

《穀梁傳》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
2008/4/09 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 5652 的「𧈧」字，Unicode 編碼是 27312，由於 XP 並不支援，仍需使用構字式「或𧈧虫」。

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：

《穀梁傳》使用舊版造字 31 個，字頻 383 次，這 31 個造字中，29 個可轉成 Windows XP 能顯示的字，字頻 380 次；另外 2 個字必須轉成構字式，字頻 3 次。

轉碼完成製作轉碼分析表，請參考附件一《穀梁傳》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. 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 2951 的「脞」字，而原書使用之字為「脞」，查詢教育部異體字字典確認「脞」並非「脞」之異體字，因此歸為錯字，以程式取代為正確字形「脞」。

附件一、《穀梁傳》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
819	候	1	8E61	202EB		彳 厶 侯	候，異體字問題。
938	燮	1	8EFA	7215	燮		燮，異體字問題。
980	犖	4	8F65	72AB	犖		
1565	絲	2	92F9	7DDC	絲		
2478	鞏	2	98DC	978C	鞏		
2651	鸚	2	99EC	9E1C	鸚		
2951	脞	1	9BDE	8140	脞		脞，錯字，以程式全部取

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							代。
3787	冢	3	8152	51A2	冢		
3791	况	3	8156	51B5	况		况，異體字問題。
3873	廩	6	81CA			广△段	廩，異體字問題。
3976	犁	1	8272	7282	犁		犁，異體字問題。
4038	竈	1	82D2	7AC8	竈		
4080	羣	3	82FC	7FA3	羣		
4091	阜	3	8348	81EF	阜		
4104	莅	8	8355	8385	莅		
4107	葱	1	8358	8471	葱		
4132		24	8371			血△丞	眾，異體字問題。
4134	衛	267	8373	885E	衛		衛，異體字問題。
4142	褒	1	837B	8943	褒		褒，異體字問題。
4176	輒	5	83BF	8F19	輒		輒，異體字問題。
4190	邳	1	83CD	90C4	邳		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4227	飡	1	83F2	98E6	飡		
4237	髡	6	83FC	9AE0	髡		
4250	鮓	1	844A	9C0C	鮓		
4469	勻	9	85AA	5304	勻		
4533	唸	1	85EA	552B	唸		
5048	婚	2	8957	36F0		女𠂇昏	婚，異體字問題。
5156	𠂇	1	89E5	353B		𠂇	
5302	昏	2	8ADA	662C	昏		昏，異體字問題。
5591	杞	18	8CC1	233CC		木𠂇巳	杞，異體字問題。
5652	蝨	2	8CFE	27312		或𠂇虫	

二、手動取代表說明：

《穀梁傳》檔案中需手動修改的情況分為以下幾種：

- (一) 《穀梁傳》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. 檔案原文摘錄：摘錄《穀梁傳》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。

附件二、《穀梁傳》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 19	目晉侯●殺	斤	
p. 25	以晉侯而●執曹伯	斤	
p. 62	公會吳于橐●	臯	

(二)《穀梁傳》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，請參考附件三。欄位說明如下：

1. 檔案原文摘錄：摘錄檔案《穀梁傳》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。

附件三、《穀梁傳》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
陳侯 <u>熹</u> 獵	陳侯 <u>憲</u> 獵	