

《史記》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
2010/05/19 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 786 的「蕩」字，Unicode 編碼是 26FC6，由於 XP 並不支援，仍需使用構字式「𠄎𠄎𠄎𠄎」。

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：

《史記》使用舊版造字 310 個，字頻 2780 次，這 310 個造字中，222 個可轉成 Windows XP 能顯示的字，字頻 2644 次；另外 88 個字必須轉成構字式，字頻 136 次。

轉碼完成製作轉碼分析表，請參考附件一《史記》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. **部份使用不同字形，手動修改**：檔案中為同一舊版造字，校對時卻發現原書使用兩個以上不同的字形，可能部份舊版造字為錯字或原書使用兩種異體字形，導致同一舊版造字轉碼時對應到兩個以上的字形；將一部份手動修改，剩餘的頻次則用程式取代成對應的 Unicode 或構字式。例：造字編號 5660 的「夔」字頻次為 5，依書將 1 字手動修改為「夔」，剩餘 4 字轉為「夔」。手動修改部份請參考附件二《史記》手動取代表。
3. **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 912 的「鷓」字，而原書使用之字為「疒△徙」，「鷓」為錯字，以程式取代為正確字形「疒△徙」。
4. **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 3935 的「槩」字，Unicode 字型呈現為「槩」，實際上仍為同一字。
5. **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 879 的「足△籍」。

附件一、《史記》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
786	礪	1	8E40	26FC6		礪	
787	臍	3	8E41	26848		臍	
790	映	1	8E44	4039		目	
792	帟	1	8E46	5E2C	帟		
793	鞮	4	8E47	97E4	鞮		
794	隋	1	8E48	9657	隋		
800	逡	1	8E4E	905D	逡		
809	擱	1	8E57	650C	擱		
810	薊	1	8E58	846A	薊		Unicode 字型 呈現差異。
811	憊	1	8E59	6178	憊		
821	脊	1	8E63	7718	脊		
825	坳	1	8E67	573D	坳		
826	癱	1	8E68	7655	癱		
828	跣	1	8E6A	8DD4	跣		
831	叻	1	8E6D	3579		叻	
835	邾	1	8E71	5380	邾		
845	鮑	1	8E7B	9F26	鮑		
847	鈺	6	8E7D	7F3B	鈺		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
848	髻	1	8E7E	9AE4	髻		
852	貉	1	8EA4	72E2	貉		
855	嬖	1	8EA7			辟△女	
856	籀	1	8EA8	25D61		𠂇△構	
857	臄	1	8EA9	26715		𠂇月丞 巳□	
858	采	3	8EAA	7F59	采		
861	藁	1	8EAD	4183		道△禾	
862	矐	1	8EAE	76AC	矐		
864	礧	1	8EB0	40F6		石△裏	
865	蚰	1	8EB1	27358		虫△幽	
866	嗅	1	8EB2	247ED		𪚩△犬	
867	磬	1	8EB3	27BC6		谷△害	
868	侏	1	8EB4	201F1		亻△未	亻△未，異體字問題。
870	躡	1	8EB6	2A60F		足△齒	
871	陬	1	8EB7	49E2		阝△區	
874	罕	1	8EBA	4350		𠂇△干	
876		1	8EBC			彳△屑	彳△屑，異體

