

「𠂇」はなぜ JIS X 0221 に含まれているのか —— Unicode 幽霊字研究 ——

安岡孝一

京都大学大型計算機センター

yasuoka@kudpc.kyoto-u.ac.jp

安岡素子

フリー

JIS X 0221 の漢字は、中国・台湾・日本・韓国の漢字コード規格から集められたものである。このうち、中国の GB 8565 からの採録とされている漢字は 290 字ある。しかし調査の結果、これら 290 字のうち 87 字は、実際には GB 8565 には含まれていないことがわかった。本稿では、これらの漢字が一体どのような経緯で JIS X 0221 に採録されることになったのかを、さまざまな漢字表を歴史順に追うことによって明らかにする。

Why is “𠂇” included in JIS X 0221?

Koichi Yasuoka

Kyoto University Data Processing Center

yasuoka@kudpc.kyoto-u.ac.jp

Motoko Yasuoka

Free Writer

CJK characters of JIS X 0221 are gathered from domestic standards of China, Taiwan, Japan, and Korea. And JIS X 0221 insists that 290 characters are from GB 8565 of China. However, in fact, 87 characters of the 290 are not included in GB 8565. In this paper, we reveal why and how these 87 characters are stuffed into JIS X 0221.

1 はじめに

JIS X 0221 の漢字部分 (4E00 ~ 9FA5) は中国・台湾・日本・韓国の 4 欄からなり、それぞれの漢字に、その漢字の採録元の規格での漢字コードが併記されている。このうち中国欄の採録元は

0: GB 2312-80	6763 字
1: GB/T 12345-90+ α	2202+150 字
3: GB 7589-87 繁体字	4835 字
5: GB 7590-87 繁体字	2842 字
7: 現代漢語通用字表	42 字
8: GB 8565-88	290 字

の 6 つ*であり、合計 17124 字が採録されている。例えば JIS X 0221 の 5520 ~ 5522 (図 1 参照) は「𠂇」が GB 2312-80 の 63-75 から、「𠂇」が GB 8565-88 の 08-85 から、「𠂇」が GB 2312-80 の 63-79 から、それぞれ採録されたということが規格票に明記されているのである。しかしこれらの漢字について、実際に採録元の規格を調べてみると「𠂇」と「𠂇」は確かに GB 2312-80 の所定の区位番号に収められているが、「𠂇」は GB 8565-88 には収められていないことがわ

*JIS X 0221 では「8: GB 8565-89」となっているが、GB 8565 には 1988 年版しかない。

かる。驚くべきことに、GB 8565-88には08区など存在しないのであり(図3参照)、「啊」がGB 8565-88の08-85から採録されることなどありえないのである。

Row/Cell	C	J	K
Hex code	G-Hanzi	T-Kanji	Hanja
085-032	唠		
5520	0-9F98 0-9379		
085-033	啊		
5521	0-9875 0-0885		
085-034	啖		
5522	0-9F8F 0-8379		

図1: JIS X 0221の5520～5522

JIS X 0221の漢字のうちGB 8565-88から採録されたことになっているものは290字ある。しかし筆者らが調査したところ、そのうち実に

4E26 並	4EB8 辨	5091 傑	50C0 働	5289 効	52F3 勳	5338 叅
540B 吋	544E 呎	5521 啊	5557 咱	5574 啖	5586 喆	577F 埒
590A 文	59B3 妳	5AAE 媪	5C05 尅	5D11 崑	5D19 崑	5DD6 巖
5E0B 帛	5EC8 廈	60AE 媪	658F 娘	65C2 旂	6607 昇	6673 皙
668E 哄	68BF 槌	691D 梨	6A11 樑	6C25 氫	6C96 冲	6D44 淨
6D6C 湮	6E1B 減	6E67 湧	6F82 激	6FEC 濬	709E 炆	71FF 耀
739A 场	74E9 阄	7526 魁	7551 畑	7652 癒	76EA 盪	778B 瞋
7B5C 筍	7B87 箇	7C50 籐	7EB5 紕	7EE4 紕	7EF9 綯	7FD9 翊
8129 脩	8460 蓓	849E 蒞	8780 蜚	88AF 袂	8955 襦	897E 所
89DA 舳	8BAC 託	8D5F 贖	8EAD 舳	8F80 舳	8F92 温	8FFA 迺
94C7 饱	94DA 姪	955A 媪	9560 媪	9575 媪	9618 媪	9895 媪
98D0 飀	998C 媪	9A94 媪	9A99 媪	9B36 媪	9E24/媪	9E40/媪
9E50 鸱	9E5D 鸱	9E60 鸱				

