

# 「市区町村が使用する外字の実態調査」

## 報告書

平成24年3月

総務省



# 「市区町村が使用する外字の実態調査」

## 報告書 目次

第一章 実態調査の概要.....	1
1. 1 目的.....	1
1. 2 文字情報基盤漢字.....	2
1. 3 調査対象の概要.....	3
(1) 文字数.....	3
(2) 提供外字の種類.....	3
1. 4 調査結果の概要.....	4
(1) 調査結果の全体.....	4
(2) 「字形一致」、「デザイン差」、「類似文字」の分類結果.....	6
(3) 「住基ネット統一文字」、「戸籍統一文字」に同定された外字の 分類結果.....	7
第二章 包摂基準.....	8
2. 1 基本的な考え方.....	8
(1) 字形一致.....	8
(2) デザイン差.....	8
(3) 類似文字.....	8
(4) 同定不可能文字.....	8
2. 2 文字構成要素について.....	10
(1) 文字構成要素の定義.....	10
(2) 文字構成要素の配置.....	10
(3) 文字構成要素の画数.....	11
2. 3 デザイン差の基準について.....	12
(1) デザイン差の基準.....	12
(2) デザイン差の基準の詳細.....	12
2. 4 文字同定作業と同定判断例.....	13
(1) 文字同定の作業手順.....	13
(2) 文字構成要素の配置.....	13
(3) 文字構成要素の形状・画数.....	14
(4) デザイン差の基準の範囲を超えている文字.....	15
2. 5 文字情報基盤漢字内に存在する字形一致/デザイン差の文字につ いて.....	16
2. 6 同定不可能文字について.....	17

2. 7	包摂基準に対する意見.....	18
第三章	調査結果の分析.....	20
3. 1	字形一致・デザイン差.....	20
3. 2	類似文字に分類された文字.....	20
	(1) 氏名に利用できるが、文字情報基盤漢字に存在しない新字体.....	20
	(2) 「誤字俗字・正字一覧表」の誤字及び簡体字.....	22
	(3) 「誤字俗字・正字一覧表」の俗字.....	23
3. 3	類似文字の調査.....	24
	(1) 「弓」へんの形状差異.....	24
	(2) つくりである「黄」の形状差異.....	25
	(3) つくりである「豕」（いのこ）の形状差異.....	25
	(4) 類似文字の調査の考察.....	26
3. 4	同定不可能文字に分類された文字.....	28
	(1) 不明文字の調査分析.....	28
	(2) 変体仮名の調査分析.....	30
	(3) 記号等.....	30
3. 5	調査結果に対する意見.....	31
第四章	市区町村向けの提供情報.....	37
4. 1	外字提供市区町村向けの成果物.....	37
	(1) 同定結果リスト.....	37
	(2) 同定結果CSV.....	41
4. 2	成果物の利用方法.....	42
	(1) 全ての市区町村での活用シーン.....	42
	(2) 外字提供市区町村での活用シーン.....	42
第五章	実態調査の考察.....	43
5. 1	実態調査の考察.....	43
	(1) 既存外字整理と今後の外字増加抑制.....	43
	(2) 外字整理に向けた課題.....	43
	(3) 記号等（約8万文字）の解消に向けた課題.....	45
	(4) 低解像度外字解消に向けた課題.....	46
	(5) 業務システムの移行や業務システム間連携の課題.....	46
	(6) 文字情報基盤漢字の網羅性向上に関する課題.....	47
5. 2	まとめ.....	49

# 第一章 実態調査の概要

## 1. 1 目的

総務省では、自治体クラウドの全国的展開に向けた取組を推進するとともに、住民サービスの向上のための電子自治体の確立に向けて、全省を挙げて取り組むために「自治体クラウド推進本部」を設置し、市区町村における取組状況の把握、課題の抽出・検討及び必要な助言を行うことを目的として、有識者懇談会を設置し、検討課題について具体的な施策を検討、推進してきたところであり、同懇談会での検討を踏まえ、平成23年度にクラウドサービス間の相互運用性の確保について取り組むこととしている。

コンピュータ処理上使用する文字については、各事業者のパッケージにおいて、住民基本台帳ネットワークシステムにおける文字コードに則した文字セットの他に、市区町村が独自に作成した外字を登録する領域が準備され、市区町村毎に多くの外字が設定されている。

このため、①市区町村の業務システムの移行に伴うデータ移行作業において、外字の同定作業やコードの割り当て作業が発生し、作業負担や移行コストが増加すること、②同一市区町村内でのシステム間連携においても、外字データの連携のための同定作業や文字コードの変換テーブル作成の作業が発生し、作業負担や移行コストが増加すること等の課題が指摘されている。

本調査は、市区町村が独自に作成、使用する外字の総数等の状況を把握するとともに、今後文字活用の共通基盤として広く普及することを目指して構築されている**文字情報基盤漢字**<sup>(※)</sup>との同定作業を実施し、市区町村外字の実態について取りまとめを行うものである。

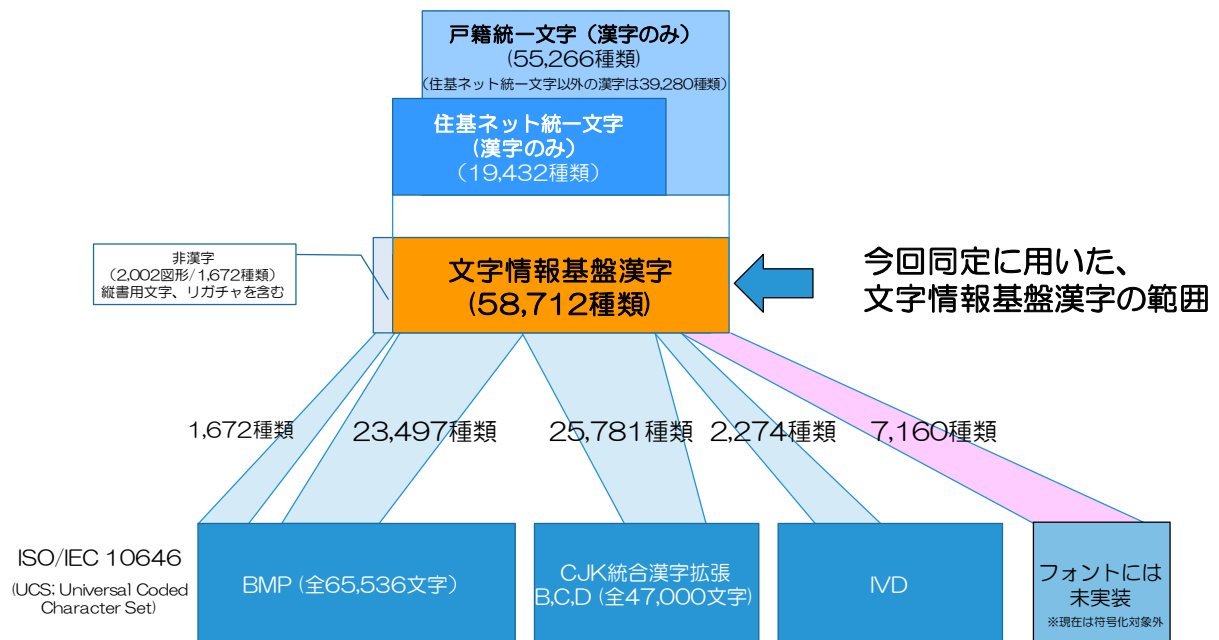
### ※ 文字情報基盤漢字

経済産業省において電子政府を推進するため、行政機関としての情報処理に必要な文字基盤を整備しており、平成22年度においては、戸籍統一文字と住民基本台帳ネットワーク統一文字の整理を目的とした文字情報基盤構築事業を行い、文字情報基盤漢字を整備している。

## 1. 2 文字情報基盤漢字

文字情報基盤漢字は、「新たな情報通信技術戦略工程表」（平成22年6月22日高度情報通信ネットワーク社会推進戦略本部決定）における文字に関する課題への取り組みの一環として、平成22年度に内閣官房(情報通信技術(IT)担当室)、総務省、法務省、経済産業省、文化庁などの関係府省や関係者が参加する文字情報基盤推進委員会を設置し、文字基盤の在り方について検討を重ね作成されたもので、戸籍統一文字や住民基本台帳ネットワークシステム統一文字（以下「住基ネット統一文字」という。）を網羅した58,712文字のフォントである。

### ◆ 文字情報基盤漢字の概要



本調査においては、平成23年10月26日に公開された正式版のフォント情報をもとに実態調査を行った。

※ 文字情報基盤のホームページ

<http://ossipedia.ipa.go.jp/ipamjfont/>

### 1. 3 調査対象の概要

総務省から全国の市区町村に対して外字情報の提出を任意に依頼し、市区町村から提供を受けた外字を対象とし調査を実施した。以下に、調査対象とした外字の詳細を報告する。

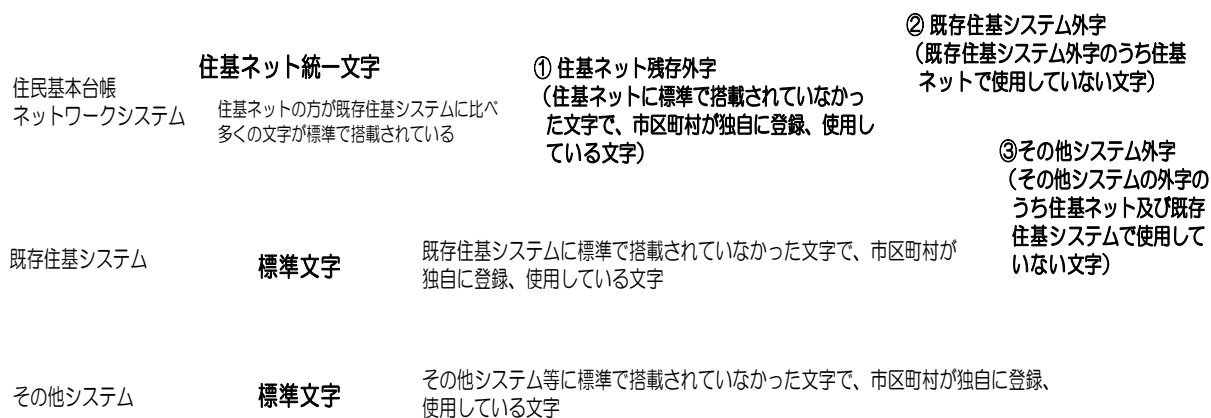
#### (1) 文字数

① 提供市区町村数	1,386 市区町村
② 総文字数	1,166,536 文字

#### (2) 提供外字の種類

① 住基ネット残存外字	1,064,660 文字
② 既存住基システム外字	69,142 文字
③ その他システム外字	32,734 文字

#### ◆ 収集した外字の種類



## 1. 4 調査結果の概要

### (1) 調査結果の全体

収集した外字 1,166,536 文字について、文字情報基盤漢字 58,712 文字に対し、包摂基準を策定し、同定した外字の総数などを以下に報告する。

同定できた外字総数 934,812 文字 (21,561 種類)

