

# 《程史》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  
2008/6/17 陳維君 製作

一、《程史》一書(程史.xml)使用舊版造字 195 個，字頻 665 次，詳如附件一。這 195 個造字中，161 個可轉成 Windows XP 能顯示的字，字頻 529 次；另外 34 個字必須轉成構字式，字頻 136 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 3269 的「憑」字，原字為「憑」。
- 2、校對問題，新版漢籍錯字，人工取代：漢籍在舊轉新的過程中，以人工修改方式取代舊漢籍造字，但是卻取代為錯誤的字形或誤刪造字，因此出現舊版漢籍沒有，卻在新版漢籍出現的錯字。處理方式為把被修改的部分，以及沒被修改到的舊漢籍造字，一併以對應的 Unicode 或構字式取代。如編號 2136 的「广△吾」字，在 xml 檔 31 頁被刪除一次造成頻次不同，需於原缺字處補上，並將其他未經修改的部分手動取代為「广△吾」字。
- 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 814 的「墓」係「蟻」字之異體，故以「蟻」字取代。又編號 3966 的「烟」字，係「煙」字之簡體字，亦以標準字的「煙」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 3846 的「壻」字，Unicode 字型呈現為「壻」，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 2136 的「唐」字，Unicode 編碼為 2222A，由於 Windows XP 並不支援，仍須使用構字式「广△吾」。

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、造字：舊版造字字形。

丙、取代：當實際使用字形與造字不同時，在此標出。

丁、檔案原文摘錄：摘錄檔案(程史.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。

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。

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

附件一、《程史》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
296	厶	9	9	FBEC			厶	
297	厶	18	18	FBED			厶	
298	厶	2	2	FBEE			厶	


