

《諸病源候論校注》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
2008/3/4 黃榮順 製作

一、《諸病源候論校注》一書(諸病源候論校注.xml)使用舊版造字 111 個，字頻 252 次，詳如附件一。這 111 個造字中，77 個可轉成 Windows XP 可顯示的字，字頻 186 次；另外 34 個字必須轉成構字式，字頻 66 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、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 566 的「尸」字，原書字形為「务」。
- 2、校對問題，新舊版缺字並存：漢籍在舊轉新的過程中，以人工修改方式取代舊漢籍造字，雖然修改的結果是正確的，但是只修改到部分造字，沒有全部修改到，導致舊漢籍造字在新版漢籍檔案中，一部分已被取代成 Unicode 或構字式，另一部分仍為舊漢籍造字的並存情形。處理方式為配合漢籍工作室的修改，將沒被修改到的舊漢籍造字，以對應的 Unicode 或構字式取代。如編號 750，舊漢籍造字為「𠂔」，比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「𠂔」者，在 xml 檔中部分已人工修改為部件外字的「𠂔」，故配合修改，將 xml 檔其餘未修改的部份也手動取代為部件外字的「𠂔」。

- 3、標記問題，新版漢籍表格內文字以圖版替代：新版漢籍的表格一律以圖版方式呈現，使得舊版漢籍表格內的文字在新版漢籍中不再以內文方式呈現。故若舊版漢籍表格內的文字出現缺字，因新版漢籍中已改用圖版呈現，而造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號 3951 的「湏」字，txt 檔出現 5 次，而 xml 檔只有 4 次，txt 檔多出的 1 次出現在原書 1464 頁的表格中，由於 xml 檔已將表格用圖版取代，所以手動將 xml 檔頻次都改為「湏」字即可。
- 4、校對問題，新版漢籍錯字，人工取代：漢籍在舊轉新的過程中，以人工修改方式取代舊漢籍造字，但是卻取代為錯誤的字形，因此出現舊版漢籍沒有，卻在新版漢籍出現的錯字。處理方式為把被修改為錯字的部分，以及沒被修改到的舊漢籍造字，一併以對應的 Unicode 或構字式取代。如編號 982 的「胃」字，xml 檔出現 2 次，比對後發現，其中 1 次原書字形為「胃」，故手動更正取代。
- 5、重碼問題，採用部件外字，不轉碼：舊漢籍造字與構字式部件外字，皆是建立在 Big5 的使用者造字區中，兩者所使用的碼區是重疊的，在同時使用構字式和舊漢籍造字的新版漢籍中，這些重疊的字碼可能被用以表達兩個不同字形，我們稱之為「重碼問題」。然而漢籍在舊轉新的過程中，若已將原舊漢籍造字的部分全數改為 Unicode 或構字式，則在 xml 出現的造字頻次皆採用部件外字，此時便不需轉碼。如編號 5075，字碼 8972，對應的舊漢籍造字為「𠂔」，部件外字為「𠂔」，一一比對原書，txt 檔出現的 2 次因用於表格文字，不需處理；xml 檔的 1 次則使用部件外字的「𠂔」，原本不需轉碼，但與書上字形略有差異，故替換為「𠂔」。又如編號 5809，字碼 8DFE，對應的舊漢籍造字為「踰」，而構字式對應的部件外字為「𠂔」，此部件漢籍用來表達未定義的字形（可能表達任何字形），若直接採用，等於沒有解決此處的缺字問題。因此在此字碼重碼的情況下，無論是採用舊漢籍造字「踰」，或採用部件外字「𠂔」，都要以人工的方式取代為原書字形。本書 xml 檔只有 1 次，採用的是部件外字「𠂔」，則以原書字形「𠂔凡」取代。

- 6、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 4134 的「衛」係「衛」字之異體，故以「衛」字取代。又如編號 4068 的「綫」係「線」字之簡體，故以「線」字取代。
- 7、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 4910 的「𤇀 𤇁」。

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 1710 的「燂」字，Unicode 編碼為 3E01，由於 Windows XP 並不支援，仍須改用構字式「火 𤇀 業」。

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

四、txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，詳細取代內容可參考附件二的手動取代表。附件二的手動取代表說明如下：