外字の実態調査結果の全体内訳を以下に示す。

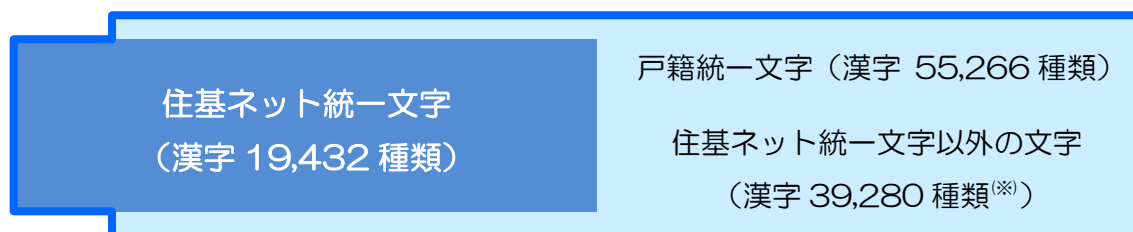
### ◆ 外字の実態調査結果 (全体)

文字種類 包摂基準	文字情報基盤漢字 58,712種類		文字情報基盤漢字に同定できなかった文字 (同定不可能文字)		
	住基ネット統一文字 19,432種類	戸籍統一文字 39,280種類 <sup>(※)</sup>	不明文字	変体仮名	記号等
字形一致 98,030文字 (12,545種類)	74,040文字 (8,856種類)	23,990文字 (3,689種類)			
デザイン差 217,313文字 (12,217種類)	166,452文字 (8,252種類)	50,861文字 (3,965種類)			
類似文字 619,469文字 (14,222種類)	579,528文字 (10,293種類)	39,941文字 (3,929種類)			
同定不可能文字 231,724文字			不明文字 52,294文字	変体仮名 97,791文字	記号等 81,639文字
	小計 820,020文字 (14,852種類)	小計 114,792文字 (6,709種類)			
総合計 1,166,536文字	合計 934,812文字 (21,561種類)		合計 52,294文字	合計 97,791文字	合計 81,639文字

※ 戸籍統一文字 39,280 種類は、戸籍統一文字の総数 55,266 種類から住基ネット統一文字に含まれる 15,986 種類を除いた数字

### ◆ 同定に用いた文字情報基盤漢字の内訳

同定に用いた文字情報基盤の文字種は、太青枠内の 58,712 種類



※ 戸籍統一文字 39,280 種類は、戸籍統一文字の総数 55,266 種類から住基ネット統一文字に含まれる 15,986 種類を除いた数字



また、同定された外字を「字形一致」、「デザイン差」、「類似文字」の3種類<sup>(※)</sup>に分類を行った結果を以下に示す。

- ① 字形一致 98,030 文字 (12,545 種類)
- ② デザイン差 217,313 文字 (12,217 種類)
- ③ 類似文字 619,469 文字 (14,222 種類、619,469 文字中、誤字又は簡体字 398,320 文字)

※ 「字形一致」、「デザイン差」、「類似文字」の定義について

- ① 字形一致  
全ての文字構成要素の配置・画数・形状が一致した文字
- ② デザイン差  
全ての文字構成要素の配置・画数が同じであるが、文字構成要素の形状の一部が、デザイン差の基準の範囲内で異なっている文字
- ③ 類似文字  
文字情報基盤漢字に類似字形が存在するが、①もしくは②の基準に当てはまらない文字

次に、同定されなかった外字（文字情報基盤漢字に類似字形が存在しない文字、変体仮名及び記号など）の総数を以下に示す。

同定不可能文字 231,724 文字

さらに、同定不可能文字 231,724 文字を、不明文字<sup>(※)</sup>、変体仮名、記号等に分類を行った結果を以下に示す。

- ① 不明文字 52,294 文字
- ② 変体仮名 97,791 文字 (94,848 文字 384 種類、2,943 文字は分類不能)
- ③ 記号等 81,639 文字

## ※ 不明文字について

漢字と思われるが文字情報基盤漢字に類似字形が存在しない文字であり、漢和辞典に掲載されている字形の文字、戸籍関連通達及び法務局通知で利用が認められた字形の文字、「誤字俗字・正字一覧表」の誤字及び簡体字の字形の文字等が含まれている。

### (2) 「字形一致」、「デザイン差」、「類似文字」の分類結果

「字形一致」、「デザイン差」、「類似文字」の分類結果を以下に示す。

#### a. 字形一致の内訳

字形一致 98,030 文字 (12,545 種類) を、住基ネット統一文字 (19,432 種類) に同定された外字と、戸籍統一文字 (住基ネット統一文字との重複 15,986 種類を除く 39,280 種類) に同定された外字に分類を行った結果を以下に示す。

- ① 住基ネット統一文字に同定された外字 74,040 文字 (8,856 種類)
- ② 戸籍統一文字に同定された外字 23,990 文字 (3,689 種類)

#### b. デザイン差の内訳

デザイン差 217,313 文字 (12,217 種類) を、住基ネット統一文字 (19,432 種類) に同定された外字と、戸籍統一文字 (住基ネット統一文字との重複 15,986 種類を除く 39,280 種類) に同定された外字に分類を行った結果を以下に示す。

- ① 住基ネット統一文字に同定された外字 166,452 文字 (8,252 種類)
- ② 戸籍統一文字に同定された外字 50,861 文字 (3,965 種類)

#### c. 類似文字の内訳

類似文字 619,469 文字 (14,222 種類) を、住基ネット統一文字 (19,432 種類) に同定された外字と、戸籍統一文字 (住基ネット統一文字との重複 15,986 種類を除く 39,280 種類) に同定された外字に分類を行った結果を以下に示す。

- ① 住基ネット統一文字に同定された外字 579,528 文字 (10,293 種類)
- ② 戸籍統一文字に同定された外字 39,941 文字 (3,929 種類)

(3) 「住基ネット統一文字」、「戸籍統一文字」に同定された外字の分類結果  
「住基ネット統一文字」、「戸籍統一文字」に同定された外字の分類結果  
を以下に示す。

a. 住基ネット統一文字に同定された外字

住基ネット統一文字 (19,432 種類) に同定された外字 820,020 文字  
(14,852 種類) を、字形一致、デザイン差、類似文字に分類を行った結果  
を以下に示す。

- ① 字形一致 74,040 文字 (8,856 種類)
- ② デザイン差 166,452 文字 (8,252 種類)
- ③ 類似文字 579,528 文字 (10,293 種類)

b. 戸籍統一文字に同定された外字

戸籍統一文字 (住基ネット統一文字との重複 15,986 種類を除く 39,280  
種類) に同定された外字 114,792 文字 (6,709 種類) を、字形一致、デザ  
イン差、類似文字に分類を行った結果を以下に示す。

- ① 字形一致 23,990 文字 (3,689 種類)
- ② デザイン差 50,861 文字 (3,965 種類)
- ③ 類似文字 39,941 文字 (3,929 種類)

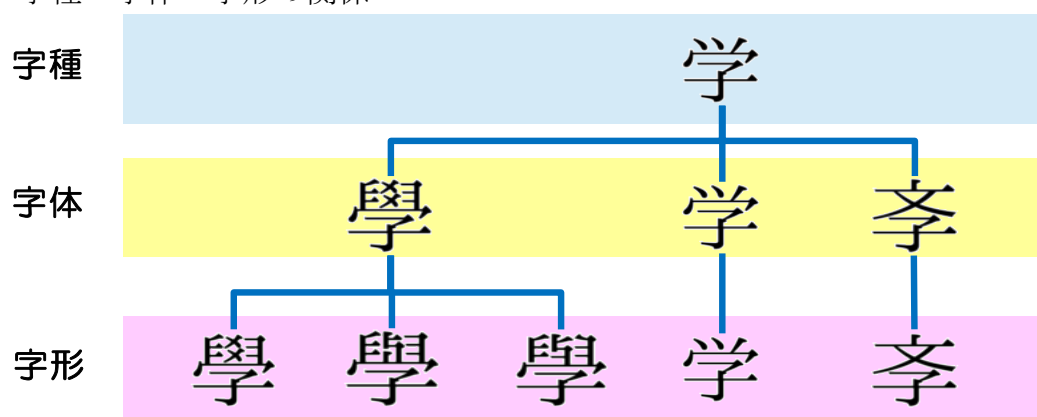
## 第二章 包摂基準

以下に包摂基準の基本的な考え方を報告する。詳細は、別添の「包摂基準書」を参照されたい。

### 2. 1 基本的な考え方

今回の同定作業においては、文字情報基盤漢字と同定しようとする文字を「字種」や「字体」のレベルではなく、「字形」レベルで同定を行っている。字種・字体・字形の関係について、以下に図示する。

