

《清實錄》（穆宗）舊版造字轉碼說明

中研院資訊所文獻處理實驗室

2007/3/22 趙苑曲

一、《清實錄》穆宗實錄部分(穆宗.xml)使用舊版造字 210 個，字頻 5810 次，詳如附件一。這 210 個造字中，132 個可轉成 Windows XP 可顯示的字，字頻 3429 次；另外 79 個字必須轉成構字式，字頻 2381 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、錯字：與原書字形不符，用正確的字取代原造字，如編號 49 的「疒△廣」字，原字應為「弊」。

2、取代：用另一個字取代原造字，如編號 5025 的「千△干」字，改用「干」字取代。

3、手動刪除：比對原書發現文本中無此字，故以手動方式刪除。如編號 785，詳見附件三的「手動刪除表」。

4、重碼造字，手動取代：同時使用舊漢籍造字及構字式，導致一個字碼代表兩種字形的重碼問題，必須一一比對原書字形，以判斷應取何者之字形，取舊漢籍造字字形者再以手動取代。如編號 297，舊漢籍造字字形為「△△」，一一比對原書字形為「△△」者再以手動取代。

5、配合修改，手動取代：xml 字頻低於 txt 字頻，經比對 txt 及 xml 檔案內容後，發現一部分原舊漢籍造字之前已被人工修改為其他字形，所以剩餘未修改的舊漢籍造字也配合修改。如編號 3883，舊漢籍造字字形為「徧」，但是比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「徧」者，在 xml 檔皆以「遍」字取代，故配合修改，將編號 2044 的「徧」字，手動取代為「遍」字。

6、txt 缺漏致頻次不符，手動全部取代：因 txt 檔案內容缺

漏，導致某舊漢籍造字 xml 字頻高於 txt 字頻，而無關重碼或人工修改因素，經確認後可以手動全部取代的方式，執行該造字的轉碼。如編號 2132 的「弼」字，雖然 txt 字頻與 xml 字頻不符，但可以直接手動全部取代為「弼」字。

- 7、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 5381 的「脛」字，Unicode 字型呈現為「脛」，實際上仍為同一字。
- 8、待造字：舊版漢籍造字字形為電腦缺字，且漢字構形資料庫尚未收錄此字形，等待補造字中，所以「造字」欄空白無法看到字形。如編號 1113 的「王 Δ 喬」。

部分「造字」欄空白者，因已使用其他字形取代，故不另外補造字。如編號 49 的「疒 Δ 廣」字，已由「弊」取代，故不另造「疒 Δ 廣」。

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 857 的「脛」字，Unicode 編碼為 26715，由於 Windows XP 並不支援，仍須改用構字式「 \square 月丞 \square 」。

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、造字分析表「備註」欄標示「手動取代」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說明如下：

- 甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。此處編號同附件一的「造字分析表」。
- 乙、造字：舊漢籍造字字形
- 丙、取代：舊漢籍造字經人工修改，在 xml 檔中取代造字的字形。
- 丁、檔案原文摘錄：摘錄檔案(穆宗.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。

戊、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。

五、造字分析表「備註」欄標示「手動刪除」者，詳細取代內容可參考附件三的手動刪除表。附件三的手動刪除表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。此處編號同附件一的「造字分析表」。

