



研究レポート

No.400 February 2013

電子行政における外字問題の解決に向けて
— 人間とコンピュータの関係から外字問題を考える —

主席研究員 榎並 利博

要旨

電子政府の推進において、外字問題は長年の懸案事項であったが、2010年度に実施された経済産業省の文字情報基盤構築プロジェクトおよび2011年度の総務省の外字の実態調査によって、その実態が明らかとなり、解決の糸口が見え始めたといえる。

しかし、解決の方向性として次のような問題がある。

- ・ 文字情報基盤(住基ネット統一文字と戸籍統一文字)を抛りどころに外字を整理し UTF-16 と IVS/IVD で対応しても、市町村に残存するそれ以外の外字(約 37,000 字)に対応できない。
- ・ 外字の存在は、日本社会全体で 30 億円くらいの経済損失をもたらしており、健全な情報化社会の発展を阻害し、日本社会全体のパフォーマンスを低下させている。文字を増やすという解決方法では、次のような理由から問題の解決には結びつかない。
 - 日本国民は、JIS 第 1 水準・第 2 水準の範囲の漢字でさえ使いこなせず、住基ネット統一文字と戸籍統一文字を含む約 60,000 字を使いこなすことは不可能である。
 - 住基ネット統一文字と戸籍統一文字を含む約 60,000 字の範囲では、文字の認識速度が低下するばかりか、誤認識が多数発生する。特に高齢者になるほど速度低下と誤認識は顕著であり、文字を増やすことは高齢社会において適さない。

このため、問題解決の方法として、コンピュータで扱う文字を増やして対応する方法ではなく、法的に外字の使用を制限する方法を提案する。

(案 1) 行政手続きで使用する漢字(氏名や地名など)を JIS 第 1 水準と JIS 第 2 水準に制限し、それ以外の漢字の使用を法律で禁止する。現状でそれ以外の漢字を使っている場合は、JIS 第 1 水準と JIS 第 2 水準の範囲内の類似した漢字に置き換え、置き換え不可能な漢字については「ひらがな」または「かたかな」に置き換えることとする。なお、氏名漢字の置き換えについては、本人の同意を得ることが望ましいが、不可能である場合は職権によって行うこととする。

(案 2) 氏名の漢字にアイデンティティとしてのこだわりを持つ人を考慮し、案 1 の適用について、戸籍を除外する。戸籍と住民基本台帳で氏名漢字が異なる場合が生じるが、個人の同一性については戸籍に住民票コード(またはマイナンバー)を記載することで確保する。

キーワード：電子政府、電子自治体、文字コード、外字、住基ネット統一文字、戸籍統一文字

目次

1. はじめに	
1. 1 本研究の問題意識…「外字」とは	1
1. 2 人名用漢字における漢字制限問題	4
2. 研究の枠組み	
2. 1 研究のプロセス	6
2. 2 本研究で扱う概念	7
3. 日本語文字コードの開発	
3. 1 日本語のコンピュータ化	8
3. 2 戸籍と住民基本台帳ネットワーク	10
4. 政府の問題認識とその対応	
4. 1 文字情報基盤構築プロジェクト	14
4. 2 総務省の外字調査について	16
4. 3 政府の解決方策の問題点	19
5. コストの視点	
5. 1 行政の実務とコスト	21
5. 2 民間の実務とコスト	23
5. 3 外字のコスト負担の問題	24
6. 文字認識の視点	
6. 1 標準的な漢字の認識能力について	26
6. 2 行政手続きで使われる漢字の認識能力について	27
6. 3 人間と機械のインタフェース	34
7. 結論	
7. 1 問題の整理	35
7. 2 問題解決のための提案1	37
7. 3 問題解決のための提案2	37
8. おわりに	38
【追補1】	40
【追補2】	41
参考文献	42
付録図表 実験票 A と実験票 B	43

1. はじめに

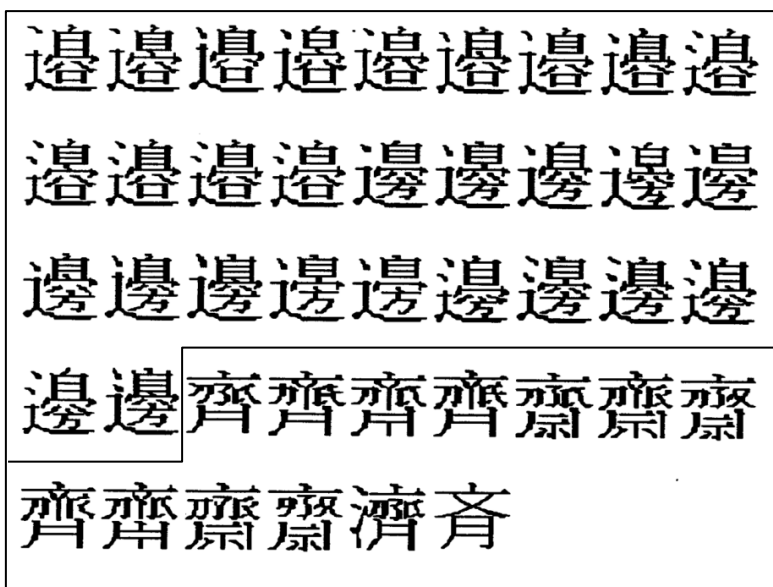
1. 1 本研究の問題意識・・・「外字」とは

「外字」という言葉は、外字新聞というように外国語の文字のことを指したり、常用漢字や人名用漢字ではない活字のことを指したり、最近ではコンピュータで扱えない文字のことを指したりと様々な意味を持っている。本研究で扱う「外字」とは、最後に示したコンピュータで扱うことができない文字という意味で使っている。

もう少し正確に定義するなら、「あるコンピュータが扱う特定の文字集合に含まれない文字」ということになるが、この表現から想定されるように、使う人によって外字のイメージが異なる。例えば、通常のパソコンの利用者から見れば、JIS 第 1 水準と第 2 水準以外の漢字が外字となる。しかし、メインフレームを使っている利用者から見れば、JIS 第 1 水準と第 2 水準およびベンダー提供漢字(富士通の場合は「拡張漢字」と呼んでいる)以外の漢字が外字となる。また、Unicode を組み込んでいるシステムであれば、その利用者はまた異なった範囲の外字のイメージを持つだろう。

本研究で扱う「外字」の定義は後述するとして、ここでは外字のイメージだけを捉えてもらうことにする。図表 1 に、ある自治体で管理している外字の例として渡辺さんの「なべ」と斉藤さんの「さい」の外字を示した。

図表 1 ある自治体で管理している外字の一例



(出所：榎並利博(2000))

日本国民、自治体の職員、そして民間企業の社員が、日常的な行政手続きをするために本当にこのような外字を使わなくては行けないのかというのが、そもそもの筆者の問題意

識である。通常見慣れた漢字をある文字セットに組み込むか否かという問題ではなく、ほとんどの人が見たこともないような、そして簡単に判別できないような漢字を厳密に識別し、使う必要があるのかという疑問である。今から 30 年以上前、筆者が自治体の現場で住民基本台帳システムの開発と導入を行っていたときからの疑問であり、その疑問はいまだに解消していない。

当時のコンピュータはメインフレームが主流であり、自治体など各行政機関は一つのコンピュータで庁内の事務処理を一手にまかっていた。それゆえ外字の問題も、その行政機関内部に閉じた問題であった。しかし、インターネットによるネットワーク化が進展し、ダウンサイジングによるコンピュータの分散化が広がってきたため、外部および内部のコンピュータ間のデータ交換が頻繁に行われるようになり、外字の問題は単に組織内の問題ではなく、社会全体の大きな問題となってきた。2000 年に出版した拙著『自治体の IT 革命』¹では、図表 1 を提示してその問題解決を訴えたのである。

日本政府は 2000 年に IT 基本法を制定し、翌年には e-Japan 戦略を発表して、世界最先端の IT 国家を目指すことを表明した。文字の問題も標準化の問題として取上げられ、ここで問題の解決が図られることが期待された。しかし、一向に問題解決の糸口は見えなかった。そのためもあり、電子政府分野における IT の利活用は思うように進まず、国連の電子政府ランキングでは 2008 年に 11 位にランクづけされたものの、2010 年は 17 位、2011 年には 18 位とランクを下げている。

一方、韓国は 2010、2011 年ともに 1 位にランクされており、日本と韓国の差がますます拡大しつつある。その日韓の比較を試みた研究²においては、さまざまな角度からその違いが論じられている。その両者の違いを原因・結果の循環図(図表 2)で表すと、韓国が正の循環になっているのに対し、日本は負の循環に陥っていることが示されている。特に、韓国が「縦割りの弊害を克服」→「重点的な戦略と IT ガバナンスの樹立」→「BPR の実施」という循環になっているのに対し、日本は「縦割り組織の未克服」→「網羅的な戦略と IT ガバナンスの欠如」→「BPR の不在」となっている。日本では、IT を活用して国全体を最適化するためには、IT が機能するよう縦割り組織の壁を打破し、IT ガバナンスを確立して BPR を実施しなければならないという認識に欠けていた。すなわち IT が組織横断的に機能するよう、縦割りになっている法律や制度を見直さなければならないというシステム的な考え方について、政府内で共通認識ができていなかったからではないかと思われる。韓国では 1960 年代に統一番号としての住民登録番号が整備され、文字も表音文字であるハングル³で統一され、戸籍制度は 2007 年の「家族関係の登録等に関する法律」によって廃止されるなど、IT が機能するよう法律や制度が見直され、基盤整備がなされている。

主要な基盤は 2 つあると思われるが、その 1 つである番号制度についてはようやく 2012

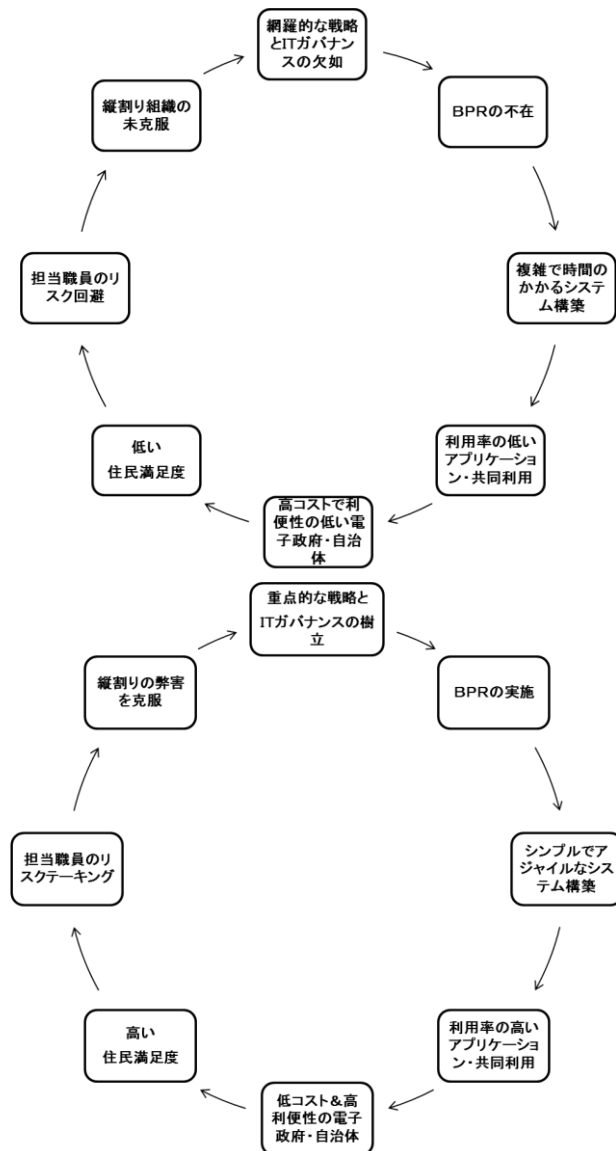
1 榎並利博(2000)

