

# 住民基本台帳ネットワーク統一文字とその問題点

## Character Code on Basic Resident Registration Network in Japan

安岡孝一  
Koichi Yasuoka

京都大学人文科学研究所附属東アジア人文情報学研究センター (〒606-8265 京都市左京区北白川東小倉町47)  
Institute for Research in Humanities, Kyoto University, Kyoto 606-8265 JAPAN  
yasuoka@kanji.zinbun.kyoto-u.ac.jp 075-753-6994

### 著者抄録

住民基本台帳ネットワーク統一文字は、当初は UCS(国際符号化文字集合) を拡張する形で設計されており、いかなるコンピュータでも使用できるオープンなシステムを目指したはずだった。しかし、UCS の基本設計に対する誤解や、その後の UCS の変化に十分追従できなかったために、住民基本台帳ネットワーク統一文字は、もはや現代の OS 上では動作しない文字コードになってしまっている。本稿では、住民基本台帳ネットワーク統一文字の問題点と、その問題点を踏まえた上での今後の方策について述べる。

### キーワード

文字コード, 国際符号化文字集合 (UCS), Unicode, 文字化け, 住民基本台帳ネットワーク, 裁判員候補者予定者名簿システム, 後期高齢者医療広域連合電算処理システム

## 1 はじめに

2012年4月6日付け最高裁判所事務総局経理局一般競争入札「裁判員候補者に対する通知書の印刷、発送及び集計等の業務一式」の官報公告には、以下の条件が付されていた。

本件業務の通知書等の印刷物は、K A J O\_\_ J 明朝フォントを利用したものを作成すること。

実は、裁判員候補者予定者名簿システムや、後期高齢者医療広域連合電算処理システムには、住民基本台帳ネットワーク(以下、住基ネット)に使われている文字コードと、同一の文字コードが用いられている。いわゆる住民基本台帳ネットワーク統一文字(以下、住基文字)と呼ばれている文字コードだ。ところが、この入札公告では、住基文字の使用を、入札条件として書いていない。その代わりに、「K A J O\_\_ J 明朝」という住基文字と互換なフォントを指定することで、結果として使用する文字コードを、住基文字に決め打ちさせているわけである。

ではなぜ、この入札公告は、住基文字の使用を、条件に書いていないのだろうか。使用する文字コードを明示せず、「K A J O\_\_ J 明朝」というフォントを指定する形となっているのは、なぜなのだろうか。付

度するに、住基文字の仕様が、現在も非公開となっているからだ。仕様が非公開のものを、入札公告の条件に書くわけにはいかず、特定のフォントを指定する形で代用しているのだ。

そもそも文字コードというものは、国際規格(ISO)に基づくオープンなものでなければならない。さもなければ、現代の国際化された OS の上で使用することができないし、システム間をまたぐ文字のやりとりができなくなってしまう。そのために、日本の文字コード関係者は、かなりの労力を割いて、日本で必要な文字を ISO に国際提案し、さらには ISO が制定している国際符号化文字集合(UCS, いわゆる Unicode)を、日本の国内規格すなわち JIS に翻訳しているのである。

住基文字も、その点を考慮して、オープン・システムとして設計されたはずの文字コードだった<sup>1)</sup>。そうすることで、住基ネットの外とのやり取りも可能とし、ゆくゆくは日本の行政を支える文字コードとなるはずだった。ところが、住基ネットやその周辺の紆余曲折<sup>2)</sup>もあって、住基文字の仕様は公開されなかった。そして住基文字は、現在では、オープン・システムにはほど遠い、かなり多くの問題を抱えた文字コードとなってしまった<sup>3,4)</sup>。本稿では、住基文字がどのよ

表 1: 住基文字の 7 つの領域と収録文字数

0020 ~ 33ff	非漢字	1,274 字
ff00 ~ fffe	互換非漢字	163 字
ac00 ~ acff	変体仮名	168 字
4e00 ~ 9fff	漢字	13,185 字
3400 ~ 4dff		806 字
aaa0 ~ abff		303 字
f900 ~ faff	互換漢字	99 字
ad00 ~ d6ff	追加漢字	5,172 字
e000 ~ f8ff	外字	-

