

《七國春秋平話》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 丁玟伶

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 1520 的「𠄎」字，Unicode 編碼是 3B20，由於 XP 並不支援，仍需使用構字式「日𠄎𠄎」。

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：

《七國春秋平話》使用舊版造字 24 個，字頻 63 次，這 24 個造字中，19 個可轉成 Windows XP 能顯示的字，字頻 55 次；另外 5 個字必須轉成構字式，字頻 8 次。

轉碼完成製作轉碼分析表，請參考附件一《七國春秋平話》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 4134 的「衛」字，是「衛」的異體字，以標準字「衛」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 3935 的「槩」字，Unicode 字型呈現為「槩」，實際上仍為同一字。

附件一、《七國春秋平話》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
297	𠄎	3	FBED			𠄎	
298	𠄏	1	FBEE			𠄏	
770	扌	1	FEEF			扌	
784	亻	2	FEFD			亻	
1067	碁	1	8FDE	68CA	碁		
1073		2	8FE4			算 𠄎 么	篡，異體字問題。
1278	碍	1	9155	788D	碍		
1520	噉	1	92CC	3B20		日 𠄎 煞	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2157	酌	1	96D5	9167	酌		
3797	凭	1	815C	51ED	凭		
3843	堦	1	81AC	5826	堦		
3846	堦	2	81AF	58FB	堦		
3876	廻	2	81CD	5EFB	廻		
3935	槩	2	8249	3BA3	槩		Unicode 字型 呈現差異。
3966	烟	7	8268	70DF	烟		
4067	綉	8	82EF	7D89	綉		
4080	羣	2	82FC	7FA3	羣		
4134	衛	4	8373	885E	衛		衛，異體字問 題。
4146	覩	9	83A1	89A9	覩		
4167	賚	1	83B6	8CEB	賚		
4184	迹	1	83C7	8FF9	迹		
4256	鷄	1	8450	9DC4	鷄		
5300	鬪	6	8AD8	9B2A	鬪		
5551	賚	3	8C77	8CF7	賚		