

# 《莊子》舊版造字轉碼說明

中研院資訊所文獻處理實驗室 2007/12/24 陳建安 製作

一、《莊子》一書(莊子.xml)使用舊版造字 87 個，字頻 1105 次，詳如附件一。這 87 個造字中，66 個可轉成 Windows XP 能顯示的字，字頻 227 次；另外 21 個字必須轉成構字式，字頻 878 次。

二、附件一的造字分析表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字

丙、字頻(txt)：造字在「.txt」文件的出現次數

丁、字頻(xml)：造字在「.xml」文件的出現次數

戊、Big5：造字的 Big5 碼

己、Unicode：造字所對應的 Unicode 碼

庚、WinXP：造字在 Windows XP 的對應字形

辛、備註凡例：

1、校對問題，舊版漢籍錯字，可用程式全部取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中；若該造字的所有頻次，皆屬於錯誤使用的錯字情形，可以用程式全部取代為正確字形。如編號 2191 的「金 $\Delta$ 戊」字，原字為「鉞」。

2、異體字問題：新版漢籍考量到使用者檢索及使用時的便利性，將用字原則改為除專詞等特殊情形之外，一律改用標準字呈現。如編號 4132 的「血 $\Delta$ 丞」係「眾」字之異體，故以「眾」字取代。

3、Unicode 字型呈現差異：Unicode 字型與舊漢籍造字有些微差異，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。

三、Unicode 目前收錄的漢字總數為 70194，分屬於三個不同區段，詳如表一。目前 Windows XP 只支援 CJK 認同表意文字區的 20902 個字，內碼為 4E00-9FFF。所以造字編 4228 號的「瀟」字，Unicode 編碼為 24162 由於 Windows XP 並不支援，仍須使用構字式「 $\Delta$ 瀟」。

表一、Unicode 的字數及編碼區段

Unicode	字集子集合	新增字數	新增編碼區段	總字數	WinXP
1.1 版	CJK 認同表意文字區	20902	4E00-9FFF	20902	支援
3.0 版	CJK 認同表意文字擴充 A 區	6582	3400-4DFF	27484	不支援
3.1 版	CJK 認同表意文字擴充 B 區	42710	20000-2A6D6	70194	不支援

四、《莊子》中未確定字形「●」共 37 處，詳細處理內容可參考附件二的未確定字形手動取代表。附件二的手動取代表說明如下：

- 甲、檔案原文摘錄：摘錄檔案(莊子.xml)中包含未確定字形「●」的文句段落，其中紅字底線的部分，為需要手動取代的字形。
- 乙、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。

#### 附件一、《莊子》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
826	癱	2	2	8E68	7655	癱		
949	柎	1	1	8F46	678F	柎		
982	胷	4	4	8F67	80F7	胷		
994	楚	8	8	8F73	8E75	楚		
1024	虵	1	1	8FB3	8675	虵		
1040	聃	45	45	8FC3	803C	聃		
1043	頰	1	1	8FC6	2945E		𠃉𠃉頁	
1044	儻	2	2	8FC7	511B	儻		
1077	稟	1	1	8FE8	7A3E	稟		

1099	睥	1	1	8FFE	66A4	睥		
1133	瓌	1	1	9061	74CC	瓌		
1383	穰	2	2	91E0	7A45	穰		
1646	脗	1	1	93AD	8117	脗		
2050	蹠	1	1	9648	8E4F	蹠		
2191	鉞	1	1	96F7			金 𠄎 戍	鉞，舊漢籍錯字，可用程式全部取代。
2722	吡	1	1	9A74	5421	吡		
2728	妲	2	2	9A7A	59B8	妲		
2768		1	1	9AC4			𦰩 𠄎 千	竿，舊漢籍錯字，可用程式全部取代。
2783	蚤	3	3	9AD3	8689	蚤		
2785	𪗇	1	1	9AD5	8C40	𪗇		
2811		2	2	9AEF			𠄎 𠄎 田	畱，舊漢籍錯字，可用程式全部取代。
2817	𦰩	4	4	9AF5	3AC4		𦰩	備，舊漢籍錯字，可用程式全部取代。
2822	愜	1	1	9AFA	6051	愜		
3787	豕	1	1	8152	51A2	豕		
3799	刼	1	1	815E	5226	刼		
3866	崐	1	1	81C3	5D10	崐		

3868	卮	4	4	81C5	5DF5	卮		
3869	币	1	1	81C6	5E00	币		
3883	徧	6	6	81D4	5FA7	徧		
3891	慙	2	2	81DC	6159	慙		
3929	黎	2	2	8243	68C3	黎		
3974	牀	1	1	8270	7240	牀		
3976	犁	1	1	8272	7282	犁		
3980	獾	5	5	8276	7328	獾		
4014	蓋	1	1	82BA	8462	蓋		
4025	礪	2	2	82C5	7934	礪		
4034	穉	1	1	82CE	7A49	穉		
4038	竈	3	3	82D2	7AC8	竈		
4055	粃	1	1	82E3	7C83	粃		
4080	羣	15	15	82FC	7FA3	羣		
4081	翱	1	1	82FD	7FFA	翱		Unicode 呈現差異。
4089	脇	1	1	8346	8107	脇		
4104	莅	1	1	8355	8385	莅		
4107	葱	1	1	8358	8471	葱		
4119		2	2	8364			𠃉 <sup>正頁已</sup> 久 <sup>久</sup>	變，異體字問題。

