

《紅樓夢校注》(標記語料)舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中央研究院語言所文獻語料小組
2012/08/30 丁致伶

轉碼主要工作是把檔案中的舊版造字轉換成 Windows XP 能支援的 Unicode 字形，Unicode 目前共收錄漢字 70194 個字，而 XP 只能支援 20902 個字（詳如表一），不支援之字將以構字式表達。例：造字編號 1429 的「登」字，Unicode 編碼是 4532，由於 XP 並不支援，仍需使用構字式「登」。

表一、Unicode 的字數及編碼區段

Unicode	新增字數	新增編碼區段	總字數	WinXP
1.1 版	20902	4E00-9FFF	20902	支援
3.0 版	6582	3400-4DFF	27484	不支援
3.1 版	42710	20000-2A6D6	70194	不支援

一、舊版造字轉碼分析：

《紅樓夢校注》(標記語料)使用舊版造字 46 個，字頻 154 次，這 46 個造字中，42 個可轉成 Windows XP 能顯示的字，字頻 140 次；另外 4 個字必須轉成構字式，字頻 14 次。

轉碼完成製作轉碼分析表，請參考附件一《紅樓夢校注》(標記語料)轉碼分析表，欄位說明如下：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 造字：舊版造字。
- (三) 頻次：舊版造字在文件的出現次數。
- (四) Big5：造字的 Big5 碼。
- (五) Unicode：造字所對應的 Unicode 碼。

(六) WinXP：造字在 Windows XP 的對應字形。

(七) 構字式：Windows XP 無法對應字形改用構字式。

(八) 備註凡例：備註欄中記錄轉碼後字形及修改原因，凡例如下：

1. 異體字問題：為了使用者查詢和使用的方便，在處理異體字時最主要的方式是以標準字取代，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 5591 的「杞」字，是「杞」的異體字，以標準字「杞」取代。
2. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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 4263 的「麩」字，Unicode 字型呈現為「麩」，實際上仍為同一字。

附件一、《紅樓夢校注》(標記語料)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
20	藪	4	FA53			ㄉㄞ ㄉㄞ 救	
946	爇	3	8F43	7216	爇		
1213	睽	1	90D3	7743	睽		
1429	登	1	924F	4532		ㄉㄥ ㄉㄥ 登	
1854	啣	2	94E0	20E95		口 ㄉㄞ 留	
2132	癩	1	96BC	3FC0		疒 ㄉㄞ 班	癩，異體字問題。
2325	睜	1	97E0	77A4	睜		
2448	隼	2	98BE	96BD	隼		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2680	狍	2	9A4A	72CD	狍		
2684	簍	6	9A4E	7C70	簍		
2999	嗵	4	9C4F	5523	嗵		
3019	腩	7	9C63	817C	腩		
3029	鏹	1	9C6D	941D	鏹		
3058	吓	1	9CAC	5413	吓		
3787	豕	1	8152	51A2	豕		
3810	匱	1	8169	5335	匱		
3826	叶	1	8179	53F6	叶		
3830	咤	5	817D	54A4	咤		
3851	姤	1	81B4	59F9	姤		
3912	拈	1	81F1	62C3	拈		
3934	榻	18	8248	69C5	榻		
3942	毡	38	8250	6BE1	毡		
3968	煨	1	826A	7145	煨		
3977	狗	1	8273	72E5	狗		
4013	盃	1	82B9	76CC	盃		
4015	眦	1	82BB	7726	眦		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4043	筭	1	82D7	7B53	筭		
4044	筍	1	82D8	7B6F	筍		
4047		2	82DB			𠄎 ^𠄎 口 乂比 ^口	筮，異體字問題。
4074	縑	14	82F6	7E27	縑		
4098	芪	2	834F	82AA	芪		
4111	蕤	1	835C	8493	蕤		
4139	袴	3	8378	88B4	袴		
4167	賁	1	83B6	8CEB	賁		
4263	麩	1	8457	9EAF	麩		Unicode 字型呈現差異。
4428	躡	2	855F	8E70	躡		
4616	壳	2	867E	58F3	壳		
4715	嘸	2	8744	5688	嘸		
4742		1	875F			衣 [△] 執	褻，異體字問題。
4791	羝	1	87B2	7F93	羝		
4800	瞪	2	87BB	5654	瞪		
5276	珣	7	8AC0	3EDE		王 [△] 扁	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5543	養	2	8C6F	9B9D	養		
5583	殮	1	8CB9	98F1	殮		
5584	悞	1	8CBA	611E	悞		
5591	杞	1	8CC1	233CC		木△巳	杞，異體字問題。

二、手動取代表說明：

《紅樓夢校注》(標記語料)檔案中出現「●」字形，共計 105 字，比對原書後以 Unicode 字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《紅樓夢校注》(標記語料)檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《紅樓夢校注》(標記語料)「●」字手動取代表

