

《南史(本紀)》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
2008/9/10 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 874 的「𠂔」字，Unicode 編碼是 4350，由於 XP 並不支援，仍需使用構字式「𠂔𠂔干」。

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：

《南史(本紀)》使用舊版造字 139 個，字頻 741 次，這 139 個造字中，122 個可轉成 Windows XP 能顯示的字，字頻 713 次；另外 17 個字必須轉成構字式，字頻 28 次。

轉碼完成製作轉碼分析表，請參考附件一《南史(本紀)》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 1195 的「卓 𠃉 盃」字，而原書使用之字為「輻」，查詢教育部異體字字典確認「輻」並非「卓 𠃉 盃」之異體字，因此歸為錯字，以程式取代為正確字形「輻」。
 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 4262 的「麪」字，Unicode 字型呈現為「麪」，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 4375 的「月 𠃉 郤」。

附件一、《南史(本紀)》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
792	帛	1	8E46	5E2C	帛		
874	𠃉	2	8EBA	4350		𠃉 𠃉 千	
938	燮	1	8EFA	7215	燮		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
981	豺	1	8F66	72B2	豺		
1044	儻	8	8FC7	511B	儻		
1048	額	2	8FCB	981F	額		
1056	陔	4	8FD3	9641	陔		
1067	綦	5	8FDE	68CA	綦		
1091	漣	2	8FF6	58CD	漣		
1097	嶠	1	8FFC	37ED		山 厶 昏	
1111	瓚	1	904B	749D	瓚		
1133	瓌	2	9061	74CC	瓌		
1195		1	90C1			車 厶 昱	輻，錯字，以程式全部取代。
1237	稍	2	90EB	77DF	稍		
1278	碍	10	9155	788D	碍		
1885	蠶	5	9540	882D	蠶		
1926	覓	1	9569	8994	覓		
1944	簪	9	957B	8A67	簪		
2191	鉞	1	96F7			金 厶 戍	鉞，異體字問題。
2199	鈺	1	9740	925F	鈺		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2207	峯	3	9748	25B57		竹 厶 峯	
2208		1	9749			形 廌 旦 鼻 鼻	廌 鼻，錯字，以程式全部取代。
2209	餽	1	974A	4B33		食 厶 半	
2210	媿	1	974B	5A6B	媿		
2305	鎖	1	97CC	93BB	鎖		
2478	鞏	1	98DC	978C	鞏		
2502	颺	1	98F4	98C8	颺		
3778	併	2	8149	5002	併		併，異體字問題。
3780	儁	6	814B	5101	儁		
3787	豕	10	8152	51A2	豕		
3789	冲	2	8154	51B2	冲		冲，異體字問題。
3801	剗	1	8160	5257	剗		
3804	効	4	8163	52B9	効		
3806	勳	3	8165	52E0	勳		
3814	却	2	816D	5374	却		
3815	登	1	816E	537A	登		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3817	廁	1	8170	53A0	廁		廁，異體字問題。
3835	噍	2	81A4	564D	噍		
3840	坂	2	81A9	5742	坂		
3846	壻	1	81AF	58FB	壻		
3849	獎	4	81B2	596C	獎		獎，異體字問題。
3858	冤	7	81BB	5BC3	冤		冤，異體字問題。
3866	崐	1	81C3	5D10	崐		
3869	币	1	81C6	5E00	币		
3874		3	81CB			忄 厶 稟	慄，異體字問題。
3877	迺	2	81CE	5EFC	迺		
3878	讐	2	81CF	8B7C	讐		
3879	讐	2	81D0	27AF7		敕 讐言	敕 讐言，錯字，以程式全部取代。
3883	徧	7	81D4	5FA7	徧		
3884	忽	4	81D5	6031	忽		
3889	慄	1	81DA	6164	慄		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3915	斃	1	8459	6585	斃		
3919	𧈧	1	81F8	23C0B		𧈧△毛	𧈧△毛，異體字問題。
3924	菴	2	81FD	671E	菴		
3932	椶	1	8246	6936	椶		
3937	檝	1	824B	6A9D	檝		
3947	舩	1	8255	4461		舟△又	
