

《荊楚歲時記》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

2008/3/10 陳建安 製作

一、《荊楚歲時記》一書(荊楚歲時記.xml)使用舊版造字 43 個，字頻 82 次，詳如附件一。這 43 個造字中，38 個可轉成 Windows XP 能顯示的字，字頻 72 次；另外 5 個字必須轉成構字式，字頻 10 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、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 884 的「冗𠂔𠂔」係「胃」字之異體，故以「胃」字取代。

2、版本不同之字形，可用程式全部取代：由於無法取得製作 txt 檔時的書籍版本，故只能以其他版本作為轉碼的參考。版本不同故有許多文字上的出入，然而這些文字出入的情形其實都是異體字的差異，並不影響文意判讀，故直接進行更改。如編號 1187 的「盖」字，轉碼時參考的版本做「蓋」，直接以「蓋」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 3978 的「猪」字，Unicode 字型呈現為「猪」，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 4626 的「粦」字，Unicode 編碼為 427D，由於 Windows XP 並不支援，仍須使用構字式「米

厶半」。

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、《荊楚歲時記》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
884		3	3	8EC4			冗厶冂	冂，異體字問題。
1187	盖	2	2	90B9	76D6	盖		盖，版本不同之字形，可用程式全部取代。
1513	粮	1	1	92C5	7CAE	粮		糧，異體字問題。
1999	贊	1	1	95D4	8CDB	贊		贊，版本不同之字形，可用程式全部取代。
2278	鎗	1	1	97B1	936E	鎗		
2415	鬪	2	2	987B	95D8	鬪		鬪，版本不同之字形，可用程式全部取代。
2520	饒	1	1	9947	993B	饒		
3805	勅	1	1	8164	52C5	勅		勅，版本不同之字形，可用程式全部取代。
3814	却	1	1	816D	5374	却		卻，版本不同之字形，可用程式全部取代。
3846	婿	1	1	81AF	58FB	婿		婿，版本不同之字形，可用程式全部取代。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3876	廻	1	1	81CD	5EFB	廻		迴，版本不同之字形，可用程式全部取代。
3966	烟	2	2	8268	70DF	烟		
3974	牀	1	1	8270	7240	牀		
3976	犁	1	1	8272	7282	犁		
3978	猪	3	3	8274	732A	猪		Unicode 呈現差異。
4012	鼓	2	2	82B8	76B7	鼓		鼓，版本不同之字形，可用程式全部取代。
4038	竈	4	4	82D2	7AC8	竈		
4058	糴	1	1	82E6	7CC9	糴		
4077	罇	1	1	82F9	7F47	罇		
4090	脚	1	1	8347	811A	脚		
4105	菓	2	2	8356	83D3	菓		
4107	葱	1	1	8358	8471	葱		蔥，版本不同之字形，可用程式全部取代。
4113		1	1	835E			𠂇 𠂇間	簡，版本不同之字形，可用程式全部取代。
4132		4	4	8371			血 𠂇 𠂇	眾，異體字問題。
4193	鈎	15	15	83D0	920E	鈎		
4195	鉞	1	1	83D2	9262	鉞		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4218	萑	2	2	83E9	97EE	萑		
4240	髯	1	1	8440	9AF4	髯		
4256	鷄	2	2	8450	9DC4	鷄		
4278	亘	1	1	8466	4E98	亘		
4502	俎	1	1	85CB	723C	俎		
4515	呪	2	2	85D8	546A	呪		
4626	粦	1	1	86AA	427D		米 𪛗 半	
4768	妬	1	1	8779	59AC	妬		妒，版本不同之字形，可用程式全部取代。
4981	鳧	1	1	88D3	9CEC	鳧		
4996	斲	1	1	88E2	65B5	斲		
5054	獾	1	1	895D	3E90		犴 𪛗 巢	
5290	携	1	1	8ACE	643A	携		攜，版本不同之字形，可用程式全部取代。
5300	鬪	8	8	8AD8	9B2A	鬪		鬪，版本不同之字形，可用程式全部取代。
5434	湟	1	1	8BC1	6E7C	湟		
5439	栢	1	1	8BC6	6822	栢		柏，版本不同之字形，可用程式全部取代。
5482	璽	1	1	8BF1	8812	璽		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5702	厨	1	1	8D71	53A8	厨		廚，版本不同之字形，可用程式全部取代。