#### ◆ 字種・字体・字形の関係



今回の同定作業における包摂基準を以下に示す。

- (1) 字形一致  
全ての文字構成要素の配置・画数・形状が一致した文字
- (2) デザイン差  
全ての文字構成要素の配置・画数が同じであるが、文字構成要素の形状の一部が、デザイン差の基準の範囲内で異なっている文字
- (3) 類似文字  
文字情報基盤漢字に類似字形が存在するが、(1) もしくは (2) の基準に当てはまらない文字
- (4) 同定不可能文字  
文字情報基盤漢字に類似字形が存在しない文字、変体仮名及び記号など

## ※ 分類した種別について

- a. 「字形一致」と「デザイン差」に同定された文字は、字形レベルで文字情報基盤漢字と一致している文字と判断した。
- b. 「類似文字」については、以下のように様々な文字が含まれているので、そのまま利用すべきか、正字等に置き換えて利用すべきか判断が必要である。
  - ① 誤字・・・戸籍では実際に利用されていない可能性の高い文字
  - ② 簡体字・・・住民基本台帳法の一部を改正する法律（平成 21 年法律第 77 号）附則第 3 条により多くが置き換えられ、実際に住民票では利用されなくなる可能性の高い文字
  - ③ 俗字など・・・現に戸籍で利用されている可能性が高い文字  
(戸籍電算化時にもそのまま利用できる文字)
- c. 「同定不可能文字」については、文字情報基盤漢字に含まれていないが漢和辞典等に含まれている文字、変体仮名及び記号等が含まれるので、このまま利用できるかどうかについては判断が必要である。

## ※ 定義で参照した参考文献

### ○デザイン差の定義

- ・「常用漢字表」（平成 22 年内閣告示第 2 号）（付）字体についての解説
- ・汎用電子情報交換環境整備プログラム「同定基準【原案】」2004. 10. 21 附属資料 4-15

### ○類似文字の定義

(新字体と旧字体の関係)

- ・昭和 58 年 3 月 22 日付け法務省民二第 1500 号通達
- ・平成 2 年 10 月 20 日付け法務省民二第 5200 号通達

(戸籍電算化時の文字取扱い)

- ・平成 6 年 11 月 16 日付け法務省民二第 7005 号、第 7006 号通達

(誤字・俗字の関係)

- ・平成 22 年 11 月 30 日付け法務省民一第 2905 号通達

## 2. 2 文字構成要素について

### (1) 文字構成要素の定義

文字構成要素とは、文字を構成する部品のことを指しており、構成要素の最小単位は、構成要素自身が文字と認識できる単位と定義した。

以下に、文字構成要素の分け方の例を示す。

#### ◆ 「松」の例

「松」は、大きく分けると「木」と「公」に分けられる。

さらに、「公」は、「八」と「ム」に分けられる。

「松」の文字構成要素は、「木」、「八」、「ム」と定義される。

### (2) 文字構成要素の配置

文字構成要素の配置とは、各文字構成要素が上下左右等にそれぞれ何処に配置されているかを定義した。「へん」、「かんむり」、「かまえ」、「つくり」等と類似したものと考えると理解しやすい。

以下に、文字構成要素の配置の例を示す。

#### ◆ 「松」の例

「松」の文字構成要素は、「木」、「八」、「ム」と定義される。

3つの構成要素の配置が異なると、字形だけではなく字体も変わってしまう。



「松」



「叢」

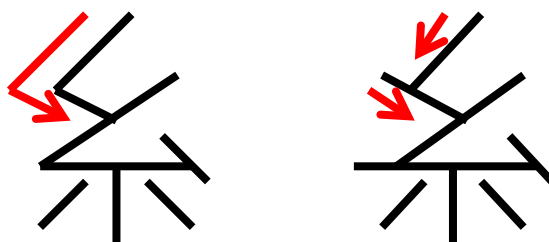
(3) 文字構成要素の画数

文字構成要素の画数とは、文字構成要素単位の画数と定義した。

以下に、文字構成要素の画数の例を示す。

◆ 「糸」の例

「糸」は6画だが、明朝体デザインで見ると1画目の曲げが2画に見える。



## 2. 3 デザイン差の基準について

以下にデザイン差の基準について考え方を示す。

### (1) デザイン差の基準

- ① 「常用漢字表（平成 22 年内閣告示第 2 号）（付）字体についての解説」と「汎用電子文字情報交換環境整備プログラムにおける文字デザインの考え方」をデザイン差の基準とし、文字構成要素の単位で形状の判断を行った。
- ② 新字体と旧字体の関係が明らかな文字については、デザイン差の基準の範囲内であっても類似文字として判断を行った。

#### ※ 新字体と旧字体とは

「国」（新字体）、「國」（旧字体）のように同じ字種であるが字体が異なる文字をさしており、昭和 24 年に「当用漢字字体表」が制定されたことにより、字体の新旧が交代し、このように表現されるようになった。

### (2) デザイン差の基準の詳細

文字構成要素の大小・高低、長短、接触非接触、接触位置、交差有無、点か棒か、傾斜方向、曲げ方折り方、曲げ方跳ね方、止め払い、止め抜き、止め跳ね、運筆方向、見かけ上の画数の違い、八屋根、筆おさえを、デザイン差の基準と設定した。以下に判断例を示す。

#### ◆ デザイン差の基準の判断例

文字構成要素の 大小・高低	長短	接触非接触	接触位置	交差有無	点か棒か	傾斜方向	曲げ方折り方
硬 ↑ 硬	雪 ↑ 雪	眞 ↑ 眞	家 ↑ 家	非 ↑ 非	帰 ↑ 帰	契 ↑ 契	頑 ↑ 頑
曲げ方跳ね方	止め払い	止め抜き	止め跳ね	運筆方向	見かけ上の画数の違い	八屋根	筆おさえ
執 ↑ 執	環 ↑ 環	耳 ↑ 耳	配 ↑ 配	考 ↑ 考	匝 ↑ 匝	八 ↑ 八	更 ↑ 更

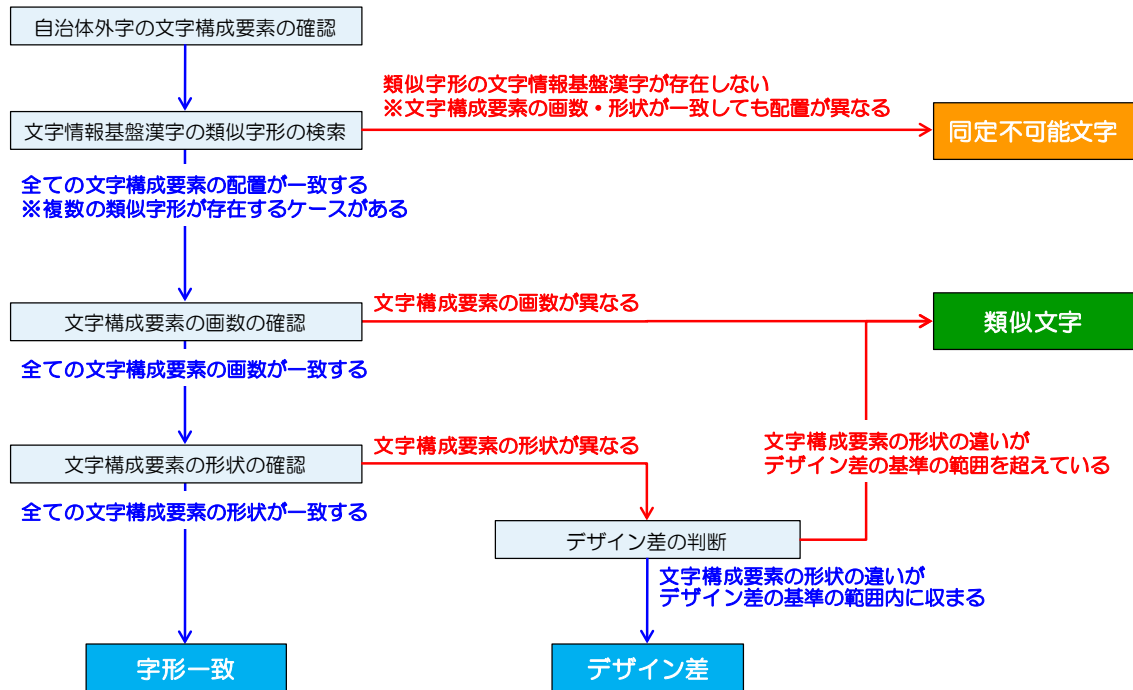


## 2. 4 文字同定作業と同定判断例

### (1) 文字同定の作業手順

包摂基準に基づき、文字同定を行った手順を以下に示す。

#### ◆ 包摂基準に基づいた文字同定の作業手順



次に、今回の同定作業における、文字同定の判断例について以下に示す。

### (2) 文字構成要素の配置

文字構成要素の画数・形状が同じであるが、配置が異なる文字は、異なる字形の文字とした。

#### ◆ 文字構成要素の配置の例

「松」と「𣎵」は、文字構成要素の画数・形状は同じであるが、配置が異なるので、同一字形ではない。




(3) 文字構成要素の形状・画数


文字構成要素の形状がデザイン差の基準の範囲内の文字は、デザイン差と判断した。また、文字構成要素の画数が異なる文字は、類似文字とした。

◆ 文字構成要素の形状・画数の例 ①

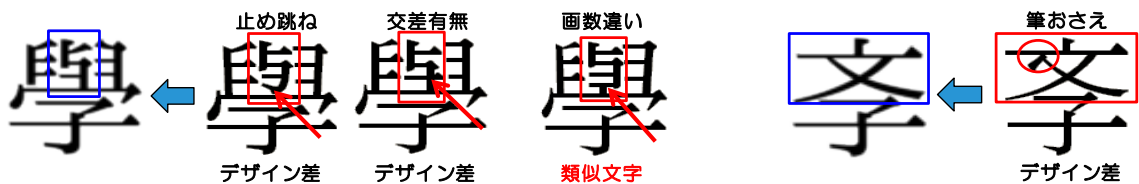
「松」は、木+八+ムから成り立っている。「木」、「八」、「ム」を文字構成要素と定義し、それぞれの要素の形状と画数を包摂基準に用いた。