3952	舩	1	825A	4470		舟△步	
3974	牀	19	8270	7240	牀		
3978	猪	1	8274	732A	猪		Unicode 字型呈現差異。
3981	珉	2	8277	73C9	珉		
3983	琕	1	8279	7415	琕		
3984	瑠	1	827A	7460	瑠		
3985	璫	1	827B	7447	璫		
3996	疎	3	82A8	758E	疎		
4010	皐	1	82B6	7690	皐		皐，異體字問題。
4011	皐	1	82B7	76A1	皐		皐，異體字問題。
4028	稟	3	82C8	7980	稟		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4035	窳	1	82CF	7A7D	窳		
4038	竈	1	82D2	7AC8	竈		
4052	簾	1	82E0	7C34	簾		竹 厶 虞，異體字問題。
4068	綫	1	82F0	7DAB	綫		
4078		1	82FA			巳 厶 升	异，異體字問題。
4080	羣	84	82FC	7FA3	羣		
4085	耻	2	8342	803B	耻		
4090	脚	5	8347	811A	脚		
4091	臯	5	8348	81EF	臯		
4107	葱	1	8358	8471	葱		
4119		1	8364			形 艹 正 頁 巳 久 口	夔，錯字，以程式全部取代。
4131	劦	1	8370	8842	劦		
4132		173	8371			血 厶 丞	眾，異體字問題。
4134	衛	93	8373	885E	衛		衛，異體字問題。
4139	袴	5	8378	88B4	袴		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4142	裛	10	837B	8943	裛		
4146	覩	4	83A1	89A9	覩		睹，異體字問題。
4155	訶	1	83AA	8B0C	訶		
4160	謫	2	83AF	8B81	謫		
4184	迹	7	83C7	8FF9	迹		
4193	鈎	1	83D0	920E	鈎		鈎，異體字問題。
4203	隣	4	83DA	96A3	隣		鄰，異體字問題。
4242	鬪	12	8442	9B2D	鬪		
4252	鰓	12	844C			魚厶眾	鰓，異體字問題。
4254	鷓	1	844E	9D7E	鷓		
4260	麇	1	8454	9E9E	麇		
4262	麪	1	8456	9EAA	麪		Unicode 字型呈現差異。
4278	亘	2	8466	4E98	亘		
4303	謗	2	84A1	470E		言厶勞	
4342	倮	1	84C8	502E	倮		
4345		2	84CB			口厶敕	待造字。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4375		1	84E9			月△卻	待造字。
4426	嶠	1	855D	21E6E		山△弁	
4448		2	8573			形糸厂 墨□	纏，錯字，以程式全部取代。
4461	勛	9	85A2	52D4	勛		
4522	瞞	6	85DF	3B13		日△票	
4535	纘	3	85EC	7E89	纘		
4684		1	86E4			衤△受	待造字。
4735	壠	2	8758	58E0	壠		
4768	妬	1	8779	59AC	妬		妒，異體字問題。
4774	鯪	1	87A1	9BA7	鯪		
4782	鞞	1	87A9	97BA	鞞		
4996	斲	1	88E2	65B5	斲		
5111	旆	2	89B8	65BE	旆		
5116	倝	1	89BD	4F37	倝		
5121		2	89C2			形忄主 方攴□	傲，錯字，以程式全部取代。
5126	暮	1	89C7	8B29	暮		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5176		3	89F9			氵 厶 元	沉，異體字問題。
5194		1	8A4C			扌 畝 且 口	擅，異體字問題。
5209	摑	3	8A5B	6386	摑		
5230		4	8A70			扌 土 畝 且 口	壇，異體字問題。
5298	昨	3	8AD6	65FF	昨		
5316	晉	1	8AE8	664B	晉		晉，異體字問題。
5319	晰	1	8AEB	6663	晰		
5396	鐸	2	8B79	938B	鐸		
5422	柿	1	8BB5	67F9	柿		
5434	湟	5	8BC1	6E7C	湟		
5437	并	41	8BC4	5E77	并		
5439	栢	3	8BC6	6822	栢		
5466	倏	1	8BE1	5010	倏		
5559	欸	2	8CA1	6B35	欸		
5666	污	3	8D4D	6C5A	污		
5697	煑	1	8D6C	7151	煑		
5704	瓊	2	8D73	6C0A	瓊		瓊，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
5705	廐	1	8D74	5ED0	廐		
5706	敍	3	8D75	654D	敍		敍，異體字問題。
5707	棧	2	8D76	6830	棧		
5709	蟀	1	8D78	872F	蟀		
5719	麇	1	8DA4	4D25		鹿△加	
5764	籊	1	8DD1	7C5E	籊		
5853		1	C6CC				《》，錯字，以程式全部取代。