

《朱子語類》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

2008/1/14 陳建安 製作

一、《朱子語類》一書(朱子語類.xml)使用舊版造字 251 個，字頻 2581 次，詳如附件一。這 251 個造字中，217 個可轉成 Windows XP 能顯示的字，字頻 2420 次；另外 34 個字必須轉成構字式，字頻 161 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 2567 的「稗」字，原字為「粹」。又編號 147，原書及造字皆無字形，手動加以刪除。
- 2、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 4904，字碼 8864，對應的舊漢籍造字為「尅」，部件外字為「三」，一一比對原書，確認是「尅」者，再以手動取代。
- 3、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 4485 的「𠩺墨」係「塵」字之異體，故以「塵」字取代。

4、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。

5、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 5749 的「𠩺庸」。

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 1337 的「慌」字，Unicode 編碼為 385B，由於 Windows XP 並不支援，仍須使用構字式「巾𠩺荒」。

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、檔案原文摘錄：摘錄檔案(朱子語類.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。

丙、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、《朱子語類》中未確定字形「𠩺」共 9 處，詳細處理內容可參考附件三的未確定字形「𠩺」手動取代表。附件三的未確定字形「𠩺」手動取代表欄位說明如下：

甲、檔案原文摘錄：摘錄檔案(朱子語類.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、《朱子語類》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
38		1	1	FA65			王𠂔已	玦，校對問題，舊版漢籍錯字，可用程式全部取代。
147		1	1	FAF4				無字形，校對問題，舊版漢籍錯字，手動刪除。
814	墓	1	1	8E5C	87C7	墓		
820	烟	1	1	8E62	70F1	烟		
832	淹	1	1	8E6E	6CB2	淹		
841	礪	6	6	8E77	7932	礪		
932	兒	1	1	8EF4	7683	兒		
982	胃	3	3	8F67	80F7	胃		
994	楚	1	1	8F73	8E75	楚		
1000		1	1	8F79			王𠂔已	玦，校對問題，舊版漢籍錯字，可用程式全部取代。
1002	鳩	1	1	8F7B	9D02	鳩		
1008	玦	1	1	8FA3	739E	玦		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1040	聃	7	7	8FC3	803C	聃		
1056	阤	1	1	8FD3	9641	阤		
1067	綦	6	6	8FDE	68CA	綦		
1076	𪛗	3	3	8FE7	5368	𪛗		Unicode 呈現差異。
1079	總	1	1	8FEA	7DEB	總		
1126	璫	1	1	905A	74B9	璫		
1147		1	1	906F			𠂔穴矣欠𠂔	竅，異體字問題。
1224	暇	1	1	90DE	7771	暇		
1226	罰	4	4	90E0	7F78	罰		
1267	𣎵	1	1	914A	6B1B	𣎵		
1337	慌	1	1	91B2	385B		巾𠂔荒	
1394	𠵽	1	1	91EB	35D6		口𠂔𠵽	𠵽，異體字問題。
1491	擷	8	8	92AF	6527	擷		
1513	糧	8	8	92C5	7CAE	糧		
1547	𠂔	2	2	92E7	4E31	𠂔		
1555	緡	3	3	92EF	7DCD	緡		
1565	緡	3	3	92F9	7DDC	緡		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1591	羗	1	1	9354	7F97	羗		
1616	逕	2	2	936D	284E6		辶厶里	
1642	脉	8	8	93A9	8109	脉		
1646	脗	1	1	93AD	8117	脗		
1739	燃	1	1	944B	3DFF		火厶僉	
1809	囧	3	3	94B3	56E6	囧		
1854	囁	3	3	94E0	20E95		口厶留	
1859	匱	2	2	94E5	5336	匱		
1893	袞	52	52	9548	886E	袞		
1925	羈	3	3	9568	8989	羈		Unicode 呈現差異。
