

《新語校注》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
2008/2/13 陳建安 製作

一、《新語校注》一書(新語校注.xml)使用舊版造字 136 個，字頻 587 次，詳如附件一。這 136 個造字中，118 個可轉成 Windows XP 能顯示的字，字頻 451 次；另外 18 個字必須轉成構字式，字頻 136 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 4910 的「𠄎𠄎𠄎」字，原字為「𠄎」。

2、重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代：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」。在使用構字式的 xml 檔中，舊漢籍造字也會呈現為重碼的部件外字，因而必須將採用舊漢籍造字部分以對應的 Unicode 或構字式取代，才能呈現正確字形；至於採用部件外字者則不需轉碼。如編號 5446，字碼 8BCD，對應的舊漢籍造字為「度𠄎𠄎」，部件外字為「𠄎」，一一比對原書，確認是「度𠄎𠄎」者，再以手動取代。

3、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「血𠄎𠄎」係「眾」字之異體，故以「眾」字取代。又編號 4507 的「吳」，係「吳」之簡體字，亦以「吳」字取代。

4、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 3935 的「槩」字，Unicode 字型呈現為「槩」，實際上仍為同一字。

5、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 5749 的「片𠄎庸」。

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 3894 的「筋」字，Unicode 編碼為 4225，由於 Windows XP 並不支援，仍須使用構字式「𠄎𠄎𠄎」。

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《新語校注》手動取代表欄位說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、取代：實際使用字形與漢籍舊版造字不同時，於此欄列出。

丙、檔案原文摘錄：摘錄檔案(新語校注.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。

丁、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。

五、《新語校注》中未確定字形「❓」共 6 處，詳細處理內容可參考附件三的未確定字形「❓」手動取代表。附件三《新語校注》未確定字形「❓」手動取代表欄位說明如下：

- 甲、 檔案原文摘錄：摘錄檔案(新語校注.xml)中包含未確定字形「」的文句段落，其中紅字底線的部分，為需要手動取代的字形。
- 乙、 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。