乙、造字：舊版漢籍造字字形

丙、檔案原文摘錄：摘錄檔案(穆宗.xml)中包含該造字的文句段落，其中紅字底線的部分，為被手動刪除的原文。

附件一、《清實錄》「穆宗.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
49		1	1	FA70			疒 △ 廣	弊，錯字。
297	△	183	1	FBED			△	重碼造字，手動取代。
299	形	484	9	FBEF			形	重碼造字，手動取代。
300	□	485	9	FBF0			□	重碼造字，手動取代。
307	○	1	1	FBF7			○	
383	灰	1	1	FCA6			灰	
465	色	1	1	FCF8			色	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
480	小	1	1	FD48			小	
520	閃	1	1	FD70			閃	
532	采	1	1	FD7C	7F59	采		
535	亾	3	3	FDA1			亾	
543	段	1	1	FDA9			段	
552	豕	1	1	FDB2			豕	
615	丩	5	5	FDF1			丩	
618	畱	1	1	FDF4			畱	
625	夫	1	1	FDFB			夫	
662	聿	1	1	FE61			聿	
675	食	1	1	FE6E			食	
681	厂	4	4	FE74			厂	厂，錯字。
705	丩	1	1	FEAE			丩	
733	ナ	1	1	FECA			ナ	
737	牛	3	3	FECE			牛	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
754	回	1	1	FEDF			回	
756	乂	1	1	FEE1			乂	
762	人	242	242	FEE7			人	
763	卩	1	1	FEE8			卩	
775	乚	2	2	FEF4			乚	
782	么	4	4	FEFB			么	
785	ノ	3	3	FEFE			ノ	手動刪除。
853	焞	8	8	8EA5	7134	焞		
857	臄	2	1	8EA9	26715		𠃉月丞巳	重碼造字，手動取代。
874	𠃉	1	1	8EBA	4350		𠃉 𠃉 干	
949	𠃉	16	16	8F46	678F	𠃉		
1022	瑤	1	1	8FB1	73E4	瑤		
1039	稟	19	19	8FC2	69C0	稟		
1067	碁	3	3	8FDE	68CA	碁		
1080	丹	2	2	8FEB	5184	丹		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1113		11	11	904D			王 厶 喬	待造字。
1175		1	1	90AD			土 厶 郎	待造字。
1177	𠵼	1	1	90AF	20C8D		口 厶 丟	
1184	狍	123	123	90B6	72D3	狍		
1186	塏	2	2	90B8	5845	塏		
1226	罰	1	1	90E0	7F78	罰		
1293	礪	2	2	9164			石 厶 畝	
1306	瀨	59	59	9171	3D6F		氵 厶 鳳	
1308	蕙	1	1	9173	85FC	蕙		
1354	禛	6	1	91C3	79A9	禛		重碼造字，手動取代。
1381	穉	3	3	91DE	7A3A	穉		
1383	糠	1	1	91E0	7A45	糠		
1407		1	1	91F8			穴 厶 雋	待造字。
1478		1	1	92A2			欠 厶 欺	待造字。
1483		1	1	92A7			土 厶 盈	待造字。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1507	盲	3	3	92BF	4EAF	盲		
1530	糺	1	1	92D6	7CFA	糺		
1603	倕	1	1	9360	5034	倕		
1635	熾	1	1	93A2	3E05		火△節	
1667		2	2	93C2			氵△窖	待造字。
1694	愆	1	1	93DD	3982		常△心	
1792	滕	1	1	94A2	8542	滕		
1830	護	1	1	94C8	8610	護		
1897	祕	1	1	954C	8890	祕		
1906	裛	40	40	9555	88E6	裛		
1956	綜	142	142	95A9	8AB4	綜		
1964	叢	77	77	95B1	8B5E	叢		
1966	翮	6	6	95B3	4397		喬△羽	
1968	窈	8	8	95B5	25945		穴△孔	
1969	昇	3	3	95B6	23179		日△宀△升	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1970		31	31	95B7	22900		興△心	待造字。
1977	奮	9	9	95BE	5964	奮		
1989	犇	1	1	95CA	246D8		牛△犇	
2050	蹠	3	3	9648	8E4F	蹠		
2066	滌	1	1	9658	3D11		氵△席	
2111	鍤	2	2	96A7	4955		金△敝	
