

# 《四庫輯本別集拾遺》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  
2008/4/21 陳建安 製作

一、《四庫輯本別集拾遺》一書(四庫輯本別集拾遺.xml)使用舊版造字 270 個，字頻 1118 次，詳如附件一。這 270 個造字中，211 個可轉成 Windows XP 能顯示的字，字頻 966 次；另外 59 個字必須轉成構字式，字頻 152 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、構字式：Windows XP 若無對應字形，則改採用構字式

壬、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 5516 的「○○束」字，原字為「棘」。
- 2、重碼問題，採用部件外字，不轉碼：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」。然而漢籍在舊轉新的過程中，若已將原舊漢籍造字的部分全數改為 Unicode 或構字式，則在 xml 出現的造字頻次皆採用部件外字，此時便不需轉碼。如編號 5797，字碼為 8DF2，對應的舊漢籍造字為「灩」，部件外字為「」，對原書發現書中皆取部件外字「」，故不需轉碼。
- 3、標記問題：標記問題主要是漢籍電子文獻在舊轉新的過程中，採用了新的標記語言，而在修改標記時，對於某些標題句做了增刪的動作，因為這些標題句的內容包含

了缺字，所以這些更改標記的動作造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號 5289 的「慤」字，xml 檔在 7 頁新增了一個標題，因而造成頻次不同，處理方式為將所有頻次皆修改為「慤」。

- 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 4448 的「𦉑 纒 厂 墨 𦉑」係「纏」字之異體，故以「纏」字取代。又編號 1513 的「糧」字，係「糧」字之簡體字，亦以標準字的「糧」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 1925 的「羈」字，Unicode 字型呈現為「羈」，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 1436 的「𦉑 昏」。

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 889 的「𦉑」字，Unicode 編碼為 476F，由於 Windows XP 並不支援，仍須使用構字式「貝 𦉑」。

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的欄位說明如下：

- 甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- 乙、取代：當實際使用字形與造字不同時，在此標出。
- 丙、檔案原文摘錄：摘錄檔案(四庫輯本別集拾遺.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。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。

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。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。

