

《雜阿含經》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 888 的「捺」字，Unicode 編碼是 3B88，由於 XP 並不支援，仍需使用構字式「木𠂔奈」。

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：

《雜阿含經》使用舊版造字 26 個，字頻 276 次，這 26 個造字中，18 個可轉成 Windows XP 能顯示的字，字頻 199 次；另外 8 個字必須轉成構字式，字頻 77 次。

轉碼完成製作轉碼分析表，請參考附件一《雜阿含經》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. 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 4495 的「𪗇」字，Unicode 字型呈現為「𪗇」，實際上仍為同一字。

附件一、《雜阿含經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
888	捺	60	8EC8	3B88		木𪗇奈	
943	禡	1	8F40	8975	禡		
1053	傲	1	8FD0	6160	傲		
1638	洩	1	93A5	3CDB		氵𪗇𪗇	
1668	臍	23	93C3	81C8	臍		
1772	藜	6	946C	4527		艹𪗇梨	
2159	𪗇	3	96D7	4D35		麥𪗇并	
2172	𪗇	2	96E4	408E		矛𪗇贊	
2278	鎗	1	97B1	936E	鎗		
2631	鴉	3	99D8	9D44	鴉		
3830	咤	2	817D	54A4	咤		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3940	鬱	50	824E	6B1D	鬱		
4206	榆	1	83DD	3BD3		木ㄩ	
4226	飡	1	83F1	98E1	飡		
4398	糞	3	8541	4D43		麥ㄩ	
4495	麩	3	85C4	9EA8	麩		Unicode 字型 呈現差異。
4716	湜	2	8745	57FF	湜		
4738	睺	92	875B	777A	睺		
4739	鞞	5	875C	97DB	鞞		
4889	毳	1	8855	3CAA		登ㄩ	
5074	𠂔	1	8971	5C2B	𠂔		
5147	恠	2	89DC	6060	恠		
5179	駟	5	89FC	99C8	駟		
5496	翳	2	8C40	7F82	翳		
5753	潜	3	8DC6	6F5C	潜		
5809	踰	2	8DFE	8E79	踰		

二、手動取代表說明：

《雜阿含經》檔案中需手動修改的情況分為以下幾種：

(一)《雜阿含經》檔案中出現「●」字形，共計 57 字，比對原書後

以 Unicond 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《雜阿含經》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、《雜阿含經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 8	阿●律陀	少△兔	
p. 32	阿●羅	少△兔	有 17 處
p. 36	洪大●直	月△庸	
p. 49	躑●懶亂想	員	
p. 61	●陸 波羅	舟△冗	待造字
p. 61	●跋滂	扱	
p. 61	尼羅●施	枳	
p. 61	拘●塢隸	閑	
p. 67	莊嚴而扱●之	飭	
p. 68	繒褥毳毼●	毛△翕	
p. 68	●直長大	月△庸	
p. 85	尊者●陀迦	足△散	有 2 處
p. 87	羅摩●閻摩	少△兔	
p. 136	●。毛悉火然	金△矛	
p. 159	及波毗瘦●	少△兔	

頁碼	檔案原文摘錄	原書字形	備註
p. 161	伊尼耶鹿●	足厶專	
p. 162	曰阿●羅陀	少厶免	
p. 163	阿●樓	少厶免	有 3 處
p. 163	●樓陀大臣	少厶免	
p. 163	次復●	扭	
p. 263	●牛四足勝	牛厶夆	待造字
p. 263	●牛資養人	牛厶夆	待造字
p. 306	身體●細	月厶庸	
p. 311	痰●膿血	癩	
p. 342	盡●其乳	殼厶牛	有 4 處
p. 344	如●牛乳頃	殼厶牛	有 7 處
p. 366	伊尼延鹿●	足厶專	有 2 處

(二)《雜阿含經》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《雜阿含經》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、《雜阿含經》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
--------	--------	----

檔案原文摘錄	手動取代結果	備註
眾寶莊嚴而 <u>校</u>	眾寶莊嚴而 <u>校</u>	
佛告 <u>#z</u> 陀迦旃延	佛告 <u>足</u> <u>△</u> 散陀迦旃延	
械枷 <u>鎚</u> 檢繫其身	械枷 <u>鑊</u> 檢繫其身	