

外字について

視覚部 一般教育等 村上佳久 鍼灸学科 伊藤隆造 森 英俊

要旨：一般に、JIS X 0208-1990 情報交換用漢字符号で制定された、いわゆる第一・第二水準以外の漢字を外字と称している。視覚部では、教育用に不可欠な医学や東洋医学関係の外字を作成してきた。情報交換用の文字コードとして新たにUNICODE が制定されようとしているが、現状では問題点も非常に多い。そこで、外字についての様々な問題を解説する。

キーワード：外字、医学用語、補助漢字、東洋医学、鍼灸、JIS、UNICODE

0. はじめに

視覚部では、手書き(墨字)・手打ち(点字)のレポート類は提出不可で、ワープロなど電子メディアを経由したものしか受け取らない教育方法を確立したが、このときに最も問題となったのは、教育用に利用できない字(外字)の存在であった。

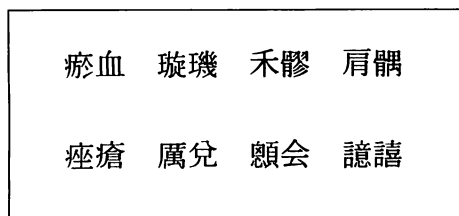


図1 医学・東洋医学で利用する外字

図1のように医学や東洋医学、鍼灸・理学療法などで利用する文字には多くの外字が存在する。また、医学史や古典などを学ぶ際にも多くの外字に遭遇する。これらは勉強の上で必要不可欠で、形が似ている漢字を無理に当てはめると、意味そのものが曖昧になるため外字の存在は不可欠である。

しかし、ワープロなどの機器には技術的な制限も多く、さらに視覚部には視覚障害補償として合成音声を用いるためにさらに技術的問題が増えることとなる。

ここでは、視覚部で教育用に利用する外字について解説し、外字の一般的な知識、さらにUNICODE(ユニコード)についても解説を行う。

1. 文字のコード化

日本語ワープロが登場したのは、1978年9月のことで、コンピュータを利用して文字を入力することが出来るようになった。同年にワープロなどのコンピュータ機器で

使用するために漢字の共通規格が制定され、よく使用される漢字として約3000文字がJIS 第一水準として、それより使用頻度の低い漢字約3400文字がJIS 第二水準として制定された。また、記号・ひらがな・カタカナなどを含めて約7000文字がJIS の規格としてコード化された。

この規格制定にあたって日本語の漢字の使用状況が調べられた。平均的に使用される文字は、一万字程度で、人名や地名などの特殊なものも考慮すると三万字程度は必要とされてきた。もちろん漢字の母国である中国では十万字以上が必要とされてきた。

しかし、それだけの文字を全て規格化してもほとんど使用されない文字が多いために、使用頻度に合わせて階級化したものである。その後、1983年、1990年、1997年と改訂され、1990年には、第一・第二水準以外の漢字として、補助漢字が約6000文字制定された。

この背景には、役所などの高度情報化により戸籍などの電算化が進み、人名の外字が必要になってきた事が挙げられる。

1997年の改訂で新しい漢字として、第三・第四水準が制定されることとなったが、全世界の文字を統一しようというUNICODE で漢字使用国の漢字を共通化する試みもあり、日本語文字コードの混乱が続いている。

2. 視覚部での対応

視覚部では、学生受入以前から教育方法開発センターで教材作成を行ってきたが、平成二年に破綻し、拡大文字の作成などについて全く考慮せず、対応出来なかったため各学科で教材作成を行うようになった。

鍼灸・理学療法の両学科で共通に利用する西洋医学関係の外字や東洋医学・鍼灸などで利用する外字を検討した。外字を利用できないと、教材や拡大教科書、レポー

ト作成などが出来ないといった問題に直面し、教育に影響を与える。

そこで両学科で平成3年度に共通外字を定めて、MS-DOS用のドットフォントとWINDOWS 3.1用のアウトラインフォント(ベクトルフォント)を作成した。

視覚部では、様々なパソコンシステムを利用しているが、全盲と弱視では利用するシステムが異なる。

全盲：MS-DOS(合成音声補償付)

軽度弱視：WINDOWS 3.1

そこで、MS-DOSとWINDOWS 3.1の両方で共通に利用できる外字が求められたが、MS-DOSとWINDOWS 3.1では利用する外字領域が異なるため、同じファイルであっても両方で共通に利用することができない。また、利用できる外字の数が異なるため、表示できない外字が利用されることになる。さらにMS-DOS上で利用する日本語入力ソフトが利用できる外字の範囲も問題となる。