2112	舩	2	2	96A8	4469		舟△古	
2131	彌	29	31	96BB	5F3B	彌		txt 缺漏致頻次不符，手動全部取代。
2230	鎰	27	27	975F	92E5	鎰		
2248	鑫	3	4	9771	9331	鑫		txt 缺漏致頻次不符，手動全部取代。
2332		1	1	97E7			𠃉 ^艹 矢隻 ^口	待造字。
2337	總	1	1	97EC	93D3	總		
2374	冢	1	1	9852	51A1	冢		
2409	闕	1	1	9875	95A7	闕		
2449		32	1	98BF			名△隹	重碼造字，手動取代。 待造字。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2520	餒	2	2	9947	993B	餒		
2680	狍	1	1	9A4A	72CD	狍		狗，錯字。
2700		1	1	9A5E			目 𠃉 王	待造字。
2701		4	4	9A5F			忄 𠃉 興	待造字。
2711	禧	112	113	9A69	3D59		氵 𠃉 喜	txt 缺漏致頻次不符，手動全部取代。
2749	枱	1	1	9AB1	67B1	枱		
2762	碩	1	1	9ABE	78D2	碩		員，錯字。
2781	蒞	1	1	9AD1	8570	蒞		
2793	鋤	2	2	9ADD	9378	鋤		
2838	曇	37	37	9B4B			晶 宀 宜	
2874	餤	3	3	9B6F	9939	餤		
2997	盥	2	2	9C4D	25081		氵 土 宀 口 皿 冫	
3117	罐	2	2	9CE7	58E6	罐		
3161	壘	1	1	9D54	58DC	壘		
3199	骸	17	17	9D7A	9ABD	骸		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3284	颯	1	1	9DF1	78B8	颯		風，錯字。
3801	剗	1	1	8160	5257	剗		
3806	勳	23	23	8165	52E0	勳		
3809	匱	4	4	8168	5333	匱		
3816	厓	9	9	816F	5393	厓		
3817	廁	3	3	8170	53A0	廁		
3835	噍	1	1	81A4	564D	噍		
3840	坂	5	5	81A9	5742	坂		
3842	堃	73	76	81AB	5803	堃		txt 缺漏致頻次不符，手動全部取代。
3866	崐	383	384	81C3	5D10	崐		txt 缺漏致頻次不符，手動全部取代。
3869	帀	14	15	81C6	5E00	帀		txt 缺漏致頻次不符，手動全部取代。
3874		9	9	81CB			忄 厶 稟	待造字。
3883	徧	182	2	81D4	5FA7	徧		遍，配合修改，手動取代。
3916	敷	10	10	81F5	656D	敷		
3923		2423	1	81FC			晶 厶 宜	疊，配合修改，手動取代。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3931	椀	1	1	8245	6900	椀		
3932	椶	6	6	8246	6936	椶		
3945		11	11	8253			氵 厶 义	待造字。
3949	涖	2	2	8257	6D96	涖		
3950	浜	5	5	8258	6D5C	浜		
3955	澗	50	50	825D	3D4E		氵 厶 間	
3959	潛	127	127	8261	6FF3	潛		
3976	犁	577	1	8272	7282	犁		犁，配合修改，手動取代。
3988	璦	1	1	827E	7482	璦		
3989	璦	2	2	82A1	74C8	璦		
3990	甄	50	50	82A2	750E	甄		
4013	盃	9	9	82B9	76CC	盃		
4014	蓋	81	1	82BA	8462	蓋		蓋，配合修改，手動取代。
4015	眦	1	2	82BB	7726	眦		txt 缺漏致頻次不符，手動全部取代。
4027	裨	9	9	82C7	7986	裨		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4035	窈	2	2	82CF	7A7D	窈		
4037	寮	1	1	82D1	7AC2	寮		寮，錯字。
4038	竈	70	1	82D2	7AC8	竈		重碼造字，手動取代。
4044	筓	2	2	82D8	7B6F	筓		
4049	籜	12	12	82DD	7BFA	籜		
4056	葬	1	1	82E4	7CA6	葬		
4068	綫	95	3	82F0	7DAB	綫		線，配合修改，手動取代。
4073	縵	37	37	82F5	7E6E	縵		
4080	羣	446	3	82FC	7FA3	羣		群，配合修改，手動取代。
4098	芪	2	2	834F	82AA	芪		
4108	莽	2	2	8359	844A	莽		
4111	蕓	20	20	835C	8493	蕓		
4134	衛	1189	12	8373	885E	衛		衛，配合修改，手動取代。
