

《九家舊晉書輯本》舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中央研究院語言所文獻語料小組
2009/01/09 丁致伶

轉碼主要工作是把檔案中的舊版造字轉換成 Windows XP 能支援的 Unicode 字形，Unicode 目前共收錄漢字 70194 個字，而 XP 只能支援 20902 個字（詳如表一），不支援之字將以構字式表達。例：造字編號 968 的「登」字，Unicode 編碼是 4741，由於 XP 並不支援，仍需使用構字式「𪛗𪛘豆」。

表一、Unicode 的字數及編碼區段

Unicode	新增字數	新增編碼區段	總字數	WinXP
1.1 版	20902	4E00-9FFF	20902	支援
3.0 版	6582	3400-4DFF	27484	不支援
3.1 版	42710	20000-2A6D6	70194	不支援

一、舊版造字轉碼分析：

《九家舊晉書輯本》使用舊版造字 218 個，字頻 964 次，這 218 個造字中，185 個可轉成 Windows XP 能顯示的字，字頻 914 次；另外 33 個字必須轉成構字式，字頻 50 次。

轉碼完成製作轉碼分析表，請參考附件一《九家舊晉書輯本》轉碼分析表，欄位說明如下：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 造字：舊版造字。
- (三) 頻次：舊版造字在文件的出現次數。
- (四) Big5：造字的 Big5 碼。
- (五) Unicode：造字所對應的 Unicode 碼。

- (六) WinXP：造字在 Windows XP 的對應字形。
- (七) 構字式：Windows XP 無法對應字形改用構字式。
- (八) 備註凡例：備註欄中記錄轉碼後字形及修改原因，凡例如下：
1. 異體字問題：為了使用者查詢和使用的方便，在處理異體字時最主要的方式是以標準字取代，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
 2. 錯字，以程式全部取代：與原書字形不符，並查詢教育部異體字字典確認非異體字後，皆歸類為錯字，以程式取代為正確字形。例：檔案中使用造字編號 3102 的「𠄎口自本𠄎」字，而原書使用之字為「噪」，因此「𠄎口自本𠄎」為錯字，以程式取代為正確字形「噪」。
 3. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。
 4. 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 1000 的「王𠄎巳」。

附件一、《九家舊晉書輯本》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
826	癱	1	8E68	7655	癱		
834	癘	2	8E70	712D	癘		
835	𠄎	1	8E71	5380	𠄎		


