

《唐語林校證》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

2008/4/10 陳建安 製作

一、《唐語林校證》一書(唐語林校證.xml)使用舊版造字 209 個，字頻 1988 次，詳如附件一。這 209 個造字中，163 個可轉成 Windows XP 能顯示的字，字頻 1324 次；另外 46 個字必須轉成構字式，字頻 664 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 4594 的「秋𪛗鳥」字，原字為「鵜」。
- 2、舊版漢籍錯字，必須人工取代：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，其中如編號 309「𪛗」，為未確定字形，無法以程式全部取代，必須人工一一比對更正。
- 3、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或誤刪造字，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 535 的「𪛗」字，在 xml 檔第 167 頁被刪除了一次，因而頻次不同，需於原缺處補上，並將其他未經修改的部分，則直接手動取代為「𪛗」字。

- 4、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，對於某些標題句做了增刪的動作，因為這些標題句的內容包含了缺字，所以這些更改標記的動作造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號 4129 的「虬」字，xml 檔在 803 頁減少了一個標題，因而造成頻次不同，處理方式為將所有頻次皆修改為「虬」。
- 5、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 1772 的「梨」係「藜」字之異體，故以「藜」字取代。又編號 3825 的「疊」字，係「疊」字之簡體字，亦以標準字的「疊」字取代。
- 6、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 1076 的「𪛗」字，Unicode 字型呈現為「𪛗」，實際上仍為同一字。
- 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 1375 的「王蒙」。

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 1384 的「𪛗」字，Unicode 編碼為 40AC，由於 Windows XP 並不支援，仍須使用構字式「石」。

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的欄位說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、取代：當實際使用字形與造字不同時，在此標出。