表 2: JIS X 0213 附属書 11 「3.3」と住基文字の差異

JIS X 0213	住基文字	JIS X 0213	住基文字
005C	↘	ff3c	2014 — 2015
00A2	℄	ffe0	2016    2225
00A3	£	ffe1	203E — ffe3
00A5	¥	005c	2212 — ff0d
00AC	⌋	ffe2	301C ~ ff5e
00D0	Ⓓ	0110	9B1D 鬚 9b1c

表 3: JIS X 0212 と JIS X 0213 とで規格票字形が異なるが UCS が衝突している住基文字

JIS X 0212 から収録	JIS X 0213 から収録	JIS X 0212 から収録	JIS X 0213 から収録	JIS X 0212 から収録	JIS X 0213 から収録
僂 ad9b	僂 50f2	瑤 b5c3	瑤 7462	葵 b9d2	葵 845c
嫚 affe	嫚 5ada	瓘 b5dc	瓘 74d8	穉 bbbe	穉 8c9b
晷 b2a2	晷 6677	瓠 b5e2	瓠 74ef	鍈 be04	鍈 9365
睽 b2aa	睽 6680	瘳 b606	瘳 7608	鍈 be05	鍈 93a1
櫟 b37b	櫟 69fe	盞 b638	盞 76d4	鍈 be19	鍈 938b
殍 b3ed	殍 6ba9	盞 b65f	盞 77a2	鍈 be93	鍈 96da
汙 b40e	汙 6c74	礪 b694	礪 7934	鍈 beb5	鍈 9755
溼 b4d5	溼 6ff9	籩 b7ed	籩 7c69	馱 c044	馱 9ce6
熉 b540	熉 71b3	絜 b817	絜 7d5c	鷲 c055	鷲 9dbf

うな問題を抱えているのか、問題を抱えるに至った経緯を明らかにするとともに、今後の方策を検討する。

## 2 住基文字の概要

住基文字は、地方自治情報センターが 2001 年 2 月 26 日に「検討版」を配布、2002 年 5 月 8 日に「確定版」21,039 字となり、2002 年 8 月 5 日に住基ネットと共に運用開始された文字コードである。当時の JIS X 0221-1『国際符号化文字集合 (UCS) 一第 1 部：体系及び基本多言語面』(2001 年 4 月 20 日制定版) の

UCS-2 に添って設計されているものの、独自の拡張を施した 2 バイトコードである。各文字には 16 進数 4 桁のコードが振られており、表 1 に示す 7 つの領域に分かれている。

住基文字の中心をなすのは、当時の JIS X 0213『7 ビット及び 8 ビットの 2 バイト情報交換用符号化拡張文字集合』(2000 年 1 月 20 日制定版) 11,123 字である。ただし、この 11,123 字がそのまま収録されているのではなく、同規格票の附属書 11「3.3 JIS X 0221 からの索引」に示された 11,319 個の UCS が、ほぼそのまま住基文字に採用されている。この中には、当

時 UCS に提案中だった 491 個のコードポイント (JIS X 0213 ではカッコ付き UCS で示されていた) も含まれているが、それらも全て住基文字となっている。なお、「確定版」住基文字においては、11,319 個のうち 12 個のコードポイントが削除され、代わりに、表 2 に示すコードポイントが使用された。

住基文字には、JIS X 0212『情報交換用符号—補助漢字』(1990 年 10 月 1 日制定版) 6,067 字も全て収録されている。うち 6,040 字は、JIS X 0221-1 の附属書 1 で示されていた UCS に収録されているが、2,879 字が JIS X 0213 とダブっており、純粋に JIS X 0212 から採録されたのは 3,161 字である。また、残り 27 字は、JIS X 0212 と JIS X 0213 とで規格票字形が異なるが UCS が衝突している漢字で、JIS X 0212 の方を追加漢字領域に収録している (表 3)。

