

《東觀漢記》舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中央研究院語言所文獻語料小組
2009/01/15 丁致伶

轉碼主要工作是把檔案中的舊版造字轉換成 Windows XP 能支援的 Unicode 字形，Unicode 目前共收錄漢字 70194 個字，而 XP 只能支援 20902 個字（詳如表一），不支援之字將以構字式表達。例：造字編號 4542 的「塏」字，Unicode 編碼是 21484，由於 XP 並不支援，仍需使用構字式「土𠩺稟」。

表一、Unicode 的字數及編碼區段

Unicode	新增字數	新增編碼區段	總字數	WinXP
1.1 版	20902	4E00-9FFF	20902	支援
3.0 版	6582	3400-4DFF	27484	不支援
3.1 版	42710	20000-2A6D6	70194	不支援

一、舊版造字轉碼分析：

《東觀漢記》使用舊版造字 122 個，字頻 582 次，這 122 個造字中，110 個可轉成 Windows XP 能顯示的字，字頻 562 次；另外 12 個字必須轉成構字式，字頻 20 次。

轉碼完成製作轉碼分析表，請參考附件一《東觀漢記》轉碼分析表，欄位說明如下：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 造字：舊版造字。
- (三) 頻次：舊版造字在文件的出現次數。
- (四) Big5：造字的 Big5 碼。
- (五) Unicode：造字所對應的 Unicode 碼。

- (六) WinXP：造字在 Windows XP 的對應字形。
- (七) 構字式：Windows XP 無法對應字形改用構字式。
- (八) 備註凡例：備註欄中記錄轉碼後字形及修改原因，凡例如下：
1. 異體字問題：為了使用者查詢和使用的方便，在處理異體字時最主要的方式是以標準字取代，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 4134 的「衛」字，是「衛」的異體字，以標準字「決」取代。
 2. 錯字，以程式全部取代：與原書字形不符，並查詢教育部異體字字典確認非異體字後，皆歸類為錯字，以程式取代為正確字形。例：檔案中使用造字編號 2738 的「广△土」字，而原書使用之字為「巡」，因此「广△土」為錯字，以程式取代為正確字形「巡」。
 3. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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。
 4. 待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 4310 的「革△健」。