- 甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- 乙、造字：舊漢籍造字。
- 丙、取代：使用字形與造字欄不同時，於此欄列出。
- 丁、檔案原文摘錄：摘錄檔案(諸病源候論校注.xml)中包含該造字的文句段落，其中紅字底線的部分，為需要手動取代的原文。文句後方標示（全部取代）者，表示執行全部取代的動作，不一一列舉所有原文。
- 戊、手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。文句後方標示（全部取代）者，表示執行全部取代的動作，不一一列舉所有原文。

附件一、「諸病源候論校注.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
298	△	316	2	FBEE			△	△，校對問題，xml 檔將表格替換為圖檔。
523	阜	2	2	FD73			阜	
565	虽	1	1	FDBF			虽	雖，異體字問題。
566	尸	1	1	FDC0			尸	务，校對問題，舊版漢籍錯字，可用程式全部取代。
595	卓	6	6	FDDD			卓	
607	陲	2	2	FDE9			陲	條，校對問題，舊版漢籍錯字，可用程式全部取代。
750	刂	5	4	FEDB			刂	刂，校對問題，新舊版缺字並存。
770	扌	14	2	FEEF			扌	扌，校對問題，xml 檔將表格替換為圖檔。
772	辶	6	1	FEF1			辶	辶，校對問題，xml 檔將表格替換為圖檔。
932	兒	1	1	8EF4	7683	兒		
982	胃	4	2	8F67	80F7	胃		胸，校對問題，xml 檔將表格替換為圖檔；異體字取代。 胃，校對問題，xml 檔其中 1 次為錯字，手動取代。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1013	𣎵	1	1	8FA8	7D5A	𣎵		
1048	頷	1	1	8FCB	981F	頷		
1056	陲	3	3	8FD3	9641	陲		
1187	蓋	3	2	90B9	76D6	蓋		蓋，校對問題，xml 檔將表格替換為圖檔；異體字取代。
1491	擷	1	1	92AF	6527	擷		
1710	爍	2	2	93ED	3E01		火𠂇業	
1897	祕	1	1	954C	8890	祕		祕，異體字問題。
2044	踈	8	7	9642	8E08	踈		疏，校對問題，xml 檔將表格替換為圖檔；異體字取代。
2074	縻	1	1	9660	2A3AD		麻𠂇黍	
2151	𠂇	7	1	96CF	961D	𠂇		𠂇，校對問題，xml 檔將表格替換為圖檔。
2193	𦍋	1	1	96F9	3CA3		叟𠂇毛	
2252		3	2	9775			舌𠂇乚	亂，校對問題，xml 檔將表格替換為圖檔；異體字取代。
2434		1	1	98B0			𠂇𠂇迨	隨，異體字問題。
2567	稗	1	1	9976	7A21	稗		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2572	銜	1	1	997B	8858	銜		
2672	𧈧	1	1	9A42			亦𧈧虫	
2797	𧈧	1	1	9AE1	939C	𧈧		
2806	𧈧	1	1	9AEA	976D	𧈧		
2893	属	2	1	9BA4	5C5E	属		屬，校對問題，xml 檔將表格替換為圖檔；異體字取代。
2996	双	4	3	9C4C	53CC	双		雙，校對問題，xml 檔將表格替換為圖檔；異體字取代。
3005	𧈧	6	6	9C55	7CCD	𧈧		
3255	善	5	5	9DD4	5584	善		
3770	个	1	1	8141	4E2A	个		
3787	豕	1	1	8152	51A2	豕		
3788	𧈧	2	2	8153	22871		𧈧𧈧心	
3789	冲	1	1	8154	51B2	冲		沖，異體字問題。
3794	𧈧	2	2	8159	51D1	𧈧		
3797	凭	3	3	815C	51ED	凭		
3816	厓	3	2	816F	5393	厓		厓，校對問題，xml 檔將表格替換為圖檔。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3831	詰	1	1	817E	5586	詰		
3864	峯	1	1	81C1	5CEF	峯		
3940	爵	1	1	824E	6B1D	爵		
3948		2	1	8256			汧  木	汧，校對問題，xml 檔將表格替換為圖檔；異體字取代。
3951	湏	5	4	8259	6E4F	湏		湏，校對問題，xml 檔將表格替換為圖檔。
3961	灾	2	1	8263	707E	灾		災，校對問題，xml 檔將表格替換為圖檔；異體字取代。
3966	烟	1	1	8268	70DF	烟		
3973		3	2	826F			𠂔  扁	𠂔  扁，校對問題，xml 檔將表格替換為圖檔。
4012	鼓	2	1	82B8	76B7	鼓		鼓，校對問題，xml 檔將表格替換為圖檔；異體字取代。
4041	豎	6	6	82D5	7AEA	豎		
4068	綫	6	6	82F0	7DAB	綫		線，異體字問題。
4069	綑	5	5	82F1	7DB3	綑		