の87字が、実際にはGB 8565-88に含まれていないことが判明した。これらの漢字は一体どのような経緯でJIS X 0221に採録されたのであろうか？

結論を先に述べると、これらの漢字は標準電碼本すなわち中国の電報コードと、現代漢語通用字表から採録されたものである。しかし本来JIS X 0221の採録元ではない標準電碼本からの採録に際しては、非常に不誠実な手段が用いられていることが明らかとなった。また現代漢語通用字表は、結局、中国のゴリ押しともいうべきやり方でJIS X 0221の採録元となったがために、その一部がGB 8565-88に含まれているかのような形でJIS X 0221に押し込まざるを得なかったのである。次章では、これらの漢字

がJIS X 0221に採録された経緯を、中国の漢字表や漢字コード規格に絡めて述べる。

2 中国の漢字表と漢字規格

中国の漢字コードの基本となる規格は、1980年に発布されたGB 2312-80である。しかし1980年以後、中国ではいくつかの重要な漢字表が発布され、それらの漢字の一部がGB 2312-80に含まれていないことが問題となった。以下、中国で1980年以後に発布された漢字表と、それに伴いGB 2312-80ではどのような漢字が不足したか、さらにそのような漢字コードの不足を最終的にUnicodeで強引に解決してしまうまでの紆余曲折について述べる。

2.1 標準電碼本と現代漢語常用字表

中華人民共和国郵電部は1983年、標準電碼本(修訂本)を発布した。図2に一部を示す。電碼とは、漢字や記号にそれぞれ4桁

5000	5001	5002	5003	5004	5005	5006	5007	5008	5009
刑	電	署				骂	罢	罹	
HKI	HKJ	HXX	HKL	HKM	HKN	HKO	HKP	HKL	HKB
5010	5011	5012	5013	5014	5015	5016	5017	5018	5019
厠	罗	累	羈	鋤	羊	羌	美		
HKSHKT	HKUHKV	HKWHKX	HKY	HKZ	HLA	HLB			
5020	5021	5022	5023	5024	5025	5026	5027	5028	5029
羴	羔		羚	羴	羴	羴	羴	羴	羴
HLC	HLD	HLE	HLF	HLG	H LH	H LI	H LJ	H LK	H LL
5030	5031	5032	5033	5034	5035	5036	5037	5038	5039
义	羯	羲	羲	羴	羴	羴	羽	翀	
HLM	HLN	HL O	HLP	HLQ	HLR	HLS	H LT	H LU	H LV
5040	5041	5042	5043	5044	5045	5046	5047	5048	5049
翁	翊	翊	翊	翊	翊	翊	翊	翊	翊
HLW	HLX	HLY	HLZ	HMA	HMB	HMC	HMD	HME	HMF
5050	5051	5052	5053	5054	5055	5056	5057	5058	5059
翠	翡		翥	翥	翥	翥	翥	翥	翥
HMG	HMH	HMI	HMJ	HMK	HML	HMM	HMN	HMO	HMP
5060	5061	5062	5063	5064	5065	5066	5067	5068	5069
翰	翳	翳	翳	翳	翳	翳	翳	翳	翳
HMQ	HMR	HMS	HMT	HMU	HMV	HMW	HMX	HMY	HMZ
5070	5071	5072	5073	5074	5075	5076	5077	5078	5079
喱	老	考	考	考	考	考	考	考	考
HNA	HNB	HNC	HND	HNE	HNF	HNG	HNH	HNI	HNJ
5080	5081	5082	5083	5084	5085	5086	5087	5088	5089
要		耐		啊	未	籽	耕	耗	耘
HNK	HNL	HNM	HNN	HNO	HNP	HNQ	HNR	HNS	HNT
5090	5091	5092	5093	5094	5095	5096	5097	5098	5099
耙	耜	耜		耜	耜	耜	耜	耜	耜
HNU	HNV	HNW	HNX	HNY	HNZ	HOA	HOB	HOC	HOD