2. 1 OS別の外字

1)MS-DOSの外字ファイル

MS-DOSでは、バージョン2.11より、日本語の利用とユーザ外字の作成が出来るようになったが、MS-DOSのシステムにある「ユーザー定義文字保守ユーティリティ(USKCGM)」を利用して作成される。これらは、「ドット」と呼ばれる小さい点の集合で、1文字は画面表示用では16×16ドット、プリンタ印字用では24×24ドットで構成される。

これらのドットファイルは、標準的には各々、USKCG16.SYS、USKCG24.SYSと言うファイル名が与えられ、ファイルに収録できる文字数は、最大188字で、JIS 7621-767E 7721-777Eのコードが割り当てられる。

これ以外に日本語を入力するための日本語入力ソフト(日本語FEP)の外字も作成する必要がある。視覚部では日本語入力に「ATOK」を利用しているが、平成三年当時は、技術的な問題から63文字しか外字を作成できなかった。「ATOK6,ATOK7」の外字が63文字しか利用できないからである。(JIS7621-765F)

平成九年に「ATOK9」に変更し、188文字まで外字が利用できるようになり、外字を63文字から156文字に拡張した。

2)WINDOWS 3.1の外字ファイル

WINDOWS 3.1のユーザ外字は、MS-DOSと異なり、一般的にShift-JISコードを用いる。

外字作成はシステムにある外字エディタを利用する。収録出来る最大文字数は1880字で、F040-F9FCのShift-JIS

コードが割り当てられる。

WINDOWS 3.1の場合は、標準でTrueTypeと呼ばれるベクトルフォントが搭載されており、外字もベクトルフォントを利用できる。しかし、システム標準で提供される外字作成ツールでは、40×40ドットで構成されるドットフォント(USERFONT.FON)しか作成できない。しかし、オプションソフトを用いると、TrueTypeフォント(USERFONT.TTF)も利用できる。ただし、WINDOWS 3.1では、書体に関係なく外字は1種類しか指定できない。このためにどのようなフォントを指定しても外字だけが特定のフォントとなる。例えば、視覚部では視覚障害を配慮して通常文章ではゴシック体を使用するが、外字ファイルが明朝体なら外字の部分だけが明朝体になってしまう。

また、WINDOWSでは日本語入力はシステムに依存するので、日本語入力用に外字フォントを作成する必要はない。

3)WINDOWS 95/NT3.51,NT4

WINDOWS 95/NTの場合は、システム標準でベクトルフォントが作成できる。(USERFONT.FON/USERFONT.TTF)さらに、明朝体なら明朝体の外字、ゴシック体の外字と言うように各書体に合わせた外字をそれぞれ作成可能である。(書体隷属外字)このために書体を変更しても外字が別の書体になるということはない。

4)Macintosh

Macintoshでは、外字の問題はさらに複雑である。漢字Talk6でシステム標準で外字エディタがあったが、漢字Talk7ではなくなった。さらにTrueTypeフォントとPostScriptフォントの一部で文字の互換性がなく、WINDOWSやMS-DOSとの互換性も低い。例えば、①(まる1)等は文字コードが異なる。視覚部では、ほとんどの場合、Macintoshを利用して日本語を操作することは非常に少ないので、Macintosh用の外字は作成しないこととした。

必要ならば、市販されている東洋医学関係の外字が利用できる。

5)UNIX

UNIXでは、文字コードは基本的にEUCコードを用いるが、JIS第1水準と第2水準ではEUCコードとJISコードで文字の違いはなく、ファイルは共通化が可能である。

しかし、外字に関しては、各メーカーのUNIXシステムに依存する場合が多く、機種によっては標準で外字エディタがなく外字を利用できない。しかし、UNIXでもX-

WINDOW化によって標準で補助漢字が利用できる環境が増えてきたので、これを使用する。

WINDOWSなどで教材作成が簡単に出来るようになってきたためと、視覚障害者の場合は端末がMS-DOSを利用していることを考えると、UNIX向けの外字作成はデータベース以外に考慮する必要は少ないと思われる。

MS-DOSとWINDOWS 3.1では、外字コードは全く異なる領域である。従って、ファイルの交換などで、共通性を持たせるためには、技術的な工夫が必要である。

WINDOWS 3.1に搭載されている外字エディタでは、MS-DOSが利用する領域の外字を作成することは出来ない。

そこで、MS-DOSとWINDOWS 3.1の両方が利用できる文字領域に外字が作成できる、WIFE FONTである、ALPS社のFONT WAVEを採用して、FONT WAVEの外字作成ツールで、外字を作成し、MS-DOSとWINDOWS 3.1で共通の外字が利用できるようになった。