4132		56	56	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	18	18	8373	885E	衛		
4140	衽	1	1	8379	88B5	衽		
4146	覩	13	13	83A1	89A9	覩		
4160	適	1	1	83AF	8B81	適		
4184	迹	20	20	83C7	8FF9	迹		
4226	凜	1	1	83F1	98E1	凜		
4227	飡	1	1	83F2	98E6	飡		
4228	滌	1	1	83F3	24162		灬 𠂇 滌	
4238	髭	1	1	83FD	9AE2	髭		
4239	髯	1	1	83FE	9AE5	髯		
4242	鬪	8	8	8442	9B2D	鬪		
4250	鱈	6	6	844A	9C0C	鱈		
4265	鼃	5	5	8459	9F03	鼃		
4267	鼈	4	4	845B	9F08	鼈		
4455	劔	1	1	857A			僉 𠂇 刀	
4458	勑	1	1	857D	52CC	勑		Unicode 呈現差異。
4562	逕	1	1	8648	28563		形 辵 四 午 𠂇	
4578		2	2	8658			形 弗 欠 非 𠂇	形 弗 欠 非 𠂇，舊漢籍錯字，可用

								程式全部取代。
4588	鮫	1	1	8662	4C54		魚厶攸	
4607	狫	1	1	8675	3E60		豸厶屯	
4817	寶	2	2	87CC	5BF3	寶		
4924		1	1	8878			冗厶曾	晉，舊漢籍錯字，可用程式全部取代。
4987	躡	1	1	88D9	4831		足厶屬	
5072	扃	1	1	896F	6244	扃		
5091	鄒	1	1	89A4	92E3	鄒		
5108	惛	2	2	89B5	60FD	惛		
5112	园	1	1	89B9	56ED	园		
5182		1	1	8A40			馬厶衣	𠂔𠂔申𠂔𠂔，舊漢籍錯字，可用程式全部取代。
5244	夔	1	1	8A7E	3738		薛厶女	
5261	挽	1	1	8AB1	639C	挽		
5302	昏	1	1	8ADA	662C	昏		
5316	晋	1	1	8AE8	664B	晋		晉，異體字問題。
5335	蚩	2	2	8AFB	459F		亡厶虫	
5371	冂	1	1	8B60	808E	冂		
5437	并	1	1	8BC4	5E77	并		

5470	棟	1	1	8BE5	68C5	棟		
5508	髻	1	1	8C4C	9B10	髻		
5585	寓	2	2	8CBB	3762		宀 厶 禹	
5739		3	3	8DB8			冗 厶 古	罍，舊漢籍錯字，可用程式全部取代。
5753	潜	2	2	8DC6	6F5C	潜		
5860	无	793	793	C6D3	65E0		无	

### 附件二、《莊子》「●」字形手動取代表

檔案原文摘錄	手動取代結果
慮嘆變●	慮嘆變執 厶 心
●然疲役而不知其所歸	尔 然疲役而不知其所歸
甕●犬癭說齊桓公	甕 央 厶 瓦 犬癭說齊桓公
相●以沫	相 形 冫 由 冫 以沫
跼●而鑑於井	跼 足 厶 鮮 而鑑於井
以死為決●潰癰	以死為決 疒 厶 丸 潰癰
以羈●	以羈 辮
蚊虻●膚	蚊虻 囁 膚
其知慳於●蠶之尾	其知慳於 厶 厶 蠶 蠶之尾
●淳散朴	冫 厶 冫 泉 淳散朴
吾以一足●踔而行	吾以一足 踔 踔而行
得水則為●	得水則為 豷
不●，久竹生青寧	不 算，久竹生青寧
以黃金注者●	以黃金注者 夕 厶 昏
吾將三月●汝	吾將三月 生 厶 豨 汝
山有●	山有 夔
无●居	无 雷 居
<qd>●</qd>	<qd> 蚤 厶 更 </qd>
●</qa>斛不敢入	缺 </qa>斛不敢入
●斛不敢入於四竟	缺 斛不敢入於四竟
墮其天●	墮其天 弋

檔案原文摘錄	手動取代結果
故以是非相●也	故以是非相[形]弟女非[口]也
日中穴●	日中穴[阨]
●內者	[券]內者
●外者	[券]外者
鏌●為下	鏌[鄒]為下
有生，●也	有生，[臧]也
●，見巧乎王	[形]扌又爪[口]，見巧乎王
雞●也	雞[广]△[雍]也
皆●可以休老	皆[女]△[威]可以休老
吾嘗食於十●	吾嘗食於十[將]△[食]
而五●先饋	而五[將]△[食]先饋
而●其所	而[形]弟女非[口]其所
夫●人特為食羹之貨	夫[將]△[食]人特為食羹之貨
子為●粉夫	子為[形]弟女非[口]粉夫
吾葬具豈不●邪	吾葬具豈不[備]邪
而不免於●斷	而不免於[魀]斷