附件一、《東觀漢記》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
793	鞮	1	8E47	97E4	鞮		
814	墓	1	8E5C	87C7	墓		
827	絃	1	8E69	9B8C	絃		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
832	沱	3	8E6E	6CB2	沱		
972	犇	4	8F5D	7287	犇		
1053	傲	1	8FD0	6160	傲		
1067	碁	1	8FDE	68CA	碁		
1069	瑄	1	8FE0	745D	瑄		
1077	稟	3	8FE8	7A3E	稟		
1091	漣	1	8FF6	58CD	漣		
1133	瓌	1	9061	74CC	瓌		
1226	罰	1	90E0	7F78	罰		
1381	穉	4	91DE	7A3A	穉		
1565	繇	1	92F9	7DDC	繇		
1925	羈	1	9568	8989	羈		Unicode 字型 呈現差異。
1981	讐	1	95C2	8B90	讐		讐，異體字問 題。
1999	贊	1	95D4	8CDB	贊		
2738		2	9AA6			广△土	巡，錯字，以 程式全部取 代。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2770	絳	1	9AC6	7D98	絳		
3787	冢	19	8152	51A2	冢		
3789	冲	4	8154	51B2	冲		冲，異體字問題。
3791	况	1	8156	51B5	况		况，異體字問題。
3792	凉	3	8157	51C9	凉		凉，異體字問題。
3796	决	4	815B	51B3	决		决，異體字問題。
3804	効	5	8163	52B9	効		
3805	勅	6	8164	52C5	勅		
3814	却	10	816D	5374	却		
3817	廁	1	8170	53A0	廁		廁，異體字問題。
3819	廡	1	8172	53AE	廡		廡，異體字問題。
3820	厦	1	8173	53A6	厦		
3826	叶	1	8179	53F6	叶		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3829	咏	1	817C	548F	咏		
3846	壻	3	81AF	58FB	壻		
3858	寃	5	81BB	5BC3	寃		寃，異體字問題。
3883	徧	2	81D4	5FA7	徧		
3889	慤	2	81DA	6164	慤		
3891	慙	17	81DC	6159	慙		
3924	菴	3	81FD	671E	菴		
3930	憵	1	8244	6901	憵		
3935	槩	1	8249	3BA3	槩		Unicode 字型呈現差異。
3966	烟	1	8268	70DF	烟		
3974	牀	11	8270	7240	牀		
3975	牕	1	8271	7255	牕		
3976	犁	2	8272	7282	犁		
3978	猪	4	8274	732A	猪		Unicode 字型呈現差異。
3996	疎	1	82A8	758E	疎		
4015	眦	1	82BB	7726	眦		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4016	着	1	82BC	7740	着		
4028	稟	2	82C8	7980	稟		
4034	穉	3	82CE	7A49	穉		
4035	穽	1	82CF	7A7D	穽		
4038	竈	8	82D2	7AC8	竈		
4041	豎	1	82D5	7AEA	豎		
4052	簾	1	82E0	7C34	簾		Unicode 字型 呈現差異。
4066	綺	1	82EE	7D5D	綺		
4080	羣	49	82FC	7FA3	羣		
4081	翱	1	82FD	7FFA	翱		
4085	耻	4	8342	803B	耻		
4091	臯	5	8348	81EF	臯		
4103	苣	1	8354	831D	苣		
4105	菓	3	8356	83D3	菓		
4107	葱	3	8358	8471	葱		
4114	藥	1	835F	8616	藥		
4132		99	8371			血会丞	眾，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4134	衛	19	8373	885E	衛		衛，異體字問題。
4139	袴	8	8378	88B4	袴		
4142	裛	12	837B	8943	裛		
4156	諂	1	83AB	8B1F	諂		
4173	駝	2	83BC	8EAD	駝		
4176	輒	1	83BF	8F19	輒		
4178	輓	1	83C1	8F2D	輓		
4184	迹	6	83C7	8FF9	迹		
4193	鈎	3	83D0	920E	鈎		鈎，異體字問題。
4203	隣	2	83DA	96A3	隣		鄰，異體字問題。
4216	鞮	1	83E7	97C8	鞮		
4222	頤	1	83ED	9847	頤		
4225	顛	1	83F0	9866	顛		
4237	髡	1	83FC	9AE0	髡		
4242	鬪	9	8442	9B2D	鬪		
4250	鱈	1	844A	9C0C	鱈		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4256	鷄	2	8450	9DC4	鷄		雞，異體字問題。
4265	龜	1	8459	9F03	龜		
4307	罵	1	84A5	99E1	罵		罵，異體字問題。
4310		1	84A8			革 Δ 健	待造字。
4342	倮	1	84C8	502E	倮		
4455	劒	7	857A			僉 Δ 刃	
4460	勅	34	85A1	52D1	勅		
4507	吳	65	85D0	5434	吳		吳，異體字問題。
4542	塿	1	85F3	21484		土 Δ 稟	
4558	厖	1	8644	53A9	厖		Unicode 字型呈現差異。
4739	鞞	1	875C	97DB	鞞		
4817	寶	1	87CC	5BF3	寶		寶，異體字問題。
4833	簿	2	87DC	25C7C		形 Δ 竹 卑 Δ 巾	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4878	宀	1	884A	5B82	宀		Unicode 字型呈現差異。
4896	繡	1	885C	42E0		纟 㒶 甫	
4897		1	885D			纟 㒶 甚	待造字。
4907	橐	1	8867	69D6	橐		
5089		1	89A2			宀 㒶 畢	畢，錯字，以程式全部取代。
5133	憇	31	89CE	6199	憇		
5144	構	1	89D9	8920	構		
5160	憇	1	89E9	6205	憇		憇，異體字問題。
5164	肇	1	89ED	8088	肇		
5190	褊	1	8A48	463F		衤 㒶 屈	
5225	邨	1	8A6B	460F		血 㒶 卩	
5290	携	3	8ACE	643A	携		
5335	蚤	1	8AFB	459F		亡 㒶 虫	
5434	湟	3	8BC1	6E7C	湟		
5437	并	1	8BC4	5E77	并		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5452	轎	1	8BD3	8F5C	轎		
5558	籠	2	8C7E	9E84	籠		
5559	欸	2	8CA1	6B35	欸		
5578	弛	1	8CB4	38AE		弓△屯	
5591	杞	1	8CC1	233CC		木△巳	杞，異體字問題。
5697	煑	1	8D6C	7151	煑		
5702	厨	3	8D71	53A8	厨		
5705	廐	2	8D74	5ED0	廐		廐，異體字問題。
5730	阨	1	8DAF	9638	阨		
5753	潜	1	8DC6	6F5C	潜		
5764	籊	1	8DD1	7C5E	籊		
5770	汧	2	8DD7			氵△开	
5783	教	1	8DE4	3A8D		孛△攴	
5850	厶	1	C6C9	53B6	厶		

二、手動取代表說明：

《東觀漢記》檔案中出現「●」字形，共計5字，比對原書後以Unicode字形或構字式手動取代，並製作「●」字手動取代表，請參

考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《東觀漢記》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《東觀漢記》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 5	胡子進狗●馬醢	月 𠂔 棠	
p. 204	滑如礪●	石 𠂔 弟	待造字
p. 259	鈎牧車屏●泥	彳 𠂔 屏	待造字
p. 364	舉手●	扌 𠂔 虎 𠂔 欠	待造字
p. 422	●目討賊	目 𠂔 棠	