


# 《紅樓夢校注》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  
2010/01/27 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》使用舊版造字 138 個，字頻 10041 次，這 138 個造字中，114 個可轉成 Windows XP 能顯示的字，字頻 9880 次；另外 24 個字必須轉成構字式，字頻 161 次。

轉碼完成製作轉碼分析表，請參考附件一《紅樓夢校注》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 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
  2. 錯字，以程式全部取代：與原書字形不符，並查詢教育部異體字字典確認非異體字後，皆歸類為錯字，以程式取代為正確字形。例：檔案中使用造字編號 3238 的「𠄎酉广艹系〇」字，而原書使用之字為「釀」，查詢教育部異體字字典確認「釀」並非「𠄎酉广艹系〇」之異體字，因此歸為錯字，以程式取代為正確字形「釀」。
  3. 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 4145 的「羈」字，Unicode 字型呈現為「羈」，實際上仍為同一字。
  4. 依原書修改未定義字形：編號 309 舊版造字「❓」代表未定義字型，依原書修改為書中字形。例：「❓」在書中的字形為「癡」，依原書修改為「癡」。

附件一、《紅樓夢校注》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
20	𠄎	4	FA53			艹 𠄎 𠄎	
296	𠄎	15	FBEC			𠄎	
297	𠄎	74	FBED			𠄎	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
298	△	11	FBEE			△	
299	形	8	FBEF			形	
300	◻	8	FBF0			◻	
309	?	1	FBF9			?	癩，依原書修改未定義字形。
565	虽	2	FDBF			虽	
658	疒	1	FE5D			疒	
717	足	2	FEBA			足	
739	勺	1	FED0			勺	
750	冫	2	FEDB			冫	
752	𠂇	3	FEDD			𠂇	
770	扌	4	FEEF			扌	
782	么	1	FEFB			么	
881	颯	1	8EC1	98C7	颯		
946	爇	3	8F43	7216	爇		
1100	身	2	9040	8EAB	身		
1213	睽	1	90D3	7743	睽		
1429	登	1	924F	4532		登	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1854	啣	2	94E0	20E95		口 㗎 留	
1859	匱	1	94E5	5336	匱		
2046	踪	26	9644	8E2A	踪		
2132	癩	1	96BC	3FC0		疒 ㄩ 班	癩，異體字問題。
2325	瞞	1	97E0	77A4	瞞		
2415	鬪	2	987B	95D8	鬪		鬪，異體字問題。
2448	隼	2	98BE	96BD	隼		
2592	臟	5	99B1	8D1C	臟		
2650	鸛	1	99EB	9E18	鸛		
2680	狍	2	9A4A	72CD	狍		
2684	簍	6	9A4E	7C70	簍		
2999	嗶	4	9C4F	5523	嗶		
3010	攷	1	9C5A	5746	攷		
3019	腩	7	9C63	817C	腩		
3020	衤	2	9C64			衤	
3029	鋤	1	9C6D	941D	鋤		
3058	吓	1	9CAC	5413	吓		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3160	坟	22	9D53	575F	坟		
3177	蟻	7	9D64	7F4E	蟻		
3238		1	9DC3			形西广 艹系	釀，錯字，以程式全部取代。
3770	个	2	8141	4E2A	个		
3787	冢	1	8152	51A2	冢		
3789	冲	36	8154	51B2	冲		沖，異體字問題。
3792	凉	28	8157	51C9	凉		涼，異體字問題。
3794	湊	71	8159	51D1	湊		湊，異體字問題。
3810	匱	2	8169	5335	匱		
3814	却	438	816D	5374	却		
3817	廁	4	8170	53A0	廁		廁，異體字問題。
3819	廝	236	8172	53AE	廝		廝，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3820	厦	20	8173	53A6	厦		
3825	叠	41	8178	53E0	叠		
3826	叶	1	8179	53F6	叶		
3829	咏	32	817C	548F	咏		
3830	咤	7	817D	54A4	咤		
3849	獎	5	81B2	596C	獎		獎，異體字問題。
3851	姹	1	81B4	59F9	姹		
3858	寃	56	81BB	5BC3	寃		寃，異體字問題。
3864	峯	9	81C1	5CEF	峯		
3876	廻	13	81CD	5EFB	廻		
3886	悞	1	81D7	609E	悞		
3912	拃	1	81F1	62C3	拃		
3934	榻	18	8248	69C5	榻		
3942	毡	38	8250	6BE1	毡		
3966	烟	251	8268	70DF	烟		
3968	煨	1	826A	7145	煨		
3976	犁	1	8272	7282	犁		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3977	狗	1	8273	72E5	狗		
3978	猪	18	8274	732A	猪		Unicode 字型 呈現差異。
3979	猫	12	8275	732B	猫		
4013	盃	1	82B9	76CC	盃		
4015	眦	1	82BB	7726	眦		
4016	着	6099	82BC	7740	着		
4028	稟	4	82C8	7980	稟		
4036	窰	13	82D0	7AB0	窰		
4041	豎	102	82D5	7AEA	豎		
4042	笋	12	82D6	7B0B	笋		
4043	筭	1	82D7	7B53	筭		
4044	筍	1	82D8	7B6F	筍		
4047		2	82DB			𠄎 <sup>竹</sup> 口 乂比 <sup>口</sup>	篋，異體字問 題。
4067	綉	34	82EF	7D89	綉		
4074	縹	15	82F6	7E27	縹		
4080	羣	51	82FC	7FA3	羣		
4085	耻	1	8342	803B	耻		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4090	脚	140	8347	811A	脚		
4098	芪	2	834F	82AA	芪		
4107	葱	13	8358	8471	葱		
4111	蕓	2	835C	8493	蕓		
4129	虬	2	836E	866C	虬		
4130	衅	1	836F	8845	衅		
4132		1130	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	3	8373	885E	衛		衛，異體字問題。
4139	袴	4	8378	88B4	袴		
4144	霸	2	837D	8987	霸		Unicode 字型呈現差異。
4145	羈	1	837E	898A	羈		Unicode 字型呈現差異。
4167	賚	1	83B6	8CEB	賚		
4173	𧣾	41	83BC	8EAD	𧣾		
4184	迹	37	83C7	8FF9	迹		
4193	鈎	13	83D0	920E	鈎		