例えば「八」が、 (2画) の場合は、「八屋根」のデザイン差の基準の範囲内と判断した。

「ム」が、 (3画) の場合は、画数が異なるので類似文字と判断した。

◆ 文字構成要素の形状・画数の例 ②



「学」の例では、「止め跳ね」、「交差の有無」、「画数の違い」、「筆おさえ」の判断によって、「デザイン差」、「類似文字」に分類した。

(4) デザイン差の基準の範囲を超えている文字

文字構成要素の画数と同じであるが、形状がデザイン差の基準の範囲を超えている文字は、類似文字とした。

◆ デザイン差の基準の範囲を超えている文字の例

豊 ≠ 豊 縄 ≠ 縄 萌 ≠ 萌

丰 ≠ 丰 小 ≠ 小 𠂇 ≠ 𠂇

椴 ≠ 椴

𠂇 ≠ 𠂇

※その他の候補

萌 MJ022239

萌 MJ022240

萌 MJ022241

## 2. 5 文字情報基盤漢字内に存在する字形一致/デザイン差の文字について

文字情報基盤漢字内の漢字（58,712種類）について、包摂基準を用いて同定した結果、「字形一致」、「デザイン差」と判断されるものが1,325文字存在した。これらの文字については、どの文字に対して同定をかけるかを決定するために、基準を設けて同定する文字の優先度を定義した。

文字情報基盤漢字内における「字形一致」、「デザイン差」の文字の優先順位は、以下のように設定した。

- ① Unicode 定義が存在する文字を優先する。  
双方に定義が存在する場合は、「UCS実装：○」<sup>(※)</sup>を優先順位とする。
- ② ①に当てはまらない場合は、住基ネット統一文字コードが存在する文字を優先する。  
どちらにも定義が存在する場合は、文字コードの若いものを優先する。
- ③ ①、②に当てはまらない場合は、戸籍統一文字番号の若いものを優先する。

### ※ 「UCS 実装：○」について

「UCS 実装：○」は、Unicode 領域に該当フォントが割り当てられており、文字情報基盤漢字（IPAmj 明朝フォント）をインストールした場合、IME パッド等で通常利用が可能な文字である。

### ◆ 文字情報基盤漢字内の字形一致/デザイン差の例

字形一致の例			字形一致とデザイン差の例			デザイン差の例	
MJ文字図形名	MJ007459	MJ057061	MJ000027	MJ000028	MJ000029	MJ007369	MJ07370
字形	切 切		襄 襄 襄			冴 冴	
UCS	U+5207	なし	U+342E(O)	U+342E	U+342E	U+51B4(O)	U+51B4
住基ネット統一	J+AF47	なし	J+342E	J+BB0D	J+BBOE	J+51B4	J+ADFB
戸籍統一	022080	026740	なし	なし	なし	019470	019480
	優先順位①に基づき、MJ007459を優先文字とした。		Unicodeは同じ定義だったが、UCS実装されていたため、MJ000027を優先文字とした。			Unicodeは同じ定義だったが、UCS実装されていたため、MJ007369を優先文字とした。	

## 2. 6 同定不可能文字について

同定できなかった文字については、「不明文字」、「変体仮名」、「記号等」に分類することとした。以下に、これらの文字の定義を示す。

① 不明文字

漢字と思われるが文字情報基盤漢字に類似字形が存在しない文字

② 変体仮名

変体仮名と思われる文字

③ 記号等

不明文字とも変体仮名とも捉えられない文字

## 2. 7 包摂基準に対する意見

包摂基準に対して、市区町村の担当者、市区町村情報基盤構築事業者及び文字有識者として意見を求めた。以下のとおり各意見を報告する。

- ◆ 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  
安岡 孝一 准教授（文字有識者）

包摂基準は概ね直感に合うものとなっており、文字コードとして実装する場合においても特に問題を生じないものと思われる。ただ、3つ以上の部品の位置関係が微妙に変わる場合を、この基準は必ずしもうまく書けない。

たとえば下図に示す「埠」の例では、下図右側の字形を MJ009081 に包摂したいとしても、それに関する基準を書くのは難しく、結局「類似字形」等で扱わざるを得ない。

埠 ← 埠  
MJ009081

もちろんこれは、元々の文字情報基盤漢字自体にそもそも内在する問題なのだが、この点に関する解決策を示せなかったのは残念だ。

- ◆ 東京都町田市（市区町村の担当者）

自治体クラウドの実現に向けて、外字コードも含めた文字コードの統一が必須条件と考えている。文字の統一を行うためには、文字同定の統一見解をもって文字同定を行う必要がある。今回の包摂基準をもとにした、統一的な文字同定に関する考え方をつくり、文字同定を行う強制力がはたらくような法整備や規格化が必要と考える。

戸籍の電算化に伴い誤字、類字はかなり正字化されているはず。今回の包摂基準にも一部組み込まれているように、戸籍と同一の見解をもとにした包摂基準がつくられるべきである。

UNICODE の内字にあるにもかかわらず作成した外字を、外字から除外する

必要がある。このような作業を行う上においても、今回の包摂基準はおおいに参考になる。

使われていない外字を除外しないと、外字領域を無駄にする可能性がある。包摂基準を作成するだけでなく、実際に自治体の中で使われていない外字を削減するための支援サービスがほしい。

◆ 東京都板橋区（市区町村の担当者）

実際の外字フォントで文字の判断例や基準を示してくれた事は、通常業務の際にも参考になる。今回の判断例を、国の方針として強く提示してもらえれば、住民への説明が非常に容易になる。

全体の画数に変更される場合や、部首や部品（=文字構成要素）の形が大きく変わる場合には、類似文字であっても、窓口で住民に納得してもらうのは難しい。

住民基本台帳、戸籍、外国人登録など、業務により文字の判断基準が異なるので、住民への説明が難しく、氏名（漢字）のデータ整合性が保てない要因となっている。

## 第三章 調査結果の分析

### 3. 1 字形一致・デザイン差

「字形一致」に分類された 98,030 文字（12,545 種類）と、「デザイン差」に分類された 217,313 文字（12,217 種類）で、合計 315,343 文字（17,252 種類）となったので報告する。

### 3. 2 類似文字に分類された文字

類似文字に分類された 619,469 文字（14,222 種類）を詳細に確認していくと以下のことが判明したので報告する。

#### (1) 氏名に利用できるが、文字情報基盤漢字に存在しない新字体

類似文字に分類された文字の中に、昭和 58 年 3 月 22 日付け法務省民二第 1500 号通達、平成 2 年 10 月 20 日付け法務省民二第 5200 号通達及び平成 6 年 11 月 16 日付け法務省民二第 7005 号・第 7006 号通達によって、「氏名に利用できるようになった新字体<sup>(※)</sup>」だが、文字情報基盤漢字に存在しない文字が見受けられた。

#### ※ 氏名に利用できるようになった新字体

戸籍に記載されている氏又は名の文字で、「**𠄎**」、「**𠄎**」、「**𠄎**」、「**𠄎**」、「**青**」を構成要素にもつ正字（旧字体）の当該部分がそれぞれ「**𠄎**」、「**𠄎**」、「**𠄎**」、「**𠄎**」、「**青**」と記載されている通用字体（新字体）は、そのまま記載する。

#### ◆ 第 1500 号/第 5200 号/第 7005 号・第 7006 号通達 新字体の例

①「しめすへん」は以下の例を含め 32 種類が確認された。

	しめすへん	上段：旧字体 / 下段：新字体
旧字体	𠄎	𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
新字体	𠄎	𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎



② 「くさかんむり」は以下の例を含め多数確認された。

	くさかんむり	上段：旧字体 / 下段：新字体
旧字体	𦰇	𦰇 𦰈 𦰉 𦰊 𦰋 𦰌 𦰍 𦰎
新字体	𦰇	𦰇 𦰈 𦰉 𦰊 𦰋 𦰌 𦰍 𦰎

③ 「しんにょう」は以下の例を含め多数確認された。

	しんにょう	上段：旧字体 / 下段：新字体
旧字体	𠂔	𠂔 𠂕 𠂖 𠂗 𠂘 𠂙 𠂚 𠂛
新字体	𠂔	𠂔 𠂕 𠂖 𠂗 𠂘 𠂙 𠂚 𠂛

④ 「しょくへん」は以下の例のとおり 5 種類が確認された。

	しょくへん	上段：旧字体 / 下段：新字体
旧字体	𦰑	𦰑 𦰒 𦰓 𦰔 𦰕
新字体	𦰑	𦰑 𦰒 𦰓 𦰔 𦰕

⑤ 「あお」は以下の例のとおり 1 種類が確認された。

	あお	上段：旧字体 / 下段：新字体
旧字体	𦰑	𦰑
新字体	𦰑	𦰑

※ 上記各表内の旧字体/新字体について

上段に配置している文字は IPAmj 明朝フォントの字形で、下段に配置している文字は各市区町村から提供のあった外字の字形である。

文字情報基盤漢字に不足する新字体の文字は、「くさかんむり」、「しんによう」については多数不足しているものがあるが、「しめすへん」については32種類、「しょくへん」については5種類、「あお」については1種類であることが分かった。

(2) 「誤字俗字・正字一覧表」の誤字及び簡体字

類似文字の中に、平成22年11月30日付け法務省民一第2905号通達による「誤字俗字・正字一覧表」に示された誤字及び簡体字が398,320文字も見受けられた。

① 誤字

戸籍の市区町村間の異動や電算化処理において正字に置き換えられるため、現在は使用されていない文字が多く存在すると推測される。

◆ 誤字の例

萩 ← 萩      義 ← 義      原 ← 𠮩  
※俗字

② 簡体字

住民基本台帳法の一部を改正する法律（平成21年法律第77号）附則第3条により、仮住民票作成時の漢字氏名に係る正字置換によって、その多くが解消されると予測される。（※一部日本人に利用できる文字も含まれているので注意が必要）