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							字問題。
877	姍	1	8EBD	36A9		女厶丹	
878	磬	1	8EBE	4D80		鼓厶合	
879		1	8EBF			足厶籍	待造字。
880	雉	1	8EC0	49F4		支厶隹	
881	颺	1	8EC1	98C7	颺		
883	陵	2	8EC3	9656	陵		
884		1	8EC4			冗厶冂	冂，異體字問題。
889	贍	1	8EC9	476F		貝厶屯	
891	標	1	8ECB			形木 𠂇 米 𠂇	
892	榭	1	8ECC	6A98	榭		
894	碟	1	8ECE	790F	碟		
896	櫛	1	8ED0	6AE9	櫛		
897	觶	1	8ED1	27917		角厶嵒	
898	牖	1	8ED2	246D1		牛厶庸	
899	馳	2	8ED3	99DE	馳		
901	崙	1	8ED5	5D2B	崙		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
902	嶠	1	8ED6	21F8A		山厶畠	
905	眊	1	8ED9	28E0F		眊	
906	睽	1	8EDA	4046		目厶奄	
907	畠	1	8EDB	21F8B		山厶畠	
909	崑	1	8EDD	5D54	崑		
912	鷓	1	8EE0	9DC0	鷓		疒厶徙，錯字，以程式全部取代。
916	駟	1	8EE4	2A001		鳥厶加	
917	駝	1	8EE5	4CD8		鳥厶我	
918		1	8EE6			玉厶鳥	瑪，異體字問題。
919	駝	1	8EE7	2A049		鳥厶交	
920	鯖	1	8EE8	2A0B4		鳥厶青	
921	磳	1	8EE9	789D	磳		
922	鰓	1	8EEA	4D0B		鳥厶畧	
924	鰓	1	8EEC	2A1AB		鳥厶渠	
925	鷓	1	8EED	2A1C5		鳥厶葳	
926	鷓	1	8EEE	4CC4		此厶鳥	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
927	浩	1	8EEF	3D46		汨厶告	
928	鮪	1	8EF0			魚厶互	
929	鱸	1	8EF1	4C9B		魚厶普	
930	鱸	1	8EF2	9C45	鱸		
931	磕	3	8EF3	791A	磕		
933	湊	1	8EF5	3D14		辶厶急	
936	鷓	3	8EF8	4D0A		義厶鳥	
949	枊	2	8F46	678F	枊		
951	斲	1	8F48	452E		ㄗ厶斯	
952	瑛	1	8F49	744C	瑛		
953	漫	3	8F4A	23E70		辶厶曼	
954	轉	1	8F4B	283A5		車厶惠	
957	遯	2	8F4E	906C	遯		
958	疆	1	8F4F	261E4		纟厶疆	
963	鮰	1	8F54	4C78		魚厶叟	
967	僂	1	8F58	50AB	僂		
970	淥	1	8F5B	6E6A	淥		
972	犇	55	8F5D	7287	犇		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
973	犮	2	8F5E	246DF		牛 ㄩ	
976	毳	2	8F61	26420		羽 ㄩ	羽 ㄩ 氏，錯字，以程式全部取代。
978	戡	2	8F63	4CD2		戈 ㄩ	
979	羴	2	8F64	56CF	羴		
980	犖	2	8F65	72AB	犖		
985	𠂔	1	8F6A	6197	𠂔		
987	褊	2	8F6C	890D	褊		
988	𠂔	5	8F6D	7F8B	𠂔		
989	𠂔	1	8F6E	89B5	𠂔		
991	阝	1	8F70	28E12		阝 ㄩ	
993	虍	2	8F72	2223B		虍	
994	楚	3	8F73	8E75	楚		
995	猯	1	8F74	734F	猯		
996	衝	6	8F75	27602		行 ㄩ	
998	潰	3	8F77	3D52		氵 ㄩ	
999	𠂔	2	8F78	5314	𠂔		