2 島田達巳・榎並利博(2010)

3 理論上は 1 万種類を超えるが、実際に使用されるのはその半分以下。

年にマイナンバー法案が国会に提出され、本人確認のための制度的な基盤構築が動き出すとともに、政府 CIO も設置された。そして、もう 1 つの重要な基盤である文字コードについても、文字情報基盤構築に向けた動きが出てきた。

図表 2 日本の負の循環（上）と韓国の正の循環（下）



(出所：島田達巳・榎並利博(2010))

しかし、文字情報基盤構築プロジェクトの資料⁴を精査してみると、コンピュータで扱う文字数を増やして対応しようという極めて技術的な見地から検討しているだけで、数万と言われる漢字をすべて扱うべきなのかという社会的文化的な問題に真正面から取り組んでいるようには思えない。筆者の考えでは、このような対応方法では日本社会全体のパフォ

⁴ 巻末参考サイト一覧の文字情報基盤構築関連のサイトを参照

パフォーマンスを低下させ続け、問題の解決にはつながらないのではないかと思う。

これまで人類の社会や文化は、技術によって大きく変わってきたという歴史がある。今、コンピュータという技術によって日本の社会や文化が変わるべきであるなら、社会や文化を然るべき方向へと変えていかなくてはならない。具体的には、日本社会全体のパフォーマンスを高める方向に変えるためには、コンピュータで扱う文字を増やすことは非合理的な選択であることを、人間と機械のインタフェースという観点、日本が直面する超高齢社会・グローバル社会という観点から示していきたい。

なお、本研究を進めるにあたって、経済産業省の文字情報基盤プロジェクトの報告書⁵や総務省の外字実態調査報告書⁶が公表されたことにより、外字に関する定量的な姿が明確になり、具体的な議論が可能となった。その意味で両省の調査プロジェクトの意義は大きいと考えている。また、政府 CIO が昨年設置されたことも、本研究の動機付けとなった。マイナンバーだけでなく、文字コード問題も省庁縦割りの弊害が大きく、この問題解決に政府 CIO の役割が大いに期待されるからだ。

なお、誤解のないよう補足しておくが、本研究で問題としているのは一般の人々が日常的な社会生活において文字を使う場面であって、古代の言語を扱う学術的な分野や文学的な表現を行う分野における文字使用を対象としているわけではない。

1. 2 人名用漢字における漢字制限問題

本研究で問題としている外字はほとんどが人名用漢字（および地名漢字）であるが、戦後新聞報道などで議論されてきた人名用漢字における漢字制限問題とは異なる。誤解を招かないよう、ここで若干説明を加えておきたい。

『人名用漢字の戦後史』⁷によれば、戦後煩雑で難しい漢字は一部エリートが使う軍国主義的象徴として排斥される傾向があり、戦後の民主化のイデオロギーによって一般大衆でも使えるよう漢字を制限する、あるいは廃止するような方向へ向かったという。そして、1946年に「当面用いる漢字」として、当用漢字表が制定されることになる。この漢字表は、日本語における漢字の使用を原則として1,850字に制限するという画期的なものであった。

しかし、子どもの名前に付ける漢字としては不足していると、名前の漢字制限に対する批判が起きてくる。そこで、政府は1951年に人名用漢字別表(92字)を制定し、さらに1976年には人名用漢字追加表(28字)を制定した。この時点で、全部で1,970字の漢字が定められた。

そして、戦後のイデオロギー論争も過去のものとなり、当用漢字表が抜本的に見直しされることになる。1981年には「常に用いる漢字」として、常用漢字表(1,945字)が制定

⁵ 経済産業省・独立行政法人情報処理推進機構(IPA)(2011)

⁶ 富士ゼロックス株式会社(2012)

⁷ 円満字二郎(2005)

されることとなった。このとき人名用漢字は、別表第二（人名用漢字）として 166 字（その内訳は、従来の 120 字+追加 54 字-常用漢字へ昇格 8 字）が定められ、制定された漢字は全部で 2,111 字となる。

その後、個人の自由・個性の時代という風潮のなかで、人名用漢字の制限に対してメディアが問題視し、人名用漢字はさらに増え続けることになる。1989 年には人名用漢字が 118 字追加されて 284 字になり、さらに直近の 2004 年には人名用漢字の第 5 次改定が行われ、983 字にまで膨れ上がる。この時点で、常用漢字と合わせて全部で 2,928 字が定められたのである。

なお、2010 年 11 月 30 日に最新の常用漢字表が内閣告示（平成 22 年内閣告示第 2 号）された。その前書きには、「この表は、法令、公用文書、新聞、雑誌、放送など、一般の社会生活において、現代の国語を書き表す場合の漢字使用の目安を示すものである」と記されており、字種は 2,136 字が指定されている。また、法令、公用文書、新聞、雑誌、放送等、一般の社会生活において、常用漢字表にない漢字を使用する場合の「字体選択のよりどころ」となることを目指して作成されているものが、「表外漢字字体表」である。これは国語審議会によって作成され、最新のものは 2000 年 12 月 8 日に文部大臣に答申したものであり、印刷標準字体（1,022 字）と簡易慣用字体（22 字）が示されている。

すなわち、常用漢字表、人名漢字表、表外漢字字体表を合わせても、一般の社会生活において使われる漢字の字種はせいぜい 4,000 字（表外漢字と人名用漢字でダブリがある）である。つまり、新聞等メディアで報じられた人名用漢字における漢字制限問題とは、文字数として JIS 第 1 水準と第 2 水準（約 6,000 字）の範囲内の問題なのである。

本研究で問題としているのは、JIS 第 1 水準と第 2 水準（約 6,000 字）の範囲内の問題ではなく、JIS 第 1 水準と第 2 水準を超えて数万字の漢字を扱うことの是非、図表 1 のような漢字を扱うことを問題としているのである。

2. 研究の枠組み

本研究のプロセスおよび扱う概念について明確にしておく。

2. 1 研究のプロセス

①日本語文字コードの開発経緯と氏名漢字に関する問題提起

日本において、日本語をコンピュータ上で表現するためにどのような取り組みを行ってきたかを概観し、現状の文字コード体系について整理する。

そして、行政手続きにおいてどのような文字コード問題が発生しているのか、政府はどのような解決策を見出そうとしているのか、2010年度に実施された経済産業省のプロジェクト報告書「文字情報基盤構築に関する研究開発事業調査報告書」⁸と2011年度の総務省の調査事業報告書「市区町村が使用する外字の実態調査報告書（平成23年度総務省請負調査）」⁹を手がかりに検証していく。

これらの政府の解決策の方向性について、欠落している視点を指摘し、その視点の重要性について次項以降で分析をしていく。

②コスト（国民の負担）の視点

欠落している一つめの視点として、国民の負担についてコストの視点から捉えていく。異なる複数の文字コード体系あるいは外字を扱わざるを得ないことによって、行政の現場ではどれくらいの経済的な損失が生じているのか、また民間企業においてはどれくらいの経済的な損失が生じているのかについて試算する。これらの試算については、ヒアリング調査を実施したうえで、全国規模の推計を行っている。また、ヒアリング調査において明らかとなったコスト以外の負担についても取上げる。

これらのコストは、行政の場合は税、民間の場合は商品価格への転嫁というかたちで国民が負担している。一部の者のために、多くの国民がこれらの負担を負うのは正しいのかという議論が必要である。また、コストとして表れない国民の心理的な負担は、健全な情報化社会の発展を阻害し、日本経済全体のパフォーマンスを低下させているのではないかという議論も必要である。

③文字認識の視点

欠落している二つめの視点として、人間による文字認識の視点が挙げられる。通常の日本人は、どれくらいの漢字を認識できるのかについてデータを示す。ここで言う「漢字の認識」とは、「漢字を理解する(正しい読み書きができる)」と「漢字を識別する(一致不一致

⁸ 経済産業省・独立行政法人情報処理推進機構(IPA)(2011)

⁹ 富士ゼロックス株式会社(2012)

を見分けられる)」の2種類を想定している。そして、識別の意味において、日本人の認識能力（認識速度および認識の正確さ）が標準的な範囲の漢字と行政手続で扱われる範囲の漢字において、どれくらい違うのかを実験データによって示す。さらに超高齢社会、グローバル社会となっている日本において、高齢者や外国人がどれくらいの範囲の漢字を認識ができるのかについても考察したい。

④新たな解決策の提案

2つの視点から得られた知見をもとに新たな解決策の提案をしていく。特に、二番目の文字認識の視点においては、人間と機械のインタフェースについての問題が存在することを論じ、技術的な解決策では問題が解決しないことを示す。具体的には、政府が示すような技術的な解決策を採用するのではなく、法制度的な解決策を採用することを提案する。

2. 2 本研究で扱う概念

本研究では「外字」という概念が重要であるが、第1章で前述したようにこの概念は多義的である。コンピュータ関連で使われる一般的な外字の概念は、「あるコンピュータが扱う特定の文字集合(文字セット)に含まれない文字」という意味であるが、本研究レポートでは次のような使い方をする。

単に「外字」という場合は、「ある文字コード体系で扱うことのできない文字」という本来の意味と、JIS第1水準・第2水準以外の文字を指す「狭義の外字」という意味を持つこととする。JIS第1水準・第2水準とはJIS X0208¹⁰の文字セットのことであり、制定年度による文字の相違まではここでは考慮しない。なお、JIS第3水準・第4水準についても近年JIS X0213として制定されたが、これは普及していない。JIS X0213はJIS X0208の上位互換であるが、すでに各ベンダーが機種依存文字として使っているJIS X0208とJIS X0212(補助漢字)の空き領域を使うよう設計されているためである。そこで、最も標準的な文字セットをJIS第1水準・第2水準(JIS X0208)と捉え、それ以外を「狭義の外字」と捉えることにする。

また、住基ネット統一文字の文字セット以外の文字については「住基ネット外字」、戸籍統一文字の文字セット以外の文字については「戸籍外字」と呼ぶことにする。「住基ネット残存外字」という言い方もあるが、これは住基ネット外字と同じ意味である。

また、外字と対になる概念として「正字」や「内字」という言葉がある。国語辞書によれば、「正字」とは「正しい書き方をした字」という意味で誤字・当て字に対して使う場合と「正式な字」という意味で略字・俗字などに対して使う場合がある。「内字」とは国語辞書には見られず、もっぱらコンピュータ用語として使われ、ある文字セットに收容されている字という意味である。しかし、一般的には「内字」化するという意味で、「正字」化す

¹⁰ 漢字だけでなく、ひらがな・カタカナを含む

るという表現が使われることが多いため、ここでも「内字」の意味で「正字」という言葉を使っていく。すなわち、「正字」とはある文字セットに収容されている字という意味であり、狭義で使う場合は JIS 第 1 水準・第 2 水準に収容されている字という意味で使っていることに留意されたい。