編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
858	窠	1	8EAA	7F59	窠		
869	縻	1	8EB5	6A5C	縻		
893	峩	2	8ECD	5CE9	峩		
899	駝	2	8ED3	99DE	駝		
910	崧	1	8EDE	5D78	崧		
931	磕	1	8EF3	791A	磕		
968	登	1	8F59	4741		𠂇厶豆	
982	胷	3	8F67	80F7	胷		胸，異體字問題。
1000		5	8F79			王𠂇巳	待造字。
1006	鯨	2	8FA1	4C50		魚𠂇孚	
1024	虵	10	8FB3	8675	虵		
1044	儻	2	8FC7	511B	儻		
1053	傲	1	8FD0	6160	傲		
1067	碁	2	8FDE	68CA	碁		
1077	稟	1	8FE8	7A3E	稟		
1088	暘	1	8FF3	757C	暘		
1091	漣	5	8FF6	58CD	漣		
1097	嶠	1	8FFC	37ED		山𠂇昏	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1133	瓌	7	9061	74CC	瓌		
1199	眊	1	90C5	7702	眊		
1237	稍	2	90EB	77DF	稍		
1350	褚	1	91BF	7987	褚		褚，異體字問題。
1381	穉	1	91DE	7A3A	穉		
1565	繇	7	92F9	7DDC	繇		
1591	羗	1	9354	7F97	羗		
1642	脉	1	93A9	8109	脉		
1781	葢	1	9475	854B	葢		
1885	逢	1	9540	882D	逢		
1886	榦	1	9541			形 車 人 木 口	
1906	裛	13	9555	88E6	裛		
1926	覓	1	9569	8994	覓		
1981	讐	1	95C2	8B90	讐		讐，異體字問題。
2162	醜	1	96DA	918E	醜		
2193	毳	1	96F9	3CA3		叟 厶 毛	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2439	礧	1	98B5	7920		石厶慈	
2502	颺	2	98F4	98C8	颺		
2524	饜	1	994B	994D	饜		
2651	鸚	1	99EC	9E1C	鸚		
2811		1	9AEF			亞厶田	畱，異體字問題。
2861		1	9B62			凹厶育	待造字。
3102		1	9CD8			形口自 本□	嗶，錯字，以程式全部取代。
3103		1	9CD9			木厶泉	椽，錯字，以程式全部取代。
3104	藪	1	9CDA	861D	藪		
3780	儁	25	814B	5101	儁		
3787	冢	28	8152	51A2	冢		
3789	冲	7	8154	51B2	冲		沖，異體字問題。
3791	况	2	8156	51B5	况		況，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3794	湊	2	8159	51D1	湊		湊，異體字問題。
3804	効	5	8163	52B9	効		
3806	勳	1	8165	52E0	勳		
3814	却	3	816D	5374	却		
3817	廁	5	8170	53A0	廁		廁，異體字問題。
3819	廝	2	8172	53AE	廝		廝，異體字問題。
3824	叙	4	8177	53D9	叙		
3829	咏	9	817C	548F	咏		
3840	坂	4	81A9	5742	坂		
3843	堦	1	81AC	5826	堦		
3846	堦	5	81AF	58FB	堦		
3849	獎	1	81B2	596C	獎		獎，異體字問題。
3855	孽	6	81B8	5B7C	孽		
3858	寃	13	81BB	5BC3	寃		寃，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3860	尠	4	81BD	5C20	尠		
3864	峯	3	81C1	5CEF	峯		
3866	崐	1	81C3	5D10	崐		
3868	卮	1	81C5	5DF5	卮		
3869	币	1	81C6	5E00	币		
3870	帀	1	81C7	5E0B	帀		
3883	徧	2	81D4	5FA7	徧		
3891	慙	14	81DC	6159	慙		
3916	敷	1	81F5	656D	敷		
3920	皓	1	81F9	6667	皓		
3921	皙	9	81FA	6673	皙		
3924	菴	2	81FD	671E	菴		
3930	憵	1	8244	6901	憵		
3937	檝	2	824B	6A9D	檝		
3964	威	1	8266	70D5	威		
3966	烟	1	8268	70DF	烟		
3974	牀	28	8270	7240	牀		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3978	猪	3	8274	732A	猪		Unicode 字型 呈現差異。
3981	珉	8	8277	73C9	珉		
3996	疎	4	82A8	758E	疎		
4000	疴	1	82AC	75B4	疴		
4016	着	4	82BC	7740	着		
4032	秣	1	82CC	2578A		禾  未	秣，異體字問 題。
4034	穉	1	82CE	7A49	穉		
4035	穽	1	82CF	7A7D	穽		
4038	竈	1	82D2	7AC8	竈		
4040	竝	23	82D4	7ADD	竝		
4041	豎	1	82D5	7AEA	豎		
4052	簾	1	82E0	7C34	簾		Unicode 字型 呈現差異。
4058	糶	1	82E6	7CC9	糶		
4070	縹	2	82F2	7DE4	縹		
4073	縊	2	82F5	7E6E	縊		
4080	羣	77	82FC	7FA3	羣		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4081	翱	1	82FD	7FFA	翱		
4090	脚	1	8347	811A	脚		
4091	臯	1	8348	81EF	臯		
4103	苙	1	8354	831D	苙		
4104	莅	6	8355	8385	莅		
4107	葱	1	8358	8471	葱		
4109	蒟	1	835A	84AD	蒟		
4115	藁	1	8360	85C1	藁		
4118	藜	1	8363	8602	藜		
4119		2	8364			形 ^々 正 頁已久 □	夔，錯字，以 程式全部取 代。
4132		38	8371			血 ^会 丞	眾，異體字問 題。
4134	衛	115	8373	885E	衛		衛，異體字問 題。
4138	裘	3	8377	88A0	裘		
