

# 《越絕書》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  
2010/07/02 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 4711 的「隧」字，Unicode 編碼是 2145E，由於 XP 並不支援，仍需使用構字式「土𠂔遂」。

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：

《越絕書》使用舊版造字 27 個，字頻 107 次，這 27 個造字中，25 個可轉成 Windows XP 能顯示的字，字頻 101 次；另外 2 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《越絕書》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 4132 的「𠄎」字，是「沉」的異體字，以標準字「沉」取代。
2. 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 3834 的「嗶」字，Unicode 字型呈現為「嗶」，實際上仍為同一字。

附件一、《越絕書》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1039	稟	1	8FC2	69C0	稟		
3787	冢	48	8152	51A2	冢		
3814	却	2	816D	5374	却		
3824	叙	5	8177	53D9	叙		
3834	嗶	10	81A3	5637	嗶		Unicode 字型呈現差異。
3840	坂	2	81A9	5742	坂		
3850	姪	1	81B3	59D9	姪		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3891	慙	1	81DC	6159	慙		
3996	疎	2	82A8	758E	疎		
4011	皞	1	82B7	76A1	皞		皞，異體字問題。
4038	竈	1	82D2	7AC8	竈		
4040	竝	3	82D4	7ADD	竝		
4045	箒	1	82D9	7B92	箒		
4068	綫	1	82F0	7DAB	綫		
4080	羣	1	82FC	7FA3	羣		
4105	菓	1	8356	83D3	菓		
4140	衽	1	8379	88B5	衽		
4146	覩	6	83A1	89A9	覩		
4226	凜	2	83F1	98E1	凜		
4237	髡	1	83FC	9AE0	髡		
4242	鬪	1	8442	9B2D	鬪		
4267	鼈	4	845B	9F08	鼈		
4608	嚙	3	8676	7581	嚙		
4711	隧	1	8740	2145E		土 厶 遂	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4865	鏗	5	87FC	4976		金厶歷	
5133	憇	1	89CE	6199	憇		
5176		1	89F9			氵厶冗	沉，異體字問題。

## 二、手動取代表說明：

《越絕書》檔案中出現「●」字形，共計2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《越絕書》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、《越絕書》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 201	巫山者，越●	鬼厶扁	
p. 327	狂●通拙	悞	