編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
299	𠄎	2	2	FBEF			𠄎	
300	𠄏	2	2	FBF0			𠄏	
304	𠄐	2	2	FBF4			𠄐	
360	皇	1	1	FC6D			皇	
527	𠄑	4	4	FD77			𠄑	
556	商	2	2	FDB6	5547	商		
752	𠄒	1	1	FEDD			𠄒	
768	𠄓	2	2	FEED			𠄓	
799	𠄔	1	1	8E4D	9F5A	𠄔		
814	墓	1	1	8E5C	87C7	墓		𡇗，異體字問題。
858	𠄕	1	1	8EAA	7F59	𠄕		
972	犇	3	3	8F5D	7287	犇		
979	𡇖	2	2	8F64	56CF	𡇖		
1010	珏	6	6	8FA5	73CF	珏		𠄖，異體字問題。
1056	阝	1	1	8FD3	9641	阝		
1075	𠄖	18	18	8FE6	9A1F	𠄖		
1133	瓌	1	1	9061	74CC	瓌		
1146	甕	1	1	906E	7506	甕		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
1205	眇	1	1	90CB	770E	眇		
1224	暇	1	1	90DE	7771	暇		
1269	𡗗	1	1	914C	21334		直会土	
1332		1	1	91AD			扌 𡗗 𡗗	攜，異體字問題。
1344		1	1	91B9			衤 𡗗 易	裼，異體字問題。
1394	啗	3	3	91EB	35D6		口 𡗗 啗	啗，異體字問題。
1416	埃	1	1	9242	7AE2	埃		
1565	絲	4	4	92F9	7DDC	絲		
1591	羗	1	1	9354	7F97	羗		
1646	脗	1	1	93AD	8117	脗		
1673	蒒	1	1	93C8	34BC		蒒	
1723		1	1	93FA			** 会 泣	蒒，異體字問題。
1806	躡	1	1	94B0	8E99	躡		
1893	袞	1	1	9548	886E	袞		袞，異體字問題。
1945	詢	2	2	957C	8A7E	詢		
1999	贊	3	3	95D4	8CDB	贊		贊，異體字問題。
2044	踈	1	1	9642	8E08	踈		
2053	躡	1	1	964B	8EA7	躡		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
2136	唐	2	1	96C0	2222A		广△吾	广△吾，校對問題，新版漢籍錯字，人工取代，見附件二。
2168	疇	3	3	96E0	91BB	疇		
2199	鈺	1	1	9740	925F	鈺		
2254	嗤	1	1	9777	20DE3		口△𩺰	嗤，異體字問題。
2327	鋒	1	1	97E2	93E0	鋒		
2409	闕	2	2	9875	95A7	闕		闕，異體字問題。
2415	鬪	1	1	987B	95D8	鬪		鬪，異體字問題。
2505	翻	1	1	98F7	98DC	翻		
2675	侶	1	1	9A45	4F40	侶		
2814	灋	1	1	9AF2	704B	灋		
2858	漣	2	2	9B5F	6EFA	漣		
2902	憖	4	4	9BAD	6138	憖		
3021	忸	1	1	9C65			忸	
3269	憑	1	1	9DE2	51F4	憑		憑，校對問題，舊版漢籍錯字，可用程式全部取代。
3780	儻	1	1	814B	5101	儻		
3789	冲	5	5	8154	51B2	冲		冲，異體字問題。
3791	况	7	7	8156	51B5	况		况，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3801	剗	1	1	8160	5257	剗		
3804	効	3	3	8163	52B9	効		
3810	匱	1	1	8169	5335	匱		
3814	却	24	24	816D	5374	却		卻，異體字問題。
3817	廁	1	1	8170	53A0	廁		廁，異體字問題。
3819	廝	2	2	8172	53AE	廝		廝，異體字問題。
3820	廈	3	3	8173	53A6	廈		廈，異體字問題。
3824	叙	6	6	8177	53D9	叙		敘，異體字問題。
3829	咏	2	2	817C	548F	咏		
3830	咤	7	7	817D	54A4	咤		
3833	嘍	7	7	81A2	5611	嘍		
3840	坂	1	1	81A9	5742	坂		
3846	墻	4	4	81AF	58FB	墻		Unicode 呈現差異。
3852	嫵	2	2	81B5	5A63	嫵		
3858	寃	7	7	81BB	5BC3	寃		冤，異體字問題。
3868	卮	4	4	81C5	5DF5	卮		
3877	迺	12	12	81CE	5EFC	迺		迺，異體字問題。
3883	徧	10	10	81D4	5FA7	徧		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3884	忽	1	1	81D5	6031	忽		
3886	悞	1	1	81D7	609E	悞		誤，異體字問題。
3890	漚	2	2	81DB	6142	漚		
3891	慙	5	5	81DC	6159	慙		慚，異體字問題。
3924	朞	1	1	81FD	671E	朞		期，異體字問題。
3935	槩	10	10	8249	69E9	槩		概，異體字問題。
3945		1	1	8253			汜 𠄎 义	汜，異體字問題。
3948		1	1	8256	6CAD	沝		
3949	涖	8	8	8257	6D96	涖		蒞，異體字問題。
3966	烟	4	4	8268	70DF	烟		煙，異體字問題。
3971	熿	2	2	826D	71BA	熿		
3974	牀	6	6	8270	7240	牀		
3977	狗	1	1	8273	72E5	狗		
3979	猫	9	9	8275	732B	猫		貓，異體字問題。
3981	珉	1	1	8277	73C9	珉		
3989	璿	1	1	82A1	74C8	璿		
3996	疎	2	2	82A8	758E	疎		
4013	盃	3	3	82B9	76CC	盃		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4016	着	1	1	82BC	7740	着		著，異體字問題。
4021	礪	2	2	82C1	78AF	礪		
4025	礪	1	1	82C5	7934	礪		
4027	裨	1	1	82C7	7986	裨		裨，異體字問題。
4028	稟	2	2	82C8	7980	稟		稟，異體字問題。
4035	窳	1	1	82CF	7A7D	窳		
4036	窳	2	2	82D0	7AB0	窳		
4038	竈	1	1	82D2	7AC8	竈		
4041	豎	2	2	82D5	7AEA	豎		豎，異體字問題。
4080	羣	46	46	82FC	7FA3	羣		群，異體字問題。
4085	耻	5	5	8342	803B	耻		
4089	脇	1	1	8346	8107	脇		
4090	脚	3	3	8347	811A	脚		腳，異體字問題。
4091	臯	3	3	8348	81EF	臯		
4093	舐	1	1	834A	445B		舌  氏	舐，異體字問題。
4094	館	1	1	834B	8218	館		
4104	蒞	1	1	8355	8385	蒞		蒞，異體字問題。
4105	菓	1	1	8356	83D3	菓		



