

《三國志平話》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
2009/12/11 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 2432 的「鞞」字，Unicode 編碼是 29350，由於 XP 並不支援，仍需使用構字式「革𠂔登」。

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：

《三國志平話》使用舊版造字 33 個，字頻 72 次，這 33 個造字中，26 個可轉成 Windows XP 能顯示的字，字頻 54 次；另外 7 個字必須轉成構字式，字頻 18 次。

轉碼完成製作轉碼分析表，請參考附件一《三國志平話》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 1073 的「算 \triangle 么」字，是「篡」的異體字，以標準字「篡」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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。

附件一、《三國志平話》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	\triangle	1	FBEC			\triangle	
297	\triangle	12	FBED			\triangle	
768	卩	1	FEED			卩	
784	彳	1	FEFD			彳	
1073		5	8FE4			算 \triangle 么	篡，異體字問題。
1808	糊	1	94B2	3BB6		木 \triangle 朔	
2157	酌	1	96D5	9167	酌		
2343	鞞	1	97F2	4A5E		革 \triangle 占	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2432	鞞	1	98AE	29350		革𠃉登	
3789	冲	1	8154	51B2	冲		沖，異體字問題。
3830	咤	1	817D	54A4	咤		
3834	嗶	1	81A3	5637	嗶		
3840	坂	10	81A9	5742	坂		
3846	壻	2	81AF	58FB	壻		
3864	峯	1	81C1	5CEF	峯		
3978	猪	1	8274	732A	猪		Unicode 字型呈現差異。
4000	疴	1	82AC	75B4	疴		
4080	羣	1	82FC	7FA3	羣		
4105	菓	1	8356	83D3	菓		
4129	虬	3	836E	866C	虬		
4146	覩	2	83A1	89A9	覩		
4167	賚	2	83B6	8CEB	賚		
4195	鉢	4	83D2	9262	鉢		
4223	顛	1	83EE	984B	顛		
4256	鷄	1	8450	9DC4	鷄		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4267	鼈	1	845B	9F08	鼈		
4515	呪	2	85D8	546A	呪		
4768	妬	1	8779	59AC	妬		
5112	园	1	89B9	56ED	园		
5300	鬪	4	8AD8	9B2A	鬪		
5551	賈	4	8C77	8CF7	賈		
5558	龕	1	8C7E	9E84	龕		
5860	无	1	C6D3	65E0	无		