

《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

中研院史語所漢籍電子文獻工作小組

2008/1/31 陳建安 製作

一、《列子集釋》一書(列子集釋.xml)使用舊版造字 205 個，字頻 1113 次，詳如附件一。這 205 個造字中，166 個可轉成 Windows XP 能顯示的字，字頻 882 次；另外 39 個字必須轉成構字式，字頻 231 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 1161 的「田~~畝~~辰」字，原字為「晨」。
- 2、舊版漢籍錯字，必須人工取代：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，然而在所有頻次中，只有一部分是錯字，另一部分是正確的，因此無法全部取代，必須人工一一比對，將錯誤的部分更正，如編號 5520「~~毋~~」，共出現 6 次，但原書字形有 1 次為「母」字，必須手動取代。
- 3、重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代：舊漢籍造字與構字式部件外字，皆是建立在 Big5 的使用者造字區中，兩者所使用的碼區是重疊的，在同時使用構字式和舊漢籍造字的新版漢籍中，這些重疊的字碼可能被用以表達兩個不同字形，我們稱之為「重碼問題」。在使用構字式的 xml 檔中，舊漢籍造字

也會呈現為重碼的部件外字，因而必須將採用舊漢籍造字部分以對應的 Unicode 或構字式取代，才能呈現正確字形；至於採用部件外字者則不需轉碼。如編號 5048，字碼 8957，對應的舊漢籍造字為「女𠂔昏」，部件外字為「𠂔」，一一比對原書，確認是「女𠂔昏」者，再以手動取代。又編號 4910，字碼 886A，對應的舊漢籍造字為「𠂔𠂔𠂔」，部件外字為「𠂔」；「𠂔𠂔𠂔」為舊版漢籍錯字，原書字形作「淫」，為重碼問題又兼有舊版漢籍錯字者，一一比對原書，將採用舊漢籍造字的部分手動改為「淫」，採用部件外字者則不需轉碼。

- 4、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「血𠂔𠂔」係「眾」字之異體，故以「眾」字取代。
- 5、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。
- 6、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 5309 的「日𠂔𠂔」。

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 790 的「映」字，Unicode 編碼為 4039，由於 Windows XP 並不支援，仍須使用構字式「目𠂔夾」。

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、txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，詳細取代內容可參考附件二的手動取代表。附件二的手動取代表欄位說明如下：

- 甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- 乙、取代：實際使用字形與漢籍舊版造字不同時，於此欄列出。
- 丙、檔案原文摘錄：摘錄檔案(列子集釋.xml)中包含該造字的文句段落，其中紅字底線的部分，為需要手動取代的原文。
- 丁、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。

五、《列子集釋》中未確定字形「」共 12 處，詳細處理內容可參考附件三的未確定字形「」手動取代表。附件三的未確定字形「」手動取代表欄位說明如下：

- 甲、檔案原文摘錄：摘錄檔案(列子集釋.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，不一一列舉所有原文。
- 丙、備註：註明其他備註事項或相同文句在檔案中出現的次數，未註明者表示只出現一次。