1926	覓	2	2	9569	8994	覓		
1934	涉	1	1	9571	23D16		氵厶𠂔	
1944	簪	1	1	957B	8A67	簪		
1981	讐	49	49	95C2	8B90	讐		
2021	趲	7	7	95EA	8D82	趲		
2152	蕳	2	2	96D0	979F	蕳		
2168	醅	1	1	96E0	91BB	醅		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2197	鉄	5	5	96FD	9244	鉄		
2408	鬧	8	8	9874	9599	鬧		
2409	関	1	1	9875	95A7	関		
2505	翻	2	2	98F7	98DC	翻		
2567	粹	1	1	9976	7A21	粹		粹，校對問題，舊版漢籍錯字，可用程式全部取代。
2592	贓	1	1	99B1	8D1C	贓		
2885	呌	5	5	9B7A	544C	呌		
2993	办	1	1	9C49	5205	办		
3021	忸	1	1	9C65			忸	
3213	塔	1	1	9DAA	366E		土𡵓答	
3770	个	24	24	8141	4E2A	个		
3787	豕	16	16	8152	51A2	豕		
3788	慙	2	2	8153	22871		殼𡵓心	慙，校對問題，舊版漢籍錯字，可用程式全部取代。
3789	冲	18	18	8154	51B2	冲		
3791	况	7	7	8156	51B5	况		
3792	凉	30	30	8157	51C9	凉		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3794	湊	2	2	8159	51D1	湊		
3796	決	2	2	815B	51B3	決		
3797	凭	1	1	815C	51ED	凭		
3801	剗	33	33	8160	5257	剗		
3804	効	9	9	8163	52B9	効		
3805	勅	5	5	8164	52C5	勅		
3807	驗	2	2	8166	9A10	驗		
3819	廝	26	26	8172	53AE	廝		
3822	叁	4	4	8175	53C1	叁		
3824	叙	21	21	8177	53D9	叙		
3825	疊	42	42	8178	53E0	疊		
3829	咏	16	16	817C	548F	咏		
3830	咤	5	5	817D	54A4	咤		
3833	嘑	1	1	81A2	5611	嘑		
3835	噍	2	2	81A4	564D	噍		
3836	喻	2	2	81A5	564F	喻		
3840	坂	5	5	81A9	5742	坂		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3843	堦	9	9	81AC	5826	堦		
3846	堦	11	11	81AF	58FB	堦		
3852	嫵	3	3	81B5	5A63	嫵		
3858	寃	13	13	81BB	5BC3	寃		
3876	廻	8	8	81CD	5EFB	廻		
3877	廻	2	2	81CE	5EFC	廻		
3883	徧	106	106	81D4	5FA7	徧		
3884	忽	5	5	81D5	6031	忽		
3886	悞	12	12	81D7	609E	悞		
3891	慙	5	5	81DC	6159	慙		
3895	淑	1	1	81E0	6E86	淑		
3909	擡	15	15	81EE	64E1	擡		
3915	斃	6	6	8459	6585	斃		
3921	晷	26	26	81FA	6673	晷		
3924	晷	7	7	81FD	671E	晷		
3929	黎	2	2	8243	68C3	黎		
3931	椀	32	32	8245	6900	椀		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3935	槩	3	3	8249	69E9	槩		Unicode 呈現差異。
3936	棹	1	1	824A	6A70	棹		
3939	櫟	3	3	824D	6AFD	櫟		
3944	氷	5	5	8252	6C37	氷		
3949	涖	4	4	8257	6D96	涖		
3961	灾	5	5	8263	707E	灾		
3966	烟	8	8	8268	70DF	烟		
3968	煨	10	10	826A	7145	煨		
3974	牀	2	2	8270	7240	牀		
3975	牕	8	8	8271	7255	牕		
3977	狗	5	5	8273	72E5	狗		
3979	猫	3	3	8275	732B	猫		
3982	珣	1	1	8278	73F7	珣		
3987	穀	3	3	827D	7474	穀		
3990	甄	2	2	82A2	750E	甄		
3991	攜	2	2	82A3	64D5	攜		
3996	疎	4	4	82A8	758E	疎		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4003	塚	1	1	82AF	585C	塚		
4008	癰	1	1	82B4	766F	癰		
4011	皞	4	4	82B7	76A1	皞		
4012	鼓	3	3	82B8	76B7	鼓		