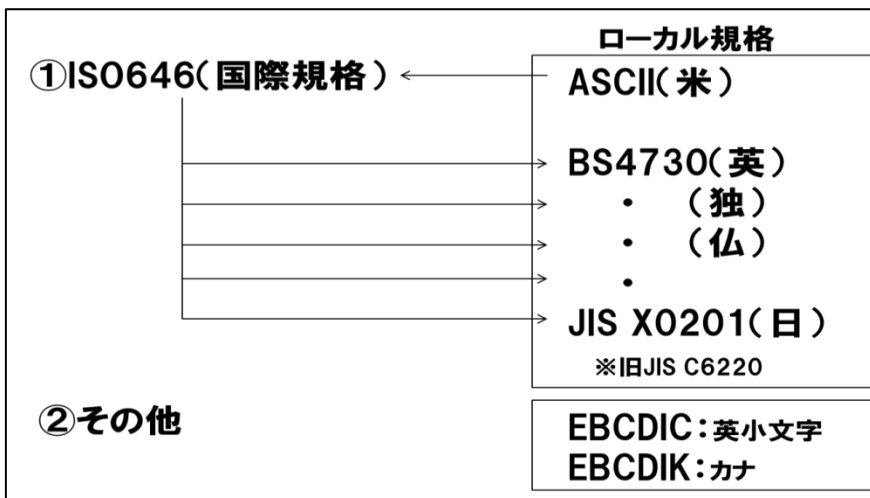
3. 日本語文字コードの開発

3. 1 日本語のコンピュータ化

はじめに、日本語をコンピュータ上で表現する文字コード開発の経緯を辿る。日本語で扱われる文字は、中国の文字を起源とした表意文字である漢字、および漢字を元に作られた表音文字であるひらがなとカタカナの 3 種類がある。ひらがなは漢字だけでは表現できない日本語を補助するために使われ、カタカナは外来語や擬態・擬音を表現するために使われる。ひらがなとカタカナの数はそれぞれ約 50 あり、漢字の数は数万とも言われている。

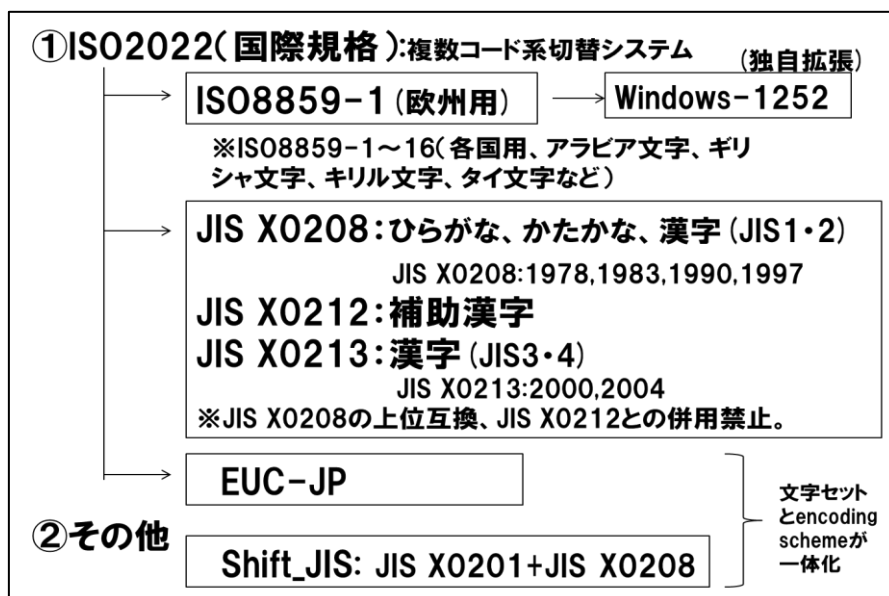
このような多くの文字を表現するためには 1 バイトでは不足し、日本では早くから複数バイトによる文字表現が求められてきた。日本における 1 バイト系文字コード(図表 3)としては ISO646 に準拠した JIS X0201 が定められ、複数バイト系文字コード(図表 4)としては ISO2022 に準拠した JIS X0208、JIS X0212、JIS X0213 が定められた。JIS X0208 は 1978 年に制定されたもので約 6,000 の漢字(JIS 第 1 水準、JIS 第 2 水準)を収容している。また、JIS X0213 はさらに約 3,700 の漢字(JIS 第 3 水準、JIS 第 4 水準)を追加したものであり、2000 年に制定された。また、文字セットとエンコーディングスキームが一体となった文字コードとして、EUC-JP(UNIX 用)および Shift_JIS(パソコン用)の規格がある。メインフレームについてはベンダーがそれぞれ独自の文字コード規格を採用している。

図表 3 1 バイト系文字コード



(出所：深沢千尋(2011)と矢野啓介(2010)より筆者作成)

図表 4 複数バイト系文字コード



(出所：深沢千尋(2011)と矢野啓介(2010)より筆者作成)

その後 1993 年に、国際的な文字コードとして Unicode(ISO10646)が制定された。ここで初めて文字セットとエンコーディングスキームを分離する考え方が採用される。Unicode の代表的な規格としては UCS-2、UCS-4、UTF-16、UTF-8、UTF-32 などがあり、UCS(Universal Coded Character Set)は文字セットとエンコーディングスキームが一体であるが、UTF(UCS Transformation Format)はエンコーディングスキームである。これらの規格においては、6.5 万から 21.5 億のコードポイントを扱うことが可能であり、世界中のすべての文字を扱うことが可能となったと言ってもよいだろう。日常的な文字だけではなく、現在では使われてない文字など学術的な研究のために広範囲な文字を扱うことが可能となった意義は大きい。

図表 5 Unicode の代表的な規格とその特徴

	文字セット	エンコーディングスキーム	ビット長	使用文字セット	コードポイント数
UCS-2	○	○	16	UCS-2	6.5 万
UCS-4	○	○	32	UCS-4	21.5 億
UTF-16	—	○	16,32	UCS-2 とサロゲートペア領域	111 万
UTF-8	—	○	8,16,24,32,40,48	UCS-4	2.5 億
UTF-32	—	○	32	UCS-4	111 万

(出所：深沢千尋(2011)と矢野啓介(2010)より筆者作成)

日本の行政の現場においては、1980年代前半にメインフレームで日本語を扱うことができるようになって以来、住民基本台帳をはじめとして様々な分野に複数バイト系の文字コードが広がっていった。自治体の現場では、JIS 第1水準とJIS 第2水準およびベンダー提供漢字で住民氏名の漢字化を行い、それ以外の漢字については自らがフォントを作成・登録して対応していった。このように各自治体では、住民氏名をせいぜい1万字未満の漢字で管理し、行政手続きを行っていたのである。

しかし、1990年代後半の戸籍システム化、そして2000年代前半の住民基本台帳ネットワーク(以降、「住基ネット」)の構築においては、1万字を大きく超える種類の漢字を扱わざるを得なくなった。当然ながら、従来の方法で文字を管理することに無理が生じてくる。次にそのあたりの事情を振り返る。

3. 2 戸籍と住民基本台帳ネットワーク

戸籍よりも住民基本台帳のコンピュータ化のほうが10年以上早かったのであるが、漢字のコンピュータ化が初めて問題となったのは戸籍である。その理由は、住民基本台帳法によって住民基本台帳の氏名漢字は戸籍と同じ文字を使うことと規定されていたため、元となっている戸籍の文字を云々することはできず、各自治体は住民基本台帳に記載されていた漢字を忠実にコンピュータ化していったためである。住民基本台帳について規定している住民基本台帳法および自治体は総務省(当時は自治省)の管轄であり、戸籍を管轄している法務省の領域には踏み込めなかったという事情がある。つまり住民基本台帳で扱う漢字の一覧表などは無く、メインフレームで用意されていない漢字については各自治体の判断で外字を作成し、住民氏名の漢字化を進めていったのである。

そして、住民基本台帳のコンピュータ化から遅れること10年、戸籍もコンピュータ化していきたいという社会的要請が高まってきた。しかし、戸籍は住民基本台帳と異なり、事細かな事務処理についても法律で規定されており、戸籍法を改正しなければ戸籍のコンピュータ化は不可能であった。例えば、戸籍を編成する用紙についても美濃判と呼ばれるコンピュータ規格外の用紙の使用が義務付けられ、縦書き・朱線抹消などコンピュータでは対応できない処理が義務付けられていた。このコンピュータ化に伴う戸籍法改正の際、戸籍で扱う文字をどうするかという問題も浮上し、誤字・俗字を職権で訂正する方針が示されることとなった。

しかし、これに国会議員等が反対し訂正できなかったという。『人名用漢字の戦後史』によれば、「法務省が戸籍法の改正案をまとめたのは、1994年2月のことであった。このとき、法務省は戸籍法の改正と合わせて、戸籍をコンピュータ化する際に誤字・俗字をすべて職権で訂正してしまう、という方針を表明したのである。それまでは、新しい戸籍を作成する場合のみの強制訂正であったのに、今度はすべてを強制訂正する、というのである。

全部で4,500万件ある戸籍のうち、十数パーセントが対象になるという試算もあった。

これには反発が強かった。折しも、結婚したときに俗字を勝手に訂正されたなどとして争われていた裁判がいくつか結審し、その中には行政側が敗訴したものもあった。さらには、自民党を中心に国会議員が『行政の都合で勝手に名前を変えられるのはおかしい』と反発した。その結果、法務省は譲歩を余儀なくされたのである。」(円満字二郎(2005)P182)

その経緯を新聞記事でたどってみると、1994年2月にまとめられた戸籍法改正案が、5月には次のように方針変更されている。

「法務省は二十五日、戸籍事務のコンピューター化のための戸籍法改正案をまとめた。今年一月の民事行政審議会（法相の諮問機関、会長・加藤一郎成城学園長）の答申で戸籍については「誤字・俗字」は職権で正字に訂正することとしていたが、戸籍が登録されている人については漢和辞典に記載されている俗字の使用を認めることにした。職権訂正に対し、自民党法務部会（下稲葉耕吉会長）などが、「行政の都合で勝手に名前を変えるのはおかしい」と反対したのを受けての措置」（1994/05/26 読売新聞）

6月にはさらに後退し、6月9日の国会質問では次のような答弁が行われた。本人からの異議があればどのような字でも認めることになり、まったく歯止めが利かなくなってしまったのである。

「九日の参院予算委員会で行われた主な質問と政府答弁の内容は次の通り。質問者は、倉田寛之、須藤良太郎、片山虎之助（いずれも自民）の三氏。

…（中略）…

〈戸籍法改正〉

——戸籍法改正案の内容は。

中井治・法相（戸籍のコンピューター化により）数十万人に（姓名の）変更をしていただかねばならない。どうしても変更したくない人は、コンピューターに載せず原簿を残す。これにより、各市町村の戸籍事務の簡素化、スピード化、人員削減に寄与したい。」

（1994/06/10 読売新聞）

「法務省は九日までに、戸籍事務のコンピューター化に伴い、辞典に載っておらず正字に変更することになっていた「誤字・俗字」であっても、本人が変更を拒否すれば、コンピューターに入れず、現在の原簿をそのまま戸籍として残すことを決めた。謄抄本はこの原簿をコピーして発行する。同日の参院予算委員会で中井治法相が明らかにした。

また、正字への変更にあたっては、「本人に通知せず、市区町村長の職権で変更する」との従来の方針を見直し、事前に本人に通知し、異議も受け付けることにした。」（1994/06/10 読売新聞）

このような経緯で1994年から戸籍のシステム化は進められていったが、文字コードの統一化には至らず、自治体およびベンダーごとに外字を作成して対応していかざるを得なかった。そして政府の規制改革の一環として、戸籍手続きオンライン化の要請が高まり、2004年に戸籍統一文字が定められることになった。

法務省は、平成2年10月20日付け民2第5200号民事局長通達「氏又は名の記載に用いる文字の取り扱いに関する通達等の整理について（通達）」及び平成16年4月1日付け民1第928号民事局長通達「電子情報処理組織による戸籍の記録事項証明等の交付請求及び戸籍の届出等の取り扱いについて」で、戸籍に用いることのできる文字及びその電子的な扱いについて示している。戸籍統一文字は、この二つの通達を前提として、「戸籍のオンライン手続に使用することを目的として整理した文字の情報」としてまとめられた文字集合で、55,267文字の漢字を含んでいる。

(経済産業省・独立行政法人情報処理推進機構(IPA)(2011)P36)

