

《秦併六國平話》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
2009/12/14 丁玫伶

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 2796 的「𠄎」字，Unicode 編碼是 295E9，由於 XP 並不支援，仍需使用構字式「金𠄎風」。

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：

《秦併六國平話》使用舊版造字 38 個，字頻 174 次，這 38 個造字中，33 個可轉成 Windows XP 能顯示的字，字頻 162 次；另外 5 個字必須轉成構字式，字頻 12 次。

轉碼完成製作轉碼分析表，請參考附件一《秦併六國平話》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。

附件一、《秦併六國平話》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	𠄎	1	FBEC			𠄎	
297	𠄎	6	FBED			𠄎	
298	𠄎	2	FBEE			𠄎	
759	广	2	FEE4			广	
1073		1	8FE4			𠄎么	篡，異體字問題。
1885	𧈧	1	9540	882D	𧈧		
1905	𧈧	1	9554	88E9	𧈧		
2796	𧈧	1	9AE0	295E9		金 𧈧 風	
2833	刃	3	9B46	5204	刃		
3792	凉	1	8157	51C9	凉		凉，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3801	剗	1	8160	5257	剗		
3826	叶	1	8179	53F6	叶		
3840	坂	7	81A9	5742	坂		
3864	峯	1	81C1	5CEF	峯		
3931	椀	2	8245	6900	椀		
3966	烟	3	8268	70DF	烟		
3977	狗	1	8273	72E5	狗		
4021	礩	2	82C1	78AF	礩		
4067	綉	9	82EF	7D89	綉		
4069	綑	1	82F1	7DB3	綑		
4080	羣	18	82FC	7FA3	羣		
4094	館	1	834B	8218	館		
4107	葱	3	8358	8471	葱		
4115	藁	1	8360	85C1	藁		
4124	蝨	1	8369	8771	蝨		
4129	虬	3	836E	866C	虬		
4134	衛	12	8373	885E	衛		衛，異體字問題。
4139	袴	1	8378	88B4	袴		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4146	覩	2	83A1	89A9	覩		
4167	賈	17	83B6	8CEB	賈		
4184	迹	2	83C7	8FF9	迹		
4226	凜	1	83F1	98E1	凜		
4231	駝	1	83F6	9A0C	駝		
4309	刼	3	84A7	523C	刼		
4460	勅	35	85A1	52D1	勅		
5008	幪	2	88EE	5E5E	幪		
5300	鬪	23	8AD8	9B2A	鬪		
5551	賈	1	8C77	8CF7	賈		

二、手動取代表說明：

《秦併六國平話》檔案中出現「●」字形，共計 2 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《秦併六國平話》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、《秦併六國平話》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 227	叫牙朱●	火△奈	
p. 227	忔●蹄番馬	忔△業	