◆ 簡体字の例



張 ← 张      馬 ← 马      原 ← 原  
※日本人に利用できる

(3) 「誤字俗字・正字一覧表」の俗字

俗字等は戸籍電算化時にもそのまま利用できる文字であるので、戸籍や住民票に利用されている可能性が高いと推測される。

※ 俗字などの中には、新字体の漢字も含まれる。

◆ 俗字などの例

駈 ≠ 駈  勤 ≠ 勤 

### 3. 3 類似文字の調査

類似文字に該当する外字情報のうち、抽出した 2,242 文字について、以下のとおり調査を行ったので報告する。

※ ( ) 内の数は類似文字の数を示す。

(1) 「弓」へんの形状差異

「弓」へんの形状差異について、「引」MJ011151 (122 文字)、「強」MJ011196 (207 文字) の類似文字を用いて調査した結果を報告する。

「弓」へんについては、「弓」に全て統一されるべきと考えるが、「弓」へんを持つ類似文字が今回の調査で、約 1 万文字強と多数存在することが分かったので、「引」MJ011151、「強」MJ011196 の類似文字を用いて調査した。

	弓	弓	弓	弓	弓
「引」	0 字	98 字	12 字	7 字	5 字
「強」	82 字	108 字	9 字	2 字	6 字
合計	82 字	206 字	21 字	9 字	11 字

「弓」へんの形状差異は、上記の 5 種類に分けられた。  
この中で、左から 2 番目の「弓」は、206 字と数多く存在した。

「弓」がどの程度存在するかどうか、また簡体字がどの程度存在するかを調査するために、さらに「張」MJ011194 (1,599 文字) の類似文字を調査した。結果は以下のとおりである。

「張」の類似文字中、「弓」は 34 文字であった。

また、簡体字である「张」が、1,549 文字と多数存在した。

(2) つくりである「黄」の形状差異

つくりである「黄」の形状差異について、「横」MJ014455 (192 文字) の類似文字を用いて調査した結果を報告する。

つくりである「黄」については、様々な文字で利用されており形状差異も複数存在するとの認識から、「横」MJ014455 の類似文字を用いて形状差異を調査した。

つくりである「黄」の形状差異は、以下のとおり。



(3) つくりである「豕」 (いのこ) の形状差異

大きなデザイン差をもつ「豕」の形状差異について、「琢」MJ030271 (117 文字) と「琢」MJ030273 (5 文字) の類似文字を用いて、調査した結果を報告する。

まず、「琢」MJ030271 と「琢」MJ030273 の、「豕」の形状差異 2 点を以下に示す。

① 「豕」左側の点の長さ

MJ030271 は、4 画目のみにかかっている

MJ030273 は、4 画目と 5 画目の両方にかかっている



② 「豕」右側の7画目の起筆位置

MJ030271 は、1画目の中央付近から起筆している

MJ030273 は、3画目の中央右側から起筆している



MJ030271



MJ030273

また、つくりである「豕」を調査した結果、形状差異の種類は以下の10種であった。



4字

点が4画のみにかかっている



5字

点が無い



9字

点が5画のみにかかっている



28字

点が4画と5画の間にかかっている



20字

左側がタのような形状になっている



5字

点が4画と5画にかかっている



8字

点が5画のみにかかっている



2字

点が4画と5画にかかっている  
加えて右側の形状が異型



2字

点が無い  
加えて右側の形状が異型



3字

点が4画と5画の間にかかっている

(4) 類似文字の調査の考察

類似文字の調査をおこなった、「弓」へん、つくりである「黄」及び「豕」について、調査の結果から今後どのように取り扱うべきか考察をしたので、以下に記す。

① 「弓」へんについて

「弓」へんについては、調査対象とした「引」及び「強」の329文字中、206字と約63%を占めた。

非常に多くの市区町村にて作成されていることが分かったが、「弓」の形状は誤字と考えられるので、戸籍の記載は正しい形状に変更されている可能性が高いと考えられる。戸籍を確認することで、外字の解消を進めることが出来ると考える。

また、簡体字である「張」については、多数存在することが判明したが、この字形は住民基本台帳法の一部を改正する法律（平成21年法律第77号）附則第3条により「張」や「𦉳」に置き換えられるので、今後解消されていくと考える。

② つくりである「黄」について

つくりである「黄」の形状差異は、多岐に渡ることが分かった。

調査した「横」については、戸籍電算化時には「横」が利用できる文字とされているので、戸籍の記載は正しい形状に変更されている可能性が高いと考えられる。戸籍を確認することで、外字の解消を進めることが出来ると考える。

また、新規に外字作成する場合においても、類似する文字を外字として作成するのではなく「横」の文字に包摂して利用されるべき<sup>(※)</sup>と考える。

※ 戸籍では、類似字形の「横」MJ014537についても利用できる文字とされているので、注意されたい。

③ つくりである「豕」について

つくりである「豕」の形状も複数の種類が確認されたが、今回の調査で確認された複数の形状差は誤字等とみられ、実際の戸籍では戸籍電算化によって「豕」または「豕」に置き換えられている可能性が高いと考えられる。戸籍を確認することで、外字の解消を進めることが出来ると考える。

### 3. 4 同定不可能文字に分類された文字

文字情報基盤漢字と同定できなかった同定不可能文字(231,724文字)を、「不明文字」、「変体仮名」、「記号等」の3つに分類した。そのうち、「不明文字」については、さらに調査分析を行ったので、結果を報告する。

#### (1) 不明文字の調査分析

不明文字(52,294文字)については、以下の観点で調査分析を行った。

#### ◆ 不明文字の分析結果

分類	分類の詳細	外字数	文字種類
①	漢和辞典に掲載されている文字	康熙字典	147文字 46種類
		新大字典	98文字 56種類
		その他の漢和辞典	555文字 92種類
②	戸籍関連通達及び法務局通知の文字	829文字	105種類
③	「誤字俗字・正字一覧表」の誤字	5,692文字	2,037種類
④	簡体字	入国管理局で利用される簡体字	2,145文字 795種類
		その他の簡体字	9,069文字 726種類
—	分類不可能だったもの	33,759文字	—
合計		52,294文字	3,857種類

#### ① 漢和辞典に掲載されている文字

戸籍の氏名に利用することが認められている文字である。市区町村で利用されている代表的な漢和辞典である「康熙字典」、「新大字典」については明確に掲載が確認できたため、それぞれ数字を示すこととした。

#### ② 戸籍関連通達及び法務局通知の文字

戸籍関連通達及び法務局通知で、氏名に利用することが認められている文字であるため、数字を示すこととした。

##### ※ 俗字等に関する通達

- ・平成6年11月16日付け法務省民二第7006号通達
- ・平成22年11月30日付け法務省民一第2905号通達

#### ③ 「誤字俗字・正字一覧表<sup>(※)</sup>」の誤字

「誤字俗字・正字一覧表」上で誤字とされている文字で、戸籍電算化等により正字に置き換えられている可能性が高い文字であるため、数字を示すこととした。



※ 「誤字俗字・正字一覧表」

「誤字俗字・正字一覧表」は、法務省から通達として各市区町村に配布されている一覧表である。

a	イ	ア	正字等
	為	威 安 嵐 愛	戸籍に記載されている文字
b	為 <sup>▲</sup>	威 <sup>▲</sup> 女 嵐 <sup>▲</sup> 威 <sup>▲</sup> 嵐	
			正字等
	域	(違) 違 倅 (爲)	戸籍に記載されている文字
	域	違 <sup>▲</sup> 倅 <sup>▲</sup> 爲 <sup>▲</sup> 倅 <sup>▲</sup> 倅 <sup>▲</sup>	

- 一 常用漢字に関するもの
- 1 常用漢字表の配列順に配列した。
  - 2 括弧が添えられていないものは、通用字体である。
  - 3 (一)内の字体は、規則別表第(四)に掲げる字体である。
  - 4 (一)内の字体は、常用漢字表において括弧が添えられている漢字のうち、子の名に用いることのできない字体である。
  - 5 (一)内の字体は、従前は子の名に用いることができた字体であるが、昭和五十六年十一月以降は子の名に用いることのできない字体である。
  - 6 (一)内の字体は、2から5までに含まれない漢字で康熙字典又は康熙字典で正字とされている字体である。
  - 7 ●の付されている字体は、別字(同字、古字又は小字を含む)であるが、誤記される例が多いので、申出がある場合には、訂正を認めるものである。
  - 8 (一)内の字体は、五(一)〇〇〇連通別表に掲げる字体である。
  - 9 ▲の付されている字体は、康熙字典に俗字として登録されている文字であり、申出がある場合には、訂正を認めるものである。

- a. 「正字等」欄の文字は、戸籍に記載できる文字が配置されている。
- b. 「戸籍に記載されている文字」欄の誤字は、戸籍に利用できない文字であり、直上の「正字等」の欄の正字等に置き換えて戸籍に記載することと、法務省より指導されている文字である。

④ 簡体字


簡体字は、住民基本台帳法の一部を改正する法律(平成21年法律第77号)附則第3条により、外国人が住民票に搭載される際に解消される文字と考えられるが、調査時点において外国人登録に利用されている文字が確認できたため、数字を示すこととした。

「入国管理局で利用される簡体字」とは、現行の外国人登録法下で利用されている、いわゆる外国人登録事務における登録証明書調製事務用特殊文字を指す。これら文字以外の簡体字については、「その他簡体字」に分類した。

## (2) 変体仮名の調査分析

変体仮名（97,791 文字）について、明確な基準の設定は行っていないが、大まかな字形の類似性をもって分類を行ったので、合計数と大まかな種類数を報告する。97,791 文字中、94,848 文字は 384 種類に分類でき、2,943 文字は分類不能であった。

### ◆ 変体仮名の例

「あ」の例 

「い」の例 

## (3) 記号等

記号等（81,639 文字）については、合計数のみを報告する。

### ◆ 記号等の例



### 3. 5 調査結果に対する意見

調査結果に対して、市区町村の担当者、市区町村情報基盤構築事業者及び文字有識者として意見を求めた。以下のとおり各意見を報告する。

#### ◆ 東京都板橋区（市区町村の担当者）

住民基本台帳、戸籍、外国人登録など、業務により文字の判断基準が異なるので、住民への説明が難しく、氏名（漢字）のデータ整合性が保てない要因となっている。そのため、管轄省庁や制度の壁を超えて、住民窓口での統一的な文字運用基準を国が示して欲しい。

