

縮退マップ利用ガイド

2015年3月31日

1. 縮退マップの目的

インターネットを使った情報公開や電子申請など、不特定多数の一般利用者を対象とするサービスでは、利用者の使う多様なデバイスで表示、入力ができるよう、使う文字の範囲を JIS X 0213 の範囲（約1万文字）とするなど、必要最小限に絞ることが求められます。組織内での日常的情報処理においても、使用する文字の範囲を不必要に大きくせず、業務に応じた範囲を設定することが、業務効率化の上で適切です。

人名を戸籍に合わせて正確に表記することが求められる業務等で用いられる情報処理システムと、上記のように、使用する文字の範囲を適切に絞ることが求められる業務で用いられる情報処理システムとの間の情報連携を効率化するためには、大規模な範囲の文字と、小規模な範囲の文字との対応関係（縮退対応の関係）を明らかにするとともに、この対応関係を統一的に管理し、組織内や組織間での対応関係に係る不整合を抑制することが必要です。

IPA の提供する「縮退マップ」は、戸籍統一文字と住民基本台帳ネットワークシステム統一文字を合わせて整理した、約6万文字の文字情報基盤の文字（以後、MJ 文字集合と呼ぶ）と、約1万文字の JIS X 0213（JIS 第1水準～第4水準）の文字（以後、JIS 文字集合と呼ぶ）の対応関係を見出す際の参考として使われることを目的としています。

2. 縮退マップとは

2.1. 概要

IPA では、現在、文字情報基盤で整理した文字の文字情報のデータベース化を進めており（2015年夏から試験公開、2016年春から本格公開予定）、その一環として、図形的類似・異体字・誤字・正字・本字等の文字の関係を整理しています。IPA の提供する「縮退マップ」は、このような関係性を、「JIS 文字集合への縮退」という視点から整理してとりまとめたものです。

IPA の提供する「縮退マップ」に示される MJ 文字集合→JIS 文字集合への

縮退対応関係はある文字について、

- 縮退先として1つの漢字が示される
とは限らず、
- 縮退先として複数の候補が示される
- IPA としては縮退先の候補を示さない
場合があります。

2.2. 縮退マップの成り立ち

IPA の提供する縮退マップは、MJ 文字集合に属する文字について、その関係性を、以下の様な根拠に基づいて JIS 文字集合への結びつきの有無を調査した結果をまとめたものです。

(ア) 図形的な類似性を調べ、JIS X 0213 に定義された「包摂規準」、および、UCS (ISO/IEC 10646) に定義された統合規則(Unification Rules) に照らして、JIS 文字集合に属する文字に包摂／統合されると判断されるもの。

(イ) 法務省が示した、戸籍法関連の通達：

誤字俗字・正字一覧表（平成一六年一〇月一四日民一第二八四二号民事局長通達）

正字・俗字等対照表（平成二年一〇月二〇日民二第五二〇二号通知別表）に示された関係性から、JIS 文字集合に属する文字へ結びつけられると判断されるもの。

(ウ) 辞書類¹に示された異体字等の関係性から、JIS 文字集合に属する文字へ結びつけられると判断されるもの。

(エ) 法務省の「在留カード等に係る漢字氏名の表記等に関する告示（平成23年法務省告示第582号）」により、JIS 文字集合に属する文字へ結びつけられると判断されるもの。