附件一、《列子集釋》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
790	𦉑	12	12	8E44	4039		目 𦉑 夾	
810	𦉑	1	1	8E58	846A	𦉑		Unicode 呈現差異。
811	𦉑	1	1	8E59	6178	𦉑		
827	𦉑	2	2	8E69	9B8C	𦉑		
829	𦉑	1	1	8E6B	81F0	𦉑		
869	𦉑	12	12	8EB5	6A5C	𦉑		Unicode 呈現差異。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
893	𪗇	2	2	8ECD	5CE9	𪗇		
917	𪗈	4	4	8EE5	4CD8		鳥𪗈我	
959	𪗉	5	5	8F50	3A49		扌𪗉翕	
972	𪗊	2	2	8F5D	7287	𪗊		
994	𪗋	1	1	8F73	8E75	𪗋		
1013	𪗌	1	1	8FA8	7D5A	𪗌		
1024	𪗍	3	3	8FB3	8675	𪗍		
1044	𪗎	2	2	8FC7	511B	𪗎		
1056	𪗏	4	4	8FD3	9641	𪗏		
1067	𪗐	2	2	8FDE	68CA	𪗐		
1097	𪗑	1	1	8FFC	37ED		山𪗑昏	
1161	𪗒	1	1	907D			田𪗒辰	晨，校對問題，舊版漢籍錯字，可用程式全部取代。
1199	𪗓	1	1	90C5	7702	𪗓		
1205	𪗔	10	10	90CB	770E	𪗔		
1226	𪗕	1	1	90E0	7F78	𪗕		
1266	𪗖	14	14	9149	784B	𪗖		
1381	𪗗	4	4	91DE	7A3A	𪗗		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1416	埃	2	2	9242	7AE2	埃		
1565	繇	1	1	92F9	7DDC	繇		
1642	脉	5	5	93A9	8109	脉		
1646	脗	1	1	93AD	8117	脗		
1718	落	7	7	93F5	83ED	落		
1871	𧈧	2	2	94F1	87C1	𧈧		
1885	逢	2	2	9540	882D	逢		
1893	袞	3	3	9548	886E	袞		
1981	讐	6	6	95C2	8B90	讐		
2042	䟽	2	2	9640	47FD		足𠂔充	
2043	䟽	2	2	9641	8E01	䟽		
2044	踈	1	1	9642	8E08	踈		
2103		6	6	967D			𠂔𠂔𠂔𠂔𠂔𠂔	還，校對問題，舊版漢籍錯字，可用程式全部取代。
2327	鋒	1	1	97E2	93E0	鋒		
2448	隼	1	1	98BE	96BD	隼		
2505	𪗇	1	1	98F7	98DC	𪗇		
2651	鸛	3	3	99EC	9E1C	鸛		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2718	僞	1	1	9A70	5097	僞		茲，校對問題，舊版漢籍錯字，可用程式全部取代。
2724	圯	2	2	9A76	5724	圯		愿，校對問題，舊版漢籍錯字，可用程式全部取代。
2744	𢇛	2	2	9AAC	6540	𢇛		夸，校對問題，舊版漢籍錯字，可用程式全部取代。
2745		1	1	9AAD			日𠂔寅	𠂔，校對問題，舊版漢籍錯字，可用程式全部取代。
2751	𣎵	1	1	9AB3	693E	𣎵		炭，校對問題，舊版漢籍錯字，可用程式全部取代。
2766	𡗗	1	1	9AC2	4ECC	𡗗		𠂔，校對問題，舊版漢籍錯字，可用程式全部取代。
2790	鉶	2	2	9ADA	91F6	鉶		麪，校對問題，舊版漢籍錯字，可用程式全部取代。
2799		1	1	9AE3			金𠂔森	𠂔，校對問題，舊版漢籍錯字，可用程式全部取代。
2801	𥓐	3	3	9AE5	947A	𥓐		𠂔，校對問題，舊版漢籍錯字，可用程式全部取代。
2811		1	1	9AEF			𠂔𠂔田	但，校對問題，舊版漢籍錯字，可用程式全部取代。
2819	及	5	5	9AF7	53CA	及		𠂔，校對問題，舊版漢籍錯字，可用程式全部取代。
2826	巡	1	1	9AFE	5EF5	巡		爛，校對問題，舊版漢籍錯字，可用

