

《燕翼詒謀錄》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
2008/5/13 陳維君 製作

- 一、《燕翼詒謀錄》一書(燕翼詒謀錄.xml)使用舊版造字 31 個，字頻 92 次，詳如附件一。這 31 個造字中，25 個可轉成 Windows XP 能顯示的字，字頻 71 次；另外 6 個字必須轉成構字式，字頻 21 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 3924 的「朞」係「期」字之異體，故以「期」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 3846 的「壻」字，Unicode 字型呈現為「壻」，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 12 的「陟」字，Unicode 編碼為 28E97，由於 Windows XP 並不支援，仍須使用構字式「卩𠂔走」。

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、《燕翼詒謀錄》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
12	陲	1	1	FA4B	28E97		阝厶走	
1069	瑄	2	2	8FE0	745D	瑄		
1226	罰	1	1	90E0	7F78	罰		罰，異體字問題。
1371	嶮	1	1	91D4	37FD		山厶票	
1572	鞞	3	3	9341			形革口王口	革厶呈，異體字問題。
2199	鉦	1	1	9740	925F	鉦		
2857	杖	1	1	9B5E			木厶犬	
3796	決	1	1	815B	51B3	決		決，異體字問題。
3804	効	3	3	8163	52B9	効		
3819	廡	1	1	8172	53AE	廡		廡，異體字問題。
3846	壻	1	1	81AF	58FB	壻		Unicode 呈現差異。
3858	寃	1	1	81BB	5BC3	寃		寃，異體字問題。
3924	朞	5	5	81FD	671E	朞		期，異體字問題。
3996	疎	1	1	82A8	758E	疎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4080	羣	4	4	82FC	7FA3	羣		群，異體字問題。
4085	耻	4	4	8342	803B	耻		
4104	莅	4	4	8355	8385	莅		
4132		14	14	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	3	3	8373	885E	衛		衛，異體字問題。
4136	衰	1	1	8375	887A	衰		
4220	穎	2	2	83EB	9834	穎		穎，異體字問題。
4242	鬪	2	2	8442	9B2D	鬪		Unicode 呈現差異。
4309	劫	1	1	84A7	523C	劫		劫，異體字問題。
4450	勑	1	1	8575	524F	勑		勑，異體字問題。
4460	勅	10	10	85A1	52D1	勅		
5225	邨	1	1	8A6B	460F		血𠂇𠂇	邨，異體字問題。
5423	楛	12	12	8BB6	6810	楛		
5483	眡	2	2	8BF2	7721	眡		
5555	迥	1	1	8C7B	9008	迥		迥，異體字問題。
5702	厨	1	1	8D71	53A8	厨		廚，異體字問題。
5706	敍	6	6	8D75	654D	敍		敘，異體字問題。