

《默記》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
2008/5/12 陳維君 製作

一、《默記》一書(默記.xml)使用舊版造字 47 個，字頻 122 次，詳如附件一。這 47 個造字中，40 個可轉成 Windows XP 能顯示的字，字頻 92 次；另外 7 個字必須轉成構字式，字 30 頻次。

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 1067 的「碁」係「棋」字之異體，故以「棋」字取代。又編號 3966 的「烟」字，係「煙」字之簡體字，亦以標準字的「煙」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 4764 的「摠」字，Unicode 字型呈現為「摠」，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 1919 的「儻」字，Unicode 編碼為 2032F，由於 Windows XP 並不支援，仍須使用構字式「亼𠂔离」。

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、《默記》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
296	𡗗	1	1	FBEC			𡗗	
1067	碁	1	1	8FDE	68CA	碁		碁，異體字問題。
1133	瓌	1	1	9061	74CC	瓌		
1153	坳	2	2	9075	362D		土 厶 幻	坳，異體字問題。
1919	儻	1	1	9562	2032F		亻 厶 离	
3160	坟	2	2	9D53	575F	坟		
3788	慤	1	1	8153	22871		殼 厶 心	慤，異體字問題。
3797	凭	1	1	815C	51ED	凭		憑，異體字問題。
3805	勅	2	2	8164	52C5	勅		
3814	却	7	7	816D	5374	却		卻，異體字問題。
3817	廁	3	3	8170	53A0	廁		廁，異體字問題。
3819	廡	1	1	8172	53AE	廡		廡，異體字問題。
3824	叙	1	1	8177	53D9	叙		敘，異體字問題。
3829	咏	1	1	817C	548F	咏		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3835	噍	1	1	81A4	564D	噍		
3843	堦	3	3	81AC	5826	堦		
3846	堦	5	5	81AF	58FB	堦		Unicode 呈現差異。
3858	寃	2	2	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。
3935	槩	2	2	8249	69E9	槩		概，異體字問題。
3945		1	1	8253			汙 𠄎 义	汙，異體字問題。
3966	烟	1	1	8268	70DF	烟		煙，異體字問題。
3974	牀	4	4	8270	7240	牀		
3996	疎	2	2	82A8	758E	疎		
4016	着	15	15	82BC	7740	着		著，異體字問題。
4069	綑	1	1	82F1	7DB3	綑		
4080	羣	6	6	82FC	7FA3	羣		群，異體字問題。
4107	葱	1	1	8358	8471	葱		蔥，異體字問題。
4108	莽	2	2	8359	844A	莽		
4115	藁	1	1	8360	85C1	藁		
4132		23	23	8371			血 𠄎 𠄎	眾，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4134	衛	1	1	8373	885E	衛		衛，異體字問題。
4184	迹	1	1	83C7	8FF9	迹		
4186	遡	1	1	83C9	9061	遡		
4188	邨	1	1	83CB	90A8	邨		
4193	鈎	2	2	83D0	920E	鈎		鈎，異體字問題。
4309	刼	1	1	84A7	523C	刼		刼，異體字問題。
4460	勅	2	2	85A1	52D1	勅		
4507	吳	8	8	85D0	5434	吳		吳，異體字問題。
4603	憇	1	1	8671	6187	憇		憇，異體字問題。
4764	摠	1	1	8775	64A7	摠		Unicode 呈現差異。
4768	妬	1	1	8779	59AC	妬		妒，異體字問題。
4839	昱	1	1	87E2	6637	昱		
5434	湟	1	1	8BC1	6E7C	湟		湟，異體字問題。
5706	敘	2	2	8D75	654D	敘		敘，異體字問題。
5852	宀	1	1	C6CB			宀	