

# 《西京雜記》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  
2010/06/07 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 5723 的「𦍋」字，Unicode 編碼是 4367，由於 XP 並不支援，仍需使用構字式「牛 𦍋羊」。

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：

《西京雜記》使用舊版造字 78 個，字頻 157 次，這 78 個造字中，66 個可轉成 Windows XP 能顯示的字，字頻 136 次；另外 12 個字必須轉成構字式，字頻 21 次。

轉碼完成製作轉碼分析表，請參考附件一《西京雜記》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 4132 的「血𠂇𠂇」字，是「眾」的異體字，以標準字「眾」取代。
2. 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 4091 的「臯」字，Unicode 字型呈現為「臯」，實際上仍為同一字。

附件一、《西京雜記》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
842	𠂇	2	8E78	3AB7		𠂇	
1056	阨	1	8FD3	9641	阨		
1067	碁	1	8FDE	68CA	碁		
1238	疾	2	90EC	77E6	疾		
1266	碇	1	9149	784B	碇		
1664	腸	1	93BF	8193	腸		
1999	贊	1	95D4	8CDB	贊		
2044	踈	1	9642	8E08	踈		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2278	銓	1	97B1	936E	銓		
2415	鬪	1	987B	95D8	鬪		
2516	餽	1	9943	9919	餽		
2650	鸛	1	99EB	9E18	鸛		
3035	賔	5	9C73	8CD4	賔		
3787	豕	3	8152	51A2	豕		
3791	況	1	8156	51B5	況		況，異體字問題。
3796	決	2	815B	51B3	決		決，異體字問題。
3805	勅	1	8164	52C5	勅		
3817	廁	3	8170	53A0	廁		
3840	坂	1	81A9	5742	坂		
3868	卮	1	81C5	5DF5	卮		
3876	廻	6	81CD	5EFB	廻		
3877	迺	1	81CE	5EFC	迺		
3883	徧	1	81D4	5FA7	徧		
3891	慙	3	81DC	6159	慙		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3974	牀	14	8270	7240	牀		
4012	鼓	3	82B8	76B7	鼓		鼓，異體字問題。
4013	盃	1	82B9	76CC	盃		
4014	蓋	7	82BA	8462	蓋		
4057	莖	7	82E5	585F	莖		
4071	纓	2	82F3	7DF5	纓		
4077	罇	1	82F9	7F47	罇		
4080	羣	6	82FC	7FA3	羣		
4085	耻	1	8342	803B	耻		
4089	脇	1	8346	8107	脇		
4090	脚	4	8347	811A	脚		
4091	臯	3	8348	81EF	臯		Unicode 字型呈現差異。
4101	珣	1	8352	3EC1		王 ㄩ 回	
4129	虬	1	836E	866C	虬		
4132		2	8371			血 ㄩ 丞	眾，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4180	鼻	1	83C3	8FA0	鼻		
4184	迹	2	83C7	8FF9	迹		
4193	鈎	1	83D0	920E	鈎		
4203	隣	2	83DA	96A3	隣		鄰，異體字問題。
4240	鬚	1	8440	9AF4	鬚		
4242	鬪	2	8442	9B2D	鬪		
4254	鷓	1	844E	9D7E	鷓		
4256	鷄	3	8450	9DC4	鷄		
4267	鼈	1	845B	9F08	鼈		
4278	亘	1	8466	4E98	亘		
4297	躔	1	8479	281E0		形足厂 墨	
4345	嗽	1	84CB			口 ㄩ 救	
4389	瑤	2	84F7	746B	瑤		
4455	劔	6	857A			僉 ㄩ 刀	
4496	段	1	85C5	53DA	段		
4502	俎	1	85CB	723C	俎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4515	呪	1	85D8	546A	呪		
4531	攢	1	85E8	6505	攢		
4552	趣	1	85FD	27F08		走△取	
4607	狫	1	8675	3E60		犛△屯	
4643	徃	7	86BB	5F83	徃		
4704	犛	1	86F8	7271	犛		
4770	鑄	2	877B	28A8B		金△壽	
4812	羴	2	87C7	26351		羴△、	
4907	橐	1	8867	69D6	橐		
5072	扁	1	896F	6244	扁		
5096	紆	2	89A9	42B8		纟△亏	
5131	窻	1	89CC	7ABB	窻		Unicode 字型 呈現差異。
5176		1	89F9			冫△冗	沉，異體字問 題。
5290	携	1	8ACE	643A	携		
5310	烏	1	8AE2	8204	烏		
5338	暎	2	8AFE	668E	暎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5439	栢	1	8BC6	6822	栢		
5533		1	8C65			穀△米	穀，異體字問題。
5559	欸	2	8CA1	6B35	欸		
5702	厨	1	8D71	53A8	厨		
5719	麈	1	8DA4	4D25		鹿△加	
5723	牂	1	8DA8	4367		牛△羊	
5764	籩	1	8DD1	7C5E	籩		

## 二、手動取代表說明：

《西京雜記》檔案中出現「●」字形，共計12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《西京雜記》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、《西京雜記》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 1074	白兪●杜	木△匊	改為[形]木勺缶[口]

頁碼	檔案原文摘錄	原書字形	備註
p. 1074	杜●桂蜀	木△匊	改為[形]木勺缶[口]
p. 1078	●徒驪山	[形]之'无[口]	改為送
p. 1081	夫目●得酒食	睂	
p. 1081	目●則呪之	睂	
p. 1082	元●曹元理	艹△兔	改為菟
p. 1083	奮皓翅之●	羽△戈	
p. 1083	●·宛脩頸	羽△戈	
p. 1084	載●載齊庶	竹△酉	
p. 1086	歸風●送		該處無字
p. 1086	倍升為●	糸△式	
p. 1086	倍●為	糸△式	