

# 《清平山堂話本》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  
2008/8/29 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 3894 的「筋」字，Unicode 編碼是 4225，由於 XP 並不支援，仍需使用構字式「𠂇𠂇𠂇𠂇」。

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：

《清平山堂話本》使用舊版造字 127 個，字頻 1806 次，這 127 個造字中，103 個可轉成 Windows XP 能顯示的字，字頻 1253 次；另外 24 個字必須轉成構字式，字頻 553 次。

轉碼完成製作轉碼分析表，請參考附件一《清平山堂話本》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 3789 的「冲」字，是「沖」的異體字，以標準字「沖」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 4409 的「回」字，而原書使用之字為「向」，查詢教育部異體字字典確認「回」並非「向」之異體字，因此歸為錯字，以程式取代為正確字形「向」。
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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。
4. 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 1064 的「王𠄎風」。

附件一、《清平山堂話本》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
79	𠄎	2	FAB0	276C2		𠄎 𠄎固	𠄎，錯字，以程式全部取代。
296	𠄎	7	FBEC			𠄎	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
297	厶	49	FBED			厶	
298	厶	12	FBEE			厶	
299	形	88	FBEF			形	
300	口	88	FBF0			口	
619	L	1	FDF5			L	
639	彳	12	FE4A			彳	
675	食	18	FE6E			食	
685	衤	1	FE78			衤	
703	口	86	FEAC			口	
711	衤	1	FEB4			衤	
717	足	1	FEBA			足	
742	夕	87	FED3			夕	
762	人	87	FEE7			人	
768	艹	2	FEED			艹	
770	扌	3	FEEF			扌	
772	辶	1	FEF1			辶	
773	厶	1	FEF2			厶	
780	彳	1	FEF9			彳	
858	采	1	8EAA	7F59	采		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1064		1	8FDB			王𠂔風	待造字。
1146	蕘	1	906E	7506	蕘		
1226	罰	2	90E0	7F78	罰		
1780	柠	1	9474	67E0	柠		
1781	蓋	4	9475	854B	蓋		
2046	踪	10	9644	8E2A	踪		
2072		2	965E			𠂔𠂔𠂔𠂔	𠂔𠂔𠂔𠂔 糜，錯字，以程式全部取代。
2157	酌	2	96D5	9167	酌		
2505	翻	1	98F7	98DC	翻		
2524	饜	1	994B	994D	饜		
3177	𧯛	1	9D64	7F4E	𧯛		
3770	个	2	8141	4E2A	个		
3789	冲	4	8154	51B2	冲		冲，異體字問題。
3791	况	27	8156	51B5	况		况，異體字問題。
3793	減	1	8158	51CF	減		減，異體字問題。
3794	湊	1	8159	51D1	湊		湊，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3796	決	12	815B	51B3	決		決，異體字問題。
3804	効	1	8163	52B9	効		
3814	却	228	816D	5374	却		
3819	廡	3	8172	53AE	廡		廡，異體字問題。
3824	叙	13	8177	53D9	叙		
3825	疊	2	8178	53E0	疊		
3830	咤	1	817D	54A4	咤		
3843	堦	14	81AC	5826	堦		
3846	壻	1	81AF	58FB	壻		
3849	獎	2	81B2	596C	獎		獎，異體字問題。
3858	寃	13	81BB	5BC3	寃		寃，異體字問題。
3864	峯	13	81C1	5CEF	峯		
3868	卮	1	81C5	5DF5	卮		
3886	悞	9	81D7	609E	悞		
3894	筋	1	81DF	4225		竹 叕 勛	
3909	擡	30	81EE	64E1	擡		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3938	橈	2	824C	6AC8	橈		
3966	烟	54	8268	70DF	烟		
3974	牀	1	8270	7240	牀		
3975	牕	1	8271	7255	牕		
3978	猪	9	8274	732A	猪		Unicode 字型 呈現差異。
3979	猫	3	8275	732B	猫		
3984	瑠	3	827A	7460	瑠		
3991	攜	1	82A3	64D5	攜		
3996	踈	6	82A8	758E	踈		
4008	癯	1	82B4	766F	癯		
4016	着	268	82BC	7740	着		
4028	稟	6	82C8	7980	稟		
4038	竈	7	82D2	7AC8	竈		
4039	竝	1	82D3	7ADA	竝		
4042	笋	5	82D6	7B0B	笋		
4043	筴	2	82D7	7B53	筴		筴，異體字問 題。
4044	筴	2	82D8	7B6F	筴		
4045	筴	4	82D9	7B92	筴		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4067	綉	23	82EF	7D89	綉		
4069	綑	1	82F1	7DB3	綑		
4074	縑	3	82F6	7E27	縑		
4080	羣	7	82FC	7FA3	羣		
4085	耻	10	8342	803B	耻		
4089	脇	1	8346	8107	脇		
4090	脚	38	8347	811A	脚		
4103	苣	1	8354	831D	苣		
4105	菓	16	8356	83D3	菓		
4107	葱	3	8358	8471	葱		
4109	葛	1	835A	84AD	葛		
4112	蔴	2	835D	8534	蔴		
4129	虬	2	836E	866C	虬		
4132		97	8371			血𠂇𠂇	眾，異體字問題。
4140	衽	1	8379	88B5	衽		
4146	覩	1	83A1	89A9	覩		睹，異體字問題。
4167	賈	5	83B6	8CEB	賈		
4184	迹	7	83C7	8FF9	迹		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4193	鈎	9	83D0	920E	鈎		鈎，異體字問題。
4226	凵	11	83F1	98E1	凵		
4232	隲	1	83F7	96B2	隲		
4256	鷄	31	8450	9DC4	鷄		雞，異體字問題。
4278	亘	1	8466	4E98	亘		
4307	罵	36	84A5	99E1	罵		罵，異體字問題。
4309	刼	1	84A7	523C	刼		
4340	撮	2	84C6	3A67		扌 ㄩ 暴	
4378	隲	1	84EC	28F65		陟 ㄩ 馬	隲，異體字問題。
4409	冏	1	854C	518B	冏		冏，錯字，以程式全部取代。
4428	蹶	1	855F	8E70	蹶		
4442	潛	6	856D	6F98	潛		
4447	捏	1	8572	63D1	捏		
4460	勅	7	85A1	52D1	勅		
4507	吳	36	85D0	5434	吳		吳，異體字問題。