4073	縑	4	4	82F5	7E6E	縑		
4085	耻	1	1	8342	803B	耻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4104	莅	1	1	8355	8385	莅		
4105	菓	5	5	8356	83D3	菓		
4134	衛	2	1	8373	885E	衛		衛，校對問題，xml 檔將表格替換為圖檔；異體字取代。
4137	袜	1	1	8376	889C	袜		
4178	輓	1	1	83C1	8F2D	輓		
4184	迹	3	3	83C7	8FF9	迹		
4226	飡	1	1	83F1	98E1	飡		
4256	鷄	9	7	8450	9DC4	鷄		雞，校對問題，xml 檔將表格替換為圖檔；異體字取代。
4265	鼃	1	1	8459	9F03	鼃		
4266	鼃	2	2	845A	9F02	鼃		
4275	义	3	2	8463	4E49	义		義，校對問題，xml 檔將表格替換為圖檔；異體字取代。
4286	从	10	9	846E	4ECE	从		从，校對問題，xml 檔將表格替換為圖檔。
4296	塢	5	4	8478	5878	塢		塢，校對問題，xml 檔將表格替換為圖檔。
4309	刼	1	1	84A7	523C	刼		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4342	倮	1	1	84C8	502E	倮		
4432		1	1	8563			𠂇 纟 厂 墨 𠂇	纏，異體字問題。
4447	捏	4	4	8572	63D1	捏		
4488	惡	1	1	85BD	6076	惡		惡，異體字問題。
4495	𪗇	6	5	85C4	9EA8	𪗇		𪗇，校對問題，xml 檔將表格替換為圖檔。
4515	呪	16	14	85D8	546A	呪		呪，校對問題，xml 檔將表格替換為圖檔。
4560	協	1	1	8646	604A	協		
4616	壳	2	2	867E	58F3	壳		
4625	𪗇	2	2	86A9	3A7B		危 𪗇 支	
4668	𪗇	2	1	86D4	5650	𪗇		𪗇，校對問題，xml 檔將表格替換為圖檔。
4680	𪗇	1	1	86E0	2680A		月 𪗇 𪗇 止	
4702	場	1	1	86F6	5872	場		
4767	懷	7	6	8778	61B9	懷		懷，校對問題，xml 檔將表格替換為圖檔。
4798		2	2	87B9			白 𪗇 夕	𪗇，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4803	脬	3	3	87BE	816D	脬		
4805	瀆	1	1	87C0	51DF	瀆		
4816	𠂇	17	1	87CB			𠂇	𠂇，校對問題，xml 檔將表格替換為圖檔。
4870	壘	1	1	8842	97F2	壘		
4908	届	1	1	8868	5C4A	届		届，異體字問題。
4910		2	2	886A			𠂇 𠂇 𠂇	待造字。
4940	韵	4	4	88AA	97F5	韵		
5005	幫	1	1	88EB	5E2E	幫		幫，異體字問題。
5013	𠂇	38	7	88F3			𠂇	𠂇，校對問題，xml 檔將表格替換為圖檔。
5056	𠂇	2	2	895F			𠂇 𠂇 𠂇 𠂇 𠂇 𠂇	
5071	𠂇	1	1	896E	5FE9	𠂇		
5075	𠂇	2	1	8972			𠂇	𠂇，重碼問題，採用部件外字，不轉碼，但因字形不同，人工取代。
5108	惛	2	2	89B5	60FD	惛		
5152	邇	1	1	89E1	28637		邇 邇 邇	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5292	𦵒	4	3	8AD0	65C9	𦵒		𦵒，校對問題，xml 檔將表格替換為圖檔。
5320	瘡	2	1	8AEC	24E07		疒△夊	疒△夊，校對問題，xml 檔將表格替換為圖檔。
5392		6	6	8B75			𠂔白攵心𠂔	𠂔，異體字問題。
5402	朮	9	1	8BA1	672F	朮		朮，校對問題，新舊版缺字並存。
5439	栢	1	1	8BC6	6822	栢		
5583	殮	2	1	8CB9	98F1	殮		殮，校對問題，xml 檔將表格替換為圖檔；異體字取代。
5591	杞	1	1	8CC1	233CC		木△巳	杞，異體字問題。
5604	𠂔	2	2	8CCE	3E26		𠂔	
5621	汨	2	2	8CDF	6CEA	汨		
5697	𦵒	4	4	8D6C	7151	𦵒		煮，異體字問題。
5705	廐	1	1	8D74	5ED0	廐		廐，異體字問題。
5753	潛	3	2	8DC6	6F5C	潛		潛，校對問題，xml 檔將表格替換為圖檔；異體字取代。
5809	踰	27	1	8DFE	8E79	踰		𠂔△凡，重碼問題，採用部件外字，不轉碼，但部件外字為𠂔，故以原書字形取代。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5846		1	1	C6C5				丿，校對問題，舊版漢籍錯字，可用程式全部取代。
5855		1	1	C6CE				广，校對問題，舊版漢籍錯字，可用程式全部取代。