住基文字の追加漢字領域には、各自治体の住民票システムで使われていたメーカー外字が、ほぼ部首画数順に収録されている。ただし、メーカー外字のうち、UCS との対応付けがうまくいった 1,201 字については、3400 ~ 9fa4 の漢字領域に収録されている。また、いわゆる IBM 外字については、互換漢字領域の fa0e ~ fa2d にも収録されている。ac00 ~ aca7 には、変体仮名 168 字が収録されている。e000 ~ f8ff は外字領域であり、ビットマップ画像と共に、住基ネットでの外字のやりとりに用いる。

なお、JIS X 0213 は 2004 年 2 月 20 日の改正で、491 個のカッコ付き UCS を全て解消し、新たな UCS を示したが、住基文字はこれに追随せず、カッコ付き UCS をそのまま使い続けている。一方、2011 年 12 月 26 日付け法務省告示第 582 号『在留カード等に係る漢字氏名の表記等に関する告示』(2012 年 7 月 9 日施行) に対しては、追加漢字領域の c109 ~ c18b に、入管漢字 131 字を新たに追加した。この結果、住基文字の総数は現在 21,170 字となっている (表 1)。

### 3 住基文字の問題点

住基文字の問題点は、それが UCS をもとに作られたものであるにもかかわらず、かなりの部分で UCS と一致していない、という点である。端的に言えば、住基文字 21,170 字中、UCS と一致しているのは 15,379 字で、残りの 5,791 字は UCS とは異なっている。5,791 字の内訳は、非漢字領域 122 字、変体仮名 168 字、漢字領域 303 字 (aaa1 ~ abbf)、互換漢字領域 26 字、追加漢字 5,172 字となっており、不一致の大半が漢字である。

UCS と一致しない漢字のうち、最も重症だと考えられるものは、漢字領域の aaa1 ~ abbf に含まれる 303 字である。これらは、JIS X 0213 のカッコ付き

表 4: 住基文字の aaa0 ~ aacf (左) と UCS の U+AAA0 ~ U+AACF (右) との比較

	aaa	aab	aac	AAA	AAB	AAC
0		劉	圻	𐄀	𐄁	𐄂
		aab0	aac0	AAA0	AAB0	AAC0
1	𠄁	𠄂	𠄃	𐄃	𐄄	𐄅
	aaa1	aab1	aac1	AAA1	AAB1	AAC1
2	𠄄	𠄅	𠄆	𐄆	𐄇	𐄈
	aaa2	aab2	aac2	AAA2	AAB2	AAC2
3	𠄇	𠄈	𠄉	𐄈	𐄉	
	aaa3	aab3	aac3	AAA3	AAB3	
4	𠄉	𠄊	𠄋	𐄉	𐄊	
	aaa4	aab4	aac4	AAA4	AAB4	
5	𠄊	𠄋	𠄌	𐄊	𐄋	
	aaa5	aab5	aac5	AAA5	AAB5	
6	𠄋	𠄌	𠄍	𐄋	𐄌	
	aaa6	aab6	aac6	AAA6	AAB6	
7	𠄌	𠄍	𠄎	𐄌	𐄍	
	aaa7	aab7	aac7	AAA7	AAB7	
8	𠄍	𠄎	𠄏	𐄍	𐄎	
	aaa8	aab8	aac8	AAA8	AAB8	
9	𠄎	𠄏	𠄐	𐄎	𐄏	
	aaa9	aab9	aac9	AAA9	AAB9	
a	𠄏	𠄐	𠄑	𐄏	𐄐	
	aaaa	aaba	aaca	AAAA	AABA	
b	𠄐	𠄑	𠄒	𐄐	𐄑	
	aaab	aabb	aacb	AAAB	AABB	
c	𠄑	𠄒	𠄓	𐄑	𐄒	
	aaac	aabc	aacc	AAAC	AABC	
d	𠄒	𠄓	𠄔	𐄒	𐄓	
	aaad	aabd	aacd	AAAD	AABD	
e	𠄓	𠄔	𠄕	𐄓	𐄔	
	aaae	aabe	aace	AAAE	AABE	
f	𠄔	𠄕	𠄖	𐄔	𐄕	
	aaaf	aabf	aacf	AAAF	AABF	