(オ) 辞書により、固有な意味を持つ文字であること等が確認され、JIS 文字集合に属する一文字への対応を付けることはできないと判断されるもの。

(カ) 上記(ア)～(オ)で処理できなかった文字について、読み、字形などから、

¹大漢和辞典（大修館書店）、大字源（角川書店）、新大字典（講談社）、日本語漢字辞典（新潮社）、大漢語林（大修館書店）

JIS 文字集合へ属する文字へ結びつけられると判断されるもの。

この調査の概要を図 1 縮退マップ作成の手順図 1 に示します。

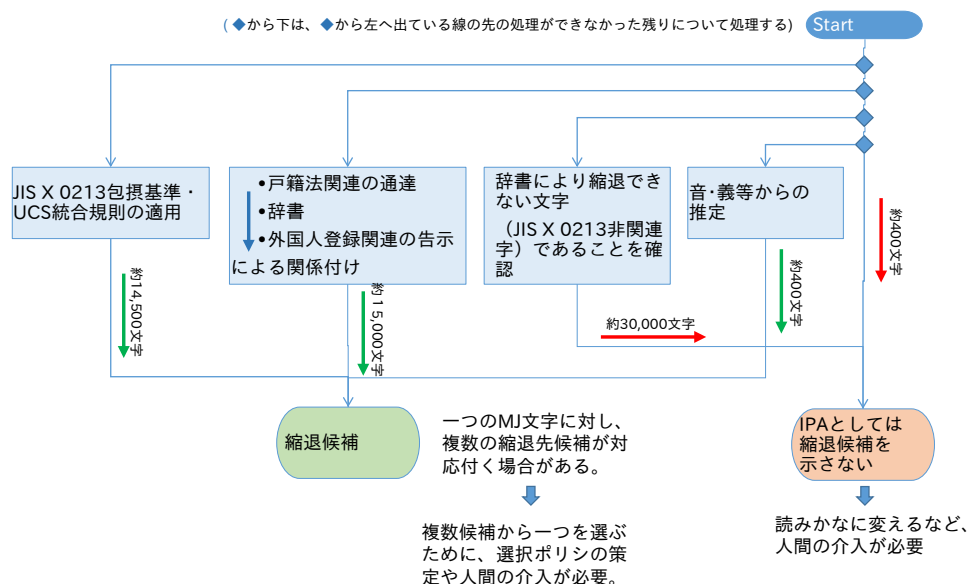


図 1 縮退マップ作成の手順

2.3. 縮退候補を複数示す場合

2.2 に示した手順により文字の関係を調べてゆくと、関連付けられた文字の複数が、JIS 文字集合に属する場合があります。さらに、法務省の告示により、複数の「正字等」のどちらへ対応付けても良いとされているものもあります。その際、候補となる複数の文字のうち、どちらの文字との結びつきがより適切であるかについて、その文字が実際に用いられる状況等に関する情報抜きで、あらかじめ客観的に判断するのが困難である場合もあります。そのような場合、IPA ではその全てを縮退の候補として示し、あえて一つだけを選ぶことは行っていません。

2.4. 縮退候補を示さない場合

(ア) 辞書の上で、固有の意味等を持つ文字であると判断され、JIS 文字集合に属する別の文字へ対応付けることは適切ではないと判断したもの。

このような文字は約 30, 000 文字ありました。その例を以下に示します。

例：𠂔→日下(くさか)、𠂔→女形(おやま)、𠂔→忌寸(いみき)

MJ057651
大漢和辞典 補231

MJ057298
大字源国字一覧 2116頁4段目6個目

MJ057347
大字源国字一覧 2116頁4段目4個目

例：𠂔→四十雀(しじゅうから)、𠂔→牛の糞(うしのあつもの)

MJ059246
大漢和辞典 補786

MJ058405
大字源国字一覧 2112頁2段目9個目

これらの約 30, 000 文字は、辞書に掲載されていたことから戸籍統一文字に収録されており、その結果 MJ 文字集合に含まれていますが、その多くは複数の文字を合成して作られた文字や、動植物名、中国古典に由来する固有名詞等、特定の意味を持つ文字であり、「日本語の表記、地名、人名などで用いられる図形文字から選定」された JIS X 0213 への対応付けができません。したがって、これらの文字の縮退先として、JIS 文字集合に属する一文字の漢字を示すことは適切では無いと判断しています。また、これらの文字が実際に日本の人名に使用されている例は極めて少ないと考えられます。

このような文字については、漢字一文字への縮退変換は行わず、読み仮名へ置き換える、あるいは、複数の漢字からなる熟語等へ置き換えるといったことが適切であると考えられます。

(イ) 辞書に見出すことができず、読み、意味等も不明な文字

このような文字は約 400 文字ありました。その例を以下に示します。

𠂔

MJ059986

𠂔

MJ059305

𠂔

MJ059350

𠂔

MJ059224

これらについては、引き続き調査を行う必要があります。

※ 現在構築中の文字情報データベースでは、こういった文字についての情報を、利用者からフィードバックしていただくための仕組みを組み込んでいます。

3. 縮退マップの活用

3.1. 活用において注意すべき点

IPA が示す縮退マップには、2.3 に示したように、ある文字について、その縮退候補が複数示されている場合があります。この場合には、縮退の変換を行う現場において、それら候補から一つを選ぶ作業が必要となります。