行政用に使用できる文字（フォントや範囲）の標準化を図り、省庁の中で統一化して、自治体に提示して欲しい。運用上でグレーゾーンが残る事で、ベンダー固有の字形（フォント）や不要な外字作成が生まれ、結果的にシステム更改時のベンダーロックやデータ移行費用を高くしている。

文字の運用に関する国の将来像が見えないため、システムのオープン化や自治体クラウドを推進するための実施計画が立てづらい状況になっている。

#### ◆ 東京都葛飾区（市区町村の担当者）

「転出前市区町村では使えた」などの理由で、外字を作成しなければならないケースが多い。文字について、戸籍については以前より通知があるが、住民基本台帳にも強制力のある統一文字などの通達が欲しい。

端末のセットアップで、各システムベンダーの外字フォントファイルが競合することがある。JIS78 や JIS90、JIS2004 など字形の相違があるフォントを使用しているため、JIS の新規格ができる度に外字対応しなければならない。今後、自治体がシステムで使用するフォントセットを指定するなど、上位機関から通達してもらうことはできないのか。

システムベンダー毎に外字を作成しているため、外字作成の費用がかさむ。各システムが連携しているため、各システムベンダーが外字を作成し、各シ

システムへ配布が終わるまで時間がかかる。(3ヶ月以上かかることもある)

今後、自治体間のデータ連携なども始まるため、全行政機関で使用すべき標準フォントセットを指定できないのか。

今回の文字調査では、各市区町村で大変な文字数を作成していることがわかったため、今後は戸籍電算化のように、住基についても文字を統一するような施策を実施していただきたい。

◆ 東京都町田市（市区町村の担当者）

戸籍の電算化に伴い誤字、類字はかなり正字化されているはず。戸籍の未電算化自治体の電算化支援を進めることにより、より外字を縮小できるのではないかと。また、戸籍の電算化はかなり進んでいるため、戸籍と住基の個人マッチングを実施すると、実用外字が見えるのではないかと。

また、UNICODEの内字にあるにもかかわらず、外字を作成している場合がある。今回の調査は、文字情報基盤文字との同定であったが、UNICODEの内字(特にCK領域)に対しても同定を行ってほしい。

自治体クラウドの実現に向けて、文字コードの統一が必須条件となる。戸籍システムは、国の拘束力が強く、業務内容が固定化されているため、他の業務システムと比較し、クラウド化の素地が整っているのではないかと。また、戸籍システムとあわせて住基システムをクラウド化すると、外字統一のクリティカルパスになるのではないかと。低コストのクラウド化のためには外字を排除し、マイナンバーとカナ文字、マイナンバーと画像文字等の組み合わせにより、クラウドサービスを構築する方法もある。

自治体の外字を単に合計しただけでは、有用な外字数は不明。住基ネット統一文字にあるにもかかわらず、外字を作成している場合がある。今回の調査の様な文字環境の見える化が重要で、これを進めるためには自治体の文字同定を支援するサービスが必要。

- ◆ 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  
安岡 孝一 准教授（文字有識者）

今回の実態調査は、あくまで文字情報基盤漢字のうち、現在 IPAmj 明朝に実装されている漢字に対して同定をおこなった。これは、もちろん当初方針としては悪くないものだったが、結果から言えば、IPAmj 明朝に実装されているのかにかかわらず、UCS に直接同定をおこなった方が良かったのではないかと個人的には思える。

たとえば、下図左の U+25313 や下図左の U+4E67 は、今回の実態調査において、かなり多くの自治体において外字登録されているのだが、IPAmj 明朝には実装されていない。ただし、いずれも UCS は割り振られており、フォントさえ準備すればすぐにでも使える状態である。一方、下図右下（「谷」の異体字）や下図右下（「邊」の異体字）は、残念ながら IPAmj 明朝にも実装されておらず、UCS も割り振られていない。このような漢字は、UCS を割り振ってもらうべく明日にでも ISO に国際提案しなければ、フォントに実装することすらできず、いつまでたっても現実に使えるようにならない。

曙      𠄎                      谷                      邊  
U+25313      U+4E67                      （「谷」の異体字）      （「邊」の異体字）

特に、今後の自治体クラウドにおけるオープン化という局面においては、このような「日本で現在使われているのに、UCS が割り振られていない字」というのが、実現のネックとなる。早急に UCS への収録を国際提案すべきである。

- ◆ 株式会社エヌ・ティ・ティ・データ（市区町村情報基盤構築事業者）

地方自治体の情報システム間の連携において、各自治体が独自に定義している外字の存在が円滑な情報連携の阻害要因となっている現状がある。システム間での文字情報交換を円滑にするためには共通の文字情報基盤を利用することが望ましく、その点で現在整備が進められている文字情報基盤はその共通基盤の候補となりうるものである。

今回の事業では自治体外字と文字情報基盤漢字との同定を行っているので、来るべき自治体クラウドの実現を見据えて、各自治体において本事業の成果を活用し、独自外字を削減する取り組みが必要であると考えます。

a. 既存外字の削減に向けた取り組み

本事業において自治体が独自に作成した外字と文字情報基盤漢字との同定を行っており、その中で文字情報基盤漢字に含まれる漢字を外字として作成しているケースが多数確認された。これらの文字は情報システムの文字基盤として文字情報基盤漢字を採用することで内字という扱いになるため、独自に外字化する必要がなくなる。また今回の調査では住基ネット統一文字に含まれている漢字を外字として作成しているケースも確認されている。これは本来住基ネット統一文字を採用したシステムであれば外字化する必要がないものであるため、文字基盤を変更することなく内字で置き換えることが可能である。したがって、本事業成果の外字同定結果を活用することで、既存外字を文字情報基盤漢字の内字で置き換えて利用することが可能であると考えます。

一方、文字情報基盤漢字と同定できなかった文字も多数存在している。これらの文字を利用する必要がある場合には依然として外字を作成する必要がある。しかし、そのような文字は利用頻度が低い誤字であることも考えられるため、その文字の利用頻度を調査した上で本当に外字化してまで使用する必要があるのか判断する必要があると考えます。このため、文字情報基盤漢字と同定できなかった外字については

- ・自治体横串での利用頻度を集計し、多くの自治体で広く利用されている文字であれば文字情報基盤漢字への追加を検討する。
- ・利用頻度の低い文字であれば、その文字が現在でもまだ利用されているのか調査を行い、すでに利用されていないのであれば削除する。

といった対応が考えられる。

なお、情報システムの文字基盤を変更したり利用可能な文字を変更・削除したりすることは稼働中システムの既存データの変更が必要となり、影響が非常に大きいため、実施に当たっては慎重に影響範囲を見極めることが必要となる。このため稼働中システムに対して変更を加えるのではなく、システムの更改や移行のタイミングで文字基盤の見直しや外字の整理を行うことが現実的な対応であると考えます。

b. 新規外字作成の抑止に向けた取り組み

情報システムで現在利用されている外字を削減するのはもちろん、それとは別に今後新たな外字が発生することを抑止するための取り組みも必要である。

そのためにはまず外字作成の起点となる自治体の窓口における、文字運用のポリシーを整備する必要があると考える。この文字運用ポリシーの策定に当たっては本事業の成果である文字の包摂基準書をベースにし、自治体の窓口において無闇に外字を作成しないようなルール作りを行う必要があると考える。

◆ IPA 独立行政法人 情報処理推進機構（文字情報基盤推進者）

文字情報基盤の確立は、地方自治体、官公庁、学校等における人名・地名を扱うシステムの構築と運用の効率化には急務といえます。本調査事業において、膨大な量の外字を統一的な視点で短期間に評価されたことは、極めて有意義な成果と言えます。

これまで、住基システム、戸籍システム等で自治体が人名・地名に使用している外字については、数の多さに対する指摘はあるものの、その実態が明確にされていませんでした。本調査事業の結果、「見える化」が図られ、これにより課題が浮き彫りになりました。この課題は国としても継続的に取り組むべき課題であると考えられます。

本調査事業では、文字情報基盤漢字との同定を行っており、ここで同定の指針としてまとめられたデザイン差の認定方法など、大変妥当なものだと思われ、包摂基準や統合規則とは別に、類似文字の概念を導入したことは、今後の類似文字検索の新しい方法への道を開く可能性を示唆しており、高く評価します。

また、文字情報基盤漢字との対応が取られないところに注目が集まりがちですが、文字情報基盤漢字と対応が取れた文字は、今後の文字図形情報の可

搬性を高める上で、各自治体にとっても、貴重な資産となるといえます。

IPA では、今後、ISO/IEC 10646 に未登録である漢字の新規符号化提案や IVD 登録を進めて行く計画ですが、このための国際提案手続き等に、本事業の成果は大変役に立つものと考えています。文字情報基盤の推進のため、国際提案活動と共に MJ 文字情報一覧表の整備、運用ガイドの整備、各種情報提供等を引き続き進めて参ります。



## 第四章 市区町村向けの提供情報

本調査で取りまとめた情報を、外字を提供頂いた市区町村向けの成果物として取りまとめた。調査結果は、同定に関する解釈の違いに関する意見も多々生じてくると思うが、本調査で用いた包摂基準とともに確認頂いた上で、「4. 2 成果物の利用方法」の中にも記述した利用例を参考にして頂き活用されたい。