附件一、《四庫輯本別集拾遺》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
74	髻	2	2	FAAB	29B06		髻△𠂔	髻，校對問題，舊版漢籍錯字，可用程式全部取代。
304	品	2	2	FBF4			品	
396	憲	1	1	FCB3			憲	
639	彳	1	1	FE4A			彳	
707	ㄩ	3	3	FEB0			ㄩ	
759	广	1	1	FEE4			广	
766	糸	1	1	FEEB			糸	
770	扌	10	10	FEEF			扌	
773	厶	2	2	FEF2			厶	
832	沱	1	1	8E6E	6CB2	沱		
841	礪	1	1	8E77	7932	礪		
881	颯	3	3	8EC1	98C7	颯		
889	𪗇	2	2	8EC9	476F		貝△𪗇	
890	𪗇	2	2	8ECA	71AD	𪗇		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
893	峩	4	4	8ECD	5CE9	峩		
896	櫚	1	1	8ED0	6AE9	櫚		
910	崧	1	1	8EDE	5D78	崧		
947	燁	1	1	8F44	7217	燁		
957	遨	1	1	8F4E	906C	遨		
1000		3	4	8F79			王𠂇巳	標記問題，新增標題句，以標準字玓取代。
1026	珣	22	22	8FB5	73E6	珣		
1079	總	5	6	8FEA	7DEB	總		總，標記問題，新增標題句，人工取代。
1133	瓌	2	2	9061	74CC	瓌		
1158	咳	1	1	907A	7561	咳		
1205	眇	3	3	90CB	770E	眇		
1299	繇	1	1	916A	2259F		彳𠂇繇	
1310	痲	1	1	9175	45EB		虫𠂇麻	
1337	慌	1	1	91B2	385B		巾𠂇荒	
1354	禊	1	1	91C3	79A9	禊		
1382	營	1	1	91DF	3F66		營𠂇田	
1416	埃	5	5	9242	7AE2	埃		
1417	凌	5	5	9243	35AB		口𠂇凌	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1436		2	2	9256			辵△昏	待造字。
1490	籛	1	1	92AE	7C5D	籛		
1513	粮	1	1	92C5	7CAE	粮		糧，異體字問題。
1542	繇	2	2	92E2	34A1		亻△繇	
1565	絲	1	1	92F9	7DDC	絲		
1642	脉	9	9	93A9	8109	脉		
1646	脗	1	1	93AD	8117	脗		
1700		2	2	93E3			竹△厶△厶△心	形竹弓心，異體字問題。
1784	萼	4	4	9478	855A	萼		
1859	匱	1	1	94E5	5336	匱		
1876	愬	3	3	94F6	6133	愬		
1891	岭	1	1	9546	37CF		山△含	
1925	羈	2	2	9568	8989	羈		Unicode 呈現差異。
1963	訖	2	2	95B0	8A5D	訖		
1983	瓊	1	1	95C4	24AE9		王△靈	
1999	贊	1	1	95D4	8CDB	贊		贊，異體字問題。
2076	穰	1	1	9662	4189		禾△罷	
2077	鞞	1	1	9663	8F4A	鞞		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2078	稕	1	1	9664	7A0F	稕		
2151	卩	4	4	96CF	961D	卩		
2168	疇	1	1	96E0	91BB	疇		
2225	緘	1	1	975A	88D3	緘		
2402	勳	1	1	986E	52EE	勳		
2409	闕	2	2	9875	95A7	闕		闕，異體字問題。
2429	秬	1	1	98AB	7A07	秬		
2432	鞞	1	1	98AE	29350		革𠂔登	
2502	颺	6	6	98F4	98C8	颺		
2572	銜	1	1	997B	8858	銜		
2602	鰐	2	2	99BB	9C10	鰐		
2834	攴	1	1	9B47	6535	攴		
2836	三	2	2	9B49			三	
2837		8	8	9B4A			一𠂔一𠂔一	三，校對問題，舊版漢籍錯字，可用程式全部取代。
2906	筭	1	1	9BB1	659A	筭		筭，異體字問題。
2938	簪	1	1	9BD1	7C2E	簪		簪，異體字問題。
3021	忄	4	4	9C65			忄	
3068	三	1	1	9CB6			三	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3071	𠄎	3	3	9CB9			𠄎	
3073	𠄎	2	2	9CBB			𠄎	
3074	𠄎	8	8	9CBC			𠄎	
3096	歲	2	2	9CD2	5D57	歲		歲，異體字問題。
3134	𠄎	1	1	9CF8	5405	𠄎		
