

《公羊傳》(標記)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
2008/4/08 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 2830 的「𠄎」字，Unicode 編碼是 4022，由於 XP 並不支援，仍需使用構字式「目𠄎矢」。

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：

《公羊傳》(標記)使用舊版造字 23 個，字頻 121 次，這 23 個造字中，16 個可轉成 Windows XP 能顯示的字，字頻 103 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《公羊傳》(標記)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，如：人名、地名等特殊情形。例：造字編號 3873 的「廢」字，是「廢」的異體字，以標準字「廢」取代。
2. 部份使用不同字形，手動修改：檔案中為同一舊版造字，校對時卻發現原書使用兩個以上不同的字形，可能部份舊版造字為錯字或原書使用兩種異體字形，導致同一舊版造字轉碼時對應到兩個以上的字形；將一部份手動修改，剩餘的頻次則用程式取代成對應的 Unicode 或構字式。例：造字編號 3976 的「犁」字頻次為 3，將人名的 2 字手動修改為「犁」，剩餘 1 字轉為「犁」。手動修改部份請參考附件二《公羊傳》手動取代表。

附件一、《公羊傳》(標記)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
297	𠃉	10	FBED			𠃉	
967	𠃉	1	8F58	50AB	𠃉		
2478	𠃉	6	98DC	978C	𠃉		
2830	𠃉	1	9B43	4022		目 𠃉 矢	
3027	𠃉	1	9C6B	3F90		疒 𠃉 列	
3814	却	2	816D	5374	却		
3873	廢	4	81CA			广 𠃉 毀	廢，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3883	徧	3	81D4	5FA7	徧		
3976	犁	3	8272	7282	犁		部份使用不同字形「犁」，手動修改。
4068	綫	1	82F0	7DAB	綫		
4080	羣	4	82FC	7FA3	羣		
4091	皐	1	8348	81EF	皐		
4142	褒	6	837B	8943	褒		褒，異體字問題。
4455	劔	3	857A			僉厶刀	
4469	勻	7	85AA	5304	勻		
4855	汩	1	87F2	3CC5		彳厶不	
4961	盼	1	88BF	80B9	盼		
5156	丕	1	89E5	353B		丕	
5173	漱	1	89F6	6F44	漱		漱，異體字問題。
5299	戟	1	8AD7	39B8		卓厶戈	戟，異體字問題。
5583	殮	2	8CB9	98F1	殮		
5591	杞	60	8CC1	233CC		木厶巳	杞，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
5652	螻	1	8CFE	27312		或𧈧虫	

二、手動取代表說明：

《公羊傳》(標記)檔案中需手動修改的情況分為以下幾種：

(一) 《公羊傳》(標記)檔案中之舊版造字，因字形部份不同而無法全部以程式取代，皆手動修改成Windows XP能顯示的字或構字式，並製作手動取代表，請參考附件二。欄位說明如下：

1. 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
2. 造字：舊版造字字形。
3. 取代：實際使用字形與舊版造字不同時，於此欄列出。
4. 標記行碼：語言所標記檔裡設定的行碼。
5. 檔案原文摘錄：摘錄檔案《公羊傳》(標記)中的文句段落，其中紅字底線的部分，為需要手動取代的原文。
6. 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。
7. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《公羊傳》(標記)手動取代表

編號	造字	取代	標記行碼	檔案原文摘錄	手動取代結果	備註
3976	犁	犁	8821	樂祁 <u>島</u> (NB1)	樂祁 <u>犁</u> (NB1)	
			9282	樂祁 <u>島</u> (NB1)	樂祁 <u>犁</u> (NB1)	

(二) 《公羊傳》(標記)檔案中出現「●」字形，共計 3 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，請參考附件三。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《公羊傳》(標記)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。

附件三、《公羊傳》(標記)「●」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. 7	1037	●(VH1) 也(T)	𠂇 𠂇 戍	
p. 35	5322	其(NH) 君(NA1) 夷●	猓	有 2 處

(三)《公羊傳》(標記)檔案中出現「？」字形，共計 5 字，比對原書後以 Unicode 字形或構字式手動取代，並製作「？」字形手動取代表，請參考附件四。欄位說明如下：

1. 標記行碼：語言所標記檔裡設定的行碼。
2. 檔案原文：摘錄《公羊傳》(標記)檔案有「？」字形的文句。
3. 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。

附件四、《公羊傳》(標記)「？」字形手動取代表

標記行碼	檔案原文	字形	備註
6650	鄭(NB3)[+prop] 公子?	鯀	
7045	鄭伯?原(NB1)[+prop]	髡	有 2 處
8590	胡子?(NB1)[+prop]	髡	
9688	樂?(NB1)[+prop]	髡	

