

《北齊書》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
2008/11/06 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 959 的「掄」字，Unicode 編碼是 3A49，由於 XP 並不支援，仍需使用構字式「扌𠂔翁」。

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：

《北齊書》使用舊版造字 165 個，字頻 1453 次，這 165 個造字中，151 個可轉成 Windows XP 能顯示的字，字頻 1371 次；另外 14 個字必須轉成構字式，字頻 82 次。

轉碼完成製作轉碼分析表，請參考附件一《北齊書》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 4901「冗𠄎且」字，而原書使用之字為「置」，查詢教育部異體字字典確認「冗𠄎且」並非「置」之異體字，因此歸為錯字，以程式取代為正確字形「置」。
 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 3985 的「璊」字，Unicode 字型呈現為「璊」，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 1036 的「王𠄎爰」。

附件一、《北齊書》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
799	𪗇	1	8E4D	9F5A	𪗇		
800	逕	1	8E4E	905D	逕		
821	脊	1	8E63	7718	脊		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
959	掄	1	8F50	3A49		扌厶翁	
972	犇	6	8F5D	7287	犇		
1000		1	8F79			王厶巳	玨，異體字問題。
1036		3	8FBF			王厶爰	待造字。
1044	儻	1	8FC7	511B	儻		
1056	陔	1	8FD3	9641	陔		
1067	碁	1	8FDE	68CA	碁		
1091	漚	2	8FF6	58CD	漚		
1133	瓌	15	9061	74CC	瓌		
1237	稍	14	90EB	77DF	稍		
1544	絙	1	92E4	7D59	絙		
1642	脉	3	93A9	8109	脉		
1755	蔘	3	945B		蔘		
1784	萼	1	9478	855A	萼		
1893	袞	2	9548	886E	袞		
1925	羈	1	9568	8989	羈		Unicode 字型