そのときの戸籍統一文字の選定基準は、下記のようになっている。

(1)戸籍統一文字選定基準

次の基準に該当する文字が戸籍統一文字に採用されています。

- ア 漢和辞典に掲載された正字等および俗字などの文字
- イ 常用漢字および人名用漢字
- ウ 規則又は通達等による俗字などの文字
- エ 規則、通達または先例などにおいて、戸籍に記載可能な文字と判断された文字。

(2)誤字の取扱い

次の例外を除き、戸籍統一文字に誤字（譌字(*1)・略字(*2)を含む)および記号などは採用されていません。

ア JIS第1水準漢字・第2水準漢字および補助漢字の文字のうち、上記(1)に含まれない文字であって誤字とされているものおよび記号。

イ 地名外字

*1 「かじ」。誤字のことで、文字の骨組みに誤りがあり、公的な字形とは認められない文字。康熙字典では「譌字」とされている。

*2 正字の字画を省いた文字。通常は誤字扱いされるが通俗的な文字として定着すると俗字となる。

(日立ホームページ「戸籍手続オンラインソリューション：戸籍統一文字とは」より

<http://www.hitachi.co.jp/Prod/comp/app/kosekionline/touitsu.html>)

このように、自治体における戸籍のコンピュータ化は戸籍統一文字を基準として進められているが、戸籍法の改正時に国会で「本人からの異議があればどのような字でも認める」こととされたため、統一文字へ収容するという強制力は持っていない。実務の実態として、このような戸籍統一文字へ収容できない漢字は、戸籍外字として各自治体が紙で管理している。つまり、戸籍のコンピュータ化が完了している自治体であっても、戸籍外字を含む戸籍謄本については電子化されておらず、従来どおり紙のまま管理されているのである。

住民基本台帳のネットワーク化については、戸籍と事情が異なる。すでに各自治体では1980年代初期から住民基本台帳のコンピュータ化を推進し、1995年時点では人口割合で

98.7%もの住民基本台帳がすでに電算システムとして稼働していたのである。このような実態を背景に、市町村や都道府県の区域を超えた本人確認システムの導入や高齢者等弱者に対するセーフティネットとして活用することを目的に、住民記録システムネットワークの構築が要請されたのである。

そして、1998年3月に「住民基本台帳法の一部を改正する法律案」が閣議決定され、1999年8月に改正住民基本台帳法が成立、2002年8月に住基ネット内部稼働、2003年8月に住基ネットが全国一斉稼働という経緯を辿った。このときに住基ネット統一文字が定められたのだが、住基ネットに関する情報が非公開のため、どのような選定基準で定められたのかは不明である。実際には、各自治体で使われていた外字（ベンダー提供漢字および自治体作成漢字）を集め、何らかの基準によって整理されたものと考えられる。

下記に記載されているように、住基ネット統一文字と戸籍統一文字の互換性は無い。そして、住基ネット統一文字に同定できない漢字（住基ネット外字、または住基ネット残存外字）については、住基ネット上ではイメージデータとして管理されている。

住基ネットで使用される文字についても「戸籍統一文字」と同様に総務省が選定した住民基本台帳ネットワーク統一文字（以下、住基ネット統一文字という）及びそれに付された統一文字コード（以下、住基コードという）の番号が存在する。住民基本台帳ネットワーク統一文字（住基ネット統一文字）の漢字数は19,432字である。

各自治体で管理している住記システムの文字コードを、自治体において、住基ネット統一文字と自治体が保有している文字を同定して、住基ネット上では住基ネット統一文字コード、もしくは住基ネット統一文字に同定できない場合には、残存外字として住基ネットではイメージ（画像）で流通している。

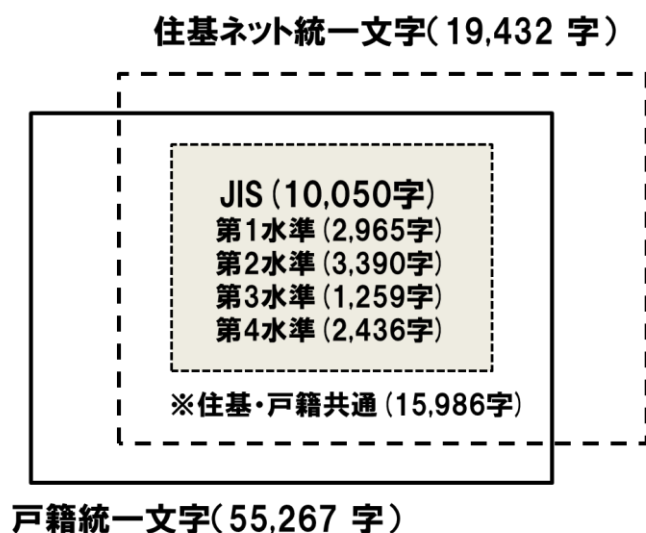
(経済産業省・独立行政法人情報処理推進機構(IPA)(2011)P45)

このように、行政手続きにおける文字セットを定義するうえでは、文字の数がJISの制定範囲を大きく超えることが要請され、住基ネット統一文字(約21,000字、漢字数は19,432字)と戸籍統一文字(約56,000字、漢字数は55,267字)が制定された。これら2つの文字セットは互換性が無いばかりではなく、実態としてすべての氏名漢字がコード化されたわけでもない。コード化できない漢字については、住基ネットではイメージデータとして管理され、戸籍においては紙のまま管理されている。

行政手続きにおいて使われるJIS、住基ネット統一文字、戸籍統一文字の3つの文字セットの関係について図表6に示した。全体では約6万字(58,713字)もの漢字が存在する。このような図表で整理されると、JIS、住基ネット統一文字、戸籍統一文字の3つの文字セットは簡単に統合できるように見えるが、そうはならない。この問題については後述するが、ある文字の同一性についてどこまでデザイン(書体)の差を許容するかという包摂の考え方、すなわち文字の概念が3つの文字セットで異なっているからだ。見方を変えると、3つの文字セットの管轄官庁が、経済産業省、総務省、法務省と異なっているからともいえる。

さらに問題を複雑にしているのは、漢字を含む国語の権威に関する管轄官庁は文部科学省（文化庁）であり、これらの文字セットには関わっていない。このような縦割り行政が根本的な原因となっている。

図表 6 JIS、住基ネット統一文字、戸籍統一文字の関係



（出所：経済産業省・独立行政法人情報処理推進機構(IPA)(2011)の「汎用電子情報交換環境整備プログラムで整理された漢字」をもとに漢字のみを対象として筆者作成）

4 政府の問題意識とその対応

4. 1 文字情報基盤構築プロジェクト

このような状況のなかで社会のネットワーク化が進展するにつれ、通常のコンピュータと行政手続用コンピュータとではデータの交換ができない、官民のデータ連携ができない、行政機関どうしのデータ交換ができない、システムの移行に大きなコストがかかるなどの問題が指摘され、政府も本腰を入れざるを得なくなってくる。

経済産業省は2010年度に文字情報統一基盤プロジェクトを実施し、上記の問題を解決するために3つの文字セットの整理を行い、その成果として文字情報基盤文字情報一覧表(MJ文字情報一覧表¹¹)を公開した。文字情報統一基盤プロジェクトの報告書¹²では主張が明確ではないが、「MJ文字情報一覧表を基に、UTF-16、IVSへの対応を促進して、約6万字を使っていく」（同報告書 P74）との記述が見られ、全体の論調は「MJ文字情報一覧表を基に各自治体は外字を同定し、UTF-16、IVSを実装していくべき」、そうすれば将来的に文字

¹¹ <http://ossipedia.ipa.go.jp/ipamjfont/mjmojiichiran/index.html>

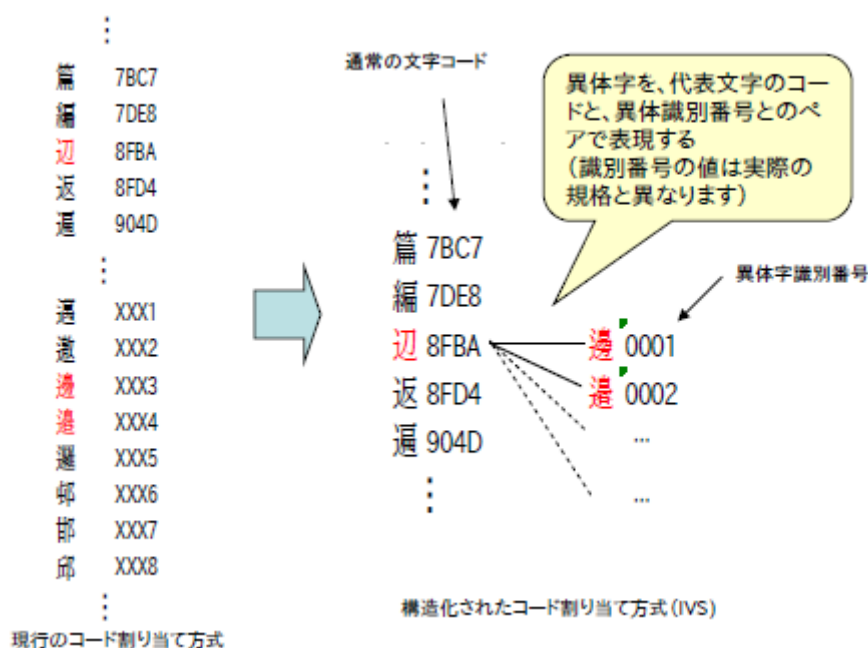
¹² 経済産業省・独立行政法人情報処理推進機構(IPA)(2011)

の問題は解消していくだろうと読み取れる。

UTF-16 は前述したとおり、111 万のコードポイントを保有し、UCS-2 とサロゲートペア¹³の領域を使うことができる。IVS について説明すると、「IVS(Ideographic Variation Sequence)は、この IVD(Ideographic Variation Database)に登録されるコレクションを構成する個々の符号位置と字形選択子の列のことを言う。(中略)文字が、形状の異なる複数の異体字を持つ場合、通常のコードスペースで符号化する文字符号は原則一つとし(以後これを「代表文字」と呼ぶ。)、個別の異体字表記については、文字コードではなく、異体字識別番号によって識別しようという考え方である。従って個々の異体字は、代表文字に与えられた文字コードと、各異体字に与えられた異体字識別番号とのペアによって識別することとなる。IVS を用いることにより、同一の符号位置に対して異なる字形イメージを特定することが可能となるが、規格上は、符号位置としての値は同値である」¹⁴(図表 7)。

IVD とは、IVS 方式による異体字識別番号を登録したデータベースであり、Unicode Consortium へ登録手続きを行う必要がある。

図表 7 IVS(Ideographic Variation Selectors/Sequences)の考え方



(出所：経済産業省・独立行政法人情報処理推進機構(IPA)(2011)P53)

経済産業省の文字情報統一基盤プロジェクトでは、UTF-16 と IVS/IVD の実装が問題解決につながるという方向性を示しているが、それでも次のような問題や課題が残る。

¹³ 日本語表記は「代用対」。16 ビット Unicode の未定義領域アドレスの前半 2 バイトと後半 2 バイトをペアにして(32 ビットで)1 文字を表す技術。

¹⁴ 経済産業省・独立行政法人情報処理推進機構(IPA)(2011)P52

- ・ 国として統一的な包摂規準(基準)がないため、IVS/IVD による類似字形の判断や登録の権限がどの官庁にあるのかが不明。
- ・ 現在、標準的な PC では扱うことができず、公的個人認証では JIS 漢字による代替文字を認めている。国民や民間に 1 万字以上の漢字を扱うことを強制できるのかという問題が残る。
- ・ (標準的な PC で扱うことができるようになったとして)そもそも行政手続きという日常的な社会生活において、国民や行政機関が 1 万字以上の漢字を正しく扱うことができるのかという問題がある。

