

# 《列子集釋》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  
2010/06/14 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 4206 的「榆」字，Unicode 編碼是 3BD3，由於 XP 並不支援，仍需使用構字式「木ㄩ翁」。

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：

《列子集釋》使用舊版造字 59 個，字頻 134 次，這 59 個造字中，51 個可轉成 Windows XP 能顯示的字，字頻 125 次；另外 8 個字必須轉成構字式，字頻 9 次。

轉碼完成製作轉碼分析表，請參考附件一《列子集釋》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. **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 4536 的「𩶑」字，而原書使用之字為「𩶑」，「𩶑」為錯字，以程式取代為正確字形「𩶑」。
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 4091 的「𩶑」字，Unicode 字型呈現為「𩶑」，實際上仍為同一字。

附件一、《列子集釋》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
827	𩶑	1	8E69	9B8C	𩶑		
869	𩶑	1	8EB5	6A5C	𩶑		
893	𩶑	2	8ECD	5CE9	𩶑		
972	𩶑	1	8F5D	7287	𩶑		
994	𩶑	1	8F73	8E75	𩶑		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1044	儻	1	8FC7	511B	儻		
1199	眊	1	90C5	7702	眊		
1205	眊	2	90CB	770E	眊		
1226	罰	1	90E0	7F78	罰		
1266	硤	2	9149	784B	硤		
1381	穉	1	91DE	7A3A	穉		
1981	讐	4	95C2	8B90	讐		
2043	蹉	1	9641	8E01	蹉		
2103		1	967D			形之口 亏	還，異體字問題。
2651	鸚	1	99EC	9E1C	鸚		
2851		1	9B58			形穴匕 示欠	櫛，錯字，以程式全部取代。
3787	冢	2	8152	51A2	冢		
3789	冲	1	8154	51B2	冲		冲，異體字問題。
3791	况	1	8156	51B5	况		况，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3820	厦	1	8173	53A6	厦		
3830	咤	1	817D	54A4	咤		
3840	坂	1	81A9	5742	坂		
3883	徧	1	81D4	5FA7	徧		
3889	慤	2	81DA	6164	慤		
3891	慙	3	81DC	6159	慙		
3930	憵	1	8244	6901	憵		
3974	牀	1	8270	7240	牀		
3980	獮	1	8276	7328	獮		
4028	稟	1	82C8	7980	稟		
4034	穉	1	82CE	7A49	穉		
4038	竈	2	82D2	7AC8	竈		
4078		1	82FA			已合併	异，異體字問題。
4080	羣	14	82FC	7FA3	羣		
4085	耻	1	8342	803B	耻		
4091	皐	3	8348	81EF	皐		Unicode 字型呈現差異。
4132		31	8371			血会爪	眾，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
4146	覩	1	83A1	89A9	覩		
4160	適	1	83AF	8B81	適		
4184	迹	2	83C7	8FF9	迹		
4190	邳	4	83CD	90C4	邳		
4206	榆	1	83DD	3BD3		木 ㄩ 翁	
4242	鬪	3	8442	9B2D	鬪		
4267	鼈	6	845B	9F08	鼈		
4393	扌	1	84FB	39D9		扌 ㄩ 必	
4507	吳	6	85D0	5434	吳		吳，異體字問題。
4536	𪗇	1	85ED	3F10		扁 ㄩ 瓜	𪗇，錯字，以程式全部取代。
4558	厩	1	8644	53A9	厩		Unicode 字型呈現差異。
4735	壠	2	8758	58E0	壠		
4768	妬	1	8779	59AC	妬		
4820	婧	1	87CF	5AA0	婧		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4904	尅	2	8864	5C05	尅		
4953	愷	1	88B7	22835		忄厶愷	
5048	婚	1	8957	36F0		女厶昏	
5149	膏	1	89DE	2676B		要厶月	
5355	羸	1	8B50	20AE4		羸	
5479	蠅	2	8BEE	9F03		圭厶電	
5703	亂	1	8D72	9F53	亂		
5706	敘	1	8D75	654D	敘		敘，異體字問題。
5770	汧	1	8DD7			氵厶汧	

## 二、手動取代表說明：

《列子集釋》檔案中需手動修改的情況分為以下幾種：

(一)《列子集釋》檔案中出現「●」字形，共計 27 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《列子集釋》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、《列子集釋》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 12	水為●	𩺰	
p. 14	●掇	鳥𩺰句	有 2 處
p. 29	事之破●	石𩺰為	
p. 39	肌色𩺰●	𩺰	有 2 處
p. 55	●骨無	骨𩺰几	
p. 55	骨無●	石𩺰為	
p. 60	仲尼曰：「●	言𩺰醫	
p. 76	●然而封戎	卜𩺰分	
p. 77	●其所患	敕𩺰非	
p. 94	右服●騶	形 <sup>艹</sup> 𩺰与日	
p. 95	造父為御，●	形𩺰𩺰𩺰	
p. 95	●為右	刀𩺰問	
p. 100	●校四時	燔	
p. 100	大火而燔●	炳	
p. 157	●俞師曠	角𩺰𩺰	
p. 167	●其肉	牙𩺰𩺰	
p. 208	墨●	尿	
p. 208	慙●	慙	
p. 209	巧之微也。●	牙𩺰琴	
p. 209	●、	卜𩺰牙	
p. 209	●極	言𩺰蹇	

頁碼	檔案原文摘錄	原書字形	備註
p. 236	筋節●急	肉厶卷	
p. 237	心●體煩	疒厶冂	
p. 255	絕塵弭●	足厶育厶攴	
p. 270	蚊蚋●膚	口厶替	

(二)《列子集釋》檔案中，對於舊版造字以外發現的錯字，以手動修改，並製作錯字修改表，請參考附件三。欄位說明如下：

1. 檔案原文摘錄：摘錄檔案《列子集釋》中的文句段落，其中紅字底線的部分，為需要手動取代的原文。
2. 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。
3. 備註：記錄修改相同詞句的次數或其他事項。

### 附件三、《列子集釋》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
<u>谷</u> 兌厶養乎	<u>此過</u> 養乎	
<u>帛</u> 厶圭厶黽為鶉	<u>若</u> 圭厶黽為鶉	
<u>水</u> 厶蓬而指	<u>擻</u> 蓬而指	
<u>褐</u> 厶謂弟子	<u>顧</u> 謂弟子	
此過歡乎？…	<u>種有幾</u>	漏句
騮而左綠耳…	<u>右驂赤驥而左白減厶木</u>	漏句