4139	袴	4	8378	88B4	袴		
4140	袵	2	8379	88B5	袵		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4142	裛	39	837B	8943	裛		
4145	羈	1	837E	898A	羈		Unicode 字型 呈現差異。
4146	覩	13	83A1	89A9	覩		睹，異體字問 題。
4150	訕	2	83A5	8A29	訕		
4154	譽	2	83A9	8AD0	譽		
4158	譌	2	83AD	8B4C	譌		
4160	適	1	83AF	8B81	適		
4167	賚	3	83B6	8CEB	賚		
4172	屨	1	83BB	8E77	屨		
4173	駝	8	83BC	8EAD	駝		
4176	輒	1	83BF	8F19	輒		
4181	辭	4	83C4	8FA5	辭		薛，異體字問 題。
4184	迹	18	83C7	8FF9	迹		
4190	邳	9	83CD	90C4	邳		
4195	鉢	2	83D2	9262	鉢		
4202	隲	1	83D9	9696	隲		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4203	隣	4	83DA	96A3	隣		鄰，異體字問題。
4209	鞞	4	83E0	9789	鞞		
4220	穎	1	83EB	9834	穎		
4221	頰	1	83EC	9833	頰		
4222	頰	1	83ED	9847	頰		
4224	頰	1	83EF	983F	頰		
4225	頰	1	83F0	9866	頰		
4232	隲	2	83F7	96B2	隲		
4237	髡	6	83FC	9AE0	髡		
4242	鬪	7	8442	9B2D	鬪		
4254	鷓	2	844E	9D7E	鷓		
4259	麪	3	8453	9E95	麪		
4262	麪	1	8456	9EAA	麪		
4275	义	1	8463	4E49	义		
4286	从	52	846E	4ECE	从		
4287	織	1	846F	7E8E	織		
4304	舩	2	84A2	26A35		舟 𠨉 行	
4309	刼	1	84A7	523C	刼		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4342	倮	1	84C8	502E	倮		
4359	棧	1	84D9	6827	棧		
4362	聊	1	84DC	35E6		口 厶 聊	
4369	晫	2	84E3	666B	晫		
4386	礩	1	84F4	7861	礩		
4388	縻	1	84F6	4328		纟 厶 縻	
4410	巖	1	854D	21F71		山 厶 巖	
4420	幟	1	8557	2213B		巾 厶 幟	
4450	勗	1	8575	524F	勗		
4469	勺	1	85AA	5304	勺		
4493	特	1	85C2	3E40		牛 厶 特	
4515	呪	3	85D8	546A	呪		
4545	踮	2	85F6	28105		足 厶 區	
4626	粦	1	86AA	427D		米 厶 半	
4671	鷓	1	86D7	4D0F		燕 厶 鳥	
4680	臙	1	86E0	2680A		月 厶 品 止	
4735	壠	2	8758	58E0	壠		
4768	妬	10	8779	59AC	妬		妒，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
4814	健	1	87C9	5FA4	健		
4822	尫	1	87D1	5C29	尫		
4836	嗜	1	87DF	20D32		口 厶 沓	
4862	柳	1	87F9	678A	柳		
4877	儼	1	8849	4A39		雨 厶 儼	
4881	羸	2	884D	283E9		羸 厶 連	
4904	尅	3	8864	5C05	尅		
4925	輔	1	8879	8F65	輔		
4934	髮	1	88A4	9B09	髮		
4950	嶮	1	88B4	5D18	嶮		
4990	巖	1	88DC	5DBB	巖		
4996	斲	1	88E2	65B5	斲		
5029	鬲	1	8944	9B32	鬲		
5034	弃	1	8949	5F03	弃		棄，異體字問題。
5057	掇	1	8960	39FB		扌 厶 豕	
5065	忼	3	8968	5FFC	忼		
5106	忼	1	89B3	5FF2	忼		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5111	旆	1	89B8	65BE	旆		
5116	佃	1	89BD	4F37	佃		
5126	暮	2	89C7	8B29	暮		
5133	憇	1	89CE	6199	憇		
5172	鐵	1	89F5	496B		金 𠂔 截	
5176		16	89F9			氵 𠂔 冗	沉，異體字問題。
5225	邨	1	8A6B	460F		血 𠂔 卩	
5251	ㄥ	3	8AA7	53B7	ㄥ		
5298	昨	1	8AD6	65FF	昨		
5302	昏	3	8ADA	662C	昏		
5315	公	1	8AE7	3CC2		氵 𠂔 公	
5318	陵	1	8AEA	4516		艹 𠂔 陵	
5338	暎	7	8AFE	668E	暎		
5355	𠂔	4	8B50	20AE4		𠂔	
5391	擡	1	8B74	648E	擡		
5398	滄	8	8B7B	6E7B	滄		
5406	尻	2	8BA5	51E5	尻		
5450	滄	1	8BD1	6E4C	滄		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5451	玳	1	8BD2	7398	玳		
5452	轎	1	8BD3	8F5C	轎		
5455	邾	1	8BD6	48DB		黍厶卩	
5463	岢	3	8BDE	65F9	岢		
5542	鑠	2	8C6E	93C1	鑠		
5549	藁	2	8C75	85F3	藁		
5583	殮	1	8CB9	98F1	殮		
5585	寓	2	8CBB	3762		宀会禹	
5588	恡	1	8CBE	6061	恡		
5597	樓	2	8CC7	28EFB		卩厶婁	
5653	暎	2	8D40	3B09		日厶奕	
5670	暈	1	8D51	66C5	暈		
5674	顛	1	8D55	294FF		須厶冉	
5703	𪗇	4	8D72	9F53	𪗇		
5705	廐	2	8D74	5ED0	廐		廐，異體字問題。
5706	敍	23	8D75	654D	敍		敍，異體字問題。
5711	駿	4	8D7A	2998E		馬厶斐	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5724	稽	1	8DA9	7A38	稽		
5742	澁	1	8DBB	6F81	澁		