これらについて検討する前に、経済産業省のプロジェクトの後に実施された総務省の外字調査プロジェクトについても概観しておきたい。

4. 2 総務省の外字調査について

この 2011 年度の調査は、「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」¹⁵と書かれているように、総務省が前年度の文字情報統一基盤プロジェクトの成果を前提に、自治体のクラウドサービス推進のために調査を行ったものである。

この調査は、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たもので、全国の 1,386 市区町村から提供された総文字数は 1,166,536 文字と 100 万を超える外字が報告された。

その結果として、934,812 文字 (21,561 種類) が文字情報基盤漢字(住基ネット統一文字と戸籍統一文字) に同定され、同定できなかった文字として不明文字が 52,294 文字、変体仮名が 97,791 文字、記号等が 81,639 文字もあることがわかった(図表 8 および図表 9)。

なお、ここでの同定作業における包摂基準は、「常用漢字表」(付)字体についての解説や法務省の通達などを参考にした独自の基準である。この包摂基準は、広範囲に包摂する JIS の包摂規準 の考え方とはかなり異なっている。そして、文字情報基盤漢字に同定されたといっても、字形が一致したのは約 12,000 種類しかなく、約 26,000 種類はデザイン差を持っていたり、類字文字(似た字であるが画数などが異なる)であったりする。デザイン差の文字や類似文字を、文字情報基盤漢字に強制的に置き換えるための法的な裏づけは無く、この時点で文字情報基盤漢字に収容できない約 26,000 字の外字が発生する可能性がある。

¹⁵ 富士ゼロックス株式会社(2012)P1

図表 8 文字情報基盤漢字に同定された外字

	文字情報基盤漢字 58,712 種類	
	住基ネット統一文字 19,432 種類	戸籍統一文字 39,280 種類※
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)
小計	820,020 文字 (14,852 種類)	114,792 文字 (6,709 種類)
合計	934,812 文字 (21,561 種類)	

※住基ネット統一文字との重複を除く

(出所：富士ゼロックス株式会社(2012)P4「外字の実態調査結果(全体)」より筆者作成)

図表 9 文字情報基盤漢字に同定できなかった外字

文字情報基盤漢字に同定できなかった文字		
不明文字	変体仮名	記号等
52,294 文字	97,791 文字	81,639 文字
231,724 文字		

(出所：富士ゼロックス株式会社(2012)P4「外字の実態調査結果(全体)」より筆者作成)

さらに、文字情報基盤漢字に同定できなかった不明文字 52,294 文字を分析したものが、次の図表 10 である。3,857 種類に分類された文字と 33,759 文字の分類不可能文字があり、少なくとも 37,000 種類以上の文字がデザイン差や類似文字としても同定できないことが明らかとなった。

図表 10 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	

(出所：富士ゼロックス株式会社(2012)P28)

報告書の「第五章 実態調査の考察 5.2 まとめ」では、次のように結論づけられている。

- ・ 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。
- ・ 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われる。
- ・ 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。

この調査では、文字情報基盤漢字(住基ネット統一文字と戸籍統一文字)だけでは、自治体の現場における行政手続きで利用する文字すべてを網羅できないことが判明しただけでなく、さらに少なくとも 37,000 種類以上の文字を扱わなければならないことが明らかとなった。報告書のまとめによれば、文字情報基盤漢字の整理・拡充を行い、文字利用技術が進展すれば問題は解決するという方向性が示されているが、本当に解決するのだろうか。

4. 3 政府の解決方策の問題点

経済産業省の文字情報統一基盤プロジェクトでは、住基ネット統一文字と戸籍統一文字の範囲だけで漢字を扱っているが、それでも IVS/IVD による類似字形の判断や登録の権限がどこにあるのか、そして日常的な社会生活において国民や行政機関が 1 万字以上の漢字を正しく扱うことができるのかという問題が残っており、単純に UTF-16 と IVS/IVD の実装で問題が解決できないのではないかという疑問を提示した。

さらに、総務省のプロジェクトでは、MJ 文字情報一覧表(住基ネット統一文字と戸籍統一文字の範囲)に同定できない不明文字(漢字のみ)が、37,000 字以上も存在していることが明らかとなった。住基・戸籍の 58,000 字と合計すれば、95,000 字以上もの漢字を扱わなければならない。さらに同定された文字のなかでも、類似文字やデザイン差として文字の置き換えができなければ、さらに 26,000 種類の文字を追加しなければならない。技術的には可能だとして、人間が正しく扱うことができるのだろうかという疑問が生じてくる。

自治体においては、業務のクラウド化を推進する上で、外字が切実な問題となってきた。自治体からの声を 2 点次に挙げたい。

自治体クラウド実現のポイントは？ ～国へのお願い～

- ・標準のパソコンにある漢字だけで業務ができるように
 - 住基ネット統一文字(約 2 万文字)、IPA の戸籍統一文字(約 6 万文字)
 - 日本だけなぜ漢字に拘^{こだわ}るのか(文化?) 日常生活では覚えられない使えない
 - これでは民間を巻き込んだワンストップサービスなどできない
 - 公的個人認証システムでは住基ネット統一文字を平易な文字に変換

(出所：奈良県河合町総務課(2012))

○共同利用における外字の処理

- ・団体毎に作成した外字の同定作業が必須であるが、ホストを使用する多くの自治体は多くの外字を作成している
 - ・メーカー毎の内字の違いも外字として登録しなおす必要がある(内字の書体はメーカー毎に異なるため、自動変換で対応できず、同定作業が必要)
 - ・共同利用では稼働後に発生する外字の統一管理が困難
 - ・新たに共同利用する団体は、外字をその都度同定するため、コスト増の要因となる
- ↓↓
- ・共同化の推進には、外字について根本的な見直し(法制度や考え方)が必要
 - ・今後使用するシステムの選択肢を多くするためにも外字を制限していく意識が必要
 - ・住基ネットワーク、広域連合ネットワークなどは同一書体文字を使用しており、市町村のシステムもメーカー毎の文字を使用するのではなく、全国共通の統一文字を活用すべき
 - ・住民や自治体の意識を変える必要あり

・戸籍以外の行政システムでは、本人が特定できれば、字形の微妙な違いは無視する意識の醸成

(出所：大分県商工労働部情報政策課(2010))

このように自治体の実務の現場からは、コンピュータで技術的に扱える外字の数を増やせば対応できるという声ではなく、問題解決のためには「外字を減らす、なくす、制限する、字形の微妙な違いは無視する」という意見があがっている。過去、自治体でシステムを開発していた筆者の経験から言っても、コンピュータで扱える外字の数を増やすべきという声を自治体職員から聞いたことはなく、外字を無くすあるいは減らすべきという意見がほとんどであった。このように経済産業省や総務省がコンピュータで技術的に扱える外字の数を増やして対処しようと考えているのに対し、自治体は外字そのものを無くす・制限すべきと考え方がまったく異なっている。

さらにもう少し広い視野で考えると、外字の数を増やして対応するという解決策は次のような問題も引き起こすだろう。

- ・ 数万という数の漢字は民間では扱いきれず、情報の官民連携ができない。
- ・ 超高齢社会である日本では、高齢者が増加の一途を辿っており、高齢者にとって画数の多い微妙な字形を扱うことは難しい。
- ・ グローバル社会の進展に伴い、外国人の数が増えてくるが、外国人にとって数万もの漢字は扱いきれない。

このような問題に加えて、筆者はこれまでの議論で欠落している二つの視点から、この問題を捉えてみたいと思う。一つめの視点はコストの視点であり、現状外字を処理するためにどれだけのコストがかかっているかを試算し、そのコストをすべての国民が負担し続けるのは正当であるかという問題を考えてみたい。例え MJ 文字情報一覧表や UTF-16 と IVS/IVD を利用しても、管轄官庁の調整と文字の登録や複雑な文字の処理に長い時間とコストがかかるばかりか、対応できない文字が 37,000 以上も残ってしまう。最悪の場合、現在よりもコストがかかるという事態にもなりかねない。これについては次章で検証する。

そして二つめは、そもそも人間は 1 万字以上の漢字を正しく扱うことができるのかという文字認識の問題である。日本の将来を考えると、高齢者や外国人も視野に入れて考えていく必要があるだろう。これは人間と人間とのインタフェースが、機械と機械のインタフェースそして人間と機械のインタフェースに代わる時、どのような問題が生じるのかというテーマでもある。これについては第 6 章で詳しく検証したい。

そのほかにも、そもそも漢字に対するコード化の考え方（字種、字体、字形、デザイン・書体）が統一されておらず、文化庁（文科省）の常用漢字表、経産省 JIS の包摂規準、法務省の通達、総務省の包摂基準がばらばらであるという問題もある。しかし、本研究では省庁ごとの文字概念の相違については踏み込まないこととする。

5. コストの視点

5. 1 行政の実務とコスト

外字の問題をコストの視点から捉える。文字の問題は極めて精神的文化的な問題で、経済合理性で判断されるものではないという主張があるが、外字を使っている国民は全体から見れば多くない(後述するが、国会議員の名簿から推測すると約5%)。つまり、残りの95%の国民が5%の国民のために、税あるいは商品価格への転嫁として余計な負担をしていることになる。この税負担が公正かつ公平なものであるかどうかという議論が必要である。なお、文字情報基盤構築プロジェクトでは、自治体アンケートによって定量的な把握をしようとしているものの、経済損失の試算までは行っていない。そこで独自に規模の異なる3つの自治体へのヒアリング調査を実施し、おおまかな経済損失を試算した。

その結果が次の表である。「規模の経済」により大きい自治体ほど効率が良く、小さい自治体ほど効率が悪くなるという傾向がある。日本の全人口を市町村数で割ると約7.3万人となるため、日本全体の年間経済損失(市町村のみ)は下記図表の約12億円から約27億円の範囲にあり、20億円前後であろうと推測される。

図表 11 市町村における外字処理コスト試算

調査自治体		C市	F市	K市
人口(万人)		95.9	13.6	5.0
コスト内訳(円)	クレーム対応等	245,000	94,500	17,500
	既存外字の確認・入力	3,733,333	126,000	175,000
	新規外字の判断・登録	1,085,000	63,000	87,500
	基本システム維持管理	710,643	1,000,000	14,000
	外部システム連携	2,450,000		26,250
	庁内システム間連携	1,087,500		750,000
	その他	61,250	70,000	0
コスト合計(円)		9,372,726	1,353,500	1,070,250
全国推計(円)		1,237,903,563	1,260,546,397	2,711,157,300

※職員時給 3,500円として計算(前提:年間労働時間2,000時間、平均年収700万円)

※全国推計は、日本の人口(1億2666万人)との比率から推計。

(出所:筆者作成)

ここで取り上げた経済損失は、コスト換算可能なものだけである。自治体の現場においては、コスト換算できないが次のような経済損失もあると指摘されており、実質的な経済損失は20億円を上回るだろうと推測される。このように外字が存在するために、日本における市町村だけで年間20億円以上もの経済的損失が生じていると推定される。

- ・ 外字がベンダー・ロックインの原因となっており、ベンダー同士の競争原理が働かないため、余計なコストがかかっている。
- ・ 職員が PC で可能な処理も、PC では外字の印字等ができないため、メインフレームのベンダーに発注しなければならないという余計なコストがかかっている。

なお、F 市は相対的にコストが小さいが、これは市役所が住民に対して正字化の指導を行っているためである。戸籍のシステム化の際、文字の正字化を行い、住民に確認のはがきを郵送して、苦情があったものだけを外字として登録している。また、婚姻届や死亡届の際に、苗字の正字化を住民に促すという対応を行っており、このような努力により外字のコストが他と比べて抑えられていると考えられる。