4013	盃	1	1	82B9	76CC	盃		
4025	礪	2	2	82C5	7934	礪		
4034	穉	1	1	82CE	7A49	穉		
4035	穽	3	3	82CF	7A7D	穽		
4036	窰	1	1	82D0	7AB0	窰		
4038	竈	27	27	82D2	7AC8	竈		
4040	竝	4	4	82D4	7ADD	竝		
4041	豎	12	12	82D5	7AEA	豎		
4042	笋	3	3	82D6	7B0B	笋		
4044	筴	1	1	82D8	7B6F	筴		
4052	簾	1	1	82E0	7C34	簾		
4055	粃	3	3	82E3	7C83	粃		
4068	綫	2	2	82F0	7DAB	綫		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4069	綑	4	4	82F1	7DB3	綑		
4080	羣	152	152	82FC	7FA3	羣		
4081	翱	10	10	82FD	7FFA	翱		Unicode 呈現差異。
4085	耻	25	25	8342	803B	耻		
4090	脚	273	273	8347	811A	脚		
4091	皐	45	45	8348	81EF	皐		
4093	舐	1	1	834A	445B		舌𠂔氏	舐，異體字問題。
4094	舘	3	3	834B	8218	舘		
4097	芑	6	6	834E			艹𠂔已	芑，校對問題，舊版漢籍錯字，可用程式全部取代。
4105	菓	2	2	8356	83D3	菓		
4112	蔴	1	1	835D	8534	蔴		
4114	藥	11	11	835F	8616	藥		
4115	藁	13	13	8360	85C1	藁		
4118	藥	2	2	8363	8602	藥		
4139	袴	2	2	8378	88B4	袴		
4140	袵	3	3	8379	88B5	袵		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4145	羈	4	4	837E	898A	羈		Unicode 呈現差異。
4146	覩	29	29	83A1	89A9	覩		
4158	譌	4	4	83AD	8B4C	譌		Unicode 呈現差異。
4172	屢	4	4	83BB	8E77	屢		Unicode 呈現差異。
4173	耽	1	1	83BC	8EAD	耽		
4176	輒	2	2	83BF	8F19	輒		
4184	迹	208	208	83C7	8FF9	迹		
4186	遡	5	5	83C9	9061	遡		
4193	鈎	6	6	83D0	920E	鈎		
4195	鉢	3	3	83D2	9262	鉢		
4198	鑛	3	3	83D5	945B	鑛		
4203	隣	7	7	83DA	96A3	隣		
4212	鞞	2	2	83E3	97B8	鞞		
4217	鞞	5	5	83E8	97E0	鞞		
4227	飡	1	1	83F2	98E6	飡		
4237	髡	6	6	83FC	9AE0	髡		
4240	髡	16	16	8440	9AF4	髡		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4242	鬪	43	43	8442	9B2D	鬪		Unicode 呈現差異。
4256	鷄	19	19	8450	9DC4	鷄		
4259	鷹	2	2	8453	9E95	鷹		Unicode 呈現差異。
4262	麪	3	3	8456	9EAA	麪		
4263	麴	1	1	8457	9EAF	麴		
4266	鼃	3	3	845A	9F02	鼃		
4267	鼈	2	2	845B	9F08	鼈		
4270	𪚩	2	2	845E	9F45	𪚩		
4278	亘	17	17	8466	4E98	亘		
4287	織	3	3	846F	7E8E	織		
4299	炳	1	1	847B	242F2		火𠂇𠂇	
4307	罵	16	16	84A5	99E1	罵		
4309	刼	24	24	84A7	523C	刼		
4366	邾	2	2	84E0	48BC		共𠂇𠂇	
4371	倜	1	1	84E5	50F4	倜		
4437	廩	3	3	8568	5EEA	廩		
4439		2	2	856A			𠂇𠂇足巳心𠂇	𠂇，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。
4447	捏	4	4	8572	63D1	捏		
4448		2	2	8573			𠂔 𠂔 墨 𠂔	纏，校對問題，舊版漢籍錯字，可用程式全部取代。
4455	劒	13	13	857A			僉 𠂔 刀	
4460	勅	43	43	85A1	52D1	勅		
4469	勻	1	1	85AA	5304	勻		
4482	廛	1	1	85B7	53AA	廛		
4485	廛	3	3	85BA	3553		𠂔 𠂔 墨	廛，異體字問題。
4515	呪	19	19	85D8	546A	呪		
4649	蕒	1	1	86C1	8552	蕒		
4679	塏	1	1	86DF	5818	塏		
4702	塲	2	2	86F6	5872	塲		
4714	廛	15	15	8743	222E8		𠂔 𠂔 墨	廛，異體字問題。
4760	𠂔	3	3	8771	89B8	𠂔		