### 4. 1 外字提供市区町村向けの成果物

#### (1) 同定結果リスト

市区町村から提供頂いた外字の文字同定結果を掲載したリストである。リストの種類は、以下の5種類である。

##### ① 全体結果

市区町村の外字と文字情報基盤漢字を同定した全ての結果である。

ファイル名は、「市区町村コード」+「市区町村名」+「システム区分」(「該当外字数」/「総外字数」文字)全体結果 となっている。

##### ② 字形一致

市区町村の外字と文字情報基盤漢字を同定した結果のうち、字形一致に分類された外字のみのリストである。

ファイル名は、「市区町村コード」+「市区町村名」+「システム区分」(「該当外字数」/「総外字数」文字)字形一致 となっている。

##### ③ デザイン差

市区町村の外字と文字情報基盤漢字を同定した結果のうち、デザイン差に分類された外字のみのリストである。

ファイル名は、「市区町村コード」+「市区町村名」+「システム区分」(「該当外字数」/「総外字数」文字)デザイン差 となっている。

④ 類似文字

市区町村の外字と文字情報基盤漢字を同定した結果のうち、類似文字に分類された外字のみのリストである。

ファイル名は、「市区町村コード」+「市区町村名」+「システム区分」(「該当外字数」／「総外字数」文字)類似文字 となっている。

⑤ 同定不可能文字

市区町村から提供頂いた外字の文字同定の結果、文字情報基盤漢字内に類似字形が存在しない文字、変体仮名及び記号などを掲載したリストである。

ファイル名は、「市区町村コード」+「市区町村名」+「システム区分」(「該当外字数」／「総外字数」文字)同定不可能文字 となっている。

同定結果リストのレイアウト及び項目について以下に示す。

<レイアウト>

(866/866文字)全体結果			(866/866文字)全体結果		
外字	MJ文字図形名	補足	外字	MJ文字図形名	補足
_1_E0DC	MJ021691	デザイン差	_1_E0E6		
𪛗	𪛗	住基: 戸籍:344780 UCS:U+82EB UCS実装:	𪛗		
_1_E0DD	MJ006366	デザイン差	_1_E0E7	MJ025036	類似文字
𪛘	𪛘	住基: 戸籍:001200 UCS:U+4E3D UCS実装:	𪛘	𪛘	住基:J+8CC0 戸籍:416510 UCS:U+8CC0 UCS実装:○
_1_E0DE			_1_E0E8		
𪛙			𪛙		
_1_E0DF	MJ009126	類似文字	_1_E0E9	MJ009401	類似文字
𪛚	𪛚	住基:J+5805 戸籍:059450 UCS:U+5805 UCS実装:○	𪛚	𪛚	住基:J+58FD 戸籍:064950 UCS:U+58FD UCS実装:○
_1_E0E0	MJ022632	デザイン差	_1_E0EA	MJ027290	類似文字
𪛛	𪛛	住基: 戸籍:358940 UCS:U+84FC UCS実装:	鐵	鐵	住基:J+9435 戸籍:464820 UCS:U+9435 UCS実装:○

_1_E0EA	MJ027290	類似文字
鐵	鐵	住基:J+9435 戸籍:464820 UCS:U+9435 UCS実装:○

## <項目の説明>

a : 外字識別番号

「市区町村コード」 + 「システム区分」 + 「外字コード」

b : 市区町村から提供された外字の字形

c : 文字情報基盤漢字の MJ 文字図形名

d : 文字情報基盤漢字の字形

e : 同定区分(字形一致/デザイン差/類似文字/表示なし)

ア. 字形一致

全ての文字構成要素の配置・画数・形状が一致した文字

イ. デザイン差

全ての文字構成要素の配置・画数が同じであるが、文字構成要素の形状の一部が、デザイン差の基準の範囲内で異なっている文字

ウ. 類似文字

文字情報基盤漢字に類似字形が存在するが、アもしくはイの基準に当てはまらない文字

エ. 表示なし(同定不可能文字)

文字情報基盤漢字に類似字形が存在しない文字、変体仮名及び記号など

f : 文字情報基盤漢字の文字属性情報

・住基 : 該当する住基ネット統一文字コード

・戸籍 : 該当する戸籍統一文字番号

・UCS : 文字情報基盤漢字に対応する Unicode (UTF16)

・UCS実装<sup>(※)</sup> : 文字情報基盤漢字のUCS実装の有無(○/表示なし)

※ 「UCS 実装 : ○」は Unicode 領域に該当フォントが割り当てられており、文字情報基盤漢字 (IPAmj 明朝フォント) をインストールした場合、IME パッド等で通常利用が可能な文字である。

(2) 同定結果 CSV

市区町村の外字と文字情報基盤漢字を同定した全ての結果を掲載した CSV ファイルである。CSV ファイルは、各 PDF リストに対応している。

CSV ファイルのレイアウト及び項目について以下に示す。

<レイアウト>

	A	B	C	D	E	F	G
1	外字	MJ文字図形名	同定区分名	住基ネット統一文字コード	戸籍統一文字番号	UCS	UCS実装
2	*****_1_E000	MJ002878	字形一致		263700	U+405C	○
3	*****_1_E001	MJ023039	デザイン差	J+C0F8	366010	U+85CD	
4	*****_1_E002	MJ022690	デザイン差		359520	U+851A	
5	*****_1_E003	MJ015127	字形一致	J+6C84	194260	U+6C84	○
6	*****_1_E004	MJ060101	デザイン差	J+B961			
7	*****_1_E005	MJ000006	デザイン差	J+3402	000690	U+3402	○
8	*****_1_E006	MJ020594	類似文字	J+7F8E	317600	U+7F8E	○
9	*****_1_E007	MJ010236	字形一致	J+5BEC	083110	U+5BEC	○
10	*****_1_E008	MJ022283	字形一致		353550	U+843D	
11	*****_1_E009	MJ022445	デザイン差		356280	U+8494	
12	*****_1_E00A	MJ023046	デザイン差		366070	U+85CF	
13	*****_1_E00B	MJ022888	類似文字		365270	U+8584	
14	*****_1_E00C	MJ026086	類似文字	J+9060	442580	U+9060	○
15	*****_1_E00D	MJ022339	デザイン差	J+C0F7	355990	U+845B	
16	*****_1_E00E	MJ025036	類似文字	J+8CC0	416510	U+8CC0	○
17	*****_1_E00F	MJ017263	字形一致		235690	U+740B	○
18	*****_1_E010	MJ013157	字形一致		154440	U+65EA	○
19	*****_1_E011		同定不可能				
20	*****_1_E012	MJ011873	類似文字	J+6176	126770	U+6176	○
21	*****_1_E013	MJ023069	デザイン差		367110	U+85DD	
22	*****_1_E014	MJ011949	類似文字	J+61B2	127190	U+61B2	○
23	*****_1_E015	MJ034907	字形一致		098200	U+21FEB	

<項目の説明>

A 列：外字

外字識別番号(市区町村コード+システム区分+外字コード)

B 列：MJ 文字図形名

文字情報基盤漢字の MJ 文字図形名

C 列：同定区分名

同定区分(字形一致/デザイン差/類似文字/同定不可能)

D 列：住基ネット統一文字コード

E 列：戸籍統一文字番号

F 列：UCS

文字情報基盤漢字に対応する Unicode (UTF16)

G 列：UCS 実装

文字情報基盤漢字の UCS 実装の有無(○/表示なし)

## 4. 2 成果物の利用方法

以下に成果物の活用シーンを一覧で記す。具体的な利用方法については、別添「外字実態調査の活用方法」を参照されたい。

### (1) 全ての市区町村での活用シーン

- ① 文字運用ポリシーの策定に活用
- ② 文字情報基盤漢字とサイトの活用

### (2) 外字提供市区町村での活用シーン

- ① 外字の整理に活用
- ② 文字環境の見える化に活用
- ③ システム間連携やシステム移行時の基準文字に活用

## 第五章 実態調査の考察

調査を踏まえて判明したことを通して、外字を整理し、自治体クラウドの全国的展開に向けた文字の課題を解決し、これらの運用負荷を低減していくために考察される事項と今後の課題を記す。

### 5. 1 実態調査の考察

#### (1) 既存外字整理と今後の外字増加抑制

今回、全国から集められた約 116 万文字の同定及び整理を行ったが、包摂基準と同定作業の手順については、文字構成要素を判断基準として位置づけたことにより、様々なデザインポリシーによって作成された文字を正確に判断し分類するための基準として、一定の成果を得ることができたと考える。

この包摂基準と同定作業の手順をもとに、各市区町村にて「文字運用ポリシー」を策定し、文字環境の見える化を行いながら運用することで、今後外字を作成する基準が定義され、不要な外字を作成することがなくなり、現状保有する外字の整理を進めることができると考える。

各市区町村で「文字運用ポリシー」を策定する際には、本調査に用いた包摂基準と同定作業手順等を参考にして頂きたい。

#### (2) 外字整理に向けた課題

##### ① 住基ネット統一文字に一致する外字（約 24 万文字）

住民基本台帳ネットワークシステムへのデータ移行では、住基ネット統一文字と既存住基システム内の現住民氏名と現住所に利用されている文字を同定し、あふれた文字のみを住基ネット残存外字として登録するとしているが、大多数の市区町村では、外字整理のリスクや移行コストの関係で、既存住基システムで利用されている外字をそのまま住基ネット残存外字として登録しているものと推測される。

住基ネット統一文字に一致した外字（字形一致及びデザイン差）は削除し、本来の住基ネット統一文字コードに変更し整理することが望まれる。整理する際には、データの変更を伴うので十分な確認が必要と考えるが、これらの外字を全て解消できると全国で約 24 万文字の外

字解消につながるので、外字を提供頂いた市区町村において、今後整理を進めて頂きたい。

② 戸籍統一文字に一致する外字（約 7 万文字）

戸籍統一文字に一致した外字（字形一致及びデザイン差）は、現に戸籍で利用されている文字であるが住基ネット統一文字に存在しないために外字を作成したと推測される。

整理には十分に戸籍を確認し注意して行う必要があるが、文字情報基盤漢字を利用する前提で考えると、全国で約 7 万文字の外字解消につながるので、今後外字を提供頂いた市区町村において検討して頂きたい。

③ 類似文字中の誤字及び簡体字（約 40 万文字）

類似文字として判定した漢字約 62 万文字中で、戸籍電算化や住民基本台帳法の一部を改正する法律（平成 21 年法律第 77 号）附則第 3 条により、正字などに置き換えられる誤字及び簡体字が約 40 万文字と多数存在した。