附件一、《新語校注》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
813	旤	1	1	8E5B	65E4	旤		
899	馳	3	3	8ED3	99DE	馳		
905	𠂔	1	1	8ED9	28E0F		𠂔	
971	𠂔	1	1	8F5C	48D9		崩𠂔𠂔	
1044	儻	2	2	8FC7	511B	儻		
1067	碁	2	2	8FDE	68CA	碁		
1077	稟	1	1	8FE8	7A3E	稟		
1097	嶠	1	1	8FFC	37ED		山𠂔昏	
1133	瓌	1	1	9061	74CC	瓌		
1205	眇	1	1	90CB	770E	眇		
1226	罰	8	8	90E0	7F78	罰		
1530	紕	1	1	92D6	7CFA	紕		
1565	繇	6	6	92F9	7DDC	繇		
1642	脉	2	2	93A9	8109	脉		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1871	蟲	1	1	94F1	87C1	蟲		
1885	蠱	1	1	9540	882D	蠱		
1906	衰	1	1	9555	88E6	衰		
1981	讐	4	4	95C2	8B90	讐		
1982	調	1	1	95C3	8B87	調		
2524	饒	2	2	994B	994D	饒		
2718	僞	2	2	9A70	5097	僞		茲，校對問題，舊版漢籍錯字，可用程式全部取代。
2806	勒	1	1	9AEA	976D	勒		駟，校對問題，舊版漢籍錯字，可用程式全部取代。
2854	捺	1	1	9B5B	6438	捺		獻，校對問題，舊版漢籍錯字，可用程式全部取代。
2865	蕝	3	3	9B66	8608	蕝		獻，校對問題，舊版漢籍錯字，可用程式全部取代。
2866	藹	3	3	9B67	8649	藹		紕 ㄞ 帶，校對問題，舊版漢籍錯字，可用程式全部取代。
2867	蟹	1	1	9B68	87B1	蟹		翽，校對問題，舊版漢籍錯字，可用程式全部取代。
3787	冢	7	7	8152	51A2	冢		
3791	况	2	2	8156	51B5	况		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3804	効	4	4	8163	52B9	効		
3810	匱	5	5	8169	5335	匱		
3814	却	4	4	816D	5374	却		
3815	𡗗	1	1	816E	537A	𡗗		
3816	厓	1	1	816F	5393	厓		
3817	厠	1	1	8170	53A0	厠		
3819	廝	1	1	8172	53AE	廝		
3820	厦	1	1	8173	53A6	厦		
3824	叙	16	16	8177	53D9	叙		
3826	叶	4	4	8179	53F6	叶		
3840	坂	3	3	81A9	5742	坂		
3860	𡗗	1	1	81BD	5C20	𡗗		
3868	卮	1	1	81C5	5DF5	卮		
3876	廻	1	1	81CD	5EFB	廻		
3882	𡗗	1	1	81D3	5F72	𡗗		
3883	徧	3	3	81D4	5FA7	徧		
3884	忽	2	2	81D5	6031	忽		
3889	𡗗	1	1	81DA	6164	𡗗		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3891	慙	7	7	81DC	6159	慙		
3894	筋	2	2	81DF	4225		𠂇𠂇筋	
3921	暫	1	1	81FA	6673	暫		
3930	憵	1	1	8244	6901	憵		
3931	椀	3	3	8245	6900	椀		
3935	槩	10	10	8249	69E9	槩		Unicode 呈現差異。
3937	檝	1	1	824B	6A9D	檝		
3969	煊	9	9	826B	714A	煊		
3974	牀	5	5	8270	7240	牀		
3976	犁	4	4	8272	7282	犁		
3977	狗	1	1	8273	72E5	狗		
3985	璫	3	3	827B	7447	璫		Unicode 呈現差異。
3996	疎	4	4	82A8	758E	疎		
4010	皐	4	4	82B6	7690	皐		Unicode 呈現差異。
4011	皞	2	2	82B7	76A1	皞		Unicode 呈現差異。
4014	蓋	1	1	82BA	8462	蓋		
4028	稟	1	1	82C8	7980	稟		
4034	穉	1	1	82CE	7A49	穉		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4038	竈	1	1	82D2	7AC8	竈		
4040	竝	12	12	82D4	7ADD	竝		
4070	縹	1	1	82F2	7DE4	縹		
4077	罇	2	2	82F9	7F47	罇		
4080	羣	61	61	82FC	7FA3	羣		Unicode 呈現差異。
4082	耇	2	2	82FE	8008	耇		
4085	耻	2	2	8342	803B	耻		
4090	脚	1	1	8347	811A	脚		
4124	螽	1	1	8369	8771	螽		
4132		96	96	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	13	13	8373	885E	衛		
4136	袞	6	6	8375	887A	袞		
4140	衽	1	1	8379	88B5	衽		
4142	袞	1	1	837B	8943	袞		
4145	羈	1	1	837E	898A	羈		Unicode 呈現差異。
4146	覩	9	9	83A1	89A9	覩		
4149	覩	1	1	83A4	89BC	覩		
4156	諂	1	1	83AB	8B1F	諂		諂，校對問題，舊版漢籍錯字，可用

