

《中本起經》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
2009/04/16 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 4856 的「𪗇」字，Unicode 編碼是 3CB2，由於 XP 並不支援，仍需使用構字式「疊𪗇毛」。

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：

《中本起經》使用舊版造字 39 個，字頻 168 次，這 39 個造字中，35 個可轉成 Windows XP 能顯示的字，字頻 164 次；另外 4 個字必須轉成構字式，字頻 4 次。

轉碼完成製作轉碼分析表，請參考附件一《中本起經》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 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 410 的「昇」字，而原書使用之字為「虵」，「昇」為錯字，以程式取代為正確字形「虵」。

附件一、《中本起經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
132	厨	2	FAE5	53A8	厨		
142	叙	5	FAEF	53D9	叙		敘，異體字問題。
147	恒	4	FAF4	6052	恒		
410	昇	4	FCC1			昇	虵，錯字，以程式全部取代。
430	仝	1	FCD5			仝	傲，錯字，以程式全部取代。
2492	類	1	98EA	983A	類		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2524	饒	1	994B	994D	饒		
2908	賊	1	9BB3	6660	賊		木 𠂇 翁，錯字，以程式全部取代。
3071	三	1	9CB9			三	艹 𠂇 汧，錯字，以程式全部取代。
3085	鮫	1	9CC7	9BCC	鮫		蹶，錯字，以程式全部取代。
3193		1	9D74			崙 𠂇 頁	協，錯字，以程式全部取代。
3497	塚	1	9F6A	585A	塚		
3805	勅	26	8164	52C5	勅		
3814	却	21	816D	5374	却		
3884	忽	1	81D5	6031	忽		
3940	鬱	9	824E	6B1D	鬱		
3972	燭	2	826E	7225	燭		
3992	嘗	1	82A4	751E	嘗		
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4028	稟	1	82C8	7980	稟		
4040	竝	1	82D4	7ADD	竝		
4041	豎	4	82D5	7AEA	豎		
4085	耻	1	8342	803B	耻		
4090	脚	1	8347	811A	脚		
4105	菓	1	8356	83D3	菓		
4134	衛	29	8373	885E	衛		衛，異體字問題。
4141	鯨	1	837A	29E49		魚 Δ 飛	裏，錯字，以程式全部取代。
4146	覩	8	83A1	89A9	覩		睹，異體字問題。
4173	躄	2	83BC	8EAD	躄		
4184	迹	6	83C7	8FF9	迹		
4195	鉢	13	83D2	9262	鉢		
4226	飡	2	83F1	98E1	飡		
4240	鬚	1	8440	9AF4	鬚		
4278	亘	1	8466	4E98	亘		
4515	呪	6	85D8	546A	呪		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4579	僂	1	8659	347A		亻厶僂	
4768	妬	3	8779	59AC	妬		妒，異體字問題。
4856	疊	1	87F3	3CB2		疊厶毛	
5133	憇	1	89CE	6199	憇		

二、手動取代表說明：

《中本起經》檔案中出現「●」字形，共計11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《中本起經》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、《中本起經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 150	不有●	恪	
p. 150	●惜名稱	恪	
p. 150	取*果●	𠄎𠄎呵	
p. 152	無以瞋怒●	扌	
p. 152	順正不阿●	扌	

頁碼	檔案原文摘錄	原書字形	備註
p. 154	茵●綻綻	禾△辱	
p. 154	幢麾羽●	羽△翳	
p. 157	●遂本志	箕	
p. 158	捐●於外	弃	改為棄
p. 159	結為●瑤	王△步	
p. 161	●橫·志行	扌	