

# 《東觀奏記》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  
2008/4/14 陳建安 製作

一、《東觀奏記》一書(東觀奏記.xml)使用舊版造字 52 個，字頻 380 次，詳如附件一。這 52 個造字中，44 個可轉成 Windows XP 能顯示的字，字頻 347 次；另外 8 個字必須轉成構字式，字頻 33 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、標記問題：標記問題主要是漢籍電子文獻在舊轉新的過程中，採用了新的標記語言，而在修改標記時，必須移動某些文句，若是有遺漏並且遺漏的內容包含了缺字，便會造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號 4507 的「吳」字，xml 檔在 93、94、95 頁各遺漏了一個註釋，因而造成頻次不同，處理方式必須補上遺漏的文句，並將未修改的造字全部取代。另外，因為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，故將異體字的「吳」字，以標準字「吳」字取代。

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 4132 的「血𠂔𠂔」係「眾」字之異體，故以「眾」字取代。又編號 4067 的「綉」字，係「繡」字之簡體字，亦以標準字的「繡」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「譎」字，



編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1010	珏	15	15	8FA5	73CF	珏		
1032	憊	1	1	8FBB	2039D		𠄎𠄎 𠄎心	
1133	瓌	3	3	9061	74CC	瓌		
1541	恫	1	1	92E1	390F		忄 𠄎冬	
1642	脉	3	3	93A9	8109	脉		
1981	讐	2	2	95C2	8B90	讐		
2157	酌	2	2	96D5	9167	酌		
3789	冲	2	2	8154	51B2	冲		沖，異體字問題。
3794	湊	1	1	8159	51D1	湊		湊，異體字問題。
3805	勅	21	21	8164	52C5	勅		
3814	却	14	14	816D	5374	却		
3826	叶	4	4	8179	53F6	叶		
3829	咏	1	1	817C	548F	咏		
3858	寃	8	8	81BB	5BC3	寃		冤，異體字問題。
3883	徧	1	1	81D4	5FA7	徧		
3891	慙	1	1	81DC	6159	慙		慚，異體字問題。
3924	朞	1	1	81FD	671E	朞		期，異體字問題。
3931	椀	2	2	8245	6900	椀		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3974	牀	1	1	8270	7240	牀		
3987	穀	2	2	827D	7474	穀		
3996	疎	1	1	82A8	758E	疎		
4015	眦	1	1	82BB	7726	眦		
4016	着	3	3	82BC	7740	着		著，異體字問題。
4041	豎	1	1	82D5	7AEA	豎		
4044	筋	1	1	82D8	7B6F	筋		
4063	藜	1	1	82EB	7CF5	藜		
4067	綉	1	1	82EF	7D89	綉		繡，異體字問題。
4080	羣	7	7	82FC	7FA3	羣		群，異體字問題。
4090	脚	2	2	8347	811A	脚		
4094	館	11	11	834B	8218	館		
4132		13	13	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	3	3	8373	885E	衛		衛，異體字問題。
4146	覩	10	10	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4184	迹	1	1	83C7	8FF9	迹		
4256	鷄	1	1	8450	9DC4	鷄		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4350	侶	2	2	84D0	5058	侶		
4507	吳	195	192	85D0	5434	吳		標記問題，手動補上遺漏文句。 吳，異體字問題。
4768	妬	1	1	8779	59AC	妬		妒，異體字問題。
4904	尅	1	1	8864	5C05	尅		
5126	暮	5	5	89C7	8B29	暮		
5133	憲	4	4	89CE	6199	憲		
5225	卹	1	1	8A6B	460F		血 $\Delta$ 卩	
5260	蝟	2	2	8AB0	29A33		出 $\Delta$ 骨	
5289	慤	1	1	8ACD	6164	慤		
5300	鬪	2	2	8AD8	9B2A	鬪		
5473	攷	1	1	8BE8	22BF1		扌 $\Delta$ 攷	
5568	开	2	2	8CAA	5F00	开		
5706	敘	2	2	8D75	654D	敘		敘，異體字問題。

附件二、《東觀奏記》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
4507	吳	頒告已數日，卻命追制	頒告已數日， <u>&lt;ge&gt;中使至鄭頒告已數日清吳抄本、山房本「中使至鄭」皆作「中使至」。</u> </ge>卻命追制
		敢不討論。 <u>〔六四〕</u> 臣尋委當	敢不討論。 <u>&lt;ge&gt;虔承聖旨敢不討論「虔」原作「據」，清吳抄本、山房本作「遽」，續粵本、宛本說郛作「虔」，登科記考卷二二同。按虔，恭敬之意，作「虔」是，據改。</u> </ge>臣尋委當
		至於聖朝， <u>〔六五〕</u> 謹專上進	至於聖朝， <u>&lt;ge&gt;自武德元年至於聖朝清吳抄本、山房本、宛本說郛「至於聖朝」皆作「至朝」。</u> 按作「至於聖朝」是。</ge>謹專上進
		<u>垂</u> （全部取代）	<u>吳</u> （全部取代）