

《五陰譬喻經》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
2009/03/24 丁玫伶

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 1533 的「劇」字，Unicode 編碼是 34DF，由於 XP 並不支援，仍需使用構字式「皮𠂔 丩」。

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：

《五陰譬喻經》使用舊版造字 3 個，字頻 7 次，這 3 個造字中，2 個可轉成 Windows XP 能顯示的字，字頻 6 次；另外 1 個字必須轉成構字式，字頻 1 次。

轉碼完成製作轉碼分析表，請參考附件一《五陰譬喻經》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 4146 的「覩」字，是「睹」的異體字，以標準字「睹」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 4561 的「𠃉〃厂曲比◻」字，而原書使用之字為「鹿」，查詢教育部異體字字典確認「𠃉〃厂曲比◻」並非「鹿」之異體字，因此歸為錯字，以程式取代為正確字形「鹿」。

附件一、《五陰譬喻經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1533	劇	1	92D9	34DF		皮𠃉 𠃉	
4146	覩	1	83A1	89A9	覩		睹，異體字問題。
4561		5	8647			𠃉〃厂 曲比◻	鹿，錯字，以程式全部取代。