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
								程式全部取代。
4158	譌	13	13	83AD	8B4C	譌		Unicode 呈現差異。
4172	厯	1	1	83BB	8E77	厯		Unicode 呈現差異。
4176	輒	1	1	83BF	8F19	輒		
4178	輓	2	2	83C1	8F2D	輓		
4180	臯	2	2	83C3	8FA0	臯		
4184	迹	5	5	83C7	8FF9	迹		
4193	鈎	2	2	83D0	920E	鈎		
4234	羸	1	1	83F9	9A58	羸		
4256	鷄	4	4	8450	9DC4	鷄		
4259	麇	1	1	8453	9E95	麇		Unicode 呈現差異。
4265	鼃	1	1	8459	9F03	鼃		
4266	鼃	4	4	845A	9F02	鼃		
4267	鼃	2	2	845B	9F08	鼃		
4286	从	10	10	846E	4ECE	从		
4307	罵	4	4	84A5	99E1	罵		
4309	刼	11	11	84A7	523C	刼		
4348	惚	5	5	84CE	506C	惚		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4437	廩	1	1	8568	5EEA	廩		
4496	段	3	3	85C5	53DA	段		
4507	吳	27	27	85D0	5434	吳		吳，異體字問題。
4515	呪	1	1	85D8	546A	呪		
4545	踰	1	1	85F6	28105		足厶區	
4629	坳	1	1	86AD	577F	坳		
4762	旃	1	1	8773	3ACB		舂厶冉	
4768	妬	2	2	8779	59AC	妬		
4776	鯈	1	1	87A3	9BB7	鯈		
4785	贗	1	1	87AC	8D0B	贗		
4894		4	4	885A			宀厶心厶冉	甯，校對問題，舊版漢籍錯字，可用程式全部取代。
4904	尅	2	2	8864	5C05	尅		
4907	橐	1	1	8867	69D6	橐		
4910		1	1	886A			辵厶畷	淫，校對問題，舊版漢籍錯字，可用程式全部取代。
4940	韵	1	1	88AA	97F5	韵		
4943	扌	1	1	88AD			扌	
4957	崑	2	2	88BB	5D53	崑		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4996	斲	3	3	88E2	65B5	斲		
5019	畧	3	3	88F9			畧	
5031	閨	3	3	8946	95B4	閨		
5034	弃	9	9	8949	5F03	弃		Unicode 呈現差異。
5133	憇	1	1	89CE	6199	憇		
5221		7	7	8A67			𠄎白米𠄎	粵，異體字問題。
5286	窻	1	1	8ACA	41AB		穴𠄎忽	
5434	湲	6	6	8BC1	5E7C		𠄎氵白工𠄎	
5438	栳	3	3	8BC5	681E	栳		Unicode 呈現差異。
5446	𡗗	5	6	8BCD	22F85		度𠄎𡗗	與𠄎重碼，重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代。
5492	握	2	2	8BFB	6943	握		
5537	离	2	2	8C69	79BC	离		
5558	龕	1	1	8C7E	9E84	龕		
5578	弛	2	2	8CB4	38AE		弓𠄎𠄎	
5588	恠	2	2	8CBE	6061	恠		
5705	廐	1	1	8D74	5ED0	廐		Unicode 呈現差異。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5706	敍	27	27	8D75	654D	敍		
5749		1	1	8DC2			片厶庸	待造字。
5752	摠	2	2	8DC5	63D4	摠		
5859		1	1	C6D2				支，校對問題，舊版漢籍錯字，可用程式全部取代。

附件二、《新語校注》手動取代表

編號	檔案原文摘錄	手動取代結果
5446	斆乃𠄎之訛	斆乃度厶文之訛
	惟其𠄎丹臚	惟其度厶文丹臚
	案斆當作𠄎	案斆當作度厶文
	𠄎，孔穎達正義本作斆	度厶文，孔穎達正義本作斆
	改周書之𠄎為斆	改周書之度厶文為斆

附件三、《新語校注》未確定字形「𠄎」手動取代表

檔案原文摘錄	手動取代結果
『寂』，古文作『𠄎』	『寂』，古文作『𠄎厶豕』
「撥」，「𠄎」之借字	「撥」，「𠄎厶止」之借字
說文𠄎部云：『𠄎，足刺𠄎也』	說文𠄎厶止部云：『𠄎厶止，足刺𠄎厶止也』
草書『存』作『𠄎』	古文字體暫維持原狀。