UCS に由来しているが、UCS の他の地域の文字とバッティングしているため (表 4)、UCS ベースのシステムでは正常な処理が期待できない。たとえば、住基文字 aabe「土」は、黒タイ文字の U+AABE「TAI VIET VOWEL AM」のコードポイントを流用しているが、「TAI VIET VOWEL AM」は非前進文字なのである。この結果、UCS ベースのシステムで aabe「土」を表示・印刷すると、それは直前の文字に重なってしまう。非前進文字として扱われるのだから仕方がない。aab0, aab2, aab3, aab4, aab7, aab8, aabf, aac1, aaeб, aaec, aaed, aaeе, aaeф, aaf5, aaf6 も同様である。

UCS と一致しない漢字のうち、これもまた重症だと考えられるものは、互換漢字領域の fa45 ~ fa5d と fa6b である。表 5 を見ればわかるとおり、UCS とは別の漢字が収録されている。端的に言えば、1 文字ズレているのだ。このようなズレ方をすれば、当然、文字化けの原因となるし、また文字化けしてもなかなか気付かない、という問題がある。このズレは、元々は JIS X 0213 のカッコ付き UCS に由来している。実は、JIS X 0213 の附属書 11「3.3 JIS X 0221 からの索引」を作成したのは筆者 (安岡孝一) であり、その点では筆者にも責任の一端があると言える。ただし、JIS X 0213 は 2004 年改正でこのズレを修正して、カッコ付き UCS を全て解消しており、住基文字がこの改正に追随しなかったところに、いまだ禍根が残っている。

追加漢字 5,172 字と変体仮名 168 字は、全て、UCS のハングルとバッティングしている。したがって、住基文字はハングルと同時に使えない。住民票には通常、ハングルは併記されないのだから、この問題が表面化する可能性は極めて低いのだが、考慮しておく必要はある。非漢字領域で、UCS と一致していない 122 字についても同様である。

#### 4 今後の方策

前節でも明らかにした通り、住基文字は UCS ベースのシステムでは、正常に扱うことができなくなっている。しかし、住基ネット内部の文字コードを変更するのは、リスクが大きい。それゆえ、入出力部分だけを UTF-16 に完全移行し、住基ネットとのインターフェース部分で住基文字との変換をおこなう、という方策を採用すべきである。すなわち、住基文字はシステム内部に閉じ込め、入出力には住基文字そのものは用いない、という方策である。以下、そのような方策を採るために、必要な手順を考えることにしよう。

まず第一に、住基文字の仕様は、公開すべきである。この 10 年の間に、住基文字の内容については、

表 5: 住基文字の fa40 ~ fa6f (左) と UCS の U+FA40 ~ U+FA6F (右) との比較

	fa4	fa5	fa6	FA4	FA5	FA6
0	懲 fa40	祐 fa50	褐 fa60	懲 FA40	祖 FA50	褐 FA60
1	敏 fa41	祖 fa51	視 fa61	敏 FA41	祝 FA51	視 FA61
2	既 fa42	祝 fa52	謁 fa62	既 FA42	禍 FA52	謁 FA62
3	暑 fa43	禍 fa53	謹 fa63	暑 FA43	禎 FA53	謹 FA63
4	梅 fa44	禎 fa54	賓 fa64	梅 FA44	穀 FA54	賓 FA64
5	概 fa45	穀 fa55	贈 fa65	海 FA45	突 FA55	贈 FA65
6	海 fa46	突 fa56	逸 fa66	渚 FA46	節 FA56	逸 FA66
7	渚 fa47	節 fa57	逸 fa67	漢 FA47	練 FA57	逸 FA67
8	漢 fa48	練 fa58	難 fa68	煮 FA48	縉 FA58	難 FA68
9	煮 fa49	縉 fa59	響 fa69	𠄎 FA49	繁 FA59	響 FA69
a	𠄎 fa4a	繁 fa5a	頻 fa6a	琢 FA4A	署 FA5A	頻 FA6A
b	琢 fa4b	署 fa5b	鬪 fa6b	碑 FA4B	者 FA5B	患 FA6B
c	碑 fa4c	者 fa5c		社 FA4C	臭 FA5C	熙 FA6C
d	社 fa4d	臭 fa5d		社 FA4D	𠄎 FA5D	館 FA6D
e	社 fa4e	𠄎 fa5e		祈 FA4E	𠄎 FA5E	
f	祈 fa4f	著 fa5f		祐 FA4F	著 FA5F	