1004	徐	3	8F7D	4FC6	徐		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1012	豸	1	8FA7	4760		豸厶原	
1015	巛	3	8FAA	5D4F	巛		
1017	肩	2	8FAC	3795		肩	
1018	暻	1	8FAD	3B08		日厶暻	
1021	嶺	1	8FB0	21E72		山厶胥	
1023	調	1	8FB2	8AB7	調		
1024	虵	2	8FB3	8675	虵		
1031	玊	1	8FBA	738F	玊		
1032	憊	1	8FBB	2039D		形彳扶 心	
1039	稟	3	8FC2	69C0	稟		
1040	聃	1	8FC3	803C	聃		
1043	頤	1	8FC6	2945E		丹厶頁	
1044	儻	1	8FC7	511B	儻		
1046	厃	3	8FC9	21C25		厃	
1048	頤	3	8FCB	981F	頤		
1050	澆	1	8FCD	23DFD		氵厶侵	
1053	傲	2	8FD0	6160	傲		
1054	那	5	8FD1	2867B		丹厶卩	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1056	阨	3	8FD3	9641	阨		
1057	夔	1	8FD4	3687	夔		
1059	𦉳	4	8FD6	51A3	𦉳		Unicode 字型呈現差異。
1060	𦉴	1	8FD7	24CF9		𦉴△一	
1061	𦉵	1	8FD8	4FC8	𦉵		
1062		2	8FD9			馬△𦉶	駟，異體字問題。
1066	頽	9	8FDD	4AC7		須△𦉷	
1067	碁	5	8FDE	68CA	碁		
1068	雷	1	8FDF	4A13		雨△回	
1072	銍	1	8FE3	4B39		倉△至	
1073		1	8FE4			算△么	篡，異體字問題。
1074	璿	2	8FE5	7449	璿		
1077	稟	6	8FE8	7A3E	稟		
1079	總	1	8FEA	7DEB	總		
1080	𦉷	62	8FEB	5184	𦉷		
1081	鶴	2	8FEC	9E16	鶴		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1082	樟	4	8FED	3BC9		木ㄩ樟	
1083	鷓	1	8FEE	9DAE	鷓		
1085		1	8FF0			崙ㄩ友	崙ㄩ友，異體字問題。
1087	黼	1	8FF2	264DE		崙ㄩ甫	
1088	暘	2	8FF3	757C	暘		
1091	漚	5	8FF6	58CD	漚		
1098	樺	2	8FFD	6A8B	樺		
1099	睺	4	8FFE	66A4	睺		
1232	曖	1	90E6	77B9	曖		
1381	穉	3	91DE	7A3A	穉		
1383	穰	10	91E0	7A45	穰		
1458	籊	1	926C	7BA3	籊		
1565	繇	8	92F9	7DDC	繇		
1773	簡	1	946D	8504	簡		𠄎ㄩ閭，錯字，以程式全部取代。
1885	蠶	6	9540	882D	蠶		
1912	襍	2	955B	894D	襍		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1937	詢	3	9574	8A3D	詢		
1944	譽	1	957B	8A67	譽		
1982	調	4	95C3	8B87	調		
2050	蹠	2	9648	8E4F	蹠		
2053	躡	4	964B	8EA7	躡		
2077	鞞	2	9663	8F4A	鞞		
2168	壽	1	96E0	91BB	壽		
2191	鉞	1	96F7			金Ⓐ戊	鉞，異體字問題。
2449	雒	1	98BF			名Ⓐ隹	雒，錯字，以程式全部取代。
2472	霤	11	98D6	9741	霤		
2478	鞏	6	98DC	978C	鞏		
2609	鯁	1	99C2	9C4F	鯁		Unicode 字型呈現差異。
2610	鯁	2	99C3	9C53	鯁		
2638	鷓	2	99DF	9D55	鷓		
