

《文子》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
2010/05/24 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 4744 的「𦉑」字，Unicode 編碼是 25FC9，由於 XP 並不支援，仍需使用構字式「糸 𠂔末」。

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：

《文子》使用舊版造字 51 個，字頻 844 次，這 51 個造字中，42 個可轉成 Windows XP 能顯示的字，字頻 801 次；另外 9 個字必須轉成構字式，字頻 43 次。

轉碼完成製作轉碼分析表，請參考附件一《文子》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 3858 的「寃」字，是「冤」的異體字，以標準字「冤」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 4910 的「𡗗」字，而原書使用之字為「淫」，查詢教育部異體字字典確認「淫」並非「𡗗」之異體字，因此歸為錯字，以程式取代為正確字形「淫」。
 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 5705 的「廐」字，Unicode 字型呈現為「廐」，實際上仍為同一字。

附件一、《文子》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	𡗗	13	FBEC			𡗗	
297	𡗘	12	FBED			𡗘	
323	𡗙	1	FC48			𡗙	
528	𡗚	2	FD78			𡗚	
755	𡗛	10	FEE0			𡗛	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
766	糸	1	FEEB			糸	
841	礮	1	8E77	7932	礮		
1205	眇	4	90CB	770E	眇		
1266	砗	1	9149	784B	砗		
1565	絲	5	92F9	7DDC	絲		
1981	讐	2	95C2	8B90	讐		
2261	勑	1	977E	42E4		素厶勺	
2936	奇	1	9BCF	7AD2	奇		奇，異體字問題。
3780	儻	4	814B	5101	儻		
3789	冲	8	8154	51B2	冲		冲，異體字問題。
3814	却	1	816D	5374	却		
3858	冤	1	81BB	5BC3	冤		冤，異體字問題。
3868	卮	1	81C5	5DF5	卮		
3883	徧	1	81D4	5FA7	徧		
3891	慙	1	81DC	6159	慙		
3980	獮	1	8276	7328	獮		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3996	疎	5	82A8	758E	疎		
4008	癯	1	82B4	766F	癯		
4010	皐	3	82B6	7690	皐		皐，異體字問題。
4028	稟	1	82C8	7980	稟		
4040	竝	2	82D4	7ADD	竝		
4080	羣	13	82FC	7FA3	羣		
4089	脇	1	8346	8107	脇		
4104	莅	4	8355	8385	莅		
4124	螽	1	8369	8771	螽		
4172	蹙	1	83BB	8E77	蹙		
4184	迹	4	83C7	8FF9	迹		
4242	鬪	7	8442	9B2D	鬪		
4267	鼈	4	845B	9F08	鼈		
4727	墜	2	8750	58C4	墜		
4744	絀	2	8761	25FC9		糸 厶 末	
4768	妬	1	8779	59AC	妬		
4910		1	886A			氵 厶 畝	滛，錯字，以

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							程式全部取代。
4955	龜	1	88B9			龜	
5029	鬲	1	8944	9B32	鬲		
5034	弃	2	8949	5F03	弃		棄，異體字問題。
5072	肩	3	896F	6244	肩		
5091	鄒	1	89A4	92E3	鄒		
5176		7	89F9			汜 厶 元	沉，異體字問題。
5186	尗	1	8A44	5C17	尗		
5568	开	1	8CAA	5F00	开		
5702	厨	1	8D71	53A8	厨		
5705	廐	1	8D74	5ED0	廐		Unicode 字型呈現差異
5724	稽	1	8DA9	7A38	稽		
5752	摠	1	8DC5	63D4	摠		摠，異體字問題。
5860	无	698	C6D3	65E0	无		