図2: 標準電碼本(修訂本)の5000～5099

の10進数を振ったもので、中国での電報に用いられるいわば電報コードである。1983年に発布された標準電碼本(修訂本)は、それまで中国で用いられてきた繁体字版の標

倭俾儆剝啞啍啞圮圮圮圮圮圮圮圮圮圮圮圮
抵捌掇咄咄咄咄咄咄咄咄咄咄咄咄咄咄咄
炮燧猱猱猱猱猱猱猱猱猱猱猱猱猱猱猱
茛菘蕘蕘蕘蕘蕘蕘蕘蕘蕘蕘蕘蕘蕘蕘蕘蕘蕘
鞴鞴饁饁饁饁饁饁饁饁饁饁饁饁饁饁饁饁饁
鞴鞴饁饁饁饁饁饁饁饁饁饁饁饁饁饁饁饁饁

の107字であることが、筆者らの調査でわかっている。

2.5 そしてUnicode

1990年、計算機と微電子発展研究中心はUnicodeの漢字部分を制定するにあたって、CCITT Chinese Primary Setの漢字7538字、標準電碼本(修訂本)の漢字7079字、現代漢語通用字表7000字の全てをUnicodeに含めるよう計画した。しかし、Unicodeの漢字の採録元が各国の国内規格に限られていたため、これらの漢字を全て採録するために計算機と微電子発展研究中心が取った方法は、非常に悪辣なものとならざるを得なかった。

CCITT Chinese Primary Setは中国の国内規格ではないため、7538字のうち国内規格にも含まれている漢字は、とりあえず国内規格経由でUnicodeに採録されている。以下に内訳を示す。

0: GB 2312-80 経由	6763 字
1: GB/T 12345-90 経由	2 字
3: GB 7589-87 繁体字経由	420 字 [¶]
5: GB 7590-87 繁体字経由	70 字 [¶]
8: GB 8565-88 経由	203 字

残る80字のうち

並傑働効勳口吋呶啞啞喆喆喆喆喆喆喆喆喆喆
嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪
嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪
嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪
嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪

の79字の採録には、非常に不誠実な方法が取られている。すなわちこれら79字は、GB 8565に含まれていないにもかかわらずGB 8565からの採録ということになっており、そこでの区位番号はCCITT Chinese Primary Setでのものでそのまま用いられているのである。なお最後の1字である「采」の中国欄は空欄となっており、この字は台

湾および日本の国内規格からの採録となっているが、この方がまだ誠実な方法であると言えよう。

標準電碼本(修訂本)の漢字7079字についても同様で、国内規格にも含まれている漢字は国内規格経由でUnicodeに採録されている。以下に内訳を示す。

0: GB 2312-80 経由	6377 字
1: GB/T 12345-90 経由	5 字
3: GB 7589-87 繁体字経由	362 字 [¶]
5: GB 7590-87 繁体字経由	70 字 [¶]
8: GB 8565-88 経由	201 字

残る64字は大きく3つに分類される。第1のグループは、GB 8565に含まれていないにもかかわらずGB 8565からの採録ということになっている漢字で

並傑働効勳口吋呶啞啞喆喆喆喆喆喆喆喆喆
嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪
嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪
嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪
嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪嬪

の58字である。このうちCCITT Chinese Primary Setに含まれている50字についてはそのままの区位番号が用いられているが、残る「匸・吋・呶・啞・文・湮・氀・西」の8字については08-83～08-87・08-90～08-92という謎の区位番号となっている。第2のグループは、Unicodeに採録されているが中国欄が空欄のもので「采・藟・氦」の3字である。このうち「氦」はGB 8565.2-88表A4の14-51に含まれているので、空欄なのは変である。第3のグループはUnicodeに採録漏れとなったもので「萑・蜚・饒」の3字である。これらはそれぞれ「萑・蜚・饒」という形でUnicodeに採録されたのだろう、と筆者らは推測している。

