

# 《平妖傳》(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  
2012/08/16 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 74 的「髻」字，Unicode 編碼是 29B06，由於 XP 並不支援，仍需使用構字式「髻𠄎𠄎」。

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：

《平妖傳》(標記語料)使用舊版造字 35 個，字頻 87 次，這 35 個造字中，32 個可轉成 Windows XP 能顯示的字，字頻 81 次；另外 3 個字必須轉成構字式，字頻 6 次。

轉碼完成製作轉碼分析表，請參考附件一《平妖傳》(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 5299 的「𢆳」字，是「𢆳」的異體字，以標準字「𢆳」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 3098 的「𢆳」字，Unicode 字型呈現為「𢆳」，實際上仍為同一字。

附件一、《平妖傳》(標記語料)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
20	𢆳	2	FA53			𢆳 𢆳	
74	𢆳	1	FAAB	29B06		𢆳 𢆳	
1170	𢆳	11	90A8	7583	𢆳		
1213	𢆳	2	90D3	7743	𢆳		
1491	𢆳	1	92AF	6527	𢆳		
1981	𢆳	1	95C2	8B90	𢆳		
2097	𢆳	3	9677	26782		月 𢆳 𢆳	
2999	𢆳	6	9C4F	5523	𢆳		
3058	𢆳	1	9CAC	5413	𢆳		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3098	揸	1	9CD4	63F8	揸		Unicode 字型 呈現差異。
3834	嗶	3	81A3	5637	嗶		Unicode 字型 呈現差異。
3846	墾	5	81AF	58FB	墾		
3934	榻	2	8248	69C5	榻		
3938	橈	4	824C	6AC8	橈		
3942	毡	2	8250	6BE1	毡		
3963	炆	2	8265	7081	炆		
3966	烟	2	8268	70DF	烟		
3970	焗	1	826C	714F	焗		
3990	甄	1	82A2	750E	甄		
3996	疎	1	82A8	758E	疎		
4013	盃	1	82B9	76CC	盃		
4036	窰	2	82D0	7AB0	窰		
4038	竈	2	82D2	7AC8	竈		
4044	筴	2	82D8	7B6F	筴		
4074	縹	3	82F6	7E27	縹		
4105	菓	2	8356	83D3	菓		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4115	藁	1	8360	85C1	藁		
4129	虬	1	836E	866C	虬		
4139	袴	1	8378	88B4	袴		
4195	鉢	8	83D2	9262	鉢		
4260	麇	1	8454	9E9E	麇		
4267	鼈	7	845B	9F08	鼈		
4533	唸	1	85EA	552B	唸		
5008	幪	2	88EE	5E5E	幪		
5299	戟	1	8AD7	39B8	卓  戈		戟，異體字問題。

## 二、手動取代表說明：

《平妖傳》(標記語料)檔案中需手動修改的情況分為以下幾種：

(一)《平妖傳》(標記語料)檔案中出現「●」字形，共計 48 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. 標記行碼：語言所標記檔裡設定的行碼。
3. 檔案原文摘錄：摘錄《平妖傳》(標記語料)檔案有「●」字形的文句。
4. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
5. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《平妖傳》(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. 3	161	袁(Nb) 公(Na) ●(VE) 著(Di) 樹梢頭	覷	
p. 19	1275	●(Na) 、 (PAUSECATEGORY) 鹿	𠂇 𠂇 巴	
p. 19	1278	忽然(Dd) 山●(Nc) 裏	𠂇	
p. 19	1293	趙壹(Nb) ●(VE)[+vrv] 定	覷	
p. 20	1325	直(VH) 到(VCL) 那 (Nh) 山●	𠂇	
p. 28	1931	要(VK) ●(VC)[+vrv] 他	𠂇 𠂇 雙	
p. 28	1968	手(Na) 劈胸(Dh) ● (VC) 那(Nh) 婆子	𠂇 𠂇 雙	
p. 28	1983	推(VC) 推(VC) ●(VC) ●	搥搥	
p. 51	3668	鴛鴦●(Na) 已(Dd) 註 定	牒	
p. 54	3858	火(Na) 燒(VC) ●● (Na)	𠂇 𠂇	
p. 67	4768	做(VC) ●●(Na) 送	𠂇 𠂇	