附件二、「諸病源候論校注.xml」手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果
298	△		作一個「 <u> </u> 凡」字	作一個「 <u>△</u> 凡」字
			字原版空門 <u> </u> 缺 (以上 xml 檔缺圖表，致頻次不符，手動全部取代)	字原版空門 <u>△</u> 缺 (以上 xml 檔缺圖表，致頻次不符，手動全部取代)
750	丿		△ <u> </u> 而休於氣	△ <u>丿</u> 而休於氣
			新食竟 罔△ <u> </u>	新食竟 罔△ <u>丿</u>
			罔△ <u> </u> 食畢	罔△ <u>丿</u> 食畢
			「新」纔，罔△ <u> </u>	「新」纔，罔△ <u>丿</u>
770	扌		<u> </u> △丸持也	<u>扌</u> △丸持也
			<u> </u> △以態詞 (以上 xml 檔缺圖表，致頻次不符，手動全部取代)	<u>扌</u> △以態詞 (以上 xml 檔缺圖表，致頻次不符，手動全部取代)
772	乚		作一個「 <u> </u> △凡」字	作一個「 <u>乚</u> △凡」字

編號	造字	取代	檔案原文摘錄	手動取代結果
982	胃	胸	三日在肌，四日在__	三日在肌，四日在 <u>胸</u>
			五日入__，入胃乃	五日入 <u>胃</u> ，入胃乃
1187	盖	蓋	__指食生葱感染而言	<u>蓋</u> 指食生葱感染而言
			__突出產後特點	<u>蓋</u> 突出產後特點
2044	踈	疏	__ (xml 檔缺圖表，致頻次不符，手動全部取代)	<u>疏</u> (xml 檔缺圖表，致頻次不符，手動全部取代)
2151	𠂔		骨蒸或五疳__ <u>𠂔</u> 介段	骨蒸或五疳 <u>𠂔</u> <u>𠂔</u> 介段
2252	舌𠂔𠂔	亂	皆成霍__	皆成霍 <u>亂</u>
			悶__冒絕	悶 <u>亂</u> 冒絕
2893	属	屬	日名，__土	日名， <u>屬</u> 土
2996	双	雙	始__𡵚足跟向	始 <u>雙</u> 𡵚足跟向
			會合__手	會合 <u>雙</u> 手
			始__豎腳跟向上	始 <u>雙</u> 豎腳跟向上
3816	厓		音俟，水 <u>𡵚</u>	音俟，水 <u>厓</u>
			地名，亦作「珠 <u>𡵚</u> 」	地名，亦作「珠 <u>厓</u> 」
3948	汧𠂔𠂔	洟	<u>緡</u> 陽、漣水以東	<u>洟</u> 陽、漣水以東
3951	湏		服之，皆 <u>純</u> 熱服	服之，皆 <u>湏</u> 熱服