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
								程式全部取代。
2835	梯	2	2	9B48	632E	梯		忙，校對問題，舊版漢籍錯字，可用程式全部取代。
2851		6	6	9B58			𠂔穴匕示欠	櫟，校對問題，舊版漢籍錯字，可用程式全部取代。
2858	漚	2	2	9B5F	6EFA	漚		碩，校對問題，舊版漢籍錯字，可用程式全部取代。
3787	豕	10	10	8152	51A2	豕		
3789	冲	19	19	8154	51B2	冲		
3791	况	20	20	8156	51B5	况		
3792	凉	4	4	8157	51C9	凉		
3794	凑	3	3	8159	51D1	凑		
3801	剗	1	1	8160	5257	剗		
3804	効	3	3	8163	52B9	効		
3805	勅	1	1	8164	52C5	勅		
3806	勳	2	2	8165	52E0	勳		
3812	𠂔	1	1	816B	20BCC		𠂔	
3814	却	18	18	816D	5374	却		
3816	厓	1	1	816F	5393	厓		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3820	厦	1	1	8173	53A6	厦		
3824	叙	4	4	8177	53D9	叙		
3825	疊	4	4	8178	53E0	疊		疊，異體字問題。
3830	咤	2	2	817D	54A4	咤		
3840	坂	3	3	81A9	5742	坂		
3846	壻	1	1	81AF	58FB	壻		Unicode 呈現差異。
3849	獎	2	2	81B2	596C	獎		
3858	寃	2	2	81BB	5BC3	寃		
3864	峯	3	3	81C1	5CEF	峯		
3868	卮	1	1	81C5	5DF5	卮		
3883	徧	9	9	81D4	5FA7	徧		
3886	悞	1	1	81D7	609E	悞		
3889	慤	4	4	81DA	6164	慤		
3891	慙	4	4	81DC	6159	慙		
3910	擘	1	1	81EF	64E5	擘		
3921	皙	3	3	81FA	6673	皙		
3925	𠂔	1	1	81FE	3775		𠂔	
3929	黎	2	2	8243	68C3	黎		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3930	攄	1	1	8244	6901	攄		
3935	槩	1	1	8249	69E9	槩		Unicode 呈現差異。
3949	洩	6	6	8257	6D96	洩		
3955	澗	1	1	825D	3D4E		澗 𣶒	
3963	炁	1	1	8265	7081	炁		
3967	炳	8	8	8269	712B	炳		
3969	煊	12	12	826B	714A	煊		
3974	牀	5	5	8270	7240	牀		
3977	狗	1	1	8273	72E5	狗		
3978	猪	2	2	8274	732A	猪		Unicode 呈現差異。
3980	獾	2	2	8276	7328	獾		
3996	疎	4	4	82A8	758E	疎		
4004	瘠	1	1	82B0	763B	瘠		
4010	皐	1	1	82B6	7690	皐		
4014	蓋	1	1	82BA	8462	蓋		
4028	稟	3	3	82C8	7980	稟		
4034	穉	3	3	82CE	7A49	穉		
4038	竈	12	12	82D2	7AC8	竈		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4040	竝	7	7	82D4	7ADD	竝		
4041	豎	7	7	82D5	7AEA	豎		
4042	笋	1	1	82D6	7B0B	笋		
4051	簌	3	3	82DF	7C35	簌		
4055	粃	2	2	82E3	7C83	粃		
4056	葬	1	1	82E4	7CA6	葬		
4078		3	3	82FA			已𠂇升	异，異體字問題。
4080	羣	103	103	82FC	7FA3	羣		Unicode 呈現差異。
4085	耻	1	1	8342	803B	耻		
4090	脚	3	3	8347	811A	脚		
4091	皐	15	15	8348	81EF	皐		Unicode 呈現差異。
4102	苣	2	2	8353	82E2	苣		
4107	葱	1	1	8358	8471	葱		
4132		108	108	8371			血𠂇𠂇	眾，異體字問題。
4138	袞	1	1	8377	88A0	袞		
4139	袴	5	5	8378	88B4	袴		
4146	覩	4	4	83A1	89A9	覩		
4158	譌	26	26	83AD	8B4C	譌		Unicode 呈現差異。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4160	適	3	3	83AF	8B81	適		
4172	蹙	6	6	83BB	8E77	蹙		Unicode 呈現差異。
4173	耽	1	1	83BC	8EAD	耽		
4184	迹	32	32	83C7	8FF9	迹		
4190	邳	12	12	83CD	90C4	邳		
4193	鈎	3	3	83D0	920E	鈎		
4206	檣	4	4	83DD	3BD3		木𠂇翕	
4235	馮	18	18	83FA	9AA9	馮		
4240	髯	5	5	8440	9AF4	髯		
4242	鬪	11	11	8442	9B2D	鬪		Unicode 呈現差異。
4252	鰓	2	2	844C			魚𩺰眾	鰓，異體字問題。