丙、檔案原文摘錄：摘錄檔案(唐語林校證.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。

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。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、《唐語林校證》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𡗗	10	10	FBEC			𡗗	
297	𡗘	29	29	FBED			𡗘	
298	𡗙	16	16	FBEE			𡗙	
299	𠂔	3	3	FBEF			𠂔	
300	𠂕	3	3	FBF0			𠂕	
309	𠂖	2	2	FBF9			𠂖	𠂔那體說文𠂕、𠂔冉體說文𠂕，舊版漢籍錯字，必須人工取代。
310	𠂗	2	2	FBFA			𠂗	几，校對問題，舊版漢籍錯字，可用程式全部取代。
500	𠂘	409	409	FD5C			𠂘	卷，異體字問題。
535	𠂙	6	5	FDA1			𠂙	誤刪，校對問題，新版漢籍錯字，人工取代。
573	𠂚	2	2	FDC7			𠂚	
615	𠂛	1	1	FDF1			𠂛	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
658	𠂇	1	1	FE5D			𠂇	
675	𠂇	6	5	FE6E			𠂇	誤刪，校對問題， 新版漢籍錯字，人 工取代。
685	𠂇	1	1	FE78			𠂇	
707	𠂇	1	1	FEB0			𠂇	
708	𠂇	4	4	FEB1			𠂇	
742	𠂇	1	1	FED3			𠂇	
762	𠂇	1	1	FEE7			𠂇	人，校對問題，舊 版漢籍錯字，可用 程式全部取代。
770	𠂇	1	1	FEEF			𠂇	
780	𠂇	3	2	FEF9			𠂇	誤刪，校對問題， 新版漢籍錯字，人 工取代。
814	𠂇	1	1	8E5C	87C7	𠂇		
985	𠂇	1	1	8F6A	6197	𠂇		
994	𠂇	9	9	8F73	8E75	𠂇		
1000		2	2	8F79			王𠂇已	玨，異體字問題。
1010	玨	24	24	8FA5	73CF	玨		
1067	𠂇	2	2	8FDE	68CA	𠂇		棋，異體字問題。
1069	𠂇	5	5	8FE0	745D	𠂇		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1076	𪛗	1	1	8FE7	5368	𪛗		Unicode 呈現差異。
1126	璿	1	1	905A	74B9	璿		
1133	瓊	37	37	9061	74CC	瓊		
1146	璿	6	6	906E	7506	璿		
1375		2	2	91D8			王𪛗蒙	待造字。
1384	碧	2	2	91E1	40AC		巩𪛗石	
1520	𪛗	1	1	92CC	3B20		日𪛗煞	
1530	紉	8	8	92D6	7CFA	紉		
1544	絙	3	3	92E4	7D59	絙		
1565	𪛗	2	2	92F9	7DDC	𪛗		
1633		1	1	937E			𪛗𪛗𪛗	待造字。
1646	𪛗	1	1	93AD	8117	𪛗		
1652	𪛗	11	11	93B3	3D29		𪛗𪛗𪛗	
1673	𪛗	1	1	93C8	34BC		𪛗	
1710	𪛗	1	1	93ED	3E01		火𪛗業	待造字。
1772	𪛗	1	1	946C			𪛗𪛗𪛗梨	𪛗，異體字問題。
1944	𪛗	1	1	957B	8A67	𪛗		
1981	𪛗	3	3	95C2	8B90	𪛗		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2050	蹠	1	1	9648	8E4F	蹠		
2056	佺	1	1	964E	4FD2	佺		
2157	酎	1	1	96D5	9167	酎		
2168	躊	1	1	96E0	91BB	躊		
2213	鉷	3	3	974E	9277	鉷		
2232		1	1	9761			𠂔厂墨卩	𠂔，異體字問題。
2244		6	6	976D			ノ厶虫	待造字。
2350	慘	19	18	97F9	50AA	慘		誤刪，校對問題， 新版漢籍錯字，人 工取代。
2377	𠂔	9	9	9855	279AC		𠂔工几言	
2505	𠂔	2	2	98F7	98DC	𠂔		
2512	𠂔	15	15	98FE	26806		月厶巽	
2516	𠂔	2	2	9943	9919	𠂔		
2520	𠂔	4	4	9947	993B	𠂔		
2567	𠂔	7	7	9976	7A21	𠂔		
2592	𠂔	2	2	99B1	8D1C	𠂔		
2602	𠂔	1	1	99BB	9C10	𠂔		
2603	𠂔	1	1	99BC	4874		車厶童	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2681	𡵗	1	1	9A4B	3C54		此𡵗夕	
2820	𡵗	1	1	9AF8	64EA	𡵗		
3021	𡵗	1	1	9C65			𡵗	
3257	𡵗	1	1	9DD6	85CA	𡵗		
3780	𡵗	2	2	814B	5101	𡵗		
3789	冲	7	7	8154	51B2	冲		冲，異體字問題。
3791	況	8	8	8156	51B5	況		況，異體字問題。
3792	凉	1	1	8157	51C9	凉		凉，異體字問題。
3796	決	5	5	815B	51B3	決		決，異體字問題。
3797	凭	1	1	815C	51ED	凭		憑，異體字問題。
3805	勅	1	1	8164	52C5	勅		
3813	𡵗	2	2	816C	536D	𡵗		
3814	却	6	6	816D	5374	却		
3822	叁	2	2	8175	53C1	叁		參，異體字問題。
3824	叙	114	114	8177	53D9	叙		敘，異體字問題。
3825	疊	14	14	8178	53E0	疊		疊，異體字問題。
3826	叶	9	9	8179	53F6	叶		
3829	咏	3	3	817C	548F	咏		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3835	噍	2	2	81A4	564D	噍		
3843	堦	4	4	81AC	5826	堦		