表 6: 住基文字と UCS の対応表案

住基文字	UCS	住基文字	UCS	住基文字	UCS
æ	00e6	一	4e00	士	571f
æ	01fd	丈	4e08	士	5721
æ	0234	丈	ad03	士	aabe
か	304b	丈	ad05	礼	793c
が	304c	丈	aaa2	礼	fa18
が	31d0	五	ad02	祖	b6a9
加	ac14	丑	4e11	祖	7956
の	ac15	丑	ad08	祖	fa51
采	ac16	丑	ad0a	雙	fa0e
あ	ac17	刃	ad07	雙	aea4
丿	303d	刃	ad09	𠄎	005c
𠄎	ad1b	刃	4e12	𠄎	ffe5

さまざまな形で非公式に言及されてきた<sup>5,6)</sup>し、本稿もその一翼をなしている。しかし、住基文字の仕様は、現在まで、公式な形で公開されたことがない。公式な形で公開されないと、一連の手続をオープンにおこなえないため、UCS すなわち国際符号化文字集合へのアクションが、十全な形で起こせないのだ。本来、文字コードというのは、プライバシーだの個人情報だのとは何の関係もない。しかも、住基文字は JIS X 0213・0212・0221-1 を元に設計した文字コードで、もともとオープンなものなのに、それを 10 年以上も非公開にしてきたという事実の方が、はっきり言って驚きだろう。また、公開をおこなうのは地方自治情報センターではなく、総務省がガバメント・アクションとしておこなうべきである。

次に、全ての住基文字について、UCS との公式な対応表を作成・公開する。もちろん、一から全てを作る必要はなく、情報処理推進機構が公開している『文字情報基盤 文字情報一覧表』<sup>7)</sup>や、あるいは JIS X 0213 の追補 1 などを利用して、対応表を作成すべきだろう。ただしそれを、公式な形で公開する、という点が重要であり、UCS にはない字は、正直に「ない」と書いてほしいのだ(表 6)。また、住基文字と UCS の

対応は、1 対 1 にすべきだ。たとえば、ad08 と ad0a はかなり字形が似ているが、ad08 に<4E11 E0102>を対応させるなら、ad0a に<4E11 E0102>をダブって対応させるべきではなく、現時点では ad0a は「ない」と言い切るべきだろう。

その上で、UCS にはない字は国際提案する。公開された住基文字の仕様と、公式な対応表を元に、国際提案する。国際提案そのものは、情報処理学会配下の ITSCJ/SC2 専門委員会に依頼することになるだろう。国際提案を通すためには、いくつか追加資料が必要になるかもしれないし、あるいは汚い手を使うことになるかもしれないが、とにかく国際提案する。

さらに、住基文字と UCS の公式な対応表を、継続的にメンテナンスして、バージョンアップし、新旧あわせて公開し続ける。最初の対応表は、もちろん間違いを含んでいるだろうし、UCS への国際提案が通れば、対応表は徐々に埋まっていくからだ。そして、対応表が全て埋まった段階になったら、改めて、住基ネットにおける文字コードをどうするか、検討すべきだろう。それまでは、住基文字という文字コードそのものは、あくまで現状維持を続けるべきだ。拙速な変更は、さらなる禍根を生むだけである。

