

《不下帶編》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
2008/5/23 葉冠孜 製作

一、《不下帶編》一書(不下帶編.xml)使用舊版造字 140 個，字頻 578 次，詳如附件一。這 140 個造字中，116 個可轉成 Windows XP 能顯示的字，字頻 506 次；另外 24 個字必須轉成構字式，字頻 72 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 3238 的「酉广艹系」字，原字為「縻」。
- 2、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 3789 的「冲」係「沖」字之異體，故以「沖」字取代。
- 3、校對問題，未確定字形，必須人工取代：在舊版漢籍電子文獻中即存在的未確定字形，持續留存在新版漢籍電子文獻中，必須人工一一比對修改，如編號 309「」，可參考附件二未確定字形「」手動取代表。
- 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 2212 的「襪」字，Unicode 編碼為 277CC，由於 Windows XP 並不支援，仍須使用構字式「衤𠂔羅」。

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、《不下帶編》一書(不下帶編.xml)中未確定字形「?

甲、檔案原文摘錄：摘錄檔案(不下帶編.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。

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、《不下帶編》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠂔	5	5	FBEC			𠂔	
297	𠂔	3	3	FBED			𠂔	
307	𠂔	2	2	FBF7			𠂔	
309	?	1	1	FBF9			?	未確定字形，人工取代，參考附件二。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
571	卩	1	1	FDC5			卩	
770	扌	1	1	FEEF			扌	
792	帛	4	4	8E46	5E2C	帛		
969	插	4	4	8F5A	7250	插		
1010	珏	1	1	8FA5	73CF	珏		
1133	瓌	4	4	9061	74CC	瓌		
1197	礪	1	1	90C3	40FA		石 厶 靡	
1201	昵	1	1	90C7	7724	昵		
1212	戠	1	1	90D2	39B7		冫 厶 戈	
1347	𡗗	1	1	91BC	21E05		冫 山 弓 女 冫	
1416	埃	6	6	9242	7AE2	埃		
1642	脉	3	3	93A9	8109	脉		
1723		1	1	93FA			艹 厶 泣	蒞，異體字問題。
1871	蟲	1	1	94F1	87C1	蟲		
1900	桂	1	1	954F	88BF	桂		
1981	讐	2	2	95C2	8B90	讐		
2021	趲	3	3	95EA	8D82	趲		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2044	踈	4	4	9642	8E08	踈		
2046	踪	1	1	9644	8E2A	踪		
2129		2	2	96B9			友彡皿	友彡皿，異體字問題。
2189	瞿	1	1	96F5	23C3D		毛彡瞿	
2190	璿	1	1	96F6	6BFA	璿		Unicode 呈現差異。
2212	禪	4	4	974D	277CC		禪 彡羅	
2350	慘	1	1	97F9	50AA	慘		
2408	鬧	2	2	9874	9599	鬧		鬧，異體字問題。
2742	抄	1	1	9AAA	6331	抄		
3054	蝮	3	3	9CA8	45B5		蝮	
3184		1	1	9D6B	6AC9	櫛		櫛，異體字問題。
3227	艫	6	6	9DB8	826A	艫		
3238	𪗇	1	1	9DC3			𪗇 西 彡 系	西 彡 𪗇，舊版漢籍錯字，可用程式全部取代。
3789	冲	2	2	8154	51B2	冲		沖，異體字問題。
3792	凉	2	2	8157	51C9	凉		涼，異體字問題。
3794	湊	1	1	8159	51D1	湊		湊，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3797	凭	1	1	815C	51ED	凭		憑，異體字問題。
3805	勅	7	7	8164	52C5	勅		敕，異體字問題。
3809	匱	2	2	8168	5333	匱		
3814	却	10	10	816D	5374	却		卻，異體字問題。
3815	𡗗	1	1	816E	537A	𡗗		
3817	廁	3	3	8170	53A0	廁		廁，異體字問題。
3819	廡	2	2	8172	53AE	廡		廡，異體字問題。
3826	叶	5	5	8179	53F6	叶		
3829	咏	44	44	817C	548F	咏		
3832	嘅	1	1	81A1	5605	嘅		Unicode 呈現差異。
3846	墻	7	7	81AF	58FB	墻		Unicode 呈現差異。
3858	寃	6	6	81BB	5BC3	寃		冤，異體字問題。
3860	𡗗	1	1	81BD	5C20	𡗗		
3870	𡗗	2	2	81C7	5E0B	𡗗		
3883	徧	4	4	81D4	5FA7	徧		
3886	悞	6	6	81D7	609E	悞		誤，異體字問題。
3890	憑	1	1	81DB	6142	憑		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3892	睺	3	3	81DD	3B0B		日 𠄎 侯	
3909	擡	2	2	81EE	64E1	擡		
3924	朞	1	1	81FD	671E	朞		