文字数は、MS-DOSの技術的限界である188字以内で作成することとした。これにより、全盲も188字以内で多くの外字が利用できる。このALPS社のFONT WAVE書体は、別途購入する必要がある。

また、合成音声ソフトウェアに登録できる外字コードもMS-DOSに依存するために、188字以内で作成する必要があり、「一太郎」などのワープロソフトでは改行・改ページマーク等に一部の外字コードを利用しているため、その部分には外字を割り当てないようにしている。

一方、WINDOWS 95/NTは、WINDOWSと名前は付いているが、文字コードに対する考え方が異なり、内部で利用するコードはUNICODEなので、WINDOWS 3.1で作成された外字の一部がWINDOWS 95/NTで利用できないといった矛盾が生じている。(MS-DOSの外字は利用できない)

したがって、広範囲なOSをサポートすることを考えると、利用できる外字をMS-DOSの限界である188字以内として、MS-DOS用ドットフォント、WINDOWS 3.1用外字フォント、WINDOWS 95/NT外字フォントをそれぞれ作成して文字コードを共有化する仕組みを考えることとした。

現在では、WINDOWS 95/NT用の外字として、TrueTypeフォントで作成し、MS-DOSとWINDOWS 95/NT用の外字コード変換ソフトを別途開発した。

3. 外字の根拠

外字を作成する場合、文字数に制限があるので全てを作成できない。そこで、教育に必要最小限度の外字を選

別する必要がある。

外字は、論文などのデータベースにも関連し、決定には十分な調査が必要である。特にその外字を必要とする根拠を明確にしなければならない。

学会レベルでも外字の問題は重視しており、日本東洋医学会や全日本鍼灸学会などでも、外字検討小委員会などで検討している。すでに日本東洋医学会の外字は1880字分が制定され、それに基づいた製品も販売されているが、全日本鍼灸学会などでは全盲などが利用する環境はMS-DOSが多いので外字を決定できていない。

そこで、視覚部独自で作成した外字を大手パソコン通信などを通じて公開し、広く意見を求めている。

外字の典拠根拠としては、「大漢和辞典」や「康熙字典」、「大字典」などに掲載されている漢字、またJISの補助漢字にある、国家試験で使用される医学・東洋医学系外字、教科書や参考書等で利用されている外字(例えば：東洋医学概論(医道の日本社編)等6冊)を検討した。

また、JISコード変更により略字体と入れ替わった物や学会などで略字として広く流通が認められている外字も含め、可能な限り正字体を採用し、略字体の方が優勢な場合は略字体を採用した。

さらにアプリケーションなどで利用されている外字

154/168	骨	骨	骨	骨
9AA8	0-3947 0-2539	1-588C 1-5676	0-397C 0-2502	0-4D89 0-4573
154/169	飮			
9AA9	E-402F E-3216			
154/170	飮	飮		
9AAA	1-6956 1-7354	1-6966 1-7370		
154/171	飮	飮		
9AAB	3-796D 3-8977	2-486F 2-4079		
154/172	飮	飮		
9AAC	3-796C 3-8976	1-6957 1-7355		
154/173	飮	飮	飮	
9AAD	3-796B 3-8975	2-486E 2-4078	0-716C 0-8178	

図2 中台日韓漢字統合の例

(例えば改行や改頁の記号など)も考慮して、平成三年に63文字、平成九年には154文字を制定し、各OSごと、各書体ごとにフォントを作成し、利用している。

4. UNICODE

UNICODEでの漢字の考え方は、漢字使用国(日本、中国、朝鮮)の同じ文字に同じコードをつけて漢字コードを減らそうと言う考え方である。

日本、中国、韓国、台湾の四カ国で漢字コードが規格化されているが、それらを全て別コードとすると膨大な量となるため、同じ形の漢字は共通として、規格化される漢字の量そのものを減らそうとするものである。

これは、一見すると合理的な考えに思えるが問題が非常の多い発想である。

漢字使用国では、同じ形の文字でも意味が異なる場合と、同じ意味でも漢字の形が異なる場合がある。例えば「骨」では前述の四カ国で形が異なる。また、中国などでは簡略体も使用されるが、台湾では正字体が主流である。

台湾では規格化が進む以前は「BIG5」と呼ばれる約五万字が利用できる状況にあった。その後、利用頻度ごとの一万字程度に規格化され、通常約二~三万字が利用される。

この量だけでも既にUNICODEの漢字領域の枠を越えている。また、中国では北京と広東など地域によって漢字が異なっていたが、香港返還を機に北京語に統一し、新たに規格化した漢字はやはり二万語以上ある。台湾と中国でも同じ漢字でも意味が異なる場合もあり、日本や韓国などとも意味の異なる場合もある。

