

《莊子集解內篇補正》舊版造字轉碼說明

中研院資訊所文獻處理實驗室 2007/12/18 陳建安 製作

一、《莊子集解內篇補正》一書(莊子集解內篇補正.xml)使用舊版造字 89 個，字頻 336 次，詳如附件一。這 89 個造字中，74 個可轉成 Windows XP 能顯示的字，字頻 522 次；另外 15 個字必須轉成構字式，字頻 461 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 2734 的「出」字，原字為「恠」。

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「血𩇑𩇑」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 4158 的「譌」字，Unicode 字型呈現為「譌」，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。所以造字編號 4607 的「狫」字，Unicode 編碼為 3E60 由於 Windows XP 並不支援，仍須使用構字式「犭厶屯」。

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、《莊子集解內篇補正》中未確定字形「●」共 51 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(莊子集解內篇補正.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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。
- 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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有修改後結果。

附件一、《莊子集解內篇補正》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
893	𦵑	5	5	8ECD	5CE9	𦵑		
932	𦵒	2	2	8EF4	7683	𦵒		
994	𦵓	8	8	8F73	8E75	𦵓		
1040	𦵔	17	17	8FC3	803C	𦵔		
1383	𦵕	8	8	91E0	7A45	𦵕		
1565	𦵖	2	2	92F9	7DDC	𦵖		
1646	𦵗	5	5	93AD	8117	𦵗		
2734	出	5	5	9AA2	5C80	出		𦵗，舊漢籍錯字，可用程式全部取代。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2771	縹	1	1	9AC7			纟 𠂔 帶	裏，舊漢籍錯字，可用程式全部取代。
2809		3	3	9AED			食 𠂔 益	欠，舊漢籍錯字，可用程式全部取代。
2851		1	1	9B58			𠂔 穴 匕 示 欠 𠂔	櫟，舊漢籍錯字，可用程式全部取代。
2875	𡗗	6	6	9B70	4EBC	𡗗		蚤，舊漢籍錯字，可用程式全部取代。
3789	冲	1	1	8154	51B2	冲		
3805	勅	1	1	8164	52C5	勅		
3814	却	1	1	816D	5374	却		
3849	獎	1	1	81B2	596C	獎		
3858	寃	1	1	81BB	5BC3	寃		
3866	崐	2	2	81C3	5D10	崐		
3868	卮	1	1	81C5	5DF5	卮		
3869	帀	1	1	81C6	5E00	帀		
3883	徧	2	2	81D4	5FA7	徧		
3889	慤	2	2	81DA	6164	慤		
3891	慤	1	1	81DC	6159	慤		
3916	敷	1	1	81F5	656D	敷		𡗗，舊漢籍錯字，可用程式全部取代。
3966	烟	3	3	8268	70DF	烟		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3969	煊	1	1	826B	714A	煊		
3974	牀	3	3	8270	7240	牀		
3976	犁	2	2	8272	7282	犁		
3980	獾	10	10	8276	7328	獾		
3984	瑠	1	1	827A	7460	瑠		
4010	皐	1	1	82B6	7690	皐		
4011	皐	4	4	82B7	76A1	皐		
4016	着	10	10	82BC	7740	着		
4025	礪	9	9	82C5	7934	礪		
4028	稟	1	1	82C8	7980	稟		
4030	秆	3	3	82CA	79C6	秆		
4038	竈	1	1	82D2	7AC8	竈		
4055	粃	10	10	82E3	7C83	粃		
4068	綫	2	2	82F0	7DAB	綫		
4070	縹	2	2	82F2	7DE4	縹		
4080	羣	23	23	82FC	7FA3	羣		
4081	翱	5	5	82FD	7FFA	翱		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4089	脇	2	2	8346	8107	脇		
4090	脚	2	2	8347	811A	脚		
4091	臯	1	1	8348	81EF	臯		
4114	藥	3	3	835F	8616	藥		
4132		106	106	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	32	32	8373	885E	衛		
4146	覩	3	3	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4160	適	4	4	83AF	8B81	適		
4184	迹	24	24	83C7	8FF9	迹		
4226	飡	3	3	83F1	98E1	飡		
4242	鬪	21	21	8442	9B2D	鬪		
4250	鰭	10	10	844A	9C0C	鰭		
4256	鷄	14	14	8450	9DC4	鷄		
4266	鼃	1	1	845A	9F02	鼃		
4269	鼃	1	1	845D	9F39	鼃		
4286	从	6	6	846E	4ECE	从		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4458	勅	1	1	857D	52CC	勅		Unicode 呈現差異。
4460	勅	1	1	85A1	52D1	勅		
4489		1	1	85BE			口𠂔𠂔	号，舊漢籍錯字，可用程式全部取代。
4496	段	9	9	85C5	53DA	段		
4507	吴	8	8	85D0	5434	吴		
4546	鷗	1	1	85F7	9D76	鷗		
4607	狔	4	4	8675	3E60		狔 𡗗 屯	
4768	妬	1	1	8779	59AC	妬		
4839	昱	1	1	87E2	6637	昱		
4870	壘	1	1	8842	97F2	壘		
5013	汙	1	1	88F3			汙	
5034	弃	2	2	8949	5F03	弃		
5091	鄒	1	1	89A4	92E3	鄒		
5112	园	4	4	89B9	56ED	园		
5131	窻	1	1	89CC	7ABB	窻		
5133	憙	1	1	89CE	6199	憙		
5162	嶺	1	1	89EB	37EA		山 𡿨 畏	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5200		1	1	8A52			𠂔ノ門米大[・]	奧，異體字問題。
5232	擊	1	1	8A72	6480	擊		擊，舊漢籍錯字，可用程式全部取代。
5250		1	1	8AA6			𢲗𢲗手	攀，舊漢籍錯字，可用程式全部取代。
5309		1	1	8AE1			日𠂔旬	昫，舊漢籍錯字，可用程式全部取代。
5335	𧈧	3	3	8AFB	459F		亡𢲗虫	
5351	𧈧	1	1	8B4C	80AC	𧈧		
5442	𧈧	1	1	8BC9	67BF	𧈧		
5546	𧈧	4	4	8C72	6A36	𧈧		
5706	𧈧	4	4	8D75	654D	𧈧		
5773	𧈧	1	1	8DDA	6FDA	𧈧		
5845		2	2	C6C4				𠂔，舊漢籍錯字，可用程式全部取代。
5846		1	1	C6C5				𠂔，舊漢籍錯字，可用程式全部取代。
5860	无	334	334	C6D3	65E0		无	

