

《大唐新語》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
2008/4/11 黃榮順 製作

一、《大唐新語》一書(大唐新語.xml)使用舊版造字 110 個，字頻 395 次，詳如附件一。這 110 個造字中，92 個可轉成 Windows XP 可顯示的字，字頻 308 次；另外 18 個字必須轉成構字式，字頻 87 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 3858 的「冤」係「冤」字之異體，故以「冤」字取代。又如編號 3966 的「烟」係「煙」字之簡體，故以「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 3846 的「壻」字，Unicode 字型呈現為「壻」，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。所以如造字編號 1600 的「獮」字，Unicode 編碼為 3E84，由於 Windows XP 並不支援，仍須改用構字式「犴 𪛗」。

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、「大唐新語.xml」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	
297	𡗘	4	4	FBED			𡗘	
299	𡗙	9	9	FBEF			𡗙	
300	𡗚	9	9	FBF0			𡗚	
519	𡗛	9	9	FD6F			𡗛	
553	𡗜	1	1	FDB3			𡗜	
766	𡗝	1	1	FEED			𡗝	
820	𡗞	3	3	8E62	70F1	𡗞		
866	𡗟	1	1	8EB2	247ED		𡗟𡗗犬	𡗟𡗗大，異體字問題。
985	𡗠	1	1	8F6A	6197	𡗠		
1014	𡗡	1	1	8FA9	73CE	𡗡		珍，異體字問題。
1053	𡗢	1	1	8FD0	6160	𡗢		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1069	瑄	4	4	8FE0	745D	瑄		
1126	璫	3	3	905A	74B9	璫		
1133	瓊	12	12	9061	74CC	瓊		
1237	稍	1	1	90EB	77DF	稍		
1490	籩	1	1	92AE	7C5D	籩		
1530	紉	1	1	92D6	7CFA	紉		
1600	猷	1	1	935D	3E84		犴 𤝵 猷	
1642	脉	4	4	93A9	8109	脉		
1893	袞	3	3	9548	886E	袞		
1944	簪	1	1	957B	8A67	簪		
1981	讐	13	13	95C2	8B90	讐		
2151	𠂔	1	1	96CF	961D	𠂔		
2402	勳	2	2	986E	52EE	勳		
2441	奢	3	3	98B7	3E19		父 𠂔 者	
2578	勛	1	1	99A3	20892		𠂔 𠂔 比力 𠂔	
2834	攵	9	9	9B47	6535	攵		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3269	憑	1	1	9DE2	51F4	憑		
3789	冲	8	8	8154	51B2	冲		沖，異體字問題。
3791	況	2	2	8156	51B5	況		況，異體字問題。
3792	凉	1	1	8157	51C9	凉		涼，異體字問題。
3796	決	1	1	815B	51B3	決		決，異體字問題。
3814	却	5	5	816D	5374	却		
3817	廁	1	1	8170	53A0	廁		廁，異體字問題。
3819	廝	1	1	8172	53AE	廝		廝，異體字問題。
3824	叙	9	9	8177	53D9	叙		敘，異體字問題。
3826	叶	5	5	8179	53F6	叶		
3840	坂	2	2	81A9	5742	坂		
3843	堦	1	1	81AC	5826	堦		
3846	堦	1	1	81AF	58FB	堦		Unicode 字型差異。
3858	冤	25	25	81BB	5BC3	冤		冤，異體字問題。
3864	峯	1	1	81C1	5CEF	峯		峰，異體字問題。
3872	鸛	1	1	81C9	4CF5		勅𪗇鳥	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3876	廻	4	4	81CD	5EFB	廻		迴，異體字問題。
3880	護	2	2	81D1	5F60	護		
3935	槩	2	2	8249	69E9	槩		概，異體字問題。
3937	檝	1	1	824B	6A9D	檝		
3966	烟	2	2	8268	70DF	烟		煙，異體字問題。
3974	牀	7	7	8270	7240	牀		
3977	狗	1	1	8273	72E5	狗		
3978	猪	3	3	8274	732A	猪		豬，異體字問題。
3996	疎	1	1	82A8	758E	疎		
4012	鼓	2	2	82B8	76B7	鼓		鼓，異體字問題。
4016	着	17	17	82BC	7740	着		著，異體字問題。
4034	穉	1	1	82CE	7A49	穉		
4041	豎	2	2	82D5	7AEA	豎		
4077	罇	1	1	82F9	7F47	罇		
4078		1	1	82FA			已会升	异，異體字問題。
4080	羣	33	33	82FC	7FA3	羣		群，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4089	脇	1	1	8346	8107	脇		
4090	脚	7	7	8347	811A	脚		腳，異體字問題。
4091	臯	1	1	8348	81EF	臯		
4094	舘	1	1	834B	8218	舘		館，異體字問題。
4105	菓	1	1	8356	83D3	菓		
4115	藁	1	1	8360	85C1	藁		
4132		35	35	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	7	7	8373	885E	衛		衛，異體字問題。
4139	袴	2	2	8378	88B4	袴		
4146	覩	3	3	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 字型差異。
4167	賫	4	4	83B6	8CEB	賫		
4169	趲	1	1	83B8	8DA6	趲		
4173	𨔵	2	2	83BC	8EAD	𨔵		
4174	𨔶	1	1	83BD	8EB6	𨔶		
4184	迹	1	1	83C7	8FF9	迹		跡，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4193	鈎	1	1	83D0	920E	鈎		
4198	鑛	1	1	83D5	945B	鑛		
4223	顙	1	1	83EE	984B	顙		
4232	隰	2	2	83F7	96B2	隰		鷺，異體字問題。
4242	鬪	4	4	8442	9B2D	鬪		
4253	鳬	1	1	844D	9CF8	鳬		
4256	鷄	6	6	8450	9DC4	鷄		雞，異體字問題。
4278	亘	1	1	8466	4E98	亘		
4309	刼	2	2	84A7	523C	刼		劫，異體字問題。
4342	倮	1	1	84C8	502E	倮		
4350	侶	1	1	84D0	5058	侶		
4426	嶺	1	1	855D	21E6E		山𡵓𡵓	
4437	廩	1	1	8568	5EEA	廩		廩，異體字問題。
4507	吳	5	5	85D0	5434	吳		吳，異體字問題。
4515	呪	2	2	85D8	546A	呪		
4768	妬	3	3	8779	59AC	妬		妒，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4779	鞣	1	1	87A6	97B2	鞣		
4834	𥇑	3	3	87DD	3DE1		𠃹目巨火𠃹	
4902	兗	1	1	8862	5156	兗		
4904	尅	1	1	8864	5C05	尅		
4927	𡵓	1	1	887B	5C9E	𡵓		
5008	幪	5	5	88EE	5E5E	幪		
5019	畧	1	1	88F9	7567	畧		略，異體字問題。
5126	暮	7	7	89C7	8B29	暮		
5133	憇	1	1	89CE	6199	憇		
5176		5	5	89F9			氵𠂔冗	沉，異體字問題。
5225	邨	1	1	8A6B	460F		血𠂔卩	
5439	栢	2	2	8BC6	6822	栢		
5482	𧈧	1	1	8BF1	8812	𧈧		
5702	厨	2	2	8D71	53A8	厨		廚，異體字問題。
5704	𧈧	1	1	8D73	6C0A	𧈧		
5705	廐	2	2	8D74	5ED0	廐		廐，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5706	敘	5	5	8D75	654D	敘		敘，異體字問題。
5720	垧	7	7	8DA5	578D	垧		