

《蘆浦筆記》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 葉冠孜 製作

一、《蘆浦筆記》一書(蘆浦筆記.xml)使用舊版造字 113 個，字頻 329 次，詳如附件一。這 113 個造字中，86 個可轉成 Windows XP 能顯示的字，字頻 254 次；另外 27 個字必須轉成構字式，字頻 75 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，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。
- 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 3864 的「峯」係「峰」字之異體，故以「峰」字取代。又編號 4068 的「綫」字，係「線」字之簡體字，亦以標準字的「線」字取代。
- 3、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 4956 的「鑲」字，Unicode

編碼為 2844A，由於 Windows XP 並不支援，仍須使用構字式「多厶農」。

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、《蘆浦筆記》一書(蘆浦筆記.xml)中未確定字形「」共出現 2 次，詳細取代內容可參考附件二未確定字形「」手動取代表。
附件二未確定字形「」手動取代表的欄位說明如下：

- 甲、檔案原文摘錄：摘錄檔案(蘆浦筆記.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、《蘆浦筆記》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	12	12	FBEC			厶	
297	厶	8	8	FBED			厶	
298	厶	1	1	FBEE			厶	
299	形	3	3	FBEF			形	
300	口	3	3	FBF0			口	
309		2	2	FBF9				未確定字形，人工取代，參考附件二。
639	;	1	1	FE4A			;	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
707	ㄩ	1	1	FEB0			ㄩ	
733	ナ	2	2	FECA			ナ	
739	勺	1	1	FED0			勺	
752	ㄨ	3	3	FEDD			ㄨ	
758	四	1	1	FEE3			四	
777	宀	2	2	FEF6			宀	
825	坳	1	1	8E67	573D	坳		
842	蚪	1	1	8E78	3AB7	蚪		
846	焯	1	1	8E7C	713B	焯		
893	峩	1	1	8ECD	5CE9	峩		峨，異體字問題。
1077	稟	1	1	8FE8	7A3E	稟		
1091	漚	1	1	8FF6	58CD	漚		漚，異體字問題。
1151	劄	1	1	9073	5273	劄		
1170	瞳	1	1	90A8	7583	瞳		
1205	眇	1	1	90CB	770E	眇		
1497		2	2	92B5			𪛗米ノ凡𪛗	𪛗，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1530	紕	1	1	92D6	7CFA	紕		
1642	脉	1	1	93A9	8109	脉		
1981	讐	8	8	95C2	8B90	讐		
2046	踪	1	1	9644	8E2A	踪		
2081	餵	6	6	9667	9926	餵		
2501	颺	1	1	98F3	98C3	颺		飄，異體字問題。
2502	颺	2	2	98F4	98C8	颺		颺，異體字問題。
2602	鰐	1	1	99BB	9C10	鰐		
2878	叟	1	1	9B73	5081	叟		
3814	却	4	4	816D	5374	却		卻，異體字問題。
3816	厓	1	1	816F	5393	厓		
3817	廁	4	4	8170	53A0	廁		廁，異體字問題。
3824	叙	3	3	8177	53D9	叙		叙，異體字問題。
3835	噍	1	1	81A4	564D	噍		
3842	莖	4	4	81AB	5803	莖		
3843	堦	1	1	81AC	5826	堦		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3864	峯	7	7	81C1	5CEF	峯		峰，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	2	2	81DC	6159	慙		慙，異體字問題。
3935	槩	2	2	8249	69E9	槩		概，異體字問題。
3939	槩	1	1	824D	6AFD	槩		槩，異體字問題。
3940	鬱	1	1	824E	6B1D	鬱		鬱，異體字問題。
3957	凜	1	1	825F	6F9F	凜		凜，異體字問題。