3846	壻	36	36	81AF	58FB	壻		Unicode 呈現差異。
3849	獎	1	1	81B2	596C	獎		獎，異體字問題。
3850	姪	2	2	81B3	59D9	姪		
3855	擘	2	2	81B8	5B7C	擘		
3858	寃	15	15	81BB	5BC3	寃		冤，異體字問題。
3860	𪛗	1	1	81BD	5C20	𪛗		
3868	卮	3	3	81C5	5DF5	卮		卮，異體字問題。
3883	徧	10	10	81D4	5FA7	徧		
3884	忽	2	2	81D5	6031	忽		
3891	慙	16	16	81DC	6159	慙		慚，異體字問題。
3909	擡	1	1	81EE	64E1	擡		
3916	𪛗	1	1	81F5	656D	𪛗		
3920	皓	2	2	81F9	6667	皓		
3931	椀	2	2	8245	6900	椀		
3935	槩	2	2	8249	69E9	槩		概，異體字問題。
3937	檝	2	2	824B	6A9D	檝		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3949	涖	1	1	8257	6D96	涖		
3966	烟	2	2	8268	70DF	烟		煙，異體字問題。
3974	牀	28	28	8270	7240	牀		
3975	牕	1	1	8271	7255	牕		Unicode 呈現差異。
3978	猪	7	7	8274	732A	猪		Unicode 呈現差異。
3987	穀	2	2	827D	7474	穀		
3990	甄	1	1	82A2	750E	甄		
3991	攜	1	1	82A3	64D5	攜		
3996	疎	8	8	82A8	758E	疎		
4010	阜	15	15	82B6	7690	阜		
4011	皐	11	11	82B7	76A1	皐		Unicode 呈現差異。
4013	盃	1	1	82B9	76CC	盃		
4015	眦	1	1	82BB	7726	眦		
4016	着	19	19	82BC	7740	着		著，異體字問題。
4038	竈	5	5	82D2	7AC8	竈		
4042	笋	3	3	82D6	7B0B	笋		
4044	筴	4	4	82D8	7B6F	筴		
4051	籐	2	2	82DF	7C35	籐		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4055	𪗇	2	2	82E3	7C83	𪗇		
4068	綫	6	6	82F0	7DAB	綫		線，異體字問題。
4077	罇	1	1	82F9	7F47	罇		
4080	羣	104	104	82FC	7FA3	羣		群，異體字問題。
4081	翱	11	11	82FD	7FFA	翱		
4089	脇	1	1	8346	8107	脇		
4090	脚	10	10	8347	811A	脚		
4091	臯	2	2	8348	81EF	臯		
4104	莅	1	1	8355	8385	莅		
4107	葱	1	1	8358	8471	葱		
4129	虬	61	59	836E	866C	虬		標記問題。
4132		74	74	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	83	83	8373	885E	衛		衛，異體字問題。
4138	褰	1	1	8377	88A0	褰		
4139	袴	3	3	8378	88B4	袴		
4140	裱	1	1	8379	88B5	裱		
4146	覩	11	11	83A1	89A9	覩		
4158	譌	10	10	83AD	8B4C	譌		Unicode 呈現差異。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4173	𪗇	3	3	83BC	8EAD	𪗇		
4176	𪗈	5	5	83BF	8F19	𪗈		
4184	迹	27	27	83C7	8FF9	迹		
4190	邳	3	3	83CD	90C4	邳		
4193	鈎	1	1	83D0	920E	鈎		
4195	鉢	5	5	83D2	9262	鉢		
4203	隣	3	3	83DA	96A3	隣		鄰，異體字問題。
4216	鞞	1	1	83E7	97C8	鞞		
4220	穎	4	4	83EB	9834	穎		穎，異體字問題。
4232	鷲	3	3	83F7	96B2	鷲		鷲，異體字問題。
4237	髡	3	3	83FC	9AE0	髡		髡，異體字問題。
4240	髯	1	1	8440	9AF4	髯		
4242	鬪	11	11	8442	9B2D	鬪		Unicode 呈現差異。
4251	鰕	1	1	844B	9C15	鰕		
4256	鷄	23	23	8450	9DC4	鷄		
4259	麇	1	1	8453	9E95	麇		Unicode 呈現差異。
4266	鼃	1	1	845A	9F02	鼃		
4267	鼈	6	6	845B	9F08	鼈		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4278	亘	1	1	8466	4E98	亘		
4309	刼	2	2	84A7	523C	刼		劫，異體字問題。
4350	侶	2	2	84D0	5058	侶		
4369	倬	1	1	84E3	666B	倬		
4384	鞞	1	1	84F2	97A7	鞞		
4450	𠂔	1	1	8575	524F	𠂔		𠂔，異體字問題。
4458	勑	2	2	857D	52CC	勑		Unicode 呈現差異。
4460	勅	1	1	85A1	52D1	勅		
4469	勻	6	6	85AA	5304	勻		
4507	吳	103	103	85D0	5434	吳		吳，異體字問題。
4515	呪	5	5	85D8	546A	呪		
4554	唬	1	1	8640	55C1	唬		
4594	鵠	1	1	8668	2A0E9		秋𪛗鳥	鵠，校對問題，舊版漢籍錯字，可用程式全部取代。
4603	憇	2	2	8671	6187	憇		
4619	但	1	1	86A3	4F39	但		
4735	壠	3	3	8758	58E0	壠		
4753		1	1	876A			𪛗𪛗比𪛗大	𪛗𪛗大，校對問題，舊版漢籍錯字，可用程式全部