4763	𠂔	1	1	8774	6586	𠂔		
4768	妬	3	3	8779	59AC	妬		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4778	鞞	5	5	87A5	97B4	鞞		
4813		5	5	87C8			𠂔 厂 墨 𠂔	灑，異體字問題。
4902	兗	12	12	8862	5156	兗		
4903	𠂔	1	1	8863	5C00	𠂔		
4904	尅	2	3	8864	5C05	尅		與三重碼，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。
4906		4	4	8866			●	示，校對問題，舊版漢籍錯字，可用程式全部取代。
4907	橐	2	2	8867	69D6	橐		
4957	崑	1	1	88BB	5D53	崑		
4996	斲	3	3	88E2	65B5	斲		
5002	鞞	1	1	88E8	4A6D		鞞	迹，校對問題，舊版漢籍錯字，可用程式全部取代。
5008	幞	29	29	88EE	5E5E	幞		
5012	嬖	1	1	88F2	36D0		女 𡥉 更	
5034	弃	3	3	8949	5F03	弃		
5043	彊	1	1	8952	5F4D	彊		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5046	彝	1	1	8955	5F5C	彝		
5066	瑣	4	4	8969	24A0F		王𠂔賁	
5074	𪔐	1	1	8971	5C2B	𪔐		
5111	旆	9	9	89B8	65BE	旆		
5115	颺	2	2	89BC	4B03		風𠂔立	
5120	虐	10	10	89C1	8657	虐		
5163	憂	1	1	89EC	621E	憂		
5176		53	53	89F9			汜 𠂔 冗	沉，異體字問題。
5225	邨	13	13	8A6B	460F		血 𠂔 卩	
5229	撐	1	1	8A6F	6490	撐		
5285	睂	1	1	8AC9	7742	睂		
5289	慤	18	18	8ACD	6164	慤		
5290	携	1	1	8ACE	643A	携		
5300	鬪	3	3	8AD8	9B2A	鬪		
5310	烏	3	3	8AE2	8204	烏		
5418	邨	1	1	8BB1	28697		冉 𠂔 卩	
5434	涅	1	1	8BC1	5E7C		𠂔 汜 白 工 𠂔	涅，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5439	栢	8	8	8BC6	6822	栢		
5466	倬	1	1	8BE1	5010	倬		Unicode 呈現差異。
5541	蠹	3	3	8C6D	8827	蠹		
5542	鑠	2	2	8C6E	93C1	鑠		
5549	藁	3	3	8C75	85F3	藁		
5551	賁	1	1	8C77	8CF7	賁		
5555	迥	4	4	8C7B	9008	迥		
5558	麓	13	13	8C7E	9E84	麓		
5559	欸	24	24	8CA1	6B35	欸		
5570	繇	2	2	8CAC	6ABE	繇		
5582	鞣	1	1	8CB8	97B5	鞣		
5583	殮	1	1	8CB9	98F1	殮		
5591	杞	13	13	8CC1	233CC		木𣎵已	杞，校對問題，舊版漢籍錯字，可用程式全部取代。
5595	塚	2	2	8CC5	579C	塚		
5697	𩇛	4	4	8D6C	7151	𩇛		Unicode 呈現差異。
5702	厨	8	8	8D71	53A8	厨		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5704	𧯛	3	3	8D73	6C0A	𧯛		
5705	廐	1	1	8D74	5ED0	廐		Unicode 呈現差異。
5706	敍	33	33	8D75	654D	敍		
5721	噸	1	1	8DA6	9814	噸		
5742	澁	18	18	8DBB	6F81	澁		
5749		1	1	8DC2			片𠂔庸	待造字。
5753	潜	6	6	8DC6	6F5C	潜		
5768	濶	29	29	8DD5	6FF6	濶		
5860	无	177	177	C6D3	65E0	无		

附件二、《朱子語類》手動取代表

編號	檔案原文摘錄	手動取代結果
147	<u>恒</u> </h>	恒</h>
4904	<u>三</u> ，曰：「五賊	<u>尅</u> ，曰：「五賊
	禮則相刑相 <u>三</u>	禮則相刑相 <u>尅</u>

附件三、《朱子語類》未確定字形「」手動取代表

檔案原文摘錄	手動取代結果
陽中陰 	陽中陰 <u>○○一△一</u>
陰中陽 	陰中陽 <u>一△○○一</u>
陰中陰 	陰中陰 <u>○○一</u>
篆文「天」作「 	篆文「天」作「  天  說文 
「而」作「 	「而」作「  而  說文 
一箇陰又生出一箇陽， 	一箇陰又生出一箇陽， <u>一△○○一</u>
一箇陽又生一箇陰， 	一箇陽又生一箇陰， <u>○○一△一</u>
一箇陰又生一箇陰，  是一象也	一箇陰又生一箇陰， <u>○○一</u> 是一象也
「扌」，如此寫「 	「扌」，如此寫「 <u>扌</u> 」