4256	鷄	2	2	8450	9DC4	鷄		
4262	麪	1	1	8456	9EAA	麪		
4263	麴	1	1	8457	9EAF	麴		
4265	鼃	3	3	8459	9F03	鼃		
4266	鼃	28	28	845A	9F02	鼃		
4267	鼃	12	12	845B	9F08	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
4299	炆	2	2	847B	242F2		火𠂇𠂇	
4316	艷	2	2	84AE	8276	艷		
4376	僂	3	3	84EA	510D	僂		
4393	𢇛	5	5	84FB	39D9		扌𠂇必	
4460	勅	2	2	85A1	52D1	勅		
4469	勾	1	1	85AA	5304	勾		
4488	惡	1	1	85BD	6076	惡		惡，異體字問題。
4496	段	11	11	85C5	53DA	段		
4507	吳	9	9	85D0	5434	吳		
4536	𪗇	3	3	85ED	3F10		扁𠂇瓜	𪗇，校對問題，舊版漢籍錯字，可用程式全部取代。
4558	𪗈	3	3	8644	53A9	𪗈		Unicode 呈現差異。
4561		1	1	8647			𪗈 ㄣ 厂 𠂇 比 𠂇	𪗈，校對問題，舊版漢籍錯字，可用程式全部取代。
4594	𪗉	1	1	8668	2A0E9		秋𠂇鳥	
4637	𪗊	1	1	86B5	65F6	𪗊		時，異體字問題。
4735	𪗋	5	5	8758	58E0	𪗋		
4737	𪗌	1	1	875A	6082	𪗌		
4768	𪗍	2	2	8779	59AC	𪗍		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4778	鞞	2	2	87A5	97B4	鞞		
4785	贗	2	2	87AC	8D0B	贗		
4820	婧	3	3	87CF	5AA0	婧		
4866	鼃	1	1	87FD	9F04	鼃		
4870	壑	1	1	8842	97F2	壑		
4878	宄	1	1	884A	5B82	宄		Unicode 呈現差異。
4902	兗	2	2	8862	5156	兗		
4904	尅	3	3	8864	5C05	尅		
4910		4	9	886A			汙 𡗗 𡗘	與𡗗重碼，重碼問題，部分採用舊漢籍造字，部分採用部件外字，且有校對問題，為舊版漢籍錯字，必須人工取代為「淫」。
4913	塿	2	2	886D	58A2	塿		
4935		2	2	88A5			𠂔 ⁺⁺ 汙 𡗗	⁺⁺ 𡗗 汙，校對問題，舊版漢籍錯字，可用程式全部取代。
4940	韵	16	16	88AA	97F5	韵		
4946	𦍋	1	1	88B0	5279	𦍋		
4953	𡗗	2	2	88B7	22835		𡗗 𡗘	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4980	𤇀	1	1	88D2			𤇀𠂇火	
4981	𤇁	1	1	88D3	9CEC	𤇁		
4996	𤇂	1	1	88E2	65B5	𤇂		
5012	𤇃	1	1	88F2	36D0		女𠂇更	
5048	婚	1	16	8957	36F0		女𠂇昏	與𠂇重碼，重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代。
5149	𤇄	3	3	89DE	2676B		要𠂇月	
5200		3	3	8A52			𠂇ノ門米大𠂇	奧，異體字問題。
5220	𤇅	6	6	8A66	6BCC	𤇅		𤇅，校對問題，舊版漢籍錯字，必須人工取代。
5244	𤇆	1	1	8A7E	3738		薛𠂇女	
5302	昏	4	4	8ADA	662C	昏		
5309		1	1	8AE1			日𠂇旬	待造字。
5310	𤇇	1	1	8AE2	8204	𤇇		
5334	𤇈	2	3	8AFA	6692	𤇈		與𠂇重碼，重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代。
5355	𤇉	5	5	8B50	20AE4		𤇉	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5381	脛	3	3	8B6A	8103	脛		Unicode 呈現差異。
5479	蠅	5	5	8BEE	9F03		圭厶𧈧	
5483	眊	13	13	8BF2	7721	眊		
5509	駢	2	2	8C4D	9A00	駢		
5540	檠	1	1	8C6C	6A53	檠		
5552	弃	1	1	8C78	5F06	弃		
5555	迥	1	1	8C7B	9008	迥		
5558	麓	1	1	8C7E	9E84	麓		
5564	槩	2	2	8CA6	6ABF	槩		
5583	殮	4	4	8CB9	98F1	殮		
5587	歛	1	1	8CBD	6B58	歛		
5588	恠	1	1	8CBE	6061	恠		
5697	𪗇	1	1	8D6C	7151	𪗇		Unicode 呈現差異。
5703	𪗈	3	3	8D72	9F53	𪗈		
5706	𪗉	31	31	8D75	654D	𪗉		
5742	澁	1	1	8DBB	6F81	澁		
5752	摠	1	1	8DC5	63D4	摠		
5770	汧	2	2	8DD7			汧 厶𧈧	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5852		2	2	C6CB				宀，校對問題，舊版漢籍錯字，可用程式全部取代。
5860	无	86	86	C6D3	65E0	无		
5861		1	1	C6D4				疒，校對問題，舊版漢籍錯字，可用程式全部取代。