また、2.4 に示したように、縮退候補を示していない文字もあります。その場合には、複数の漢字からなる熟語等に置き換えることや、読み仮名へ置き換えるといった対応が必要となります。人名の読み方については、必ずしも辞書に示されている通りでない場合も多く、また、辞書に複数の読みが示されている場合にはそのどれであるかを判断する必要があり、機械的な処理は困難です。

したがって、MJ 文字集合が用いられたデータを JIS 文字集合の文字だけからなるデータへ変換する作業の場面において、あるいはその変換のためにコンピュータへセットする「変換テーブル」を作成する場面において、実際の縮退先を判断するために現場毎の人の判断が必要となることがあります。

図 2 に、縮退マップの活用イメージを示します。

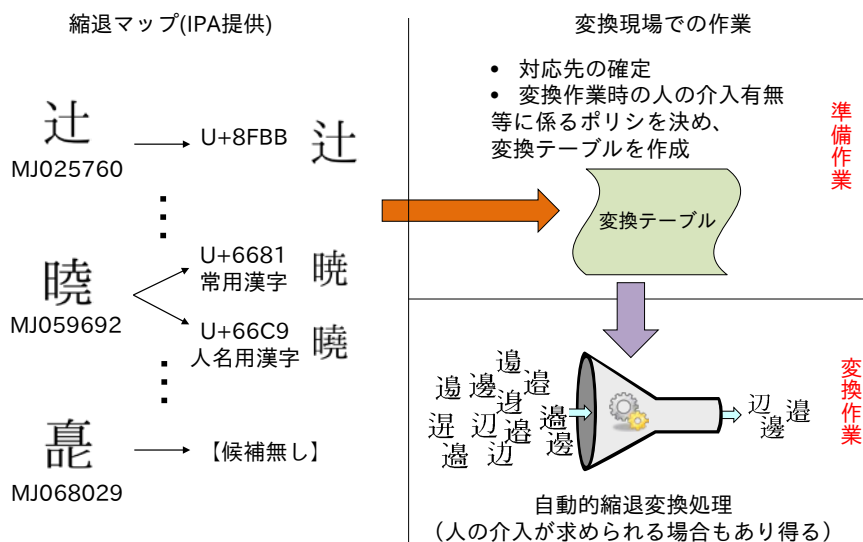


図 2 縮退マップの活用イメージ

3.2. 縮退変換処理の例

(ア)変換テーブルの準備

ここで、「変換テーブル」とは、IPA の提供する「縮退マップ」を参照し、MJ 文字集合が用いられたデータを JIS 文字集合の文字だけからなるデータへ変換する為にコンピュータにセットするためのテーブルを指します。

縮退マップに縮退先として複数の候補が示されている文字については、現場において、その実際の縮退先を選択する必要があります。

複数候補の中から一つを選ぶ方法として、「常用漢字に属するもの」「画数の小さい物」「JIS コードの最も若いもの」といった規則を決め、それに従って変換テーブルを作成することなどが考えられます、この手がかりになる情報は、IPA の提供する文字情報一覧表及び文字情報データベースから得ることができます。

縮退マップに縮退先の候補が示されていない文字については、その読み仮名や、複数漢字からなる熟語等へ変換するよう、テーブルを構成することが考えられます。

したがって、変換テーブルは、漢字一文字を別の漢字一文字に変換するものとは限らず、漢字一文字を、複数文字からなる文字列に置き換えることもあることを念頭にシステムを設計することが求められます。

また、変換不能であることを示す特殊な文字（例えば「≡」等）に置きかえるように変換テーブルを構成しておき、その処理を先送りするようにシステムを構築することも考えられます。

この場合、「≡」等に加え、オリジナルな文字の MJ 文字図形名あるいは UCS コードを、データの中に記録として残すよう、変換テーブル及びデータ構造を構成しておく、後の処理を容易にすることができます。

(イ)データ変換作業

上記(ア)で準備した変換テーブルを用い、これをデータ変換用のコンピュータへセットするなどしてデータを変換する作業を行います。

「≡」など、変換テーブルにおいて変換不能であることを示す文字に行き当たった場合、人の介入を求めることが適当である場合が考えられます。それに備えた適切なユーザインタフェースを考慮してシステムを設計しておく必要があります。

漢字一文字を複数文字の漢字または仮名の列に置き換えるように変換テーブルが構成されている場合には、変換前と変換後で、文字数が変化することを考慮の上、そのデータを用いるシステムを設計しておく必要があります。