

《關尹子》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 5306 的「簡」字，Unicode 編碼是 25CD1，由於 XP 並不支援，仍需使用構字式「竹𠂇間」。

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：

《關尹子》使用舊版造字 27 個，字頻 68 次，這 27 個造字中，25 個可轉成 Windows XP 能顯示的字，字頻 64 次；另外 2 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《關尹子》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 5034 的「弃」字，Unicode 字型呈現為「弃」，實際上仍為同一字。

附件一、《關尹子》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1067	碁	1	8FDE	68CA	碁		
1658	函	1	93B9	5705	函		
3047		1	9CA1			𠂇𠂇 𠂇	眾，異體字問題。
3792	涼	2	8157	51C9	涼		涼，異體字問題。
3817	廁	1	8170	53A0	廁		廁，異體字問題。
3883	徧	1	81D4	5FA7	徧		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3931	椀	1	8245	6900	椀		
3944	氷	1	8252	6C37	氷		
3963	炁	17	8265	7081	炁		
3996	疎	1	82A8	758E	疎		
4004	瘿	1	82B0	763B	瘿		
4012	鼓	8	82B8	76B7	鼓		鼓，異體字問題。
4080	羣	2	82FC	7FA3	羣		
4102	苜	2	8353	82E2	苜		
4118	藜	1	8363	8602	藜		
4132		13	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	1	8373	885E	衛		衛，異體字問題。
4146	覩	1	83A1	89A9	覩		
4242	鬪	1	8442	9B2D	鬪		
4251	鰕	2	844B	9C15	鰕		
4267	鼈	1	845B	9F08	鼈		
4515	呪	1	85D8	546A	呪		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4525	哏	1	85E2	54EF	哏		
4904	尅	1	8864	5C05	尅		
5034	弃	1	8949	5F03	弃		Unicode 字型 呈現差異。
5306	簡	3	8ADE	25CD1		竹 ㄉ 間	
5739		1	8DB8			冗 ㄉ 古	罍，異體字問 題。