この 3 市へのヒアリングから、外字処理に関する現場の実態について下記に示したい。

①C 市の現場ヒアリングから

- ・ 以前はメインフレームで処理をしていたが、コストダウンのため現在ではオープン系システムへ移行している。各オープン系システムでは外字対応をしておらず、これほどの手書き作業が発生していることは今回初めて知った。
- ・ 外字の作成については、調達仕様の作成や納品のチェックおよび校正の作業でかなりの時間がかかっている。
- ・ 介護保険では、介護給付審査等のデータ及び出力帳票の一部を C 県国民健康保険団体連合会のシステムで出力を行っている。国保連合会のシステムには、C 市で作成した外字情報を移行する仕組みがなく、C 市において外字登録している文字であっても、国保連合会のシステムに登録されていなければ該当の文字は空白で出力されるため、手書きで補完する等の作業が発生する。
- ・ 住民税関係の給与支払報告書および異動情報（民間企業からのデータ）については、データをホストに取り込む際にカナ氏名と生年月日で外部データとホストのデータとをマッチングをしているため、そのデータのマッチングを行う時点においては外字に関する作業はないが、その後納税通知書を作成（印字）する際、外字宛名等の手書き作業が発生する。
- ・ 税業務や福祉業務では、外字の連携ができていますが、一部未登録外字が発生しており、この場合のみ内容を確認して手書き作業が発生している。特に税業務に関しては、未登録の外字を住記端末で確認するとともに、外字表(外字を管理している冊子で、約 2cm × 2cm の外字が手書きで書かれている)を参照しながら正確に手書きをするため、かなり時間がかかっている。

②F 市の現場ヒアリングから

- ・ 戸籍に関して、親子で苗字の漢字が異なるといったケースのトラブルが月に数件発生する（例えば、「深」と「深の木が米になっている漢字」など）。
- ・ メインフレームからオープン系へシステムを移行したとき、外字の変換等の費用を抑え

るため、コード変換プログラムを標準装備しているメインフレームと同じベンダーを選択した。

- ・ 基幹系システムの外字登録領域と他のシステムの外字登録領域の数に違いがあるため、現在もすべての外字が他システムに移行できていない状況である。
- ・ 福祉システムに外字データを渡しているが、福祉の業務では字形についてあまりこだわっていない。(理由については後述)
- ・ 外字を持った人が転入してくると、即時に住民票、印鑑登録証明、保険証等の発行ができない。即時発行が必要な場合、転出証明書の外字を切り貼りして対応している。

③K市の現場ヒアリングから

- ・ ある業務でベンダーが変わる場合など、外字の移行でかなりのコストがかかっている。それだけでなく、システム移行や新システム設計時における外字処理の検討作業が発生し、目に見えないコストもかかっている。
- ・ 広域連合との情報連携については、外字作成時に外字のデータを FPD で広域連合へ送付して対応している。
- ・ 外字を持った人が転入してきた時、(外字が作成・登録されるまで) 外字は空白のまま処理するため、転出元市町村とのやり取りが発生している。

5. 2 民間の実務とコスト

それでは民間企業では、外字を処理するためにどれだけのコストをかけているのだろうか。富士通株式会社人事総務サービスセンターおよび実務を担当する株式会社富士通 HR プロフェッショナルズの協力を得て、富士通グループ全体の従業員氏名の実務に関するヒアリング調査およびコスト試算を行った。

富士通グループの従業員数は約 17.3 万人¹⁶であり、従業員氏名については 2 種類の氏名を管理している。一つは社内氏名であり、これは JIS 第 1 水準・第 2 水準の範囲で管理し、範囲外の漢字については JIS 第 1 水準・第 2 水準内の当て字に押し込めるという運用を行っている。そしてどうしても押し込められないものはカタカナで表記している。もう一つは戸籍氏名であり、これは戸籍と同等の文字を使い、JEF¹⁷ (JIS 第 1 水準・第 2 水準+拡張漢字+ユーザ定義文字) の文字コード体系の中で管理している。

このように従業員氏名は二重管理されており、社内事務処理はすべて社内氏名で処理し、行政など外部との連携のときに戸籍氏名を使うという運用を行っている。つまり、民間企業では行政との情報連携のために、わざわざ外字を管理しているのである。連携といっても系統的に連携しているわけではない。自治体とデータ連携する場合は、取り決めに

¹⁶ 2012 年 3 月末現在の公表された数字であり、コスト試算については国内従業員のみを対象とした別の数字を使っている。

¹⁷ 富士通メインフレーム系の文字コード体系

よって外字を●や□に変更して送信している。あくまでも「行政手続き用の正式な氏名」の確認が必要になったときに、使っているのである。

一般的な外字保有比率(5%)から言えば数千人以上が外字を保有しており、強制的に JIS 第 1 水準・第 2 水準の範囲内に正字化することは大きな抵抗がありそうだが、フォントのデザイン差などで社員からクレームがくるのはせいぜい年に 1 件くらいとのことである。

新入社員の登録時には外字を保有した社員が入社してくるため、既存の外字で対応できない社員のために、年に数件程度新規に外字を作成している。このように民間企業でも戸籍氏名で外字を管理しているため、自治体と同様にコストがかかっており、新規外字の同定・登録作業、システム移行に伴う外字データ移行費用、システム連携のための外字テーブルの作成と同定作業、外字処理ソフトの導入、文字連携トラブルの対応などのコストを負担している。

これらのコストを積算し、富士通グループの従業員と日本の就業人口との比率から全国推計すると、全体で約 3 億円のコストになる。流通業など個人の顧客を抱える企業では特に顧客に気を使って氏名を外字管理しているためコストがもっとかかっている、あるいは大企業の場合は規模の経済が働いて効率が良いが中小企業は効率が悪くもっとコストがかかっているという考え方もあれば、零細企業・自営業などは紙で処理している場合が多くコストはほとんどかかっていないという考え方もある。一民間企業の試算から全国推計することは乱暴であると承知しているが、自治体の約 20 億円に対して、民間企業の約 3 億円という数字はあながち的外れでもないと考える。

今回試算できなかった外字が関係する行政機関としては、都道府県、後期高齢者医療などの広域連合、国の行政機関（税務署、ハローワーク、日本年金機構など）、国および自治体（公務員の人事管理）、地方自治情報センター（住基ネットの運用）など多数ある。また一般国民の立場としても、外字があるために正確な情報が入手できなかったり、煩雑な処理をしなければならなかったりすることがある。これらの行政機関や一般国民が負担するコストを合計すれば、少なく見積もっても日本全体で年間 30 億円くらいのコストがかかっていると見てよいだろう。

5. 3 外字のコスト負担の問題

この年間約 30 億円というコストが大きいのか小さいのかという議論は別として、外字のコストを誰が負担すべきなのかについて考えなくてはならない。外字を保有している人がどれくらいいるのか、確かな調査データは無いが、国会議員名簿から推定される比率は約 5%である。(図表 12)

図表 12 国会議員における外字(JIS 第 1 水準・第 2 水準以外)の比率

	議員数	外字数	比率
衆議院 2010.6.17	480	24	5.0%
衆議院 2013.1.16.	480	33	6.9%
参議院 2010.8.16	242	10	4.1%
参議院 2012.12.17	242	13	5.4%

(注)衆議院は 2012 年 12 月の選挙結果により名簿が大幅に変わっている。

(出所：衆議院および参議院のホームページより筆者作成)

さらに、外字を保有している人のなかで、「正字に変更することは絶対に許さない」と強硬に主張する人がどれだけいるのだろうか。国民を対象とした外字アンケート調査などのデータが無いので、定量的なデータとして示すことはできないが、外字の使用を強硬に主張する人は限りなく少ないのではないだろうか。民間企業の事例で見たように、社内で外字を JIS 第 1 水準・第 2 水準の範囲に正字化して使っている、大きな問題は起こっていない。実際に外字を保有している人の意見を聞くと下記のような返答であり、外字の使用を強硬に主張するような人に遭遇したことはない。

- ・ 自分の字が外字だとは知らなかった。運転免許証を作ったときに指摘されて初めて知った。
- ・ 今まで社会生活上正字を使っていて問題なかったため、自分の苗字が外字であることを、戸籍謄本を入手するまで知らなかった。
- ・ 以前は正字であったはずなのに、知らないうちに戸籍が外字になっていた。

実際問題として、戸籍の転記ミスによって「知らないうちに戸籍が外字になっていた」ケースが大半だと思われる。すなわち、外字を保有する人たちの多くは、現状の外字を継続して使用することを必ずしも主張しているわけではない。外字を継続して使用することを主張する人たちの理由とは、文献調査やヒアリング調査から伺える傾向として、「(先祖代々使ってきた字を)正字化されると自分のアイデンティティが失われる、あるいは損なわれる」ことを恐れているようである。

問題を立て直すと、「日本人の 5%のなかの、ほんのわずかな人々のアイデンティティを保護するために、行政機関や民間企業を含め全国民がそのコストを負担すべきだろうか」という問題になる。漢字は日本の精神や文化だという主張は良くわかるが、外字に関していえば、極めて個人的なテーマだといえるだろう。

6. 文字認識の視点

6. 1 標準的な漢字の認識能力について

本章では、そもそも人間が数万という漢字を扱うことができるのかという、人間の漢字認識能力について考えていく。文字情報統一基盤プロジェクトや外字調査プロジェクトでは、外字を整理し技術の進展によって扱う文字の数を増やしていけば外字問題は解消するとしているが、それだけの文字数を人間が正しく認識できなければ問題の解決は不可能である。その理由は、人間対人間では問題にならなかった事象が、機械対機械や人間対機械という関係では大きな問題となって立ちはだかるからだ。

第 1 章で述べたように、我が国における「一般の社会生活において、現代の国語を書き表す場合の漢字使用の目安を示すもの」として、文部科学省文化庁から常用漢字表が公開されている。この字種は 2,136 字であり、表外漢字字体表（常用漢字表に掲載されていない漢字を使う場合の目安）を合わせても、せいぜい 3,000 字である。すなわち、一般の社会生活においては約 3,000 字があれば十分であるとされている。

まず、日本人の標準的な漢字の認識能力(読み書きの能力)を知るために、財団法人日本漢字能力検定協会のデータを調査した。財団法人日本漢字能力検定協会では、年に 3 回の漢字検定を実施している。1 級から 10 級まであり、1 級が JIS 第 1 水準と JIS 第 2 水準の範囲の漢字(約 6,000 字)を使いこなすことができ、準 1 級が JIS 第 1 水準の範囲の漢字(約 3,000 字)を使いこなすことができるという基準になっている。そこで、漢字検定の結果を使って、どれだけの日本人が PC に標準搭載されている JIS 第 1 水準と JIS 第 2 水準の範囲の漢字約 6,000 字を理解し、使いこなすことができる(読み書きができる)のかについてデータを示したい。

