

《公羊傳》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》使用舊版造字 28 個，字頻 435 次，這 28 個造字中，24 個可轉成 Windows XP 能顯示的字，字頻 429 次；另外 4 個字必須轉成構字式，字頻 6 次。

轉碼完成製作轉碼分析表，請參考附件一《公羊傳》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. 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。
 3. 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 5156 的「丕」字，而原書使用之字為「平」，查詢教育部異體字字典確認「平」並非「丕」之異體字，因此歸為錯字，以程式取代為正確字形「平」。

附件一、《公羊傳》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
967	僮	1	8F58	50AB	僮		
2478	鞏	6	98DC	978C	鞏		
2830	眈	1	9B43	4022		目𠃉𠃉	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3814	却	2	816D	5374	却		
3873	廡	4	81CA			广△段	廡，異體字問題。
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	臯	1	8348	81EF	臯		
4104	莅	4	8355	8385	莅		
4132		26	8371			血△丞	眾，異體字問題。
4134	衛	281	8373	885E	衛		衛，異體字問題。
4142	裛	6	837B	8943	裛		裛，異體字問題。
4176	輒	5	83BF	8F19	輒		輒，異體字問題。
4237	髡	5	83FC	9AE0	髡		
4250	鱒	1	844A	9C0C	鱒		
4259	麇	1	8453	9E95	麇		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4455	劒	3	857A			僉△刀	
4469	勻	7	85AA	5304	勻		
4855	汩	1	87F2	3CC5		汩△不	
4961	矜	1	88BF	80B9	矜		
5156	平	1	89E5	353B		平	平，錯字，以程式全部取代。
5173	漱	1	89F6	6F44	漱		漱，異體字問題。
5299	戟	1	8AD7	39B8		卓△戈	戟，異體字問題。
5583	殮	2	8CB9	98F1	殮		
5591	杞	62	8CC1	233CC		木△已	杞，異體字問題。 部份頻次多餘，手動修改。
5652	蝥	1	8CFE	27312		或△虫	

二、手動取代表說明：

《公羊傳》檔案中需手動修改的情況分為以下幾種：

- (一) 《公羊傳》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. 檔案原文摘錄：摘錄檔案《公羊傳》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。文句後方標示（全部取代）者，表示執行全部取代的動作，不一一列舉所有原文。
5. 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。文句後方標示（全部取代）者，表示執行全部取代的動作，不一一列舉所有原文。
6. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《公羊傳》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果	備註
3976	犁	犁	宋樂祁 <u>島</u>	宋樂祁 <u>犁</u>	
			執宋行人樂祁 <u>島</u>	執宋行人樂祁 <u>犁</u>	
5591	杞		<u>唐</u> 宋公使華元來聘	宋公使華元來聘	
			夏 <u>唐</u> 公圍成	夏公圍成	

(二)《公羊傳》檔案中出現「●」字形，共計 13 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. 檔案原文摘錄：摘錄《公羊傳》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。

附件三、《公羊傳》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 7	曷為以二日卒之・●也	𠄎 𠄎 戍	
p. 7	●・何以書	虫 𠄎 眾	
p. 15	大瘠者何・●也	疒 𠄎 列	
p. 24	八月・●	虫 𠄎 眾	
p. 31	遂在外也・○●	虫 𠄎 眾	
p. 35	晉趙盾弑其君夷●	隼	有 2 處
p. 37	秋・八月●	虫 𠄎 眾	
p. 39	楚子伐宋・ 秋●	虫 𠄎 眾	
p. 47	八月・●	虫 𠄎 眾	
p. 67	冬・十有二月・●・	虫 𠄎 眾	
p. 67	九月・●	虫 𠄎 眾	
p. 67	十有二月・●	虫 𠄎 眾	

(三)《公羊傳》檔案中，對於舊版造字以外發現的錯字，以手動修改，並製作錯字修改表，請參考附件四。欄位說明如下：

1. 檔案原文摘錄：摘錄檔案《公羊傳》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。
2. 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。
3. 備註：記錄修改相同詞句的次數或其他事項。

附件四、《公羊傳》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
晉殺其大夫 <u>平</u> 鄭父	晉殺其大夫 <u>丕</u> 鄭父	