2651	鸚	3	99EC	9E1C	鸚		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3770	个	1	8141	4E2A	个		
3787	冢	49	8152	51A2	冢		
3801	剗	1	8160	5257	剗		
3804	効	1	8163	52B9	効		
3806	勳	4	8165	52E0	勳		
3814	却	1	816D	5374	却		
3819	廡	1	8172	53AE	廡		廡，異體字問題。
3826	叶	1	8179	53F6	叶		
3830	咤	1	817D	54A4	咤		
3833	嘑	2	81A2	5611	嘑		
3835	噍	3	81A4	564D	噍		
3836	喻	2	81A5	564F	喻		
3840	坂	2	81A9	5742	坂		
3846	壻	8	81AF	58FB	壻		
3855	孽	14	81B8	5B7C	孽		
3868	卮	3	81C5	5DF5	卮		
3869	币	2	81C6	5E00	币		
3873		1	81CA			广△殷	廡，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3877	迺	2	81CE	5EFC	迺		
3882	彫	1	81D3	5F72	彫		
3883	徧	41	81D4	5FA7	徧		
3884	忽	2	81D5	6031	忽		
3887	息	2	81D8	60A4	息		Unicode 字型 呈現差異。
3888	惠	5	81D9	60B3	惠		
3889	慤	1	81DA	6164	慤		
3891	慙	28	81DC	6159	慙		
3910	擘	2	81EF	64E5	擘		
3915	斃	1	8459	6585	斃		
3920	皓	7	81F9	6667	皓		
3921	皙	6	81FA	6673	皙		
3924	朞	6	81FD	671E	朞		
3929	黎	1	8243	68C3	黎		
3930	挀	10	8244	6901	挀		
3935	槩	8	8249	3BA3	槩		Unicode 字型 呈現差異。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3955	澗	2	825D	3D4E		氵 ㄩ 間	
3965	裁	7	8267	70D6	裁		
3974	牀	13	8270	7240	牀		
3976	犁	26	8272	7282	犁		
3980	獾	3	8276	7328	獾		
3985	璫	5	827B	7447	璫		Unicode 字型 呈現差異。
3996	疎	1	82A8	758E	疎		
4015	眦	5	82BB	7726	眦		
4018	矚	3	82BE	77AF	矚		
4029	褱	1	82C9	799D	褱		
4033	稽	1	82CD	7A2D	稽		
4038	竈	20	82D2	7AC8	竈		
4043	筭	3	82D7	7B53	筭		
4048		1	82DC			形 竹 广 虎 口	麓，錯字，以 程式全部取 代。
4066	綺	2	82EE	7D5D	綺		
4070	縹	12	82F2	7DE4	縹		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4071	纓	1	82F3	7DF5	纓		
4072	縹	3	82F4	7DE5	縹		
4080	羣	284	82FC	7FA3	羣		
4081	翹	4	82FD	7FFA	翹		Unicode 字型 呈現差異。
4090	脚	1	8347	811A	脚		
4097	芑	1	834E			艹 厶 巳	
4103	苙	5	8354	831D	苙		
4104	苙	4	8355	8385	苙		
4115	藁	1	8360	85C1	藁		
4124	蟲	1	8369	8771	蟲		
4132		388	8371			血 厶 𠂇	眾，異體字問題。
4134	衛	537	8373	885E	衛		衛，異體字問題。
4139	袴	4	8378	88B4	袴		
4140	裱	1	8379	88B5	裱		
4142	裊	48	837B	8943	裊		
4144	霸	1	837D	8987	霸		Unicode 字型