頁碼	標記行碼	檔案原文摘錄	原書字形	備註
		(VD) 老(VH) 師太		
p. 79	5684	單(A) 槳兒(Na) ● (VC) 著	撻	
p. 79	5687	依先(Dd) ●(VC) 著 (Di) 渡船	撻	
p. 80	5711	●(VC) 了(Di) 渡船	撻	
p. 80	5738	又(Dl) ●(VC)[+vrv] 去(VCL)[+vrr] 了	撻	
p. 91	6483	不(Dc) 論(VE) 老(VH) 的(DE) ●(VH) 的	嫩	
p. 91	6529	用(VC) 力(Na) 一 (Neu) ●	扌△雙	
p. 113	8179	兩(Neu) 個(Nf) 大 (VH) ●●	饞饞	
p. 117	8496	一眼(Dh) ●(VC) 著 (Di) 蛋子和尚	睇	
p. 118	8520	那(Nh) 婆子(Na) ● (VC) 我	睇	
p. 120	8772	萬(Neu) 僧(Na) 齋● (Na) 之(DE) 費	貝△親	
p. 129	9444	●●(VH) 的(T4) 說	𠵼𠵼	

頁碼	標記行碼	檔案原文摘錄	原書字形	備註
		(VE) 話(Na)		
p. 155	11436	雷(Nb) 太監(Na) 家 (Nc) 沒瞅沒●	係	
p. 159	11820	將(P) 妾(Nh) 魂● (Na) 送(VD) 豐都	牒	
p. 173	12865	來(T7) 個(Nf) 大(VH) ●●	饅饅	
p. 176	13135	望(P) 上(Ng) 一(Dd) ●	扌△雙	
p. 178	13270	把(P) 頭(Na) 一(Dd) ●	扌△雙	
p. 206	15642	員外(Na) 一(Dd) ● (VC)	扌△雙	
p. 247	18742	是(SHI) 個(Nf) 唸● (VH) 的(DE) 人	噫	
p. 250	18944	文●(Na)	牒	有 2 處
p. 253	19205	替(P) 他(Nh) ●(VC) 了(Di) 魚	犀△司	
p. 257	19578	面面相●(VA)	覷	
p. 259	19677	賣(VD) ●●(Na) 的 (DE) 喚	餽蝕	

頁碼	標記行碼	檔案原文摘錄	原書字形	備註
p. 290	22147	瘡●(Na)	厭△肉	有 5 處
p. 344	26252	肱●(A) 一(Neu) 聲 (Nf)	月△察	
p. 349	26630	只(Daa) 見(VK) 旌● (Na) 焜耀	从△童	

(二) 《平妖傳》素語料中之「●」字，為日後線上檢索方便，在標記語料多修正為古今通用字，與文獻處理實驗室較以原書為準修改其轉碼後素語料檔案的「●」字，兩者用字略有不同，其中多為異體字差異，特製作用字差異記錄表，以供未來參考。請參考附件三《平妖傳》素語料與標記語料「●」字差異記錄表。欄位說明如下：

1. 頁碼：修改字形所在位置之原書頁碼。
2. 素語料原文摘錄：摘錄《平妖傳》素語料檔案的文句段落，其中紅字部分，是經由文獻處理實驗室轉碼修改後的「●」字。
3. 標記語料原文摘錄：摘錄《平妖傳》標記語料檔案的文句段落，其中藍字的部分，為標記語料修正素語料中之「●」字形。
4. 備註：記錄修改相同辭句的次數或其他事項。

### 附件三、《平妖傳》素語料與標記語料

#### 「●」字差異記錄表

頁碼	素語料原文摘錄	標記語料原文摘錄	備註
p. 6	覷著許多儒字	覷(VE) 著(Di) 許多 (Neqa) 儒(Na) 字	