編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4195	鉢	5	83D2	9262	鉢		
4256	鷄	82	8450	9DC4	鷄		
4262	麪	22	8456	9EAA	麪		Unicode 字型 呈現差異。
4263	麩	1	8457	9EAF	麩		Unicode 字型 呈現差異。
4278	亘	3	8466	4E98	亘		
4287	織	1	846F	7E8E	織		
4307	罵	140	84A5	99E1	罵		罵，異體字問 題。
4309	刼	15	84A7	523C	刼		
4376	傻	65	84EA	510D	傻		
4428	躡	3	855F	8E70	躡		
4432		16	8563			𠄎 纒 厂 墨 𠄎	纏，異體字問 題。
4447	捏	12	8572	63D1	捏		
4616	壳	2	867E	58F3	壳		壳，異體字問 題。
4715	嘍	2	8744	5688	嘍		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4742		2	875F			衣△執	褻，異體字問題。
4750	卅	4	8767	5344	卅		卅，錯字，以程式全部取代。
4784	鞞	1	87AB	9793	鞞		
4791	羝	1	87B2	7F93	羝		
4796		1	87B7			形疒四 彳	療，異體字問題。
4800	噎	2	87BB	5654	噎		
4816	彳	1	87CB			彳	
4904	尅	1	8864	5C05	尅		
4940	韵	68	88AA	97F5	韵		
5013	彳	3	88F3			彳	
5075	彳	2	8972			彳	
5176		32	89F9			彳△冗	沉，異體字問題。
5218		31	8A64			形宀寬 、	寬，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5276	璠	7	8AC0	3EDE		王厶扁	
5289	慤	1	8ACD	6164	慤		
5290	携	41	8ACE	643A	携		
5300	鬪	41	8AD8	9B2A	鬪		
5543	養	2	8C6F	9B9D	養		
5583	殮	1	8CB9	98F1	殮		
5584	慥	1	8CBA	611E	慥		
5590	朶	9	8CC0	6736	朶		
5591	杞	1	8CC1	233CC		木厶巳	杞，異體字問題。
5621	泪	28	8CDF	6CEA	泪		
5666	污	12	8D4D	6C5A	污		污，異體字問題。
5706	敘	57	8D75	654D	敘		敘，異體字問題。
5768	潤	8	8DD5	6FF6	潤		

## 二、手動取代表說明：

《紅樓夢校注》檔案中需手動修改的情況分為以下幾種：

(一)《紅樓夢校注》檔案中出現「●」字形，共計 1 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》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》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 264	卍●各種花樣	#圖	

(二)《紅樓夢校注》檔案中，對於舊版造字以外發現的錯字，以手動修改，並製作錯字修改表，請參考附件三。欄位說明如下：

1. 檔案原文摘錄：摘錄檔案《紅樓夢校注》中的文句段落，其中紅字底線的部分，為需要手動取代的原文。
2. 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。
3. 備註：記錄修改相同詞句的次數或其他事項。

### 附件三、《紅樓夢校注》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
<u>虫</u> □的過他	<u>牛</u> □的過他	
或卍 <u>形一口回</u> □	或卍 <u>#圖</u>	