二、手動取代表說明：

《九家舊晉書輯本》檔案中出現「●」字形，共計 23 字，比對原書後以 Unicode 字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《九家舊晉書輯本》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《九家舊晉書輯本》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 18	次●鼓中道	木厶罝	
p. 86	或受●於拙目	山厶欠	
p. 86	宣風聲於不●	泯	改為泯
p. 87	代大匠●者	斲	
p. 92	邇陌如矢·●	總	
p. 96	顧請旋於●汎	亻厶霍	
p. 97	●·太乙巖嶇	山厶薛	改為薛

頁碼	檔案原文摘錄	原書字形	備註
p. 98	●白刃以萬舞	搆	
p. 101	貫鰓●尾	𠂇△勺	
p. 102	●以鐵鑠機關	𠂇△勺	
p. 103	●以長壑	戾△彳	
p. 104	●梁為礪	𠂇△勺	
p. 149	孟●未達時	日△永	改為昶
p. 149	●于籬間窺之	日△永	改為昶
p. 234	東宮馬子莫聾●	啞	有 2 處
p. 291	稻不可以為●	𠂇 ⁺⁺ 𠂇川一韭 𠂇	
p. 357	●城登梯	扌△屬	
p. 388	孝宗穆皇帝諱●	目△丹	待造字
p. 451	趨●勇果	走△巢	
p. 477	須得●蛇膽為藥	蚘	
p. 532	琇上夫人●子帔	鼠△單	待造字
p. 541	高平李●女	日△桓	待造字