

《鬼谷子》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
2010/06/10 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 2042 的「䟽」字，Unicode 編碼是 47FD，由於 XP 並不支援，仍需使用構字式「𠂔𠂔𠂔」。

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：

《鬼谷子》使用舊版造字 14 個，字頻 43 次，這 14 個造字中，12 個可轉成 Windows XP 能顯示的字，字頻 36 次；另外 2 個字必須轉成構字式，字頻 7 次。

轉碼完成製作轉碼分析表，請參考附件一《鬼谷子》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「血𠂇𠂇」字，是「眾」的異體字，以標準字「眾」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 4081 的「翺」字，Unicode 字型呈現為「翺」，實際上仍為同一字。

附件一、《鬼谷子》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2042	疏	4	9640	47FD		足𠂇𠂇	
2044	踈	3	9642	8E08	踈		
2947	聰	1	9BDA	8066	聰		
3804	効	1	8163	52B9	効		
3814	却	4	816D	5374	却		
3996	踈	4	82A8	758E	踈		
4080	羣	1	82FC	7FA3	羣		
4081	翺	1	82FD	7FFA	翺		Unicode 字型 呈現差異。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4132		5	8371			血𠂇𠂇	眾，異體字問題。
4190	邳	6	83CD	90C4	邳		
4240	鬚	2	8440	9AF4	鬚		
4242	鬪	6	8442	9B2D	鬪		
4551	嶠	3	85FC	21F0F		山𠂇間	
5860	无	2	C6D3	65E0	无		

二、手動取代表說明：

《鬼谷子》檔案中出現「●」字形，共計1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《鬼谷子》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、《鬼谷子》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 48	熒●之處安在	火𠂇惑	