4136	衰	2	2	8375	887A	衰		
4142	裛	16	17	837B	8943	裛		txt 缺漏致頻次不符，手動全部取代。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4184	迹	12	1	83C7	8FF9	迹		跡，配合修改，手動取代。
4198	鑛	1	1	83D5	945B	鑛		
4202	隲	4	4	83D9	9696	隲		
4212	鞞	1	1	83E3	97B8	鞞		
4223	顛	1	1	83EE	984B	顛		
4234	羸	47	47	83F9	9A58	羸		
4254	鷓	6	6	844E	9D7E	鷓		
4259	麇	214	214	8453	9E95	麇		Unicode 字型呈現差異。
4267	鼈	4	4	845B	9F08	鼈		
4295	澣	2	2	8477	3D72		灑 △ 舞	
4348	惚	1	1	84CE	506C	惚		
4370	德	9	9	84E4	50E1	德		
4378	隲	1	1	84EC	28F65		陟 △ 馬	
4409	回	4	4	854C	518B	回		
4440	峇	1	1	856B	37E7		山 △ 号	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4469	勻	3	3	85AA	5304	勻		
4554	唬	4	4	8640	55C1	唬		
4566	礪	1	1	864C	78F5	礪		
4623	玩	1	1	86A7	5743	玩		
4636	坩	1	1	86B4	576E	坩		
4655	坩	2	2	86C7	57BB	坩		
4657	竣	2	2	86C9	57C8	竣		
4710	潭	1	1	86FE	58B0	潭		
4712	灑	1	1	8741	7059	灑		
4766	揜	1396	1403	8777	6366	揜		txt 缺漏致頻次不符，手動全部取代。
4799	顛	1	1	87BA	9856	顛		
4812	羗	3	3	87C7	26351		羗△、	
4816	亻	1	1	87CB			亻	
4822	尪	1	1	87D1	5C29	尪		
4910		1	1	886A			亻 △ 畷	待造字。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4933	尙	26	26	88A3	5CDD	尙		
4934	鬢	2	2	88A4	9B09	鬢		
4967		1	1	88C5			山厶桂	待造字。
4992		4	4	88DE			然厶土	待造字。
5011	驚	1	1	88F1	9DD4	驚		
5025		1	1	8940			千厶干	𠂇，取代。
5048	婚	17	17	8957	36F0		女厶昏	
5086		1	1	897D			冗厶卓	冈厶卓，取代。
5132	愷	7	7	89CD	6198	愷		
5173	漱	27	27	89F6	6F44	漱		
5189	宀	5	5	8A47	5B82	宀		
5219	妃	18	6	8A65			女厶巳	妃，配合修改，手動取代。
5222	亘	1	1	8A68	4EB6	亘		
5223		880	886	8A69			𠂇扌宀人夕𠂇	txt 缺漏致頻次不符，手動全部取代。待造字。
5306	簡	770	771	8ADE			𠂇厶閒	txt 缺漏致頻次不符，手動全部取代。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5310	烏	1	1	8AE2	8204	烏		
5311	眇	1	1	8AE3	6644	眇		
5344	瞖	48	48	8B45	55F8	瞖		
5381	脍	1	1	8B6A	8103	脍		Unicode 字型呈現差異。
5398	涸	102	102	8B7B	6E7B	涸		
5402	朮	4	4	8BA1	672F	朮		
5410	柎	6	6	8BA9	67AC	柎		
5422	栝	1	1	8BB5	67F9	栝		
5437	并	1	1	8BC4	5E77	并		
5445	覘	4	4	8BCC	6898	覘		
5493	之	2	2	8BFC			之	
5505	耆	1	1	8C49	264BE		老 叺 占	
5516		1	1	8C54			〇〇束	待造字。
5549	藁	16	16	8C75	85F3	藁		
5556	隸	2	2	8C7C	96B7	隸		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5562	稞	21	21	8CA4	7A09	稞		
5573	舉	12	12	8CAF	399B		與厶心	興厶心，錯字。
5582	鞣	3	3	8CB8	97B5	鞣		
5729	滌	2	2	8DAE	6F34	滌		
5742	澁	1	1	8DBB	6F81	澁		
5753	潛	2	3	8DC6	6F5C	潛		txt 缺漏致頻次不符，手動全部取代。
5764	籊	3	3	8DD1	7C5E	籊		
5854	么	2	2	C6CD	5E7A	么		