・アプリケーションの対応

WINDOWS 95/NT用のUNICODE対応のアプリケーションでは利用できる外字領域が、7F21-927E(JIS CODE)とIBM拡張漢字領域 9321-972C(JIS CODE)に限られる。従ってWINDOWS 3.1のようなDOSとの互換性がなくなる。同じWINDOWS 95でもソフトウェアによって利用できる外字領域が異なるという問題も発生する。したがって様々な種類の外字を作成し、その外字のコードを制定し、互換性を持たせることは、非常に難しい状況である。

例えば、「一太郎」という代表的な日本語ワープロがあるが、バージョンによって外字の利用領域は異なる。MS-DOS版のV4.3とWINDOWS 3.1版のV5、V6とWINDOWS 95/NT用のV7とUNICODE対応のV8では、利用できる外字領域は全く同じではなく、WINDOWS版のV5~V8の一部の領域の外字が共通化できるだけで、視

覚部で全盲が主として利用するV4とは互換性は全くない。

・漢字の形

「意味が同じであれば、多少形が異なっても問題はない」という考え方は、情報処理技術者の間には多く見られるが、人文科学的観点から言えば、「文字の形が異なることは意味やニュアンスが異なることであり、文章の意味そのものにも影響を与えかねないので、漢字の形そのものが正確に記録されなければならない」という事になり、学問そのものの問題に発展する。特に東洋医学などでは古典が重要視されることが多いが、この場合でも古代から受け継がれてきている文字も模写している間にその形が変化していることも十分に考えられるが、中国などで見られる古い文献や石版などに見られる文字の形を正確に記録しておくことは非常に重要である。

従って、少なくとも可能な限り、漢字の形を残すことは文化財としての意味をも含めて意義のあることである。

以上のことからUNICODEには問題が多く、日本の現状では、第一水準と第二水準を中心とした文字が必要最小限度として利用され、無理にUNICODE統一しなくとも現状のままで第三・第四水準漢字を追加していく方法が互換性の点から言っても問題が少なく、それ以外は外字対応とし、個別に対応すべきと考える。

視覚障害者のように利用できる文字に制限のある利用者もいることを忘れてはならない。

5. おわりに

技術的・歴史的意義などを考慮しない人は、UNICODEで全世界の文字を統一すべきだという考えに到達するかもしれない。一つだけの利点を強調して様々な不備な点を切り捨てる発想には同意できない。筑波技術短期大学は、視覚障害者と聴覚障害者を対象とした教育機関でありながら視覚障害者が利用できない環境を押し進めようとする人もいる。

教育と研究、技術と歴史などが互いに重なり合って環境を整備していく発想が重要と思われる。

参考事項

視覚部で作成されたファイル類は次の通り

- ・MS-DOS用システム外字：USKCG16.SYS、USKCG24.SYS
- ・一太郎V4用外字：JFGAIJ.UFO

- ・ ATOK用 外字辞書登録ファイル：ATOK.TXT
- ・ 合成音声ソフトVDM100用外字：VDM100.TXT
- ・ WINDOWS用外字
 - FontWave書体(WINDOWS 3.1用)
 - TrueType書体(WINDOWS 95/NT用)
 - JTC細角ゴシック外字
 - JTC中太角ゴシック外字
 - JTC太角ゴシック外字
 - JTC極太角ゴシック外字
 - JTC細丸ゴシック外字
 - JTC中太丸ゴシック外字
 - JTC太丸ゴシック外字
 - 平成明朝体W4外字
 - 平成ゴシック体外外字
 - セイビ楷書体外字
 - イワタ正楷書体外字 (FontWaveのみ)
 - NTC正楷書体外字 (FontWaveのみ)

- ・ 番号が一部飛んでいる理由：
 - 一太郎V4.3が外字コードの一部を使用しているため
(改行・改頁・大文字・修飾・罫線・半角罫線等)
 - 合成音声使用時の音化けを防ぐため

- ・ 整列基準：
 - 7621-765F JIS X 0212-1990の順番 非掲載文字が最後
 - 7660-7764 JIS X 0212-1990及び「大漢和辞典」の番号
順 (おおよそ)

参考文献：

- 大漢和辞典 諸橋てつ次 大修館書店
- 康熙字典
- 大字典 上田万年 他 講談社
- いま日本語が危ない 太田昌孝 丸山学芸図書

漢字が足りない 萩原基弘ほか, 医道の日本, p174-185
第625号 平成8年9月号,1996, 医道の日本社
パソコンにない漢字の取り扱いについて 北川 毅, 医道の
日本, p186-197 第625号 平成8年9月号,1996, 医道の日
本社

