

《邵氏聞見後錄》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 陳維君 製作

一、《邵氏聞見後錄》一書(邵氏聞見後錄.xml)使用舊版造字 118 個，字頻 612 次，詳如附件一。這 118 個造字中，96 個可轉成 Windows XP 能顯示的字，字頻 514 次；另外 22 個字必須轉成構字式，字頻 98 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 5218 的「一寬、」係「寬」字之異體，故以「寬」字取代。又編號 4016 的「着」字，係「著」字之簡體字，亦以標準字的「著」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 4773 的「姉」字，Unicode 字型呈現為「姉」，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 998 的「潰」字，Unicode 編碼為 3D52，由於 Windows XP 並不支援，仍須使用構字式「費」。

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、《邵氏聞見後錄》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
10	辵	1	1	FA49	8FA4	辵		
296	厶	7	7	FBEC			厶	
297	厶	17	17	FBED			厶	
299	𠂇	1	1	FBEF			𠂇	
300	𠂇	1	1	FBF0			𠂇	
336	畝	1	1	FC55			畝	
379	支	1	1	FCA2			支	
768	𠂇	1	1	FEED			𠂇	
771	𠂇	6	6	FEF0			𠂇	
780	𠂇	4	4	FEF9			𠂇	
781	𠂇	1	1	FEFA			𠂇	
881	颺	2	2	8EC1	98C7	颺		
884		1	1	8EC4			𠂇厶員	胃，異體字問題。
887	鷺	2	2	8EC7	9E0E	鷺		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
893	峩	3	3	8ECD	5CE9	峩		峨，異體字問題。
972	犇	2	2	8F5D	7287	犇		
998	潰	1	1	8F77	3D52		氵 ㄩ 費	
1024	虵	4	4	8FB3	8675	虵		
1076	高	1	1	8FE7	5368	高		Unicode 呈現差異。
1092	瑞	1	1	8FF7	747A	瑞		
1126	璿	3	3	905A	74B9	璿		
1133	瓌	1	1	9061	74CC	瓌		
1146	蕘	1	1	906E	7506	蕘		
1240		1	1	90EE			𠄎革卅 𠄎伐 𠄎	革 ㄩ 蕘，異體字問題。
1649		1	1	93B0			氵 ㄩ 𧈧	氵 ㄩ 𧈧，異體字問題。
1710	爍	1	1	93ED	3E01		火 ㄩ 業	
1994	達	1	1	95CF	6FBE	達		
2021	趁	1	1	95EA	8D82	趁		
2402	勳	1	1	986E	52EE	勳		
2448	隼	1	1	98BE	96BD	隼		
2520	饊	4	4	9947	993B	饊		
2834	攵	5	5	9B47	6535	攵		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3204	體	1	1	9DA1	8EC6	體		
3255	善	96	96	9DD4	5584	善		
3789	冲	3	3	8154	51B2	冲		沖，異體字問題。
3791	況	6	6	8156	51B5	況		況，異體字問題。
3792	凉	2	2	8157	51C9	凉		涼，異體字問題。
3796	决	1	1	815B	51B3	决		決，異體字問題。
3797	凭	2	2	815C	51ED	凭		憑，異體字問題。
3804	効	7	7	8163	52B9	効		
3805	勅	2	2	8164	52C5	勅		
3814	却	1	1	816D	5374	却		卻，異體字問題。
3819	廡	1	1	8172	53AE	廡		廡，異體字問題。
3846	壻	5	5	81AF	58FB	壻		Unicode 呈現差異。
3858	寃	12	12	81BB	5BC3	寃		冤，異體字問題。
3877	迺	2	2	81CE	5EFC	迺		迺，異體字問題。
3883	徧	2	2	81D4	5FA7	徧		
3891	慙	7	7	81DC	6159	慙		慚，異體字問題。
3924	朞	3	3	81FD	671E	朞		期，異體字問題。
3935	槩	2	2	8249	69E9	槩		概，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3944	冰	2	2	8252	6C37	冰		冰，異體字問題。