3931	椀	11	11	8245	6900	椀		
3932	椶	1	1	8246	6936	椶		
3942	毡	1	1	8250	6BE1	毡		
3961	灾	2	2	8263	707E	灾		
3966	烟	6	6	8268	70DF	烟		煙，異體字問題。
3974	牀	9	9	8270	7240	牀		
3975	牕	3	3	8271	7255	牕		Unicode 呈現差異。
3996	疎	11	11	82A8	758E	疎		
4008	癯	2	2	82B4	766F	癯		
4010	臯	1	1	82B6	7690	臯		
4016	着	4	4	82BC	7740	着		著，異體字問題。
4028	稟	1	1	82C8	7980	稟		稟，異體字問題。
4034	穉	1	1	82CE	7A49	穉		
4036	窰	2	2	82D0	7AB0	窰		窯，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4038	竈	2	2	82D2	7AC8	竈		灶，異體字問題。
4039	竚	1	1	82D3	7ADA	竚		
4042	笋	1	1	82D6	7B0B	笋		
4066	綺	1	1	82EE	7D5D	綺		
4068	綫	3	3	82F0	7DAB	綫		線，異體字問題。
4078		1	1	82FA			巳𠂇卅	异，異體字問題。
4080	羣	22	22	82FC	7FA3	羣		群，異體字問題。
4090	脚	9	9	8347	811A	脚		腳，異體字問題。
4091	臯	2	2	8348	81EF	臯		
4107	葱	4	4	8358	8471	葱		
4114	藜	1	1	835F	8616	藜		
4115	藁	1	1	8360	85C1	藁		
4129	虬	1	1	836E	866C	虬		
4132		17	17	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	3	3	8373	885E	衛		衛，異體字問題。
4139	袴	2	2	8378	88B4	袴		
4145	羈	1	1	837E	898A	羈		Unicode 呈現差異。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4146	覩	6	6	83A1	89A9	覩		
4173	駝	2	2	83BC	8EAD	駝		
4176	輒	1	1	83BF	8F19	輒		
4178	輓	3	3	83C1	8F2D	輓		
4184	迹	2	2	83C7	8FF9	迹		
4188	邨	19	19	83CB	90A8	邨		
4193	鈎	5	5	83D0	920E	鈎		
4194	鈐	1	1	83D1	9241	鈐		
4195	鉢	2	2	83D2	9262	鉢		
4203	隣	11	11	83DA	96A3	隣		鄰，異體字問題。
4220	穎	1	1	83EB	9834	穎		穎，異體字問題。
4234	羸	1	1	83F9	9A58	羸		
4237	髡	1	1	83FC	9AE0	髡		髡，異體字問題。
4242	鬪	2	2	8442	9B2D	鬪		Unicode 呈現差異。
4259	麇	1	1	8453	9E95	麇		Unicode 呈現差異。
4267	鼈	4	4	845B	9F08	鼈		
4271	齧	2	2	845F	9F69	齧		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4278	亘	1	1	8466	4E98	亘		
4307	罵	1	1	84A5	99E1	罵		罵，異體字問題。
4309	刼	1	1	84A7	523C	刼		劫，異體字問題。
4369	晬	2	2	84E3	666B	晬		
4437	廩	1	1	8568	5EEA	廩		廩，異體字問題。
4447	捏	1	1	8572	63D1	捏		捏，異體字問題。
4450	狝	2	2	8575	524F	狝		創，異體字問題。
4485	廛	1	1	85BA	3553		厂 厶 墨	廛，異體字問題。
4507	吳	77	77	85D0	5434	吳		吳，異體字問題。
4603	憇	1	1	8671	6187	憇		
4785	贗	1	1	87AC	8D0B	贗		贗，異體字問題。
4805	瀆	2	2	87C0	51DF	瀆		瀆，異體字問題。
4902	兗	3	3	8862	5156	兗		
4910		1	1	886A			彳 厶 畚	淫，異體字問題。
4940	韻	20	20	88AA	97F5	韻		韻，異體字問題。
4996	斲	3	3	88E2	65B5	斲		
5019	畧	8	8	88F9			畧	略，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5046	彝	3	3	8955	5F5C	彝		
5176		9	9	89F9			氵 厶 冫	沉，異體字問題。
5184	增	1	1	8A42	71B7	增		
5225	邨	3	3	8A6B	460F		血 厶 卩	
5241	鷺	28	28	8A7B	9D5E	鷺		鵠，異體字問題。
5290	携	7	7	8ACE	643A	携		攜，異體字問題。
5405	鞞	1	1	8BA4	6A37	鞞		Unicode 呈現差異。
5409	盞	1	1	8BA8	76CB	盞		
5439	栢	2	2	8BC6	6822	栢		
5470	榑	1	1	8BE5	68C5	榑		
5551	賈	1	1	8C77	8CF7	賈		
5559	欸	5	5	8CA1	6B35	欸		款，異體字問題。
5703	𪗇	1	1	8D72	9F53	𪗇		
5704	𪗈	3	3	8D73	6C0A	𪗈		
5742	澁	1	1	8DBB	6F81	澁		澀，異體字問題。
5753	潜	2	2	8DC6	6F5C	潜		
5767	濶	1	1	8DD4	24050		氵 厶 閣	

附件二、《不下帶編》未確定字形「」手動取代表

檔案原文摘錄	手動取代結果
皆因誤從  作從卅也。	皆因誤從  作從卅也。