

# 《孟子》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  
2010/04/23 丁致伶

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 917 的「𪗇」字，Unicode 編碼是 4CD8，由於 XP 並不支援，仍需使用構字式「鳥𪗇我」。

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：

《孟子》使用舊版造字 36 個，字頻 75 次，這 36 個造字中，34 個可轉成 Windows XP 能顯示的字，字頻 72 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《孟子》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 2809 的「鎡」字，而原書使用之字為「鎡」，「鎡」為錯字，以程式取代為正確字形「鎡」。

附件一、《孟子》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
917	𪗇	2	8EE5	4CD8		鳥 𪗇 我	
994	𪗈	1	8F73	8E75	楚		
1044	𪗉	1	8FC7	511B	儻		
1226	𪗊	1	90E0	7F78	罰		
1565	𪗋	1	92F9	7DDC	絲		
2155	𪗌	4	96D3	9D83	駝		
2809	鎡	1	9AED			食 𪗍 益	鎡，錯字，以程式全部取代。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3787	冢	1	8152	51A2	冢		
3833	嘍	1	81A2	5611	嘍		
3883	徧	3	81D4	5FA7	徧		
3921	皙	4	81FA	6673	皙		
3924	菴	1	81FD	671E	菴		
3930	惇	3	8244	6901	惇		
3974	牀	1	8270	7240	牀		
4018	矚	2	82BE	77AF	矚		
4080	羣	1	82FC	7FA3	羣		
4104	莅	1	8355	8385	莅		
4114	藥	1	835F	8616	藥		
4132		11	8371			血𩺰𩺱	眾，異體字問題。
4134	衛	10	8373	885E	衛		衛，異體字問題。
4146	覩	1	83A1	89A9	覩		
4184	迹	1	83C7	8FF9	迹		
4227	鈇	1	83F2	98E6	鈇		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4237	髡	1	83FC	9AE0	髡		
4242	鬪	3	8442	9B2D	鬪		
4252	鰓	1	844C			魚𩺰眾	鰓，異體字問題。
4265	鼃	3	8459	9F03	鼃		
4267	鼈	3	845B	9F08	鼈		
4455	劔	1	857A			僉𠄎刃	
4507	吳	1	85D0	5434	吳		吳，異體字問題。
4760	覲	1	8771	89B8	覲		
4996	斲	1	88E2	65B5	斲		
5108	惛	1	89B5	60FD	惛		
5169	危	1	89F2	6239	危		
5302	昏	2	8ADA	662C	昏		
5583	殮	2	8CB9	98F1	殮		

## 二、手動取代表說明：

《孟子》檔案中出現「●」字形，共計1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《孟子》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、《孟子》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 299	●不欲有夫妻	岂	