3946	汹	1	1	8254	6C79	汹		
3949	涖	1	1	8257	6D96	涖		
3974	牀	9	9	8270	7240	牀		
3979	猫	1	1	8275	732B	猫		貓，異體字問題。
3996	疎	18	18	82A8	758E	疎		
4015	眦	1	1	82BB	7726	眦		
4016	着	12	12	82BC	7740	着		著，異體字問題。
4028	稟	1	1	82C8	7980	稟		稟，異體字問題。
4035	窳	1	1	82CF	7A7D	窳		
4038	竈	5	5	82D2	7AC8	竈		
4040	竝	1	1	82D4	7ADD	竝		
4042	笋	1	1	82D6	7B0B	笋		
4068	綫	1	1	82F0	7DAB	綫		線，異體字問題。
4080	羣	28	28	82FC	7FA3	羣		群，異體字問題。
4081	翱	3	3	82FD	7FFA	翱		
4090	脚	4	4	8347	811A	脚		腳，異體字問題。
4091	臯	29	29	8348	81EF	臯		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4115	藁	7	7	8360	85C1	藁		
4118	藜	1	1	8363	8602	藜		
4129	虬	1	1	836E	866C	虬		
4132		35	35	8371			血𠄎𠄎	眾，異體字問題。
4134	衛	16	16	8373	885E	衛		衛，異體字問題。
4142	褻	3	3	837B	8943	褻		
4145	羈	1	1	837E	898A	羈		Unicode 呈現差異。
4146	覩	7	7	83A1	89A9	覩		
4150	訕	2	2	83A5	8A29	訕		
4165	獾	2	2	83B4	8C9B	獾		
4184	迹	28	28	83C7	8FF9	迹		
4220	穎	8	8	83EB	9834	穎		穎，異體字問題。
4226	凜	1	1	83F1	98E1	凜		
4240	髯	3	3	8440	9AF4	髯		
4242	鬪	15	15	8442	9B2D	鬪		Unicode 呈現差異。
4265	鼃	1	1	8459	9F03	鼃		
4278	亘	1	1	8466	4E98	亘		
4307	罵	6	6	84A5	99E1	罵		罵，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4309	刼	10	10	84A7	523C	刼		劫，異體字問題。
4450	翔	7	7	8575	524F	翔		Unicode 呈現差異。
4460	勅	5	5	85A1	52D1	勅		
4461	勛	1	1	85A2	52D4	勛		
4507	吳	43	43	85D0	5434	吳		吳，異體字問題。
4546	鷓	1	1	85F7	9D76	鷓		
4603	憇	1	1	8671	6187	憇		
4702	場	1	1	86F6	5872	場		場，異體字問題。
4773	姉	1	1	877E	59C9	姉		Unicode 呈現差異。
4805	瀆	2	2	87C0	51DF	瀆		瀆，異體字問題。
4816	彳	3	3	87CB			彳	
4907	橐	1	1	8867	69D6	橐		橐，異體字問題。
4961	盼	1	1	88BF	80B9	盼		
4996	斲	2	2	88E2	65B5	斲		
5008	幞	1	1	88EE	5E5E	幞		
5028	昇	2	2	8943	662A	昇		
5074	尫	1	1	8971	5C2B	尫		尫，異體字問題。
5103	忍	2	2	89B0	43F0		月𠂔忍	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5176		2	2	89F9			氵厶元	沉，異體字問題。
5218		7	7	8A64			宀宀寬、宀	寬，異體字問題。
5225	卹	3	3	8A6B	460F		血厶卩	卹，異體字問題。
5392	息	1	1	8B75	60A4	息		
5402	朮	1	1	8BA1	672F	朮		
5422	梯	3	3	8BB5	67F9	梯		
5434	湟	1	1	8BC1	6E7C	湟		湟，異體字問題。
5447	躡	1	1	8BCE	8E83	躡		躡，異體字問題。
5542	鑠	2	2	8C6E	93C1	鑠		
5549	藁	2	2	8C75	85F3	藁		
5559	欸	1	1	8CA1	6B35	欸		
5591	杞	1	1	8CC1	233CC		木厶巳	杞，異體字問題。
5705	廐	1	1	8D74	5ED0	廐		廐，異體字問題。
5706	敘	10	10	8D75	654D	敘		敘，異體字問題。