附件二、《列子集釋》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
4910	淫	汎 <u>𣎵</u> 兮無	汎 <u>淫</u> 兮無
		一作沉 <u>𣎵</u>	一作沉 <u>淫</u>
		侵尋即浸 <u>𣎵</u> 也	侵尋即浸 <u>淫</u> 也
		亦言浸 <u>𣎵</u> 轉大也	亦言浸 <u>淫</u> 轉大也
5048		人不 <u>𡗗</u> 宦	人不 <u>女𡗗</u> 昏宦
5220		字誤作 <u>友</u>	字誤作 <u>毋</u>
5334		日晡時天 <u>主</u> 晏	日晡時天 <u>暝</u> 晏
		<u>主</u> 即晴	<u>暝</u> 即晴

附件三、《列子集釋》未確定字形「❓」手動取代表

檔案原文摘錄	手動取代結果	備註
篆文又作❓	篆文又作 <u>𠄎</u> 又 <u>𠄎</u> 說文❓	
為古文作❓(全部取代)	為古文作 <u>𠄎</u> 為 <u>𠄎</u> 說文古文❓(全部取代)	2
丙 <u>𠄎</u> 支字作❓	丙 <u>𠄎</u> 支字作 <u>𠄎</u> 更 <u>𠄎</u> 集成 4341❓	
字林云隱作❓	字林云隱作 <u>𠄎</u> 一ノ口𠄎❓	
拜當作❓	拜當作 <u>𠄎</u> 升 <u>𠄎</u> 說文❓	
❓又作〇〇手	<u>𠄎</u> 升 <u>𠄎</u> 說文❓又作〇〇手	
❓，竦手也	<u>𠄎</u> 升 <u>𠄎</u> 說文❓，竦手也	
「❓」字形相近	「 <u>𠄎</u> 心 <u>𠄎</u> 說文❓」字形相近	
如五作❓	如五作 <u>三</u> <u>𠄎</u> <u>𠄎</u>	
「言」古文「❓」	無法處理的古文字形，暫時維持原狀。	
乙鼓魚 <u>𠄎</u> 𠄎字省從❓		