3238		5	5	9DC3			𠄎西广艹糸	𠄎，校對問題，舊版漢籍錯字，可用程式全部取代。
3789	冲	17	18	8154	51B2	冲		標記問題，新增標題句，以標準字冲取代。
3791	况	2	2	8156	51B5	况		况，異體字問題。
3792	凉	2	2	8157	51C9	凉		凉，異體字問題。
3793	減	2	2	8158	51CF	減		減，異體字問題。
3796	決	4	4	815B	51B3	決		決，異體字問題。
3805	勅	1	1	8164	52C5	勅		
3810	匱	2	2	8169	5335	匱		
3814	却	58	58	816D	5374	却		
3816	厓	7	7	816F	5393	厓		
3817	厠	2	2	8170	53A0	厠		厠，異體字問題。
3820	廈	8	8	8173	53A6	廈		廈，異體字問題。
3824	叙	11	11	8177	53D9	叙		叙，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3826	叶	7	7	8179	53F6	叶		
3829	咏	6	6	817C	548F	咏		
3830	咤	2	2	817D	54A4	咤		
3840	坂	2	2	81A9	5742	坂		
3843	堦	4	4	81AC	5826	堦		
3846	墀	2	2	81AF	58FB	墀		Unicode 呈現差異。
3849	獎	2	2	81B2	596C	獎		獎，異體字問題。
3852	嫵	4	4	81B5	5A63	嫵		
3860	尠	1	1	81BD	5C20	尠		
3866	崐	1	1	81C3	5D10	崐		
3868	卮	5	5	81C5	5DF5	卮		卮，異體字問題。
3877	迺	8	8	81CE	5EFC	迺		迺，異體字問題。
3883	徧	8	8	81D4	5FA7	徧		
3884	忽	4	4	81D5	6031	忽		
3886	悞	1	1	81D7	609E	悞		
3891	慙	10	10	81DC	6159	慙		慙，異體字問題。
3895	淑	1	1	81E0	6E86	淑		
3918	擅	1	1	81F7	65DC	擅		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3931	椀	3	3	8245	6900	椀		
3933	桀	10	10	8247	6998	桀		
3937	檝	1	1	824B	6A9D	檝		
3940	鬱	1	1	824E	6B1D	鬱		
3944	冰	9	9	8252	6C37	冰		冰，異體字問題。
3948	沝	2	2	8256	6CAD	沝		
3949	洑	6	6	8257	6D96	洑		
3957	凜	1	1	825F	6F9F	凜		
3966	烟	19	19	8268	70DF	烟		煙，異體字問題。
3968	煨	2	2	826A	7145	煨		
3974	牀	16	16	8270	7240	牀		
3976	犁	1	1	8272	7282	犁		
3979	猫	1	1	8275	732B	猫		
3981	珉	1	1	8277	73C9	珉		
3988	璦	1	1	827E	7482	璦		
3989	璿	2	2	82A1	74C8	璿		
3996	疎	2	2	82A8	758E	疎		
4000	疴	2	2	82AC	75B4	疴		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4008	癯	5	5	82B4	766F	癯		
4012	鼓	2	2	82B8	76B7	鼓		
4013	盃	2	2	82B9	76CC	盃		
4015	眦	1	1	82BB	7726	眦		
4019	矻	1	1	82BF	77F4	矻		
4025	礪	2	2	82C5	7934	礪		
4028	稟	11	11	82C8	7980	稟		稟，異體字問題。
4034	穉	3	3	82CE	7A49	穉		
4035	穽	5	5	82CF	7A7D	穽		
4038	竈	10	10	82D2	7AC8	竈		
4039	竚	4	4	82D3	7ADA	竚		
4041	豎	1	1	82D5	7AEA	豎		
4042	笋	3	3	82D6	7B0B	笋		
4045	筵	1	1	82D9	7B92	筵		
4055	粃	1	1	82E3	7C83	粃		
4067	綉	3	3	82EF	7D89	綉		繡，異體字問題。
4068	綫	1	1	82F0	7DAB	綫		線，異體字問題。
4077	罇	4	4	82F9	7F47	罇		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4081	翱	3	3	82FD	7FFA	翱		
4085	耻	29	29	8342	803B	耻		
4089	脇	1	1	8346	8107	脇		
4090	脚	16	16	8347	811A	脚		