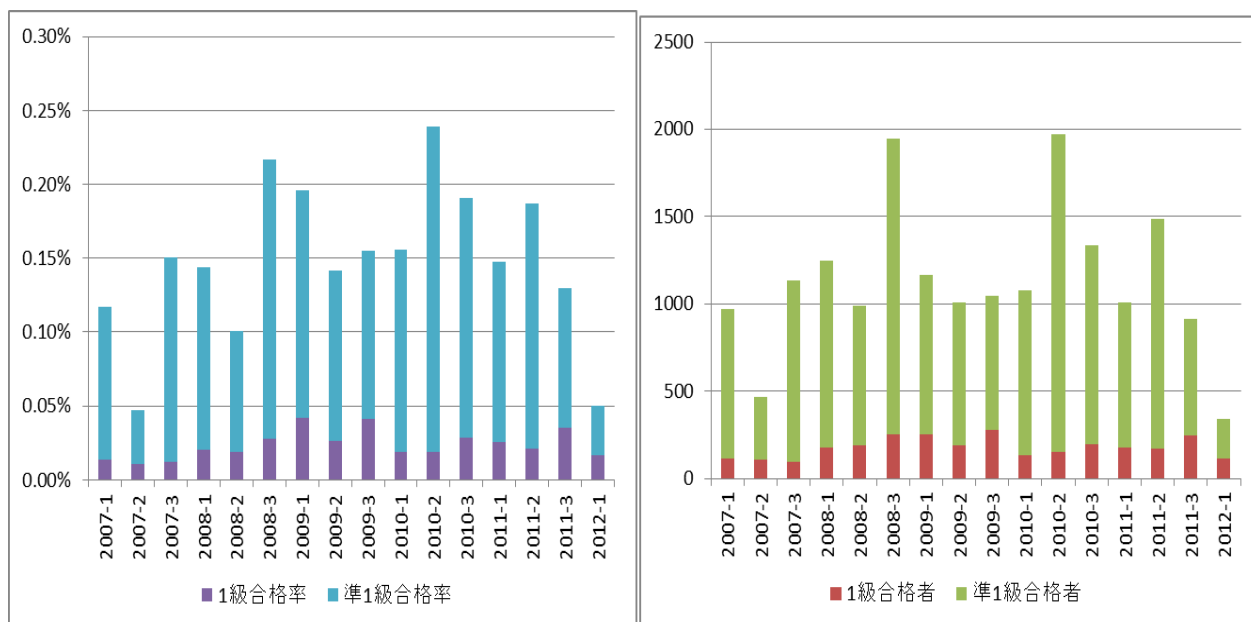
この漢字検定の第 1 回(2007 年 1 回目)から現在までの合格率(受験者に対する 1 級合格率と準 1 級合格率)と合格者数を図表 13 に示す。この結果が示唆することは、1 級と準 1 級を合わせても合格率が 0.25%以下、つまり漢字に対して関心を持って学習している者を母集団としても、JIS 第 1 水準(約 3,000 字)の漢字を使いこなすことができるのは 0.25%以下しか存在しないということである。さらに 1 級の合格率は 0.05%以下、すなわち JIS 第 1 水準と JIS 第 2 水準(約 6,000 字)の漢字を使いこなすことができるのは 0.05%以下しか存在しないのである。

そして、これまでの 1 級合格者と準 1 級合格者を合計しても約 1.8 万人であり、この数字は日本人の人口のわずか 0.014%(1 万人に 1 人)に過ぎない。すなわち、ほとんどの日本人は JIS 第 1 水準(約 3,000 字)の漢字でさえ使いこなすことができず、まして JIS 第 1 水準と JIS 第 2 水準(約 6,000 字)の漢字を使いこなすことができる日本人は奇跡的存在(10 万人に 2 人)といえる。

この事実は重要である。JIS 第 1 水準と JIS 第 2 水準だけでは漢字が不足しているという理由で、近年になって JIS 第 3 水準と JIS 第 4 水準が制定されたが、日本人は JIS 第 1

水準でさえ使いこなすことができないのである。わずか 3,000 字さえ使いこなせない人間が、どうして数万もの漢字を扱うことができるのだろうか。

図表 13 1 級および準 1 級の合格率[母数は全受験者](左)と 1 級および準 1 級の合格者数(右)



(出所：財団法人日本漢字能力検定協会のデータを基に筆者作成)

6. 2 行政手続きで使われる漢字の認識能力について

次に、日本人が行政手続きで扱われる約 60,000 字もの漢字(住基ネット統一文字および戸籍統一文字を含む範囲)に対する認識能力(識別能力)をどれだけ持っているかを確認するため、実験票を用いた実験を行った。実験票は 2 種類を用意し、実験票 A(図表 14 の左)は JIS 第 1 水準の範囲、実験票 B(図表 14 の右)は住基ネット統一文字および戸籍統一文字を含む範囲としている。実験内容は、2 分間で 2 列に並んでいる合計 24 種類の漢字を同定(1 文字あたり 5 秒)するもので、A と B でどれくらい認識率が異なるのかを明らかにした。

なお、この実験票における漢字の大きさは 28 ポイント(約 9.85mm×9.85mm)相当としており、通常の事務で扱う 10.5 ポイントや 12 ポイントよりもかなり大きい。付録図表に原寸大の実験票を添付したが、この大きさの漢字の同定作業で誤認識が起きるのであれば、通常の事務作業ではまず同定が不可能だと言ってもよいだろう。

被験者は自治体職員、民間企業、一般市民の 3 グループを対象として設定した。具体的には、自治体職員は F 市の事務職員、民間企業は(株)富士通総研の幹部社員を対象とした。一般市民については、日本人の平均寿命が女性 85 歳、男性 79 歳であり、今後も高齢者が増加すると予想されているため、平均的日本人としてシニア層を対象とした。公益財団法人川崎市生涯学習財団の協力を得て、生涯学習講座に定期的に参加している市民に被験者にな

ってもらった。それぞれのグループは、一般的日本人の平均以上の教育レベルおよび能力があると想定される集団である。なお、10未満のサンプルは参考値(#)として扱うこととする。

図表 14 実験票 A (左) と実験票 B (右)

巖 ○ 議 ①	顧 ○ 釀 ①	齋 ○ 齋 ①	邊 ○ 邊 ①
懸 ○ 耀 ②	轟 ○ 馨 ②	齋 ○ 齋 ②	邊 ○ 邊 ②
欄 ○ 襲 ③	露 ○ 麟 ③	齋 ○ 齋 ③	邊 ○ 邊 ③
灘 ○ 讓 ④	鰯 ○ 鐘 ④	齋 ○ 齋 ④	邊 ○ 邊 ④
艦 ○ 纏 ⑤	馨 ○ 魔 ⑤	齋 ○ 齋 ⑤	邊 ○ 邊 ⑤
纏 ○ 欄 ⑥	鐘 ○ 驚 ⑥	齋 ○ 齋 ⑥	邊 ○ 邊 ⑥
纂 ○ 巖 ⑦	響 ○ 露 ⑦	齋 ○ 齋 ⑦	邊 ○ 邊 ⑦
議 ○ 艦 ⑧	釀 ○ 騰 ⑧	齋 ○ 齋 ⑧	邊 ○ 邊 ⑧
競 ○ 懸 ⑨	麟 ○ 轟 ⑨	齋 ○ 齋 ⑨	邊 ○ 邊 ⑨
襲 ○ 灘 ⑩	魔 ○ 響 ⑩	齋 ○ 齋 ⑩	邊 ○ 邊 ⑩
讓 ○ 競 ⑪	騰 ○ 顧 ⑪	齋 ○ 齋 ⑪	邊 ○ 邊 ⑪
耀 ○ 纂 ⑫	驚 ○ 鰯 ⑫	齋 ○ 齋 ⑫	邊 ○ 邊 ⑫

(出所：筆者作成)

図表 15 被験者グループのプロフィール(#参考値)

グループ	人数	性別			国籍			
		男性	女性	不明	日本	外国	不明	備考
①自治体	96	80	15	1#	94	0	2#	
②民間企業	60	54	6#	0	59	1#	0	ドイツ
③一般市民	47	40	7#	0	47	0	0	
④合計	203	174	28	1#	200	1#	2#	
グループ	年齢							
	20代	30代	40代	50代	60代	70代	80代	不明
①自治体	5#	16	26	45	3#	0	0	1#
②民間企業	0	3#	35	17	5#	0	0	0
③一般市民	0	1#	1#	2#	24	15	3#	1#
④合計	5#	20	62	64	32	16	3#	1#

(出所：筆者作成)

①自治体グループを対象とした実験結果

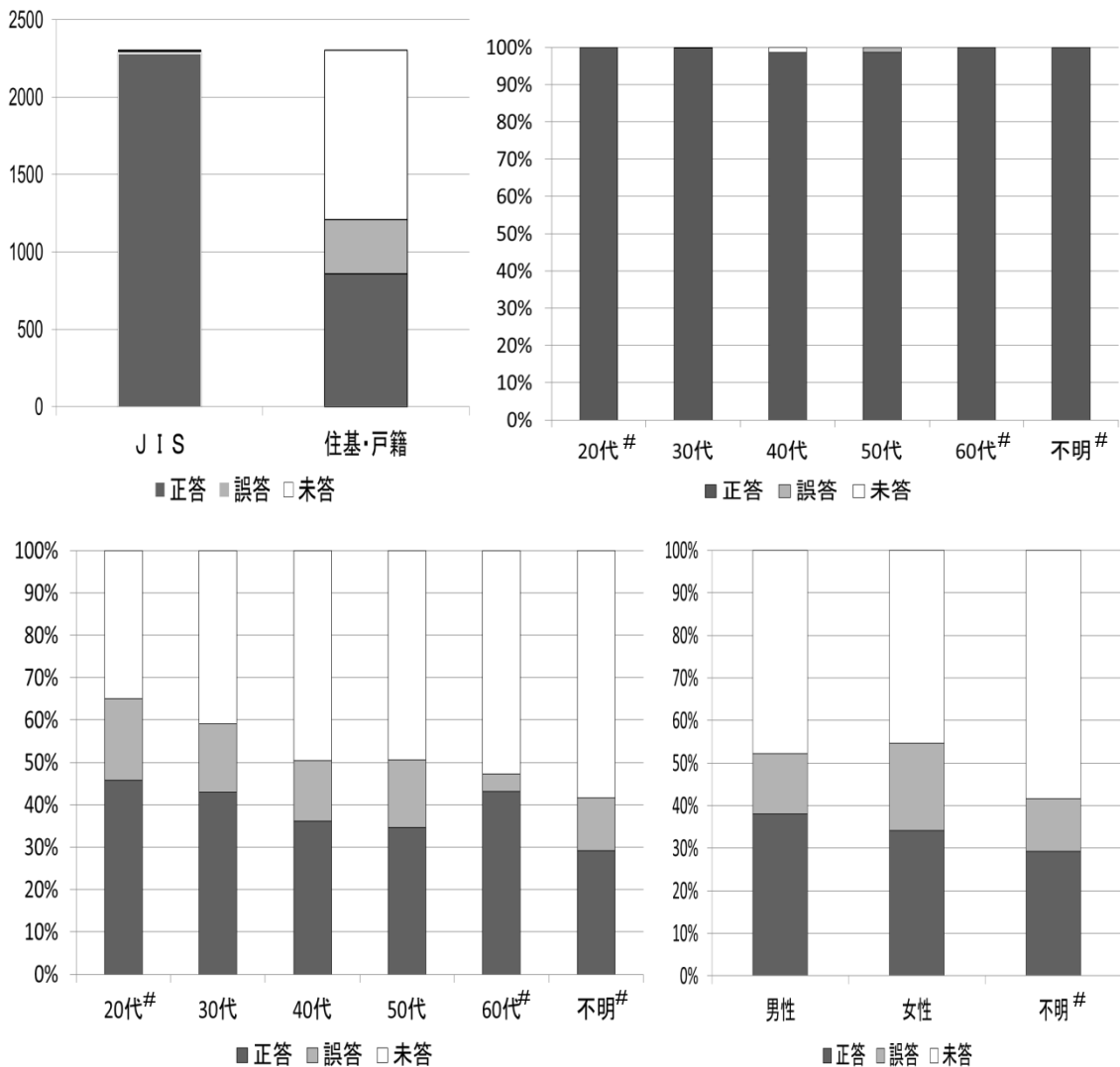
自治体グループを対象とした実験結果を下記に示す。JIS(実験票 A)ではほぼ 100%正確な認識ができているのに対し、住基・戸籍(実験票 B)ではかなり状況が異なっている。半数近くの 47%が時間不足で回答できず、正答が 37%、誤答が 15%と回答の約 1/3 が間違っている。つまり、住基・戸籍(実験票 B)では文字の誤認識が多数発生することを意味している。