## 5 おわりに

住民基本台帳ネットワーク統一文字について、現状の問題点と、今後おこなうべき方策を述べた。現代の OS は、UCS をベースに設計されており、そのような OS 上では、住基文字はほぼ確実に文字化けする。この状態を脱却するためには、入出力に住基文字を使わないようにするしかない。

以前、筆者は、本誌 2007 年 5 月号に『ケータイの絵文字と文字コード』<sup>8)</sup>と題する小論を発表した。これが導火線となって、ケータイ絵文字の UCS 化がおこなわれたが、最終的に国際規格となったのは 2011 年 3 月 15 日で、3 年と 10 ヶ月を要した。一方、筆者は、住基文字に対しても、2010 年 10 月に警告を発した<sup>3)</sup>つもりだったのだが、関係各部署(住基ネット

を担当する総務省、裁判員候補者予定者名簿システムを担当する最高裁判所、後期高齢者医療広域連合電算処理システムを担当する厚生労働省)からは何の音沙汰もなく、また、何のアクションもない。

住基文字の開発と保守を担ってきた地方自治情報センターは、2013 年 3 月で廃止が予定されている。その際に住基文字がどうなるのか、その処遇は現時点では不明である。その意味では、今、住基文字の仕様を公開しておかなければ、住基ネットを継続するにしろ、マイナンバーへの移行をおこなうにしろ、全くの手詰まりとなってしまう可能性が高い。国民全員と日本在住の外国人全てを記述するための文字コードである住基文字が、水泡と帰す危険性すらあるのだ。関係諸氏の速やかな対応を、強く望むところである。

## 参考文献

- 1) 統一文字仕様書。地方自治情報センター，2002。
- 2) 榎並利博。住基ネットはなぜ『悪者』となったのか。富士通総研 (FRI) 経済研究所，研究レポート。2011，no.368。 <http://jp.fujitsu.com/group/fri/report/research/2011/report-368.html> (2012-11-01 閲覧)。
- 3) 安岡孝一。失われた文字コード。漢字文献情報処理研究。2010，vol.11，p.76-81。  
<http://kanji.zinbun.kyoto-u.ac.jp/~yasuoka/publications/JAET2010.pdf> (2012-11-01 閲覧)。
- 4) 津田博。自治体の情報システムにおける外字処理のあり方。生産管理。2011，vol.18，no.1，p.196-202。
- 5) 汎用電子情報交換環境整備プログラム成果報告書別冊。日本規格協会，2009。
- 6) 安岡孝一。UCS にない住民基本台帳ネットワーク統一文字。東洋学へのコンピュータ利用，第 23 回研究セミナー。2012，p.33-74。  
<http://kanji.zinbun.kyoto-u.ac.jp/~yasuoka/publications/2012-03-16.pdf> (2012-11-01 閲覧)。
- 7) 文字情報基盤 文字情報一覧表 Ver.002.02。情報処理推進機構，2012。  
<http://ossipedia.ipa.go.jp/ipamjfont/mjmojiichiran/> (2012-11-01 閲覧)。
- 8) 安岡孝一。ケータイの絵文字と文字コード。情報管理。2007，vol.50，no.2，p.67-73。  
<http://dx.doi.org/10.1241/johokanri.50.67> (2012-11-01 閲覧)。

## Author Abstract

The character code on Basic Resident Registration Network in Japan was designed as an extension of UCS (Universal Coded Character Set). And it was planned to be used as an "open" character code. But it now becomes a "Galapagos" code because of its misunderstanding of UCS. In this paper the author warns that the character code on Basic Resident Registration Network is already uncomformable to UCS and he shows the way to break away from it.

## Key words

character code, Universal Coded Character Set (UCS), Unicode, mojibake, Basic Resident Registration Network, Saiban-in (Lay Judge) System, Long-life Medical Care System