編號	造字	取代	檔案原文摘錄	手動取代結果
			便應 <u>純</u> 依土俗	便應 <u>湏</u> 依土俗
			但 <u>純</u> 微發驚	但 <u>湏</u> 微發驚
			體疼。所以 <u>純</u> 解其	體疼。所以 <u>湏</u> 解其
3961	灾	災	一種致 <u>鮮</u> 兇神	一種致 <u>災</u> 兇神
3973	衤 𠂔 扁		原作「 <u>貫</u> 」，形近之誤	原作「 <u>衤 𠂔 扁</u> 」，形近之誤
			一寸而色白，形小 <u>貫</u>	一寸而色白，形小 <u>衤 𠂔 扁</u>
4012	鼓	鼓	身體虛腫，腹內 <u>三</u> 脹	身體虛腫，腹內 <u>鼓</u> 脹
4134	衛	衛	<u>贗</u> 氣之行	<u>衛</u> 氣之行
4256	鷄	雞	<u>雥</u> (xml 檔缺圖表，致頻次不符，手動全部取代)	<u>雞</u> (xml 檔缺圖表，致頻次不符，手動全部取代)
4275	义	義	<u>𩚑</u> 。然胞衣不出	<u>義</u> 。然胞衣不出
			更有臨床指導意 <u>𩚑</u>	更有臨床指導意 <u>義</u>
4286	从		<u>𩚑</u> (xml 檔缺圖表，致頻次不符，手動全部取代)	<u>从</u> (xml 檔缺圖表，致頻次不符，手動全部取代)
4296	塹		不治，乃至 <u>𩚑</u>	不治，乃至 <u>塹</u>
			<u>𩚑</u> (ou 嘔)	<u>塹</u> (ou 嘔)
			「 <u>𩚑</u> 」，隆起之沙堆	「 <u>塹</u> 」，隆起之沙堆

編號	造字	取代	檔案原文摘錄	手動取代結果
			「 <u>蛸</u> ，沙堆。」	「 <u>塹</u> ，沙堆。」
4495	麩		亦可服大麥 <u>麩</u>	亦可服大麥 <u>麩</u>
			大麥 <u>麩</u> (chao 炒)良	大麥 <u>麩</u> (chao 炒)良
			作「大麥 <u>麩</u> 方」	作「大麥 <u>麩</u> 方」
			「大麥 <u>麩</u> 」	「大麥 <u>麩</u> 」
			食 <u>麩</u> 自愈	食 <u>麩</u> 自愈
4515	呪		<u>婁</u> (xml 檔缺圖表，致頻次不符，手動全部取代)	<u>呪</u> (xml 檔缺圖表，致頻次不符，手動全部取代)
4668	噐		築者，直舂之 <u>王</u>	築者，直舂之 <u>噐</u>
4767	懷		<u>尢</u> (xml 檔缺圖表，致頻次不符，手動全部取代)	<u>懷</u> (xml 檔缺圖表，致頻次不符，手動全部取代)
4816	彳		作「解 <u>彳</u> 厶亦」	作「解 <u>彳</u> 厶亦」
5013	彳		<u>月</u> (全部取代)	<u>彳</u> (全部取代)
5075	彳	丰	「文厶 <u>丰</u> 」	「文厶 <u>丰</u> 」
5292	𦵏		已瘥未健，詣華 <u>堊</u>	已瘥未健，詣華 <u>𦵏</u>
			華 <u>堊</u> (敷)	華 <u>𦵏</u> (敷)
			<u>堊</u> 曰：雖瘥尚虛	<u>𦵏</u> 曰：雖瘥尚虛

編號	造字	取代	檔案原文摘錄	手動取代結果
5320	瘡		除 <u>𤑔</u>	除 <u>疔</u> <u>△</u> <u>奎</u>
5402	朮		所用之 <u>𣎵</u> 散及	所用之 <u>朮</u> 散及
5583	飧	飧	逆之則傷肺，冬為 <u>𤑔</u> 泄	逆之則傷肺，冬為 <u>飧</u> 泄
5753	潛	潛	而胃內 <u>𣎵</u> 冷	而胃內 <u>潛</u> 冷
			龍 <u>𣎵</u> 庵編著	龍 <u>潛</u> 庵編著
5809	踰		字元作 <u>𠂇</u>	字元作 <u>𠂇</u> <u>△</u> <u>凡</u>