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
								取代。
4768	妬	10	10	8779	59AC	妬		妒，異體字問題。
4817	寶	15	15	87CC	5BF3	寶		寶，異體字問題。
4902	兗	4	4	8862	5156	兗		
4903	𪛗	1	1	8863	5C00	𪛗		
4904	尅	2	2	8864	5C05	尅		
5008	幪	5	5	88EE	5E5E	幪		
5013	彳	2	2	88F3			彳	
5028	昇	1	1	8943	662A	昇		
5075	彳	1	1	8972			彳	
5109	櫛	1	1	89B6	6A10	櫛		
5126	暮	12	12	89C7	8B29	暮		
5131	窻	2	2	89CC	7ABB	窻		Unicode 呈現差異。
5192	𨾏	1	1	8A4A	969F	𨾏		
5225	邨	1	1	8A6B	460F		血𨾏𠂔	
5290	携	1	1	8ACE	643A	携		
5299	戟	7	7	8AD7	39B8		卓𨾏戈	戟，異體字問題。
5437	并	15	15	8BC4	5E77	并		并，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5439	栢	1	1	8BC6	6822	栢		
5463	岢	1	1	8BDE	65F9	岢		
5542	鑠	8	8	8C6E	93C1	鑠		
5544	硎	1	1	8C70	2543B		石𠩺冈	厭，校對問題，舊版漢籍錯字，可用程式全部取代。
5549	藁	7	7	8C75	85F3	藁		
5551	賁	4	4	8C77	8CF7	賁		
5555	迥	1	1	8C7B	9008	迥		
5558	麓	1	1	8C7E	9E84	麓		
5583	殮	2	2	8CB9	98F1	殮		
5587	歛	2	2	8CBD	6B58	歛		
5590	朶	2	2	8CC0	6736	朶		朶，異體字問題。
5591	杞	18	18	8CC1	233CC		木𠩺巳	杞，異體字問題。
5697	𪗇	1	1	8D6C	7151	𪗇		煮，異體字問題。
5702	厨	1	1	8D71	53A8	厨		廚，異體字問題。
5704	𪗇	1	1	8D73	6C0A	𪗇		
5705	廐	10	10	8D74	5ED0	廐		廐，異體字問題。
5706	敘	88	88	8D75	654D	敘		敘，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5720	埇	23	23	8DA5	578D	埇		
5721	頓	14	14	8DA6	9814	頓		
5753	潛	1	1	8DC6	6F5C	潛		潛，異體字問題。
5769	邠	2	2	8DD6	90CD	邠		
5770	汧	3	3	8DD7			汧 𣶒 汧	潛，異體字問題。
5860	无	2	2	C6D3	65E0	无		

附件二、《唐語林校證》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
309	𠂔那體說文	「那」篆文作「𠂔」	「那」篆文作「𠂔那體說文」
	𠂔冉體說文	「𠂔」「舟」形近	「𠂔冉體說文」「舟」形近
535		<qe>失之	<qe>𠂔失之
		𠂔 (全部取代)	𠂔 (全部取代)
675		<qe> 崔頭	<qe>𠂔 崔頭
		𠂔 (全部取代)	𠂔 (全部取代)
780		助<qe> 助	助<qe>𠂔 助
		𠂔 (全部取代)	𠂔 (全部取代)
2350		<qe>貞元六年進士	<qe>𠂔貞元六年進士
		𠂔 (全部取代)	𠂔 (全部取代)
4129		𠂔 (全部取代)	𠂔 (全部取代)