

# 《永樂大典戲文三種》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  
2009/11/23 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 572 的「囟」字，Unicode 編碼是 342B，由於 XP 並不支援，仍需使用構字式「囟」。

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：

《永樂大典戲文三種》使用舊版造字 39 個，字頻 402 次，這 39 個造字中，17 個可轉成 Windows XP 能顯示的字，字頻 203 次；另外 22 個字必須轉成構字式，字頻 199 次。

轉碼完成製作轉碼分析表，請參考附件一《永樂大典戲文三種》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. 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 541 的「內」字，而原書使用之字為「內」，查詢教育部異體字字典確認「內」並非「內」之異體字，因此歸為錯字，以程式取代為正確字形「內」。

附件一、《永樂大典戲文三種》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	厶	9	FBEC			厶	
297	厶	93	FBED			厶	
298	厶	10	FBEE			厶	
299	形	6	FBEF			形	
300	口	6	FBF0			口	
432	業	11	FCD7			業	
434	鼠	2	FCD9			鼠	
541	內	1	FDA7			內	內，錯字，以程式全部取代。
572	囟	1	FDC6	342B		囟	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
639	丿	1	FE4A			丿	
685	禘	2	FE78			禘	
717	疋	2	FEBA			疋	
746	𠂇	8	FED7			𠂇	
752	𠂇	3	FEDD			𠂇	
766	𠂇	1	FEEB			𠂇	
768	𠂇	5	FEED			𠂇	
770	扌	16	FEEF			扌	
772	辶	7	FEF1			辶	
777	宀	1	FEF6			宀	
784	亻	3	FEFD			亻	
899	駝	1	8ED3	99DE	駝		
3801	剗	4	8160	5257	剗		
3805	勅	3	8164	52C5	勅		
3846	壻	21	81AF	58FB	壻		
3851	姪	1	81B4	59F9	姪		
3966	烟	21	8268	70DF	烟		
3996	疎	3	82A8	758E	疎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4016	着	134	82BC	7740	着		
4039	疇	1	82D3	7ADA	疇		
4107	葱	1	8358	8471	葱		
4146	靚	3	83A1	89A9	靚		
4167	賚	4	83B6	8CEB	賚		
4173	躡	2	83BC	8EAD	躡		
4184	迹	1	83C7	8FF9	迹		
4342	倮	1	84C8	502E	倮		
5034	弃	10	8949			弃	
5075	豸	1	8972			豸	
5439	栢	1	8BC6	6822	栢		
5726	呂	1	8DAB			呂	呂，錯字，以程式全部取代。