附件二、《清實錄》「穆宗.xml」手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果
297	厶		皇__書門而敬度。	皇厶書門而敬度。
299	形		__山戍不 (全部取代)	形山戍不 (全部取代)
300	口		山戍不__ (全部取代)	山戍不口 (全部取代)
857	臚		莊贖。鞠__陳辭。	莊贖。鞠形月丞巳口陳辭。
1354	禊		深仁。謳祝長騰乎千__。	深仁。謳祝長騰乎千禊。

2131	彌		<u>囿</u> (全部取代)	<u>彌</u> (全部取代)
2248	鑫		<u> </u> (全部取代)	<u>鑫</u> (全部取代)
2449	錐		演 <u> </u> 書五十六字而探微	演 <u>名</u> <u>錐</u> 書五十六字而探微
2711	禧		<u> </u> (全部取代)	<u> </u> <u> </u> <u>禧</u> (全部取代)
3842	堃		<u>益</u> (全部取代)	<u>堃</u> (全部取代)
3866	崐		<u>崐</u> (全部取代)	<u>崐</u> (全部取代)
3869	市		<u>市</u> (全部取代)	<u>市</u> (全部取代)
3883	徧	徧	懷柔 <u>徧</u> 及於垓埏。	懷柔 <u>徧</u> 及於垓埏。
			炊 <u>徧</u> 窮檐之	炊 <u>徧</u> 窮檐之
3923		疊	敬畏 <u>疊</u> 形諸篇詠。	敬畏 <u>疊</u> 形諸篇詠。
3976	犁	犁	筆帖式升任伊 <u>犁</u> 理事同知	筆帖式升任伊 <u>犁</u> 理事同知
4014	蓋	蓋	奇徵萃與 <u>贊</u> 之祥。	奇徵萃與 <u>蓋</u> 之祥。
4015	眦		<u>连</u> (全部取代)	<u>眦</u> (全部取代)
4038	竈		治嚴宮府。誅煬 <u>畢</u> 之貂璫。	治嚴宮府。誅煬 <u>竈</u> 之貂璫。
4068	綫	線	<u>責</u> (全部取代)	<u>線</u> (全部取代)
4080	羣	群	<u>仓</u> (全部取代)	<u>群</u> (全部取代)
4134	衛	衛	<u> </u> (全部取代)	<u>衛</u> (全部取代)
4142	裊		<u> </u> (全部取代)	<u>裊</u> (全部取代)
4184	迹	跡	彰施之 <u>壻</u> 。舉昭	彰施之 <u>迹</u> 。舉昭

4766	揜		<u>龜</u> (全部取代)	<u>揜</u> (全部取代)
5219	妃	妃	<u>亼</u> (全部取代)	<u>妃</u> (全部取代)
5223			<u>夾</u> (全部取代)	<u>𠂔</u> 扌 穴 人 夕 <u>𠂔</u> (全部取代)
5306	簡		<u>劍</u> (全部取代)	<u>竹</u> <u>間</u> (全部取代)
5753	潛		<u>旱</u> (全部取代)	<u>潛</u> (全部取代)

附件三、《清實錄》「穆宗.xml」手動刪除表

編號	造字	檔案原文摘錄
785	ノ	勦賊出力。前已有旨以總 <u> </u>