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							呈現差異。
4146	覩	11	83A1	89A9	覩		
4155	訶	1	83AA	8B0C	訶		
4158	譌	2	83AD	8B4C	譌		Unicode 字型 呈現差異。
4172	屨	7	83BB	8E77	屨		
4174	躑	2	83BD	8EB6	躑		
4180	臯	15	83C3	8FA0	臯		
4184	迹	22	83C7	8FF9	迹		
4185	邊	1	83C8	9037	邊		
4190	邳	8	83CD	90C4	邳		
4201	陘	1	83D8	967B	陘		
4209	鞞	1	83E0	9789	鞞		
4234	羸	2	83F9	9A58	羸		
4242	鬪	65	8442	9B2D	鬪		
4250	鱒	3	844A	9C0C	鱒		
4252	鰈	4	844C			魚𩺰眾	鰈，異體字問題。
4266	鼉	38	845A	9F02	鼉		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4267	鼈	6	845B	9F08	鼈		
4342	倮	4	84C8	502E	倮		
4469	勻	7	85AA	5304	勻		
4476		210	85B1			良 Δ P	即，異體字問題。
4482	廛	2	85B7	53AA	廛		
4533	唵	1	85EA	552B	唵		
4576	誣	2	8656	567D	誣		
4629	坳	1	86AD	577F	坳		
4723	塶	1	874C	58C3	塶		
4735	壠	1	8758	58E0	壠		
4768	妬	2	8779	59AC	妬		
4801	斌	1	87BC	5A2C	斌		
4990	截	1	88DC	5DBB	截		
5034	弃	196	8949	5F03	弃		Unicode 字型呈現差異。
5065	忼	5	8968	5FFC	忼		
5133	憇	2	89CE	6199	憇		
5169	戇	12	89F2	6239	戇		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5211	駢	3	8A5D	99F5	駢		
5212	𠂔	2	8A5E	27BAC		谷厶凵	
5213		1	8A5F			𠂔女𠂔 且◻	嬪，異體字問題。
5214		1	8A60			𠂔魚𠂔 且◻	鱣，異體字問題。
5215	𦍋	1	8A61			龜厶免	𦍋，異體字問題。
5216	鞞	1	8A62	97CA	鞞		
5218		3	8A64			𠂔宀覓 、◻	寬，異體字問題。
5219	妃	4	8A65			女厶巳	
5220	𠂔	4	8A66	6BCC	𠂔		
5221		1	8A67			𠂔白米 𠂔◻	粵，異體字問題。
5222	𠂔	1	8A68	4EB6	𠂔		
5230		2	8A70			𠂔土𠂔 且◻	壇，異體字問題。
5245	擻	1	8AA1			掉厶戈	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5282	敷	3	8AC6	656B	敷		
5302	昏	3	8ADA	662C	昏		
5319	晰	1	8AEB	6663	晰		
5374	盼	1	8B63	670C	盼		
5385	漚	2	8B6E	3D4C		氵厶循	
5388	黜	2	8B71	4D63		黑厶旦	
5389	竿	1	8B72	7B07	竿		
5428	礪	1	8BBB	27BCB		礪谷門 牙	
5430	塹	1	8BBD	2141B		土厶斬	
5438	桀	2	8BC5	681E	桀		
5439	栢	1	8BC6	6822	栢		
5450	滄	2	8BD1	6E4C	滄		
5451	玳	1	8BD2	7398	玳		
5454	榭	1	8BD5	68B8	榭		
5455	邴	2	8BD6	48DB		黍厶卩	
5564	槩	1	8CA6	6ABF	槩		
5583	殮	3	8CB9	98F1	殮		
5591	杞	17	8CC1	233CC		木厶巳	杞，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
5630	愚	1	8CE8	22838		單△心	
5660	燙	5	8D47	3DED		逢△火	部份使用不同字形「燙」，手動修改。
5666	污	2	8D4D	6C5A	污		污，異體字問題。
5691	鴛	1	8D66	6A22	鴛		异，錯字，以程式全部取代。
5697	糞	2	8D6C	7151	糞		
5705	廐	1	8D74	5ED0	廐		Unicode 字型呈現差異。
5722	朥	1	8DA7	671E	朥		走△畢，錯字，以程式全部取代。
5723	样	2	8DA8	4367		牛△羊	
5724	稽	1	8DA9	7A38	稽		
5730	阨	13	8DAF	9638	阨		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5766	灩	1	8DD3			灩 灩	灩，異體字問題。
5808	賀	1	8DFD	27D4D		𠄎	
5860	无	1	C6D3	65E0	无		

二、手動取代表說明：

《史記》檔案中之舊版造字，因字形部份不同而無法全部以程式取代，皆手動修改成Windows XP能顯示的字或構字式，並製作手動取代表，請參考附件二。欄位說明如下：

1. 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
2. 造字：舊版造字字形。
3. 取代：實際使用字形與舊版造字不同時，於此欄列出。
4. 檔案原文摘錄：摘錄檔案《史記》中的文句段落，其中紅字底線的部分，為需要手動取代的原文。
5. 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。
6. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《史記》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果	備註
5660	𡗗	𡗗	<u>𡗗</u> 大鼓，有寇	<u>遂</u> 大鼓，有寇	