

《黃帝內經素問譯解》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
2008/1/21 陳建安 製作

一、《黃帝內經素問譯解》一書(黃帝內經素問譯解.xml)使用舊版造字 26 個，字頻 55 次，詳如附件一。這 26 個造字中，23 個可轉成 Windows XP 能顯示的字，字頻 50 次；另外 3 個字必須轉成構字式，字頻 5 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、備註凡例：

1、校對問題，舊版漢籍錯字，可用程式全部取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中；若該造字的所有頻次，皆屬於錯誤使用的錯字情形，可以用程式全部取代為正確字形。如編號 3120 的「白△久」字，原字為「囟」。

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 4345 的「口△敕」係「嗽」字之異體，故以「嗽」字取代。

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 2260 的「癩」字，Unicode 編碼為 3FD7，由於 Windows XP 並不支援，仍須使用構字式「疒△類」。

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、《黃帝內經素問譯解》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
844	膠	4	4	8E7A	9ACE	膠		
852		1	1	8EA4	72E2	貉		
1223	瘥	1	1	90DD	75D3	瘥		
2260	癩	3	3	977D	3FD7		疒△類	
2415	鬪	1	1	987B	95D8	鬪		
3064	鯛	1	1	9CB2	9C02	鯛		
3120		1	1	9CEA			白△久	囟，校對問題，舊版漢籍錯字，可用程式全部取代。
3794	湊	4	4	8159	51D1	湊		
3796	决	3	3	815B	51B3	决		
3797	凭	1	1	815C	51ED	凭		
3858	寃	1	1	81BB	5BC3	寃		
3966	烟	5	5	8268	70DF	烟		
3967	炳	4	4	8269	712B	炳		
4015	眦	8	8	82BB	7726	眦		
4115	藁	1	1	8360	85C1	藁		
4240	髯	1	1	8440	9AF4	髯		
4345		1	1	84CB			口△敕	嗽，異體字問題。
4767	懷	1	1	8778	61B9	懷		
4817	寶	3	3	87CC	5BF3	寶		
4951	礪	1	1	88B5	7930	礪		
5452	轡	2	2	8BD3	8F5C	轡		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5497	宧	1	1	8C41	8873	宧		裏，校對問題，舊版漢籍錯字，可用程式全部取代。
5621	汨	3	3	8CDF	6CEA	汨		
5724	稽	1	1	8DA9	7A38	稽		
5742	澁	1	1	8DBB	6F81	澁		
5752	摠	1	1	8DC5	63D4	摠		