

《泊宅編》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
2008/6/23 葉冠孜 製作

一、《泊宅編》一書(泊宅編.xml)使用舊版造字 80 個，字頻 336 次，詳如附件一。這 80 個造字中，72 個可轉成 Windows XP 能顯示的字，字頻 281 次；另外 8 個字必須轉成構字式，字頻 55 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、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 3791 的「况」係「況」字之異體，故以「況」字取代。又編號 5034 的「弃」字，係「棄」字之簡體字，亦以標準字的「棄」字取代。
- 2、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。

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 5695 的「勝」字，Unicode 編碼為 3B3A，由於 Windows XP 並不支援，仍須使用構字式「朕△巾」。

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、《泊宅編》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	厶	1	1	FBEC			厶	
297	厶	14	14	FBED			厶	
770	扌	2	2	FEEF			扌	
813	旣	1	1	8E5B	65E4	旣		
860	裊	1	1	8EAC	890F	裊		
899	馳	2	2	8ED3	99DE	馳		
1067	碁	5	5	8FDE	68CA	碁		碁，異體字問題。
1077	稟	3	3	8FE8	7A3E	稟		
1126	璿	9	9	905A	74B9	璿		
1893	袞	1	1	9548	886E	袞		
2044	踈	2	2	9642	8E08	踈		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2046	踪	1	1	9644	8E2A	踪		
2151	𠃉	3	3	96CF	961D	𠃉		
2213	銛	1	1	974E	9277	銛		
2829	月	1	1	9B42			月	
2885	𠃉	1	1	9B7A	544C	𠃉		
2988	𧄸	1	1	9C44	9C6C	𧄸		
3791	况	1	1	8156	51B5	况		况，異體字問題。
3804	効	2	2	8163	52B9	効		效，異體字問題。
3814	却	1	1	816D	5374	却		卻，異體字問題。
3816	厓	1	1	816F	5393	厓		
3843	堦	1	1	81AC	5826	堦		
3860	𧄸	2	2	81BD	5C20	𧄸		
3876	廻	2	2	81CD	5EFB	廻		迴，異體字問題。
3883	徧	6	6	81D4	5FA7	徧		
3891	慙	1	1	81DC	6159	慙		慚，異體字問題。
3931	椀	6	6	8245	6900	椀		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3935	槩	1	1	8249	69E9	槩		概，異體字問題。
3966	烟	4	4	8268	70DF	烟		煙，異體字問題。
3974	牀	13	13	8270	7240	牀		
3975	牕	2	2	8271	7255	牕		Unicode 呈現差異。
3979	猫	3	3	8275	732B	猫		貓，異體字問題。
3996	疎	1	1	82A8	758E	疎		
4010	阜	11	11	82B6	7690	阜		
4016	着	1	1	82BC	7740	着		著，異體字問題。
4025	礪	2	2	82C5	7934	礪		
4040	竝	11	11	82D4	7ADD	竝		
4044	筋	2	2	82D8	7B6F	筋		
4055	粃	2	2	82E3	7C83	粃		
4080	羣	12	12	82FC	7FA3	羣		群，異體字問題。
4090	脚	3	3	8347	811A	脚		腳，異體字問題。
4115	藁	1	1	8360	85C1	藁		
4131	劔	1	1	8370	8842	劔		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4132		33	33	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	3	3	8373	885E	衛		衛，異體字問題。
4140	衽	1	1	8379	88B5	衽		
4146	覩	11	11	83A1	89A9	覩		
4158	譌	2	2	83AD	8B4C	譌		Unicode 呈現差異。
4184	迹	5	5	83C7	8FF9	迹		跡，異體字問題。
4195	鉢	2	2	83D2	9262	鉢		
4203	隣	1	1	83DA	96A3	隣		鄰，異體字問題。
4213	鞞	1	1	83E4	97BE	鞞		
4262	麪	15	15	8456	9EAA	麪		麵，異體字問題。
4267	鼈	1	1	845B	9F08	鼈		
4278	亘	2	2	8466	4E98	亘		
4342	倮	3	3	84C8	502E	倮		
4455	劒	1	1	857A			僉𠂇刀	劍，異體字問題。
4458	勑	3	3	857D	52CC	勑		Unicode 呈現差異。
4461	勑	9	9	85A2	52D4	勑		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4469	勻	2	2	85AA	5304	勻		
4507	吳	54	54	85D0	5434	吳		吳，異體字問題。
4515	呪	1	1	85D8	546A	呪		
4735	壠	6	6	8758	58E0	壠		
4772	輝	1	1	877D	9BF6	輝		
4773	姉	1	1	877E	59C9	姉		Unicode 呈現差異。
4957	崑	11	11	88BB	5D53	崑		
5031	閨	1	1	8946	95B4	閨		閨，異體字問題。
5034	弃	2	2	8949	5F03	弃		棄，異體字問題。
5075	彡	1	1	8972			彡	
5095	來	1	1	89A8	6765	來		來，異體字問題。
5131	窻	4	4	89CC	7ABB	窻		Unicode 呈現差異。
5133	憇	1	1	89CE	6199	憇		
5241	鶯	1	1	8A7B	9D5E	鶯		鶯，異體字問題。
5327	嶂	1	1	8AF3	5D2A	嶂		
5582	鞣	1	1	8CB8	97B5	鞣		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。
5693	𪗇	2	2	8D68	9E7B	𪗇		
5695	𪗈	2	2	8D6A	3B3A		𪗈△巾	
5706	𪗉	4	4	8D75	654D	𪗉		𪗉，異體字問題。
5707	𪗊	3	3	8D76	6830	𪗊		