現代漢語通用字表の7000字については、国内規格経由で採録されているのは

0: GB 2312-80 経由	6593 字
1: GB/T 12345-90 経由	2 字
3: GB 7589-87 繁体字経由	271 字
8: GB 8565-88 経由	66 字

[¶]「俞・嶇」の採録元はそれぞれ5-1731・3-2580となるべきだが、規格票では3-1893・5-3123と誤記されている。他にも区位番号のズレが散見される。

である。残る 68 字のうち

𧣾 𧣿 𧤀 𧤁 𧤂 𧤃 𧤄 𧤅 𧤆 𧤇 𧤈 𧤉 𧤊 𧤋 𧤌 𧤍 𧤎 𧤏 𧤐 𧤑 𧤒 𧤓 𧤔 𧤕 𧤖 𧤗 𧤘 𧤙 𧤚 𧤛 𧤜 𧤝 𧤞 𧤟 𧤠 𧤡 𧤢 𧤣 𧤤 𧤥 𧤦 𧤧 𧤨 𧤩 𧤪 𧤫 𧤬 𧤭 𧤮 𧤯 𧤰 𧤱 𧤲 𧤳 𧤴 𧤵 𧤶 𧤷 𧤸 𧤹 𧤺 𧤻 𧤼 𧤽 𧤾 𧤿

の 27 字は、GB 8565 に含まれていないにもかかわらず GB 8565 からの採録ということになっている漢字であり、これらは全て CCITT Chinese Primary Set に含まれている。最後の 41 字すなわち

𧤶 𧤷 𧤸 𧤹 𧤺 𧤻 𧤼 𧤽 𧤾 𧤿 𧥀 𧥁 𧥂 𧥃 𧥄 𧥅 𧥆 𧥇 𧥈 𧥉 𧥊 𧥋 𧥌 𧥍 𧥎 𧥏 𧥐 𧥑 𧥒 𧥓 𧥔 𧥕 𧥖 𧥗 𧥘 𧥙 𧥚 𧥛 𧥜 𧥝 𧥞 𧥟 𧥠 𧥡 𧥢 𧥣 𧥤 𧥥 𧥦 𧥧 𧥨 𧥩 𧥪 𧥫 𧥬 𧥭 𧥮 𧥯 𧥰 𧥱 𧥲 𧥳 𧥴 𧥵 𧥶 𧥷 𧥸 𧥹 𧥺 𧥻 𧥼 𧥽 𧥾 𧥿

は、Unicode の採録元として

7: 現代漢語通用字表 42 字

を作り、それを經由して無理矢理採録された漢字である。ここでの区位番号は 01-01 ~ 01-09・01-11 ~ 01-42 となっている¹¹⁾が、現代漢語通用字表の漢字にはいかなる番号も付与されておらず、どういう理由でこの区位番号が付けられたのかは謎である。

3 まとめ

ここまでをまとめると、GB 8565 に含まれていないにもかかわらず GB 8565 から Unicode に採録されたことになっている漢字 87 字すなわち

4E26 並 4EB8 𧣾 5091 傑 50CD 働 52B9 効 52F3 勳 5338 𧣾 540B 吋 544E 呎 5521 𧣾 5557 𧣾 5574 𧣾 5586 𧣾 577F 𧣾 590A 文 59B3 妳 5AAE 媪 5C05 尅 5D11 崑 5D19 崑 5DD6 巖 5E0B 帛 5EC8 廈 60AE 悞 658F 煨 65C2 旂 6607 昇 6673 哲 668E 暎 68BF 槌 691D 梨 6A11 樑 6C25 氫 6C96 沖 6D44 淨 6D6C 湮 6E1B 滅 6E67 湧 6F82 激 6FEC 潛 709E 炆 71FF 燿 739A 𧣾 74E9 𧣾 7526 甦 7551 烟 7652 瘡 76EA 盪 778B 瞋 7B5C 筓 7B87 箇 7C50 籐 7E8B 纒 7EE4 紿 7EF9 綯 7FD9 𧣾 8129 脩 8460 葆 849E 蒞 8780 蜚 88AF 裒 8955 裒 897E 𧣾 89D4 舫 8BAC 託 8D5F 贊 8EAD 𧣾 8F80 輶 8F92 輶 8FFA 迺 94C7 𧣾 94DA 𧣾 955A 𧣾 9560 𧣾 9575 𧣾 9618 𧣾 9895 𧣾 98D0 𧣾 998C 𧣾 9A94 𧣾 9A99 𧣾 9B36 𧣾 9E24 𧣾 9E40 𧣾 9E50 𧣾 9E60 𧣾