編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4515	呪	2	85D8	546A	呪		
4519	咲	1	85DC	54B2	咲		
4657	竣	1	86C9	57C8	竣		
4762	旃	1	8773	3ACB		从△冉	
4768	妬	2	8779	59AC	妬		妒，異體字問題。
4805	瀆	1	87C0	51DF	瀆		瀆，異體字問題。
4817	寶	38	87CC	5BF3	寶		寶，異體字問題。
4903	吋	1	8863	5C00	吋		
4904	尅	2	8864	5C05	尅		
5119	博	1	89C0	613D	博		
5123	愧	1	89C4	6130	愧		
5173	漱	1	89F6	6F44	漱		
5176		30	89F9			氵△冗	沉，異體字問題。
5218		14	8A64			宀△寬 、□	寬，異體字問題。
5281	餓	1	8AC5	994A	餓		
5338	暎	1	8AFE	668E	暎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5478	飴	2	8BED	4D77	飴		
5558	龕	1	8C7E	9E84	龕		
5587	歛	1	8CBD	6B58	歛		
5591	杞	1	8CC1	233CC		木𠂇巳	杞，異體字問題。
5697	煑	1	8D6C	7151	煑		Unicode 字型呈現差異。
5702	厨	4	8D71	53A8	厨		
5797	灑	4	8DF2	7054	灑		

## 二、手動取代表說明：

《清平山堂話本》檔案中出現「●」字形，共計 3 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《清平山堂話本》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、《清平山堂話本》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 15	懊惱碎●懷	情	

頁碼	檔案原文摘錄	原書字形	備註
p. 161	話本說●	徹	
p. 254	推●在中心裡	開	