

《佛說罵意經》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
2009/05/20 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 1758 的「火𠄎貴」字，Unicode 編碼是 243F3，由於 XP 並不支援，仍需使用構字式「火𠄎貴」。

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：

《佛說罵意經》使用舊版造字 8 個，字頻 9 次，這 8 個造字中，7 個可轉成 Windows XP 能顯示的字，字頻 7 次；另外 1 個字必須轉成構字式，字頻 2 次。

轉碼完成製作轉碼分析表，請參考附件一《佛說罵意經》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 4768 的「妬」字，是「妒」的異體字，以標準字「妒」取代。
 2. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，但只是字體風格差異，實際上仍為同一個字，因此仍取 Unicode 字型。如編號 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。
 3. 待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 1758 的「火𤇀貴」。

附件一、《佛說罵意經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1758		2	945E	243F3		火𤇀貴	待造字。
3143	潑	1	9D42	6FFD	潑		
3814	却	1	816D	5374	却		
3978	猪	1	8274	732A	猪		Unicode 字型呈現差異。
4256	鷄	1	8450	9DC4	鷄		雞，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4267	鼈	1	845B	9F08	鼈		
4768	妬	1	8779	59AC	妬		妒，異體字問題。
4775	姝	1	87A2	59B7	姝		

二、手動取代表說明：

《佛說罵意經》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《佛說罵意經》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、《佛說罵意經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 534	當●脹壞爛	臄	
p. 534	尻肉血兩脛●	脛	
p. 534	脛●確正	脛	