編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4107	葱	1	1	8358	8471	葱		蔥，異體字問題。
4114	蘼	1	1	835F	8616	蘼		
4115	藁	3	3	8360	85C1	藁		
4129	虬	1	1	836E	866C	虬		
4131	劬	2	2	8370	8842	劬		
4132		55	55	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	13	13	8373	885E	衛		衛，異體字問題。
4138	裘	2	2	8377	88A0	裘		
4140	衽	1	1	8379	88B5	衽		
4146	覩	4	4	83A1	89A9	覩		
4154	譽	1	1	83A9	8AD0	譽		
4156	謠	3	3	83AB	8B1F	謠		
4158	譌	2	2	83AD	8B4C	譌		Unicode 顯示差異。
4167	賈	1	1	83B6	8CEB	賈		
4176	輒	3	3	83BF	8F19	輒		輒，異體字問題。
4184	迹	15	15	83C7	8FF9	迹		
4193	鈎	9	9	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	1	1	83D2	9262	鉢		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4203	隣	1	1	83DA	96A3	隣		鄰，異體字問題。
4220	穎	2	2	83EB	9834	穎		穎，異體字問題。
4221	頰	1	1	83EC	9833	頰		頰，異體字問題。
4232	隲	1	1	83F7	96B2	隲		鷲，異體字問題。
4235	鬣	1	1	83FA	9AA9	鬣		鬣，異體字問題。
4237	髡	2	2	83FC	9AE0	髡		髡，異體字問題。
4240	髯	1	1	8440	9AF4	髯		
4242	鬪	7	7	8442	9B2D	鬪		Unicode 呈現差異。
4256	鷄	2	2	8450	9DC4	鷄		
4267	鼈	7	7	845B	9F08	鼈		
4278	亘	2	2	8466	4E98	亘		
4299	焯	1	1	847B	242F2		火𠂇𠂇	
4307	罵	1	1	84A5	99E1	罵		罵，異體字問題。
4308	劫	1	1	84A6	5227	劫		劫，異體字問題。
4309	劫	4	4	84A7	523C	劫		劫，異體字問題。
4371	倜	1	1	84E5	50F4	倜		
4450	狷	1	1	8575	524F	狷		Unicode 呈現差異。
4458	勸	2	2	857D	52CC	勸		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4469	勻	4	4	85AA	5304	勻		
4495	尪	1	1	85C4	9EA8	尪		
4496	段	2	2	85C5	53DA	段		段、段，異體字問題，必須人工取代，見附件二。
4507	吳	49	49	85D0	5434	吳		吳，異體字問題。
4533	唸	1	1	85EA	552B	唸		
4603	憇	8	8	8671	6187	憇		
4735	壠	4	4	8758	58E0	壠		壠，異體字問題。
4778	鞞	2	2	87A5	97B4	鞞		
4782	鞞	1	1	87A9	97BA	鞞		
4817	寶	1	1	87CC	5BF3	寶		寶，異體字問題。
4836	嗜	2	2	87DF	20D32		口 ㄩ 沓	
4878	冗	1	1	884A	5B82	冗		冗，異體字問題。
4902	充	2	2	8862	5156	充		充，異體字問題。
4910		3	3	886A			彳 ㄩ 畷	淫，異體字問題。
5008	幪	1	1	88EE	5E5E	幪		
5013	彳	1	1	88F3			彳	
5031	閨	1	1	8946	95B4	閨		閨，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5111	旆	3	3	89B8	65BE	旆		旆，異體字問題。
5135	迂	1	1	89D0			迂△亏	迂，異體字問題。
5176		5	5	89F9			沉△冗	沉，異體字問題。
5192	隙	1	1	8A4A	969F	隙		
5225	卹	2	2	8A6B	460F		血△卩	卹，異體字問題。
5290	携	1	1	8ACE	643A	携		
5299	戟	1	1	8AD7	39B8		卓△戈	戟，異體字問題。
5302	昏	2	2	8ADA	662C	昏		昏，異體字問題。
5312		1	1	8AE4			形女〃冈比	媿，異體字問題。
5433	爰	1	1	8BC0	590B	爰		
5483	眦	2	2	8BF2	7721	眦		
5515	軌	1	1	8C53			車△丸	軌，異體字問題。
5542	鑠	1	1	8C6E	93C1	鑠		
5549	薰	1	1	8C75	85F3	薰		
5558	龕	1	1	8C7E	9E84	龕		
5559	欸	1	1	8CA1	6B35	欸		欸，異體字問題。
5578	弛	5	5	8CB4	38AE		弓△也	弛，異體字問題。
5583	殮	1	1	8CB9	98F1	殮		殮，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5587	歛	1	1	8CBD	6B58	歛		
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。
5595	朶	3	3	8CC5	579C	朶		朶，異體字問題。
5697	煑	1	1	8D6C	7151	煑		煮，異體字問題。
5702	厨	1	1	8D71	53A8	厨		廚，異體字問題。
5703	𪗇	1	1	8D72	9F53	𪗇		
5704	𪗈	1	1	8D73	6C0A	𪗈		𪗈，異體字問題。
5705	廡	2	2	8D74	5ED0	廡		廡，異體字問題。
5706	敍	4	4	8D75	654D	敍		敍，異體字問題。
5713	臧	2	2	8D7C	481E		足△臧	臧，異體字問題。
5717	陘	1	1	8DA2	967F	陘		
5764	籊	3	3	8DD1	7C5E	籊		
5792	籊	1	1	8DED	7C55	籊		籊，異體字問題。
5797	灑	1	1	8DF2	7054	灑		

附件二、《程史》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果
2136	广△吾		蓋𪗇亭在南，岵臺在北	蓋广△吾亭在南，岵臺在北
			<qe>亭在南岵臺在北	<qe>广△吾亭在南岵臺在北

編號	造字	取代	檔案原文摘錄	手動取代結果
4496	段		乃 <u>𠄎</u> 借字	乃 <u>段</u> 借字
		段	易以 <u>𠄎</u> 文昌之作	易以 <u>段</u> 文昌之作