

《穀梁傳》(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
2008/4/09 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「蝥」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》(標記)使用舊版造字 18 個，字頻 63 次，這 18 個造字中，16 個可轉成 Windows XP 能顯示的字，字頻 60 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》(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「廐」字，是「廐」的異體字，以標準字「廐」取代。

附件一、《穀梁傳》(標記)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
980	犖	4	8F65	72AB	犖		
1565	繇	2	92F9	7DDC	繇		
2478	鞏	2	98DC	978C	鞏		
2651	鸛	2	99EC	9E1C	鸛		
3787	豕	3	8152	51A2	豕		
3873	廐	6	81CA			广△段	廐，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	臯	3	8348	81EF	臯		
4469	勺	9	85AA	5304	勺		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4533	唸	1	85EA	552B	唸		
5048	婚	2	8957	36F0		女𡗗昏	婚，異體字問題。
5133	憇	1	89CE	6199	憇		
5156	𠂔	1	89E5	353B		𠂔	
5302	昏	2	8ADA	662C	昏		昏，異體字問題。
5591	杞	18	8CC1	233CC		木𡗗巳	杞，異體字問題。
5652	螽	2	8CFE	27312		或𡗗虫	

二、手動取代表說明：

《穀梁傳》(標記)檔案中需手動修改的情況分為以下幾種：

(一) 《穀梁傳》(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，請參考附件二。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《穀梁傳》(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。

附件二、《穀梁傳》(標記)「？」字形手動取代表

標記行碼	檔案原文	字形	備註
3997	弑(VC1) 其(NH) 君(NA1) ?	髡	

標記行碼	檔案原文	字形	備註
3998	日(VAN)[+nv] ?(NB1)[+prop]	髡	
5109	井竈(NA2)[+co] ?韭	葱	
5752	鄭(NB3)[+prop] 公子?(NB1)	鮪	
7300	歎(VC1) ?粥	飡	

(二)《穀梁傳》(標記)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. 標記行碼：語言所標記檔裡設定的行碼。
2. 檔案原文摘錄：摘錄檔案《穀梁傳》(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。
3. 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。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件三、《穀梁傳》(標記)缺字外錯字修改

標記行碼	檔案原文摘錄	手動取代結果	備註
2719	齊(NB4)[+prop] <u>_(VAN)</u> 盟(NI)	齊(NB4)[+prop] <u>菝</u> (VAN) 盟(NI)	
2720	<u>_(NI)</u> 者(T)	<u>菝</u> (NI) 者(T)	
4239	莒(NB4)[+prop] <u>_(VAN)</u> 盟(NI)	莒(NB4)[+prop] <u>菝</u> (VAN) 盟(NI)	
6051	以(P) <u>_(VAN)</u> 祭祀	以(P) <u>菝</u> (VAN) 祭祀	
6949	齊(NB4)[+prop] <u>_(VAN)</u> 盟	齊(NB4)[+prop] <u>菝</u> (VAN) 盟	
6950	<u>_(NI)</u>	<u>菝</u> (NI)	
6952	謂(VF) 之(NH) <u>_(NI)</u>	謂(VF) 之(NH) <u>菝</u>	

標記行碼	檔案原文摘錄	手動取代結果	備註
		(NI)	
8122	<u>一</u> (VAN) 盟(NI)	<u>莅</u> (VAN) 盟(NI)	
4334	我(NH) <u>之</u> (VJ) 之	我(NH) <u>褒</u> (VJ) 之	
7569	<u>一</u> 宛(NB1)[+prop]	<u>邳</u> 宛(NB1)[+prop]	