頁碼	標記行碼	檔案原文摘錄	原書字形	備註
p. 47	1943	大紅洋緞窄●襖	褙	
p. 49	2142	玻璃●(Na)	台△皿	
p. 52	2341	憊●	賴△心	
p. 68	2886	胭脂(Na) ●	疒△計	

p. 119	4726	●(VC) 舌(Na) 咂	磻	
p. 123	4826	挽(VC) 著(Di) ●兒	髟△贊	有 2 處
p. 134	5649	混(Dh) ●	心	
p. 140	5791	●兒(Na)	髟△贊	有 2 處
p. 156	6663	覷●(VH)	覷	
p. 159	6883	管(VE) 你(Nh) ●●	毛△几 毛△巴	
p. 159	6888	●喝(VE)	口△么	有 17 處
p. 192	8592	●拉拉(VH)	喙	
p. 201	8961	上來(VCL) ●扶	摺	
p. 205	9249	婆娘(Na) ●	唢	
p. 225	10083	紅(VH) ●帶	鞞	
p. 232	10619	●(VC) 在(P) 自己	扌△塞	
p. 239	10829	●鴿香串(Na)	鵲	
p. 255	11735	奇(VH) 花(Na) ●灼	烟	
p. 261	12205	藿(Na) ●(Na)	蒟	
p. 264	12474	卍(Na) ●(Na)	#圖	有 2 處
p. 299	13618	茗煙(Nb) ●●(Dh)	出△欠 出△欠	
p. 308	14380	●(VC)[+vrv] 上	扌△層	
p. 317	14823	嬾嬾(Na) ●(VC)	聞	
p. 318	14898	忽(Dd) 聽(VE) ●	唢	

p. 322	15264	青(VH) ●(Na) 披風	欣	
p. 330	15702	雞聲鵝●	聞	
p. 332	15845	毒(Na) 盡(VH) ●	癩	
p. 340	16151	我(Nh) ●●(VH) 的	口△邦 口△邦	
p. 384	18376	個(Nf) ●(Na)	髟△贊	
p. 408	19707	文章(Na) ●灼	炯	
p. 443	21895	一(Neu) 根(Nf) ●●	毛△几 毛△巴	
p. 478	23850	鴻●	涑△鳥	
p. 583	30204	窗檻(Na) 上(Ng) ●	爪△甲	
p. 613	31326	卻(Dl) ●(Dh)	赳	
p. 617	31618	那裏(Nc) 鐵●	鉞	
p. 617	31619	那裏(Dj) ●	强△牛	改為犖
p. 632	32426	把(P) 皮(Na) ●	𠃉竹一韭 𠃉	
p. 636	32697	●(Na) 爬(Na) 罕	分△瓜	
p. 636	32704	點犀●	喬△皿	
p. 636	32705	斟(VC) 了(Di) 一 (Neu) ●	喬△皿	
p. 636	32721	一(Neu) 個(Nf) 大 (VH) ●	台△皿	

p. 667	34508	街(Nc) 就(D1) ●	走△真	
p. 704	36610	稟性(Na) 愚●	亻△強	
p. 710	37181	看(VE) 你(Nh) ●	走△僉	
p. 725	38259.	(PARENTHESISCATEGORY) ●(Dh)	啞	
p. 747	39362	●●(VH) 笑(VA)	出△欠 出△欠	有 2 處
p. 752	39764	窄●小袖	梢	
p. 752	39765	狐●(Na)	歎	
p. 755	39960	鐵絲●(Na)	纟△蒙	
p. 768	40319	笑●●	出△欠 出△欠	
p. 768	40319	●(VC) 了(Di) 一 (Neu) 枝	虔△力	
p. 791	41201	溫(VHC) 水(Na) ●	形: 竹 吳 馬□	
p. 791	41201	一(Neu) ●(VC)	形: 竹 吳 馬□	
p. 815	42617	實實(Dbb) ●	撐	
p. 864	45497	他(Nh) 一(Dd) ●	強△牛	改為犂
p. 906	48123	有(V_2) ●(VC) 樹	烏△卩	
p. 929	49570	眼(Na) ●摔(VC)	躑	

p. 932	49799	大●小喝	口△么	
p. 968	52273	嘴(Na) 裏(Ng) 汗●	火△敝	
. p979	52689	酞●(Na)	糸△式	
p. 983	52959	茶●花	艹△糜	
p. 988	53359	●(Na) 來	髟△贊	
p. 1089	59232	諸(Nes) 務(Na) ●集	犴	
p. 1157	63241	釵(Na) ●(VH) 鬢	髟	
p. 1190	65257	●(VC) 了(Di) 腿	足△歪	
p. 1265	68890	●嚟(VC)	口△床	
p. 1286	69993	蓼●(Nc) 一帶	涂△支	改為澈
p. 1306	71344	●呼(VA) 聲兒	口△么	
p. 1338	73388	●子	鈔	有 4 處
p. 1370	75158	●喇喇(VH)	唸	
p. 1377	75729	骨●●(Dh)	口△彖 口△彖	
p. 1387	76061	別(Dc) 逼●	木△靠	
p. 1512	84008	鬢(Na) ●(VH)	髟	
p. 1549	86157	只(Daa) 聽(VE) ●	唸	