の素性は大きく 4 つに分類されることがわかる。第 1 のグループは、標準電碼本(修訂本)から直接 Unicode に採録された漢字で

¹¹⁾さらに 01-43 という区位番号で、現代漢語通用字表に含まれていない「𧣾」が Unicode に追加されている。これを含めて「7: 現代漢語通用字表」からの採録は 42 字ということになっている。

5338 𧣾 540B 吋 544E 呎 5521 𧣾 590A 文 6D6C 湮 74E9 𧣾 897E 𧣾

の 8 字である。第 2 のグループは、標準電碼本(修訂本)から CCITT Chinese Primary Set 經由で Unicode に採録された漢字で

4E26 並 5091 傑 50CD 働 52B9 効 52F3 勳 5557 𧣾 5586 𧣾 577F 𧣾 59B3 妳 5AAE 媪 5C05 尅 5D11 崑 5D19 崑 5DD6 巖 5E0B 帛 5EC8 廈 60AE 悞 658F 煨 65C2 旂 6607 昇 6673 哲 668E 暎 691D 梨 6A11 樑 6C25 氫 6C96 沖 6D44 淨 6E1B 滅 6E67 湧 6F82 激 6FEC 潛 709E 炆 71FF 燿 7526 甦 7551 烟 7652 瘡 76EA 盪 778B 瞋 7B87 箇 7C50 籐 8129 脩 8460 葆 849E 蒞 89D4 舫 8BAC 託 8EAD 𧣾 8FFA 迺 94C7 𧣾 9895 𧣾 9A94 𧣾

の 50 字である。第 3 のグループは、現代漢語通用字表から CCITT Chinese Primary Set 經由で Unicode に採録された漢字で

4EB8 𧣾 5574 𧣾 739A 𧣾 7B5C 筓 7EE4 紿 7EE4 紿 7EF9 綯 7FD9 𧣾 88AF 裒 8955 裒 8D5F 贊 8F80 輶 8F92 輶 94DA 𧣾 955A 𧣾 9560 𧣾 9575 𧣾 9618 𧣾 98D0 𧣾 998C 𧣾 9A99 𧣾 9B36 𧣾 9E24 𧣾 9E40 𧣾 9E50 𧣾 9E60 𧣾

の 27 字である。最後のグループは、何らかの間違いで CCITT Chinese Primary Set に紛れ込んだ漢字が Unicode に採録されたもので「𧣾 𧣾 8780 蜚」の 2 字である。

4 おわりに

Unicode は 1993 年に国際規格の ISO/IEC 10646-1:1993 となり、さらに 1995 年に JIS X 0221 となったため、これら 87 字は、採録元に存在しないという異様な形で JIS X 0221 に採録されることとなった。元々は採録元でないはずの標準電碼本(修訂本)や CCITT Chinese Primary Set まで Unicode に盛り込みたい、という中国の熱意はわからなくもないが、それならば現代漢語通用字表のように採録元としてねじこむ方がまだマシだったと言えよう。今後 JIS X 0221 を改訂する際には、これらの漢字の出所を何らかの形ではっきりさせておく必要があると考えられる。

なおこの研究の結果得られた、Unicode・GB 2312-80・GB 8565.2-88 表 A4・CCITT Chinese Primary Set・標準電碼本(修訂本)の全コード対応表を、<ftp://ginkaku.kudpc.kyoto-u.ac.jp/CJKtable/Uni2GB.Z> においてフリーで公開している。せいぜい利用されたい。