

《生經》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
2009/03/30 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 5364 的「𠄎」字，Unicode 編碼是 239D1，由於 XP 並不支援，仍需使用構字式「歹𠄎凶」。

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：

《生經》使用舊版造字 58 個，字頻 390 次，這 58 個造字中，48 個可轉成 Windows XP 能顯示的字，字頻 367 次；另外 10 個字必須轉成構字式，字頻 23 次。

轉碼完成製作轉碼分析表，請參考附件一《生經》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 5104 的「礼」字，而原書使用之字為「愨」，「礼」為錯字，以程式取代為正確字形「愨」。
3. 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 5131 的「窻」字，Unicode 字型呈現為「窻」，實際上仍為同一字。

附件一、《生經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3		1	FA42			𠄎; 完 文 𠄎	鑲，錯字，以程式全部取代。
131	𪗇	1	FAE4	9F08	𪗇		
147	恒	9	FAF4	6052	恒		
195	勢	2	FB65	20859		教 𠄎 力	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
230	踰	1	FBAA	8E79	踰		
325	蓄	1	FC4A			蓄	木 ㄩ 奈，錯字，以程式全部取代。
414	会	1	FCC5			会	形 亻 夂 心 口，錯字，以程式全部取代。
972	犇	3	8F5D	7287	犇		
1133	瓌	1	9061	74CC	瓌		
1513	粮	1	92C5	7CAE	粮		糧，異體字問題。
1642	脉	1	93A9	8109	脉		
2162	碱	1	96DA	918E	碱		
2505	翻	1	98F7	98DC	翻		
2908	曷	1	9BB3	6660	曷		木 ㄩ 翁，錯字，以程式全部取代。
3054	蚰	5	9CA8	45B5	蚰		麥 ㄩ 黃，錯字，以程式全部取代。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3266		1	9DDF			鹿厶谷	忄厶龍，錯字，以程式全部取代。
3269	憑	1	9DE2	51F4	憑		火厶霍，錯字，以程式全部取代。
3497	塚	1	9F6A	585A	塚		
3724	爦	2	A0D2	2435C		火厶參	
3804	効	2	8163	52B9	効		
3805	勅	14	8164	52C5	勅		
3808	勾	9	8167	5303	勾		
3814	却	6	816D	5374	却		
3824	叙	5	8177	53D9	叙		
3860	尠	1	81BD	5C20	尠		
3883	徧	1	81D4	5FA7	徧		
3884	忽	2	81D5	6031	忽		
3897	慇	2	81E2	615C	慇		
3940	鬱	3	824E	6B1D	鬱		
3944	冰	1	8252	6C37	冰		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3979	猫	12	8275	732B	猫		
3992	嘗	1	82A4	751E	嘗		
3996	疎	2	82A8	758E	疎		
4041	豎	5	82D5	7AEA	豎		
4085	耻	1	8342	803B	耻		
4090	脚	11	8347	811A	脚		
4134	衛	68	8373	885E	衛		衛，異體字問題。
4141	鯨	6	837A	29E49		魚𩺰飛	裏，錯字，以程式全部取代。
4146	覩	35	83A1	89A9	覩		睹，異體字問題。
4176	輒	20	83BF	8F19	輒		
4184	迹	7	83C7	8FF9	迹		
4195	鉢	26	83D2	9262	鉢		
4203	隣	13	83DA	96A3	隣		
4237	髡	2	83FC	9AE0	髡		
4267	鼈	30	845B	9F08	鼈		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4278	亘	4	8466	4E98	亘		
4455	劒	1	857A			僉  刃	劒，異體字問題。
4515	呪	28	85D8	546A	呪		
4768	妬	3	8779	59AC	妬		妒，異體字問題。
4773	姉	6	877E	59C9	姉		Unicode 字型呈現差異。
4904	尅	1	8864	5C05	尅		
5104	礼	2	89B1	793C	礼		愬，錯字，以程式全部取代。
5131	窻	1	89CC	7ABB	窻		Unicode 字型呈現差異。
5133	憇	3	89CE	6199	憇		
5364	殂	8	8B59	239D1		歹  凶	
5785	快	1	8DE6	225D7		忄  犬	
6078	鬪	10	C851	9B2A	鬪		
6173	躡	2	C8D2	8E83	躡		

二、手動取代表說明：

《生經》檔案中出現「●」字形，共計43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《生經》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、《生經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 73	中●惜	悒	
p. 73	●惜如此	悒	
p. 74	吾為得此●	輦	
p. 75	●海水	漣△升	
p. 76	即器●水	漣△升	
p. 76	若今●水	漣△升	
p. 76	節即問夫●	輦	
p. 76	獼猴誘●我夫	言△求	
p. 76	其●瞻勞	輦	
p. 77	五●截耳割	木△瓦	
p. 81	賢者牛●	伺	
p. 81	身出光●	火△僉	

頁碼	檔案原文摘錄	原書字形	備註
p. 82	牛●·爾之所說	𠂔	
p. 82	至於鞭●	革𠂔亢	待造字
p. 82	鞭●之惱	革𠂔亢	待造字
p. 83	光●之	火𠂔僉	
p. 84	一時佛遊●	少𠂔兔	
p. 84	檀●國	少𠂔兔	
p. 84	●偈陀	𠂔𠂔𠂔隋	
p. 84	摩呵曼那●陀	少𠂔兔	
p. 84	彌檀●南模	少𠂔兔	
p. 85	●摩丘	金𠂔畀	待造字
p. 85	耶勿遮坻●	金𠂔畀	待造字
p. 85	移阿●	金𠂔畀	待造字
p. 85	阿迦羅 ●羅	金𠂔畀	待造字
p. 85	莫迦垣羅●	颯	
p. 85	羅●提摩那	颯	
p. 86	謂所●釋子曰	欣	
p. 87	●陀和黎	颯	
p. 88	便角●眼	目𠂔翕	
p. 88	●眼視吾	目𠂔翕	
p. 91	方便·權●	扌𠂔骨	

頁碼	檔案原文摘錄	原書字形	備註
p. 97	令其女●止毒	聶	
p. 97	及散久●	迥	
p. 99	●化亂意	糞	
p. 101	不●萬民	枉	
p. 102	名曰●蔗	甘	
p. 103	福光從●入	亦△肉	
p. 104	將無見●	枉	
p. 104	所見侵●	枉	
p. 106	●不敬重之	聶	
p. 106	見●心異	聶	
p. 106	●便聽之	聶	