- JISハンドブック情報処理 用語・コード編 日本規格協
会
- JIS C 6226-1978 情報交換用漢字符系
- JIS C 6226-1983 情報交換用漢字符系
- JIS X 0208-1983 情報交換用漢字符号(1987改訂 JIS
C6226-1883より名称変更)

- JIS X 0208-1990 情報交換用漢字符号
- JIS X 0212-1990 情報交換用漢字符号ー補助漢字
- JIS X 0208:1997 7ビット及び8ビットの2バイト情報交
換用符号化漢字集合
- JIS X 0221 国際符号化文字集合(UCS)第1部体系及び基
本多言語面

The Unicode standard -Worldwide Character Encoding-
Version 1.0, Volume 2
The Unicode Consortium, Addison-Wesley Publishing
company.

インターネットとWINDOWSと日本語 ASCII, Vol.21,
#6 June 1997 p340-343.
「UNICODE」上田純美礼, SuperASCII, Oct 1995 - Mar
1996.

外字一覧表

(7621-765F)旧外字、(7660-7764)追加外字、13/July/1997 制定

上段：JISコード 下段：Shift-JISコード

(JIS) 76 21 22 23 24 25 26 27 28 29 2A 2B 2C 2D 2E 2F
 (S-JIS)F0 40 41 42 43 44 45 46 47 48 49 4A 4B 4C 4D 4E
 厲 嶮 嶮 嶮 嶮 嶮 嶮 嶮 嶮 嶮 嶮 嶮 嶮 嶮 嶮
 30 31 32 33 34 35 36 37 38 39 3A 3B 3C 3D 3E 3F
 4F 50 51 52 53 54 55 56 57 58 59 5A 5B 5C 5D 5E
 灑 炷 燾 牖 璇 璣 瘞 痧 痧 瘞 瘞 瘞 瘞 瘞 瘞
 40 41 42 43 44 45 46 47 48 49 4A 4B 4C 4D 4E 4F
 5F 60 61 62 63 64 65 66 67 68 69 6A 6B 6C 6D 6E
 癩 癩 疔 腕 膜 臉 膾 瀆 芎 芎 蔞 蠱 祛 諳 附 躄
 50 51 52 53 54 55 56 57 58 59 5A 5B 5C 5D 5E 5F
 6F 70 71 72 73 74 75 76 77 78 79 7A 7B 7C 7D 7E
 軀 邈 邈 鉞 鉞 鑣 頤 頤 飡 駝 髭 髭 髭 髭 髭
 60 61 62 63 64 65 66 67 68 69 6A 6B 6C 6D 6E 6F
 80 81 82 83 84 85 86 87 88 89 8A 8B 8C 8D 8E 8F
 必 呃 响 响 响 嘈 噓 噓 咿 啞 啞 凶 唾 豈 巩 忡
 70 71 72 73 74 75 76 77 78 79 7A 7B 7C 7D 7E
 90 91 92 93 94 95 96 97 98 99 9A 9B 9C 9D 9E
 征 忡 忡 忡 擗 擗 擗 哺 喝 欸 淖 溱 溱 溱 溱

(JIS) 77 21 22 23 24 25 26 27 28 29 2A 2B 2C 2D 2E 2F
 (S-JIS)F0 9F A0 A1 A2 A3 A4 A5 A6 A7 A8 A9 AA AB AC AD
 澣 溢 澣 炗 焯 焯 焯 痾 痾 痾 痾 痾 痾 痾 痾
 30 31 32 33 34 35 36 37 38 39 3A 3B 3C 3D 3E 3F
 AE AF B0 B1 B2 B3 B4 B5 B6 B7 B8 B9 BA BB BC BD
 痾 痾 痾 痾 眈 眈 眈 眈 眈 眈 眈 眈 眈 眈 眈
 40 41 42 43 44 45 46 47 48 49 4A 4B 4C 4D 4E 4F
 BE BF C0 C1 C2 C3 C4 C5 C6 C7 C8 C9 CA CB CC CD
 臊 臄 臄 臄 胫 胫 胫 臄 臄 臄 苕 萋 藿
 50 51 52 53 54 55 56 57 58 59 5A 5B 5C 5D 5E 5F
 CE CF D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 DA DB DC DD
 蹠 蹠 蹠 蹠 . . 銜 銜 雞 鞭 頰 頰 頰 頰 頰 頰
 60 61 62 63 64
 DE DF E0 E1 E2
 骸 骸 骸 骸 骸