

## &lt;解説&gt;

## ●文字コードと規格化

コンピュータでは、英文字、かな、漢字等の文字に「文字コード」という番号をつけ、その番号を使って情報の処理や共有を行っています。もし、このコードがシステム毎に個別につけられていたら、同じ値の文字コードであっても、異なるシステムでは、異なる文字を指すことになってしまいます。よって仮にこれらのシステム間で電子文書をやり取りした場合、いわゆる「文字化け」が起きてしまいます。

どのシステムをどこから調達しても、また、どこで作った文書をどこへ伝送しても、文字化けを起こさずに電子文書を処理できるようにするためには、文字コードは国際的に統一されている必要があります。

その実現に向けて行われたのが ISO/IEC 10646 国際符号化文字集合 Universal Coded Character Set という国際規格です（以下、これを UCS と呼ぶ）。

一般に、「ユニコード」規格と呼ばれることもある UCS には、世界の多様な文字が収録されています。日本語に関わる文字としては、日本工業規格（JIS）で定められた、JIS 第 1～第 4 水準（JIS X 0213）の漢字約 1 万文字が収録されています。

漢字や仮名文字等、日本語で用いる文字について、わが国初の標準規格が作られたのは、1978 年のことでした。このとき、約 6,000 種の漢字について文字コードが決められました。また、世界初の日本語ワードプロセッサが発売されたのはその翌年の 1979 年でした。以後、急速に日本語の電算処理が拡大します。

しかし、電算化が進むにつれ、約 6,000 種の漢字では人名や地名の表記には不足していることが問題となり、工業規格（JIS）の立場、戸籍の電算化の立場、住民台帳の電算化の立場等から各所で文字種の拡張への取り組みが行われてきました。

## ●電子政府における文字フォント等のこれまでの取り組み

1997 年改定の閣議決定「行政情報化推進基本計画」では、文字種が個別に拡張されることによる相互運用性の問題を懸念し、「外字」について、交換ルールを策定すべきとうたわれました。

さらに 2002 年、経済産業省は、調査事業「汎用電子情報交換環境整備プログラム」<sup>(1)</sup>を開始し、戸籍や住民基本台帳等で必要とされる文字の整理に取り組みました。2009 年度までに渡る 8 年間の事業でした。

2010 年度、この調査で得られた約 6 万文字の漢字を、実際に計算機で使用可能とするた

<sup>(1)</sup> 財団法人日本規格協会（現一般財団法人 日本規格協会）、独立行政法人国立国語研究所（現 大学共同利用機関法人 人間文化研究機構 国立国語研究所）、社団法人情報処理学会（現一般社団法人 情報処理学会）への委託事業として実施。

め、文字フォント等の整備を行う「文字情報基盤構築に関する研究開発事業」を IPA が経済産業省の委託事業として着手しました。これが、本日、国際規格化が完了した文字情報基盤整備事業の始まりでした。

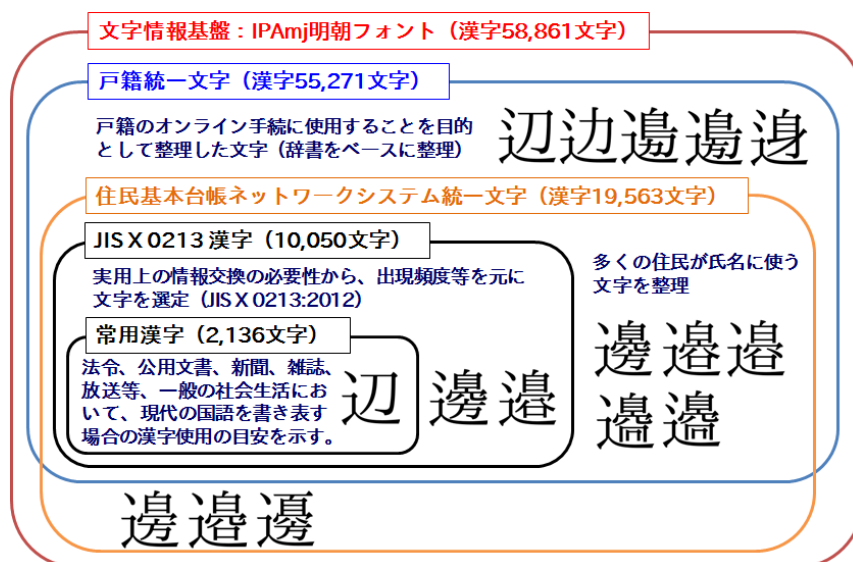
#### ●文字情報基盤 関連事業

また、IPA では、これら約 6 万文字の漢字について、文字コードや文字の関連性等の文字情報を収録した一覧表、データベース、縮退マップ等の開発を進めてきました。国際規格化に当たっては、文字の国際規格を担当する ISO/IEC JTC1 SC2<sup>(2)</sup>に対応する国内委員会である、一般社団法人 情報処理学会 情報規格調査会 SC2 専門委員会とともに、それまでの規格では扱えなかった漢字を国際規格に追加するための作業を進めてきました。

文字情報基盤で整備した漢字は、JIS 第 1 水準～第 4 水準、戸籍統一文字、住民基本台帳ネットワーク統一文字を包含し、それぞれの文字と一対一の対応関係を持つように整備されています(図参照)。

今回の ISO/IEC 10646 第 5 版において、提案した漢字の全てが収録され、文字情報基盤で整備した全ての文字と、国際規格化された文字との対応関係が確定しました。

IPA での 2010 年の事業開始から 7 年間、その基礎となった調査事業「汎用電子情報交換環境整備プログラム」を入れると、当該事業は延べ 15 年間に渡りました。



図：人名等に用いられる漢字の例

<sup>(2)</sup> 国際標準化機構 (ISO) と国際電気標準会議 (IEC) による第 1 合同技術委員会 (Joint Technical Committee 1) に設置された第 2 分科会(Subcommittee)