4091	臯	18	19	8348	81EF	臯		臯，標記問題，新增標題句，人工取代。
4097	芑	2	2	834E			艹 厶 巳	芑，異體字問題。
4107	葱	9	9	8358	8471	葱		
4114	藜	1	1	835F	8616	藜		
4115	藁	17	17	8360	85C1	藁		
4118	藜	3	3	8363	8602	藜		
4129	虬	3	4	836E	866C	虬		虬，標記問題，新增標題句，人工取代。
4131	劬	3	3	8370	8842	劬		
4135	衙	2	2	8374	885C	衙		
4138	袞	1	1	8377	88A0	袞		
4139	袴	3	3	8378	88B4	袴		
4140	袵	1	1	8379	88B5	袵		
4145	羈	6	6	837E	898A	羈		Unicode 呈現差異。
4146	覩	16	16	83A1	89A9	覩		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4149	靚	1	1	83A4	89BC	靚		
4150	訥	1	1	83A5	8A29	訥		
4167	賚	1	1	83B6	8CEB	賚		
4172	屨	1	1	83BB	8E77	屨		Unicode 呈現差異。
4184	迹	33	33	83C7	8FF9	迹		
4186	遡	6	6	83C9	9061	遡		
4193	鈎	12	12	83D0	920E	鈎		
4195	鉢	4	4	83D2	9262	鉢		
4203	隣	3	3	83DA	96A3	隣		鄰，異體字問題。
4213	鞞	1	1	83E4	97BE	鞞		
4221	頰	2	2	83EC	9833	頰		
4226	飡	1	1	83F1	98E1	飡		
4232	隲	1	1	83F7	96B2	隲		隲，異體字問題。
4237	髡	3	3	83FC	9AE0	髡		髡，異體字問題。
4240	髯	1	1	8440	9AF4	髯		
4242	鬪	17	17	8442	9B2D	鬪		Unicode 呈現差異。
4251	鰕	3	3	844B	9C15	鰕		
4256	鷄	14	14	8450	9DC4	鷄		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4263	麩	2	2	8457	9EAF	麩		
4278	亘	6	6	8466	4E98	亘		
4287	織	5	5	846F	7E8E	織		織，異體字問題。
4307	罵	2	2	84A5	99E1	罵		罵，異體字問題。
4348	惚	2	2	84CE	506C	惚		
4353	菟	1	1	84D3	83AC	菟		
4363	彭	2	2	84DD	291D5		青 厶 彡	
4368	僨	2	2	84E2	50D9	僨		
4399	塼	1	1	8542	213D6		土 厶 盍	
4428	躡	1	1	855F	8E70	躡		躡，異體字問題。
4437	廩	7	7	8568	5EEA	廩		廩，異體字問題。
4442	潛	1	1	856D	6F98	潛		Unicode 呈現差異。
4448		1	1	8573			形 纟 厂 墨 口	纏，異體字問題。
4455	劒	1	2	857A			僉 厶 刃	僉 厶 刃，標記問題，新增標題句，人工取代。
4482	厖	3	3	85B7	53AA	厖		
4510	𠵼	4	4	85D3	5493	𠵼		
4542	塼	1	1	85F3	21484		土 厶 稟	塼，異體字問題。
4566	礪	1	1	864C	78F5	礪		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4603	憇	4	4	8671	6187	憇		
4610	筭	2	2	8678	41FE		竹 讠 言	
4643	徃	3	3	86BB	5F83	徃		
4735	壠	4	4	8758	58E0	壠		
4737	慳	1	1	875A	6082	慳		
4763	敦	6	6	8774	6586	敦		
4768	妬	5	5	8779	59AC	妬		妒，異體字問題。
4805	瀆	1	1	87C0	51DF	瀆		瀆，異體字問題。
4816	亻	1	1	87CB			亻	
4817	寶	2	2	87CC	5BF3	寶		寶，異體字問題。
4844		1	1	87E7			灃 讠 凡	灃，異體字問題。
4878	宀	1	1	884A	5B82	宀		Unicode 呈現差異。
4884		1	1	8850			𠂔 足 厂 墨 𠂔	躔，異體字問題。
4901		1	1	8861			冗 讠 且	置，異體字問題。
4907	橐	1	1	8867	69D6	橐		
4913	墮	1	1	886D	58A2	墮		
4948	𦉳	10	10	88B2	2308F		𦉳 斗	
4957	崑	4	4	88BB	5D53	崑		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4961	盼	2	2	88BF	80B9	盼		
