戈將響樂	
		或手執戈 _𪗇	或手執戈卓 _𪗇 戈	
		漏定 _𪗇 杖齊	漏定卓 _𪗇 戈杖齊	