

《大樓炭經》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
2009/03/19 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 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：

《大樓炭經》使用舊版造字 36 個，字頻 230 次，這 36 個造字中，32 個可轉成 Windows XP 能顯示的字，字頻 224 次；另外 4 個字必須轉成構字式，字頻 6 次。

轉碼完成製作轉碼分析表，請參考附件一《大樓炭經》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 4561 的「𠄎」字，而原書使用之字為「鹿」，查詢教育部異體字字典確認「𠄎」並非「鹿」之異體字，因此歸為錯字，以程式取代為正確字形「鹿」。
 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 4495 的「𪗇」字，Unicode 字型呈現為「𪗇」，實際上仍為同一字。

附件一、《大樓炭經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
888	捺	1	8EC8	3B88		木𠄎奈	
1513	糧	1	92C5	7CAE	糧		糧，異體字問題。
1772	藜	2	946C	4527		𠄎𠄎梨	
2539	𠄎	1	995A	43F6		月𠄎𠄎	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2699	埤	5	9A5D	5813	埤		
3804	効	2	8163	52B9	効		
3805	勅	4	8164	52C5	勅		
3814	却	2	816D	5374	却		
3843	堦	1	81AC	5826	堦		
3883	徧	1	81D4	5FA7	徧		
3940	鬱	29	824E	6B1D	鬱		
3944	氷	7	8252	6C37	氷		
3966	烟	8	8268	70DF	烟		
3992	嘗	1	82A4	751E	嘗		
4012	鼓	2	82B8	76B7	鼓		鼓，異體字問題。
4021	磻	2	82C1	78AF	磻		
4041	豎	3	82D5	7AEA	豎		
4089	脇	4	8346	8107	脇		
4090	脚	4	8347	811A	脚		
4105	菓	2	8356	83D3	菓		
4134	衛	10	8373	885E	衛		衛，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4193	鈎	1	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	26	83D2	9262	鉢		
4256	鷄	1	8450	9DC4	鷄		雞，異體字問題。
4455	劔	13	857A			僉厶刀	劍，異體字問題。
4495	麩	1	85C4	9EA8	麩		Unicode 字型呈現差異。
4515	呪	3	85D8	546A	呪		
4528	嶼	1	85E5	5D60	嶼		
4561		1	8647			𠃉 厶 厂 𠃉 比 𠃉	麤，錯字，以程式全部取代。
4768	妬	7	8779	59AC	妬		妒，異體字問題。
4775	姝	1	87A2	59B7	姝		
4856	疊	2	87F3	3CB2		疊厶毛	
5147	恠	2	89DC	6060	恠		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5300	鬪	41	8AD8	9B2A	鬪		
5548	楸	35	8C74	683F	楸		
5697	煑	3	8D6C	7151	煑		

二、手動取代表說明：

《大樓炭經》檔案中出現「●」字形，共計 81 字，比對原書後以 Unicod 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《大樓炭經》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、《大樓炭經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 277	真珠車●	𠃉王巨木𠃉	有 2 處
p. 277	車●壁一切	𠃉王巨木𠃉	有 3 處
p. 277	車●欄楯	𠃉王巨木𠃉	
p. 277	車●柱楸	𠃉王巨木𠃉	有 5 處
p. 277	車●*刀	𠃉王巨木𠃉	
p. 277	車●*樹車	𠃉王巨木𠃉	
p. 277	樹車●根莖	𠃉王巨木𠃉	

頁碼	檔案原文摘錄	原書字形	備註
p. 277	●馬璫	𠃉王巨木𠃉	
p. 278	車●馬璫	𠃉王巨木𠃉	有 18 處
p. 278	車●陸七寶枕	𠃉王巨木𠃉	
p. 279	車●馬璫	𠃉王巨木𠃉	有 2 處
p. 279	車●欄楯	𠃉王巨木𠃉	
p. 279	車●交露	𠃉王巨木𠃉	有 3 處
p. 279	車●樹者	𠃉王巨木𠃉	有 4 處
p. 279	車●根	𠃉王巨木𠃉	有 4 處
p. 279	車●陸一切	𠃉王巨木𠃉	
p. 281	●壁者一切	𠃉王巨木𠃉	
p. 281	真珠枕·車●	𠃉王巨木𠃉	
p. 281	珞車●交	𠃉王巨木𠃉	
p. 286	大呼●*喚	口𠃉弟	
p. 286	十斛四升●	篇	
p. 287	焦腹中腸●	月𠃉胃	
p. 288	●壁一切	𠃉王巨木𠃉	
p. 291	皆具●好	月𠃉庸	
p. 291	皆白●好	月𠃉庸	
p. 292	車●樹	𠃉王巨木𠃉	有 3 處
p. 292	車●葉車	𠃉王巨木𠃉	
p. 292	葉車●	𠃉王巨木𠃉	

頁碼	檔案原文摘錄	原書字形	備註
p. 292	華車●實	𠃉王巨木◻	
p. 292	車●莖	𠃉王巨木◻	
p. 292	瑚虎魄車●	𠃉王巨木◻	
p. 292	車●壁水精壁	𠃉王巨木◻	
p. 292	珊瑚車●	𠃉王巨木◻	
p. 292	珊瑚壁車●	𠃉王巨木◻	有 2 處
p. 292	車●欄楯	𠃉王巨木◻	有 2 處
p. 292	虎魄車●	𠃉王巨木◻	
p. 292	車●欄	𠃉王巨木◻	有 2 處
p. 292	虎魄樹車●	𠃉王巨木◻	
p. 293	●樹者車	𠃉王巨木◻	
p. 293	車●莖根	𠃉王巨木◻	
p. 294	車●壁者	𠃉王巨木◻	
p. 301	見●披大樹	𠃉	
p. 305	●消散微盡	漂△寸	有 2 處