4981	鳧	3	3	88D3	9CEC	鳧		
4996	斲	2	2	88E2	65B5	斲		
4999	巘	2	2	88E5	5DD7	巘		
5019	畧	3	3	88F9			畧	略，異體字問題。
5025		2	2	8940			千𠂇千	𠂇，異體字問題。
5034	弃	5	5	8949	5F03	弃		Unicode 呈現差異。
5046	彝	17	17	8955	5F5C	彝		
5066	瑣	1	1	8969	24A0F		王𠂇賁	瑣，異體字問題。
5111	旆	3	3	89B8	65BE	旆		旆，異體字問題。
5126	暮	1	1	89C7	8B29	暮		
5131	窻	1	1	89CC	7ABB	窻		Unicode 呈現差異。
5150	伏	1	1	89DF	3432		亻𠂇大	伏，異體字問題。
5163	憂	1	1	89EC	621E	憂		
5176		20	20	89F9			氵𠂇冗	沉，異體字問題。
5194		2	2	8A4C			扌𠂇且	擅，異體字問題。
5195		1	1	8A4D			木𠂇且	檀，異體字問題。
5214		1	1	8A60			魚𠂇且	鱣，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5230		2	2	8A70			𠄎土亩且𠄎	壇，異體字問題。
5289	慤	2	3	8ACD	6164	慤		慤，標記問題，新增標題句，人工取代。
5290	携	14	14	8ACE	643A	携		
5293		2	2	8AD1			冗厶幕	幕，異體字問題。
5300	鬪	3	3	8AD8	9B2A	鬪		
5305	晒	4	3	8ADD	661E	晒		晒，標記問題，新增標題句，人工取代。
5338	暎	2	2	8AFE	668E	暎		
5483	眡	2	2	8BF2	7721	眡		
5484	楛	1	1	8BF3	6973	楛		
5499	梏	1	1	8C43	6947	梏		
5507	怵	2	2	8C4B			忄 厶木	
5512	𪗇	1	1	8C50	8C3A	𪗇		
5515	軌	2	2	8C53			車 厶 丸	軌，異體字問題。
5516		1	1	8C54			∞ 束	棘，校對問題，舊版漢籍錯字，可用程式全部取代。
5542	鑠	1	1	8C6E	93C1	鑠		
5549	藁	1	1	8C75	85F3	藁		
5551	賈	1	1	8C77	8CF7	賈		
5555	迥	5	5	8C7B	9008	迥		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5558	龕	1	1	8C7E	9E84	龕		
5578	弛	2	2	8CB4	38AE		弓厶屯	
5581	啻	3	3	8CB7	20F2A		口厶移	
5583	殮	4	4	8CB9	98F1	殮		
5586	歛	2	2	8CBC	6B57	歛		
5587	歛	1	1	8CBD	6B58	歛		
5588	恡	1	1	8CBE	6061	恡		
5621	泪	1	1	8CDF	6CEA	泪		淚，異體字問題。
5693	𪗇	1	1	8D68	9E7B	𪗇		
5702	厨	2	2	8D71	53A8	厨		廚，異體字問題。
5704	𪗇	4	4	8D73	6C0A	𪗇		
5705	廐	1	1	8D74	5ED0	廐		廐，異體字問題。
5706	敍	76	77	8D75	654D	敍		標記問題，新增標題句，以標準字敘取代。
5707	棧	1	1	8D76	6830	棧		
5717	陘	1	1	8DA2	967F	陘		
5720	垧	1	1	8DA5	578D	垧		
5729	滌	2	2	8DAE	6F34	滌		
5752	摠	1	1	8DC5	63D4	摠		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5764	籊	1	1	8DD1	7C5E	籊		
5767	濶	1	1	8DD4	24050		灬 厶 閣	闊，校對問題，舊版漢籍錯字，可用程式全部取代。
5768	濶	1	1	8DD5	6FF6	濶		闊，異體字問題。
5797	灩	4	45	8DF2	7054	灩		與厶重碼，重碼問題，採用部件外字，不轉碼。

附件二、《四庫輯本別集拾遺》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
1000	玨	__ (全部取代)	<u>玨</u> (全部取代)
1079		__ (全部取代)	<u>總</u> (全部取代)
3789	沖	𠂇 (全部取代)	<u>沖</u> (全部取代)
4091		𠂇 (全部取代)	<u>皇</u> (全部取代)
4129		𠂇 (全部取代)	<u>虬</u> (全部取代)
4455		𠂇 (全部取代)	<u>僉厶刃</u> (全部取代)
5289		𠂇 (全部取代)	<u>慤</u> (全部取代)
5305		𠂇 (全部取代)	<u>晒</u> (全部取代)
5706	敘	𠂇 (全部取代)	<u>敘</u> (全部取代)