3963	炆	1	1	8265	7081	炆		
3974	牀	7	7	8270	7240	牀		
3978	猪	1	1	8274	732A	猪		豬，異體字問題。
3990	甄	13	13	82A2	750E	甄		
3996	疎	5	5	82A8	758E	疎		
4014	蓋	9	9	82BA	8462	蓋		蓋，異體字問題。
4034	穉	1	1	82CE	7A49	穉		
4068	綫	1	1	82F0	7DAB	綫		線，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4080	羣	8	8	82FC	7FA3	羣		群，異體字問題。
4090	脚	1	1	8347	811A	脚		腳，異體字問題。
4100	苳	1	1	8351	82FD	苳		
4115	藁	1	1	8360	85C1	藁		
4129	虬	8	8	836E	866C	虬		
4132		12	12	8371			血𠂇𠂇	眾，異體字問題。
4146	覩	2	2	83A1	89A9	覩		
4158	譌	30	30	83AD	8B4C	譌		Unicode 呈現差異。
4184	迹	5	5	83C7	8FF9	迹		
4186	遡	2	2	83C9	9061	遡		
4190	邳	1	1	83CD	90C4	邳		
4221	頰	1	1	83EC	9833	頰		
4237	髡	2	2	83FC	9AE0	髡		髡，異體字問題。
4240	髯	3	3	8440	9AF4	髯		
4242	鬪	4	4	8442	9B2D	鬪		Unicode 呈現差異。
4254	鷓	1	1	844E	9D7E	鷓		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4262	麪	2	2	8456	9EAA	麪		麵，異體字問題。
4272	龐	1	1	8460	9F8E	龐		
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。
4428	躡	1	1	855F	8E70	躡		躡，異體字問題。
4455	劒	5	5	857A			僉 刀	劍，異體字問題。
4461	劬	2	2	85A2	52D4	劬		
4469	勻	1	1	85AA	5304	勻		
4482	廛	1	1	85B7	53AA	廛		廛，異體字問題。
4507	吳	33	33	85D0	5434	吳		吳，異體字問題。
4603	憇	1	1	8671	6187	憇		
4643	徃	1	1	86BB	5F83	徃		
4653	魯	2	2	86C5	9C81	魯		魯，異體字問題。
4817	寶	1	1	87CC	5BF3	寶		寶，異體字問題。
4904	尅	1	1	8864	5C05	尅		尅，異體字問題。
4910		1	1	886A			氵 畧	淫，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4940	韻	1	1	88AA	97F5	韻		韻，異體字問題。
4956	鑲	4	4	88BA	2844A		多 𠂇 農	
5013	汙	1	1	88F3			汙	
5098	寤	1	1	89AB	7AB9	寤		
5218		1	1	8A64			𠂇 宀 覓、𠂇	寬，異體字問題。
5289	慤	9	9	8ACD	6164	慤		慤，異體字問題。
5306		1	1	8ADE			𠂇 𠂇 閒	簡，異體字問題。
5319	晰	1	1	8AEB	6663	晰		
5437	并	1	1	8BC4	5E77	并		并，異體字問題。
5522	喋	1	1	8C5A	5DAB	喋		
5549	藁	1	1	8C75	85F3	藁		
5556	隸	1	1	8C7C	96B7	隸		隸，異體字問題。
5558	龕	2	2	8C7E	9E84	龕		
5559	欸	1	1	8CA1	6B35	欸		款，異體字問題。
5565	斛	3	3	8CA7	469F		角 𠂇 斗	
5568	开	1	1	8CAA	5F00	开		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5583	殮	1	1	8CB9	98F1	殮		
5585	禹	1	1	8CBB	3762		宀△禹	字，異體字問題。
5586	歛	1	1	8CBC	6B57	歛		
5587	歛	2	2	8CBD	6B58	歛		
5633	眙	1	1	8CEB			日△台	
5654	籊	1	1	8D41	7C4B	籊		
5694		1	1	8D69			𠄎自无木𠄎	概，異體字問題。
5697	煑	1	1	8D6C	7151	煑		煮，異體字問題。
5706	敘	5	5	8D75	654D	敘		敘，異體字問題。
5742	澁	1	1	8DBB	6F81	澁		澀，異體字問題。

附件二、《蘆浦筆記》未確定字形「❓」手動取代表

檔案原文摘錄	手動取代結果
右❓土△井❓ナシ；	右害土△井 ^レ ナシ；