誤字については、平成 6 年 11 月 16 日付け法務省民二第 7005 号、第 7006 号通達等により、戸籍電算化の進展により多くの戸籍氏名に記載されていた誤字は既に解消されていると考えられる。

しかし、住民票の最新氏名の記載は戸籍氏名と同様に更正されているが、過去の履歴内に外字が残っているケース、本籍が住所地以外にあることから住民票の氏名更正そのものが行われていないケース、又は戸籍そのものが未電算となっており氏名に誤字が残っているケースなどにより、住基ネット残存外字に誤字が残っていると推測する。

よって、ここで発生している文字の多くは、戸籍電算化の進展と電算化戸籍の氏名に合わせた文字へ住民票の氏名が更正され、さらに保存期間経過やシステム更新によって住民票が改製されることにより、多くが解消されるであろうと考察する。



また簡体字については、住民基本台帳ネットワークシステムに搭載されているデータは日本人のみであるが、既存住基システムでは外国人との混合世帯の表記などのために外国人氏名が搭載されているケースや、既存住基システムの宛名などにも利用されているケースがあることから、これらで利用されている外字がそのまま住基ネット残存外字へ移行されたと推測される。

住民基本台帳法の一部を改正する法律（平成 21 年法律第 77 号）により、外国人住民が住民票に記載されるにあたり、外国人の氏名について住基ネット統一文字に包摂して利用することになるので、これを機に各市区町村において、簡体字について整理・解消されることを望む。

これらの誤字及び簡体字の整理が完了すると、全国で約 40 万文字もの外字解消につながるので、外字を提供頂いた市区町村において、今後の対応を検討頂き、整理を進めて頂きたい。

外字整理の中で、仮に文字情報基盤漢字を採用したとすると、住基ネット統一文字に一致する外字約 24 万文字の整理、戸籍統一文字に一致する外字約 7 万文字の整理並びに誤字及び簡体字に該当する約 40 万文字の整理が完了し、全国から集められた約 116 万文字の約 61%にあたる、合計約 71 万文字の外字を解消することができると思う。

### （3）記号等（約 8 万文字）の解消に向けた課題

記号などの中には、既存住基システム等で利用する証明書の印影を分割した図形や、表組み用の罫線などが確認できたが、これらについては外字として利用するのではなく、例えば、表であれば帳票ツールの罫線機能で作成、グラフであればグラフィイメージ処理用のライブラリを用いてプログラミングするなど、その他のコンピュータ処理技術を用いて解決することができるので検討されることを望む。

また、①、②などの記号も見受けられたが、これらは JIS 規格にも登録されているので外字として作成するのではなく内字の利用を第一優先とし、

行政事務において多用されている、郵便カスタマバーコードや○付き文字については、文字情報基盤へ追加の検討や、自治体クラウド内の環境下においても引き続き外字として管理すべきか検討が必要である。

#### (4) 低解像度外字解消に向けた課題

市区町村から提出された外字を一覧化する作業を行った中で、残存外字フォントについて、デザインポリシーやデザイン解像度が市区町村毎に違いがあることはもちろんのこと、同一市区町村の中でもフォントの作成時期等により大きく変動している例が見受けられた。外字の作成については、非常に労度のかかる作業であり、市区町村の方々の大変なご苦勞が改めて感じられた。

今後も新規での外字フォント作成や既存外字フォントの見直しを進められていくことと思われるが、本調査で利用した IPAmj 明朝フォントに存在する文字であれば、そのまま外字エディタ等にコピーをして利用することができる。低解像度の外字が解消されることにより、証明書における文字トラブルや、窓口における確認間違い等の住民とのトラブルが軽減されるだけでなく、外字作成の労度も軽減されるので、別添「外字実態調査の活用方法」の参照をお勧めする。

#### (5) 業務システムの移行や業務システム間連携の課題

現在多くの市区町村では、業務システム毎の文字コードや外字の割り当てが異なることによって、市区町村の業務システムの移行に伴うデータ移行作業において、外字の同定作業やコードの割り当て作業が発生し、作業負担や移行コストの増加が発生している。また、同一市区町村内でのシステム間連携においても、外字データの連携のための同定作業や文字コードの変換テーブル作成の作業が発生し、作業負担や移行コストが増加していること等の課題が指摘されている。

本調査の成果物である同定結果リスト及び CSV には、文字情報基盤漢字の MJ 文字図形名及び UCS を掲載しているので、これらの情報を用いて各種業務システムにおける文字環境の見える化を行い、各種業務システムの文

字を文字情報基盤漢字に関連付けすることにより、文字情報基盤漢字をシステム間連携やシステム移行の標準文字として活用することができる。

文字情報基盤漢字と各種業務システム文字コードを関連付けすることで、システム間連携の標準文字コードとして文字情報基盤漢字コードを利用することができるようになり、システム間連携構築の際のベンダーロックを解消することができる。

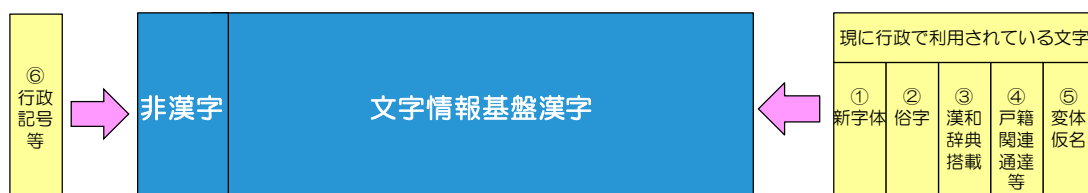
また、既存業務システムの文字コードと文字情報基盤漢字の対応付けを行った上で、別途総務省がまとめたデータ移行用の中間標準レイアウトと合わせて活用することで、システム移行時に固有文字コードや固有移行レイアウト等によるベンダーロックを解消することが期待できるので、各市町村において活用願いたい。

#### (6) 文字情報基盤漢字の網羅性向上に関する課題

文字情報基盤漢字同定リスト及び不明文字リストを確認していくと、下記の①～⑤のように、現状の氏名に利用されていると考えられる文字でありながら、文字情報基盤漢字に搭載されていない文字が存在することが分かった。

全国から集められた約 116 万文字の約 39%にあたる約 46 万文字（類似文字中の誤字及び簡体字を除く約 22 万文字、不明文字約 6 万文字、変体仮名約 10 万文字、記号約 8 万文字）の外字をさらに分類、分析し、文字情報基盤漢字に追加拡充することにより、外字の発生しない文字情報基盤を実現できると考える。

#### ◆ 文字情報基盤漢字の拡充



##### ① 新字体

「3. 2 (1) 氏名に利用できるが文字情報基盤漢字に存在しない新字体」を参照

##### ② 俗字

「3. 2 (2) 『誤字俗字・正字一覧表』の誤字及び簡体字」及び「3. 2 (3) 『誤字俗字・正字一覧表』の俗字」を参照

- ③ 漢和辞典搭載、④ 戸籍関連通達等  
「3. 4 (1) 不明文字の調査分析-①漢和辞典に掲載されている文字」及び「3. 4 (1) 不明文字の調査分析-②戸籍関連通達及び法務局通知の文字」を参照
- ⑤ 変体仮名  
「3. 4 (2) 変体仮名の調査分析」を参照
- ⑥ 行政記号等  
「3. 4 (3) 記号等」及び「5. 1 (3) 記号等 (約8万文字) の解消に向けた課題」を参照

これらの文字をそのまま文字情報基盤漢字に追加していくべきかどうかについては、各市区町村での利用状況等や、戸籍や住民票の氏名に利用できるかどうかについて様々な辞典や文献を確認した上でなければ結論を出すことはできないが、「行政で利用する文字を全て網羅し、業務システム運用上で外字の発生がないような文字情報基盤」を実現するために、次回以降の文字情報基盤漢字への文字追加検討の際には、今回の調査結果を参考にして頂きたい。

## 5. 2 まとめ

今回の調査では、1,386市区町村という多くの市区町村のご協力のもとに、本報告に記したような大きな成果を得られたところである。今後、各市区町村が自治体クラウドへの円滑な移行を目指すためには、移行に際して重要な要素の一つである文字課題を解決するために、各市区町村にて、既存外字の見える化及び整理を行い、大幅に外字を削減した上で、新たな外字を増やさないために文字運用ポリシーを作成して運用することが望まれる。

また、今回の調査では、文字情報基盤漢字のあるべき姿である「行政で利用する文字を全て網羅し、業務システム運用上で外字の発生がないような文字情報基盤」とするための課題も判明したので、今後、文字情報基盤の整理・拡充において本調査結果を活用されることが望まれる。

仮に、文字情報基盤漢字の整理・拡充が進み、行政に利用できる全ての文字を正確に運用するための基盤として完成されたならば、自治体クラウドへの円滑な移行が可能になるばかりでなく、自治体クラウドを運用する上における外字の管理という大きな課題の一つを解決することができると思う。

さらに、この整備された文字情報基盤漢字を、自治体クラウドにおいて利用されるであろう、①電子申請システム、②教育・健康・医療クラウド、③マイナンバーにて結合される様々な業務システムとそれを活用した各種ポータル、④国民一人一人が利用するパソコンやスマートデバイス、⑤各企業が運用する各種業務システムにまで普及することができたと仮定するならば、官から民まで全てのコンピューティング環境が、国際整合性のとれた統一された文字コード及び文字字形で運用され、非常に運用性の高い、国民本位の電子行政が実現されると考える。

ここまで到達するためには、本調査の結果を踏まえ、戸籍を含めたさらなる文字環境の調査をもとにした文字情報基盤漢字の整理・拡充や、Unicode上の多くの文字を取り扱うための文字利用技術の進展が必要であるが、コンピュータ技術の急速な進展や自治体業務システムにおけるクラウド化という電子自治体の流れから考えれば、技術的にも制度的にも実現可能な範囲であると考えられる。

最後に、本調査を契機として、さらなる自治体クラウドの実現に向けた研究が進められるとともに、自治体クラウドの全国的展開に向けた文字に関する技術的、制度的な課題の解決が進展し、自治体クラウドへの円滑な移行が実現することを切に望むところである。