附件二、《莊子集解內篇補正》未確定字形手動取代表

檔案原文摘錄	手動取代結果
「培」當為「培」之●	「培」當為「培」之誤
●淳散朴	汙 𡵓 𡵓 淳散朴
●也。●，由救反	𠂔 𡵓 由也。𠂔 𡵓 由，由救反
塊，●或體	塊，𠂔或體
从●者今文	从𠂔者今文
見其成功，●然	見其成功，𠂔 𡵓 然
盧文弨云：「●，當作	盧文弨云：「𠂔 𡵓，當作
●，釋文「乃結	𠂔 𡵓，釋文「乃結
刃若新●研	刃若新磨 𡵓 研
熒，●之借字	熒，𠂔 𡵓 之借字
●，惑也	𠂔 𡵓，惑也
心●體煩	心𠂔 𡵓 體煩
蚤音文。本或作●	蚤音文。本或作𠂔 𡵓
眴、恂，並●之	眴、恂，並𠂔之
●，驚辭也	𠂔，驚辭也
甕●大癭說齊桓公	甕𠂔 𡵓 大癭說齊桓公

檔案原文摘錄	手動取代結果
甕●，大癭貌	甕 <u>央△瓦</u> ，大癭貌
●，烏葬反	<u>央△瓦</u> ，烏葬反
別入跟中，上●內	別入跟中，上 <u>𩇛</u> 內
●萬物而不為義	<u>形𠂔次韭</u> ●萬物而不為義
●即嚴殺之義	<u>形𠂔次韭</u> ●即嚴殺之義
「與」當是●之借字	「與」當是 <u>走△與</u> 之借字
●，安行也	<u>走△與</u> ，安行也
李謂為●借	李謂為 <u>走△與</u> 借
相●以沫	相 <u>形𠂔𠂔𠂔</u> 以沫
●，本又作濡	<u>形𠂔𠂔𠂔</u> ，本又作濡
是與欽●殺祖江	是與欽 <u>丕△鳥</u> 殺祖江
古駮、●本一字	古駮、 <u>丕△鳥</u> 本一字
</qn>𨔵●	</qn>𨔵 <u>足△鮮</u>
𨔵●，曳疾貌	𨔵 <u>足△鮮</u> ，曳疾貌
●，悉田反	<u>足△鮮</u> ，悉田反
以死為決●潰癰	以死為決 <u>疒△丸</u> 潰癰
●，胡亂反	<u>疒△丸</u> ，胡亂反
若●癰決潰	若 <u>疒△丸</u> 癰決潰

檔案原文摘錄	手動取代結果
<u>𠂔</u> 萬物而不為義	<u>形</u> 𠂔 <u>次</u> 𠂔萬物而不為義
<u>𠂔</u> ，碎也	<u>形</u> 𠂔 <u>次</u> 𠂔，碎也
說文作 <u>𠂔</u> ，亦作𠂔。隸省作 <u>𠂔</u>	說文作 <u>形</u> 𠂔 <u>次</u> 𠂔，亦作𠂔。隸省作 <u>形</u> 𠂔 <u>次</u> 𠂔
<u>𠂔</u> ，子兮反	<u>形</u> 𠂔 <u>次</u> 𠂔，子兮反
<u>𠂔</u> 萬物	<u>形</u> 𠂔 <u>次</u> 𠂔萬物
雞 <u>𠂔</u> 也	雞 <u>广</u> 𠂔也
致肌色𠂔 <u>𠂔</u>	致肌色𠂔 <u>𠂔</u>
<u>𠂔</u> ，說文作 <u>𠂔</u>	<u>山</u> 𠂔 <u>罪</u> ，說文作 <u>山</u> 𠂔 <u>皇</u>
故以 <u>𠂔</u> 乎形容之	故以 <u>山</u> 𠂔 <u>罪</u> 乎形容之
以 <u>𠂔</u> 乎形容之	以 <u>山</u> 𠂔 <u>罪</u> 乎形容之
<u>𠂔</u> ，山貌	<u>山</u> 𠂔 <u>罪</u> ，山貌
如作「 <u>𠂔</u> 」，則無所取義矣	如作「 <u>山</u> 𠂔 <u>罪</u> 」，則無所取義矣
正字通弟作 <u>𠂔</u>	正字通弟作 <u>𠂔</u>
<u>𠂔</u> 然而封戎	<u>𠂔</u> 𠂔 <u>分</u> 然而封戎