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							呈現差異。
1944	簪	4	957B	8A67	簪		
1981	簪	16	95C2	8B90	簪		讎，異體字問題。
2305	鎖	1	97CC	93BB	鎖		
2502	颯	2	98F4	98C8	颯		
3778	併	1	8149	5002	併		
3779	霰	1	814A	290F1		雨 ㇀ 沱	
3780	儁	64	814B	5101	儁		
3783	律	1	814E	4F93	律		
3785	𠂔	1	8150	355E		𠂔	
3786	証	1	8151	27976		言 ㇀ 王	
3787	冢	18	8152	51A2	冢		
3788	慤	1	8153	22871		殼 ㇀ 心	慤，異體字問題。
3789	冲	8	8154	51B2	冲		冲，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3791	况	2	8156	51B5	况		况，異體字問題。
3792	凉	3	8157	51C9	凉		凉，異體字問題。
3796	决	1	815B	51B3	决		决，異體字問題。
3800		1	815F			杀 厶 丿	刹，異體字問題。
3803	餽	2	8162	4B54		食 厶 追	
3804	効	22	8163	52B9	効		
3805	勅	8	8164	52C5	勅		
3806	勳	2	8165	52E0	勳		
3807	驗	2	8166	9A10	驗		
3814	却	16	816D	5374	却		
3817	廁	3	8170	53A0	廁		廁，異體字問題。
3819	廝	7	8172	53AE	廝		廝，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3820	厦	1	8173	53A6	厦		
3822	叁	2	8175	53C1	叁		參，異體字問題。
3826	叶	5	8179	53F6	叶		
3829	咏	1	817C	548F	咏		
3830	咤	1	817D	54A4	咤		
3840	坂	3	81A9	5742	坂		
3846	壻	6	81AF	58FB	壻		
3849	獎	20	81B2	596C	獎		獎，異體字問題。
3858	冤	10	81BB	5BC3	冤		冤，異體字問題。
3864	峯	3	81C1	5CEF	峯		
3866	崐	1	81C3	5D10	崐		
3868	卮	3	81C5	5DF5	卮		
3874		2	81CB			忄 ㄇ 稟	慄，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3876	廻	9	81CD	5EFB	廻		
3883	徧	2	81D4	5FA7	徧		
3884	忽	6	81D5	6031	忽		
3886	悞	1	81D7	609E	悞		
3889	慤	9	81DA	6164	慤		
3915	斆	1	8459	6585	斆		
3931	椀	1	8245	6900	椀		
3944	氷	1	8252	6C37	氷		
3948	沝	1	8256	6CAD	沝		
3974	牀	21	8270	7240	牀		
3977	狗	2	8273	72E5	狗		
3978	猪	4	8274	732A	猪		
3981	珉	3	8277	73C9	珉		
3985	璊	1	827B	7447	璊		Unicode 字型 呈現差異。
3990	甄	1	82A2	750E	甄		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3991	攜	3	82A3	64D5	攜		
3996	疎	2	82A8	758E	疎		
4015	眦	1	82BB	7726	眦		
4016	着	8	82BC	7740	着		
4021	礪	1	82C1	78AF	礪		
4028	稟	13	82C8	7980	稟		
4035	窳	1	82CF	7A7D	窳		
4039	竚	1	82D3	7ADA	竚		
4041	豎	3	82D5	7AEA	豎		
4044	筋	2	82D8	7B6F	筋		
4055	粃	1	82E3	7C83	粃		
4080	羣	106	82FC	7FA3	羣		
4085	耻	2	8342	803B	耻		
4090	脚	8	8347	811A	脚		
4091	臯	2	8348	81EF	臯		
4094	舘	1	834B	8218	舘		
4103	苙	2	8354	831D	苙		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4104	莅	6	8355	8385	莅		
4109	蒟	1	835A	84AD	蒟		
4114	藜	1	835F	8616	藜		
4115	藁	2	8360	85C1	藁		
4129	虬	2	836E	866C	虬		
4131	劔	2	8370	8842	劔		
4132		240	8371			血𠄎𠄎	眾，異體字問題。
4134	衛	135	8373	885E	衛		衛，異體字問題。
4137	袜	1	8376	889C	袜		
4139	袴	5	8378	88B4	袴		
4140	衽	1	8379	88B5	衽		
4145	羈	3	837E	898A	羈		Unicode 字型呈現差異。
4146	覩	4	83A1	89A9	覩		睹，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4171		1	83BA			𠂔足厂 封	蹶，異體字問題。
4172	蹶	1	83BB	8E77	蹶		
4173	躄	1	83BC	8EAD	躄		
4176	輒	2	83BF	8F19	輒		
4184	迹	22	83C7	8FF9	迹		
4185	邊	2	83C8	9037	邊		
4193	鈎	2	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	9	83D2	9262	鉢		
4197	鍔	1	83D4	936B	鍔		
4203	隣	14	83DA	96A3	隣		鄰，異體字問題。
4207	霧	1	83DE	9736	霧		
4220	穎	1	83EB	9834	穎		
4222	頰	1	83ED	9847	頰		
4226	凜	1	83F1	98E1	凜		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4234	羸	1	83F9	9A58	羸		
4237	髡	2	83FC	9AE0	髡		
4242	鬪	16	8442	9B2D	鬪		
4252	鰓	2	844C			魚厶眾	鰓，異體字問題。
4256	鷄	4	8450	9DC4	鷄		雞，異體字問題。
4267	鼈	1	845B	9F08	鼈		
4278	亘	1	8466	4E98	亘		
4432		7	8563			形糸厂 墨口	纏，異體字問題。
4455	劔	1	857A			僉厶刀	
4460	勅	134	85A1	52D1	勅		
4515	呪	1	85D8	546A	呪		
4528	嶼	1	85E5	5D60	嶼		
4558	厩	1	8644	53A9	厩		廐，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4603	憇	1	8671	6187	憇		
4706	澄	1	86FA	96A5	澄		
4707	悛	56	86FB	3944		忄 厶 夂	
4708	擲	5	86FC	22DCB		標 厶 寸	
4716	塗	1	8745	57FF	塗		
4768	妬	10	8779	59AC	妬		妒，異體字問題。
4854	凝	7	87F1			冫 厶 疑	
4870	壘	1	8842	97F2	壘		
4876	羴	1	8848	437D		羊 厶 歷	
4901		1	8861			冗 厶 且	置，錯字，以程式全部取代。
4961	盼	1	88BF	80B9	盼		
5074	𠂔	2	8971	5C2B	𠂔		
5111	旆	1	89B8	65BE	旆		
5133	憇	1	89CE	6199	憇		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5225	邨	1	8A6B	460F		血厶卩	
5290	携	1	8ACE	643A	携		
5300	鬪	1	8AD8	9B2A	鬪		
5360	鄺	1	8B55	913D	鄺		
5437	并	138	8BC4	5E77	并		
5439	栢	10	8BC6	6822	栢		
5466	倏	1	8BE1	5010	倏		
5514	塾	1	8C52	25095		執厶皿	
5541	蠹	1	8C6D	8827	蠹		蠹，異體字問題。
5556	隸	2	8C7C	96B7	隸		
5559	欸	20	8CA1	6B35	欸		
5587	歛	1	8CBD	6B58	歛		
5591	杞	1	8CC1	233CC		木厶巳	杞，異體字問題。
5666	汚	2	8D4D	6C5A	汚		
5697	煑	2	8D6C	7151	煑		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5702	厨	2	8D71	53A8	厨		
5704	𩇛	3	8D73	6C0A	𩇛		𩇛，異體字問題。
5705	廡	4	8D74	5ED0	廡		廡，異體字問題。
5706	敍	8	8D75	654D	敍		敍，異體字問題。
5742	澁	2	8DBB	6F81	澁		
5753	潜	2	8DC6	6F5C	潜		
5799	蜀	1	8DF4	262FA		𠄎 𠄎 𠄎 𠄎	