年代別の状況を見ても、住基・戸籍(実験票 B)の場合は年代が上がるにつれ、認識速度が遅くなり、かつ正答率も下がっている。

男女別では、女性は男性に比べ、認識速度が速いわりに誤回答が多くなっている。

図表 16 自治体グループの実験結果 (#参考値)

JIS(実験票 A)と住基・戸籍(実験票 B)の結果比較[左上]、実験票 A の年代別結果[右上]、実験票 B の年代別結果[左下]、実験票 B の男女別結果[右下]



(出所：筆者作成)

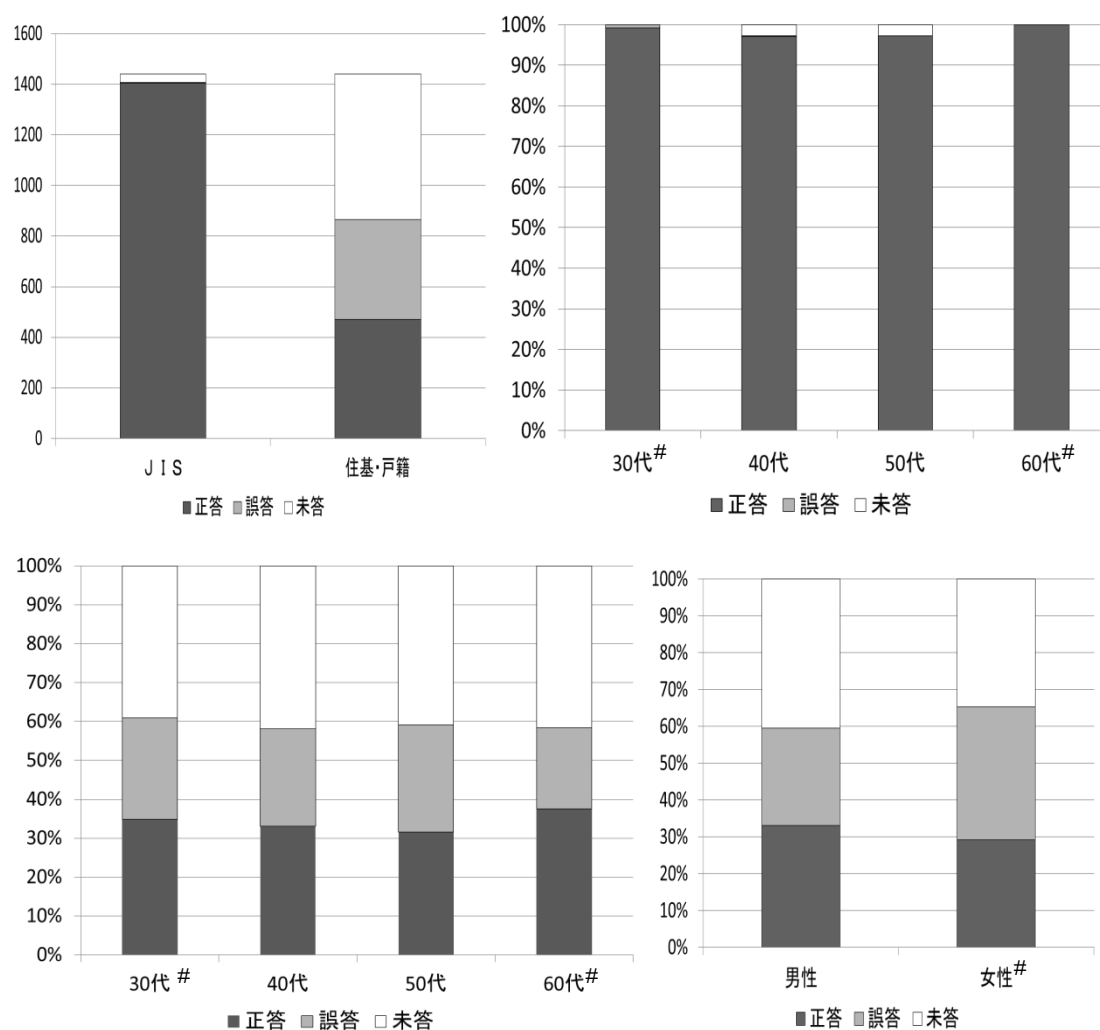
②民間企業グループを対象とした実験結果

民間企業グループを対象とした実験結果を下記に示す。JIS(実験票 A)ではほぼ 100%正確な認識ができているのに対し、自治体グループと同様住基・戸籍(実験票 B)ではかなり状況が異なっている。40%が時間不足で回答できず、正答が 33%、誤答が 27%と回答の半数近くが間違っている。つまり、住基・戸籍(実験票 B)では約半数の誤認識が発生することを意味している。自治体に比べて速度は速いが、認識が粗いといえる。

住基・戸籍(実験票 B)において、年代別の状況では自治体グループと異なり、差はあまり見られない。男女別では、女性は男性に比べ、認識速度が速いわりに誤回答が多くなっているのは自治体と同じである。

図表 17 民間企業グループの全体結果 (#参考値)

JIS(実験票 A)と住基・戸籍(実験票 B)の結果比較[左上]、実験票 A の年代別結果[右上]、実験票 B の年代別結果[左下]、実験票 B の男女別結果[右下]



(出所：筆者作成)

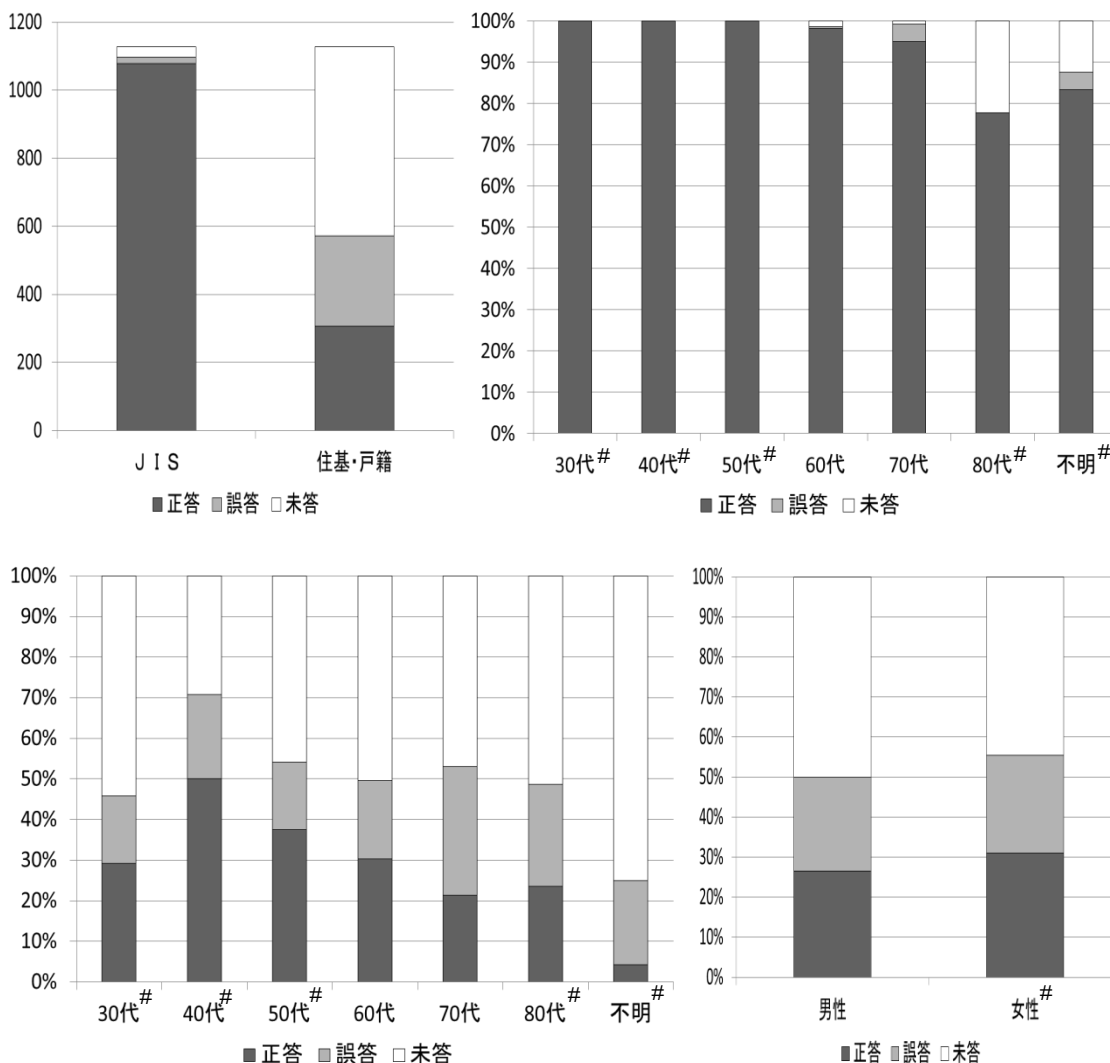
③一般市民グループを対象とした実験結果

一般市民グループを対象とした実験結果を下記に示す。JIS(実験票 A)では、70代や80代では速度が落ちたり、誤認識が発生したりしているものの、全体としてはほぼ100%近くが認識できている。住基・戸籍(実験票 B)では、49%が時間不足で回答できず、正答が27%、誤答が24%と回答の半数が間違っている。つまり、住基・戸籍(実験票 B)では半数の誤認識が発生することを意味している。

住基・戸籍(実験票 B)において、年代別の状況では年代が上がるにつれ、認識速度の低下や誤認識が認められる。

図表 18 一般市民グループの実験結果 (#参考値)

JIS(実験票 A)と住基・戸籍(実験票 B)の結果比較[左上]、実験票 A の年代別結果[右上]、実験票 B の年代別結果[左下]、実験票 B の男女別結果[右下]



(出所：筆者作成)

④3 グループの実験結果のまとめ

3 グループの実験結果を図表 19 に示す。この結果からわかることは、実験票 A の範囲ではほとんどの被験者が 100% 近く正解を記入できるのに対し、実験票 B ではまったく異なった結果が得られている。時間不足のため約 4 割以上が未回答であり、記入された回答についても半分近くが誤って回答されている。つまり、実験票 A を完璧に回答できる日本人でも、実験票 B においては認識の速度が落ちるとともに、誤認識が多くなるという事実が明確になった。つまり、住基ネット統一文字および戸籍統一文字を通常の PC やコンピュータで扱えるようにしても、文字の認識速度が極端に低下し、そして半数近くを誤認識してしまうということが明らかになったのである。

さらに付け加えるなら、今回の実験では 28 ポイント相当の文字を使って同定作業をしてもらった。通常の事務で扱う文字の大きさは 10.5 ポイントや 12 ポイントであり、このような文字を使って実験をすれば、今回の実験よりもさらに認識速度は低下し、誤認識率は飛躍的に高くなるだろう。つまり、住基ネット統一文字および戸籍統一文字を通常の PC やコンピュータで扱えるようにすると、ある条件下では文字の誤認識が頻繁に発生し、社会的混乱を招くことにもつながるのである。

ここで 3 グループの特徴を分析しておく、自治体グループはさすがに行政事務を担っているだけあって、速度が遅くなっても正確に同定しようとしている。民間企業グループでは正確さよりも速度を重視した同定のしかたである。市民グループは平均年代が高いこともあり、認識速度が遅くなり、約半分を誤認識している。

図表 19 3 グループの実験結果

グループ	被験者数 (平均年代)	実験票 A の結果 (%)			実験票 B の結果 (%)		
		正答	誤答	未答	正答	誤答	未答
自治体	96 (40 代)	99	1	0	37	15	47
民間企業	60 (50 代)	98	0	2	33	27	40
市民	47 (60 代)	96	2	3	27	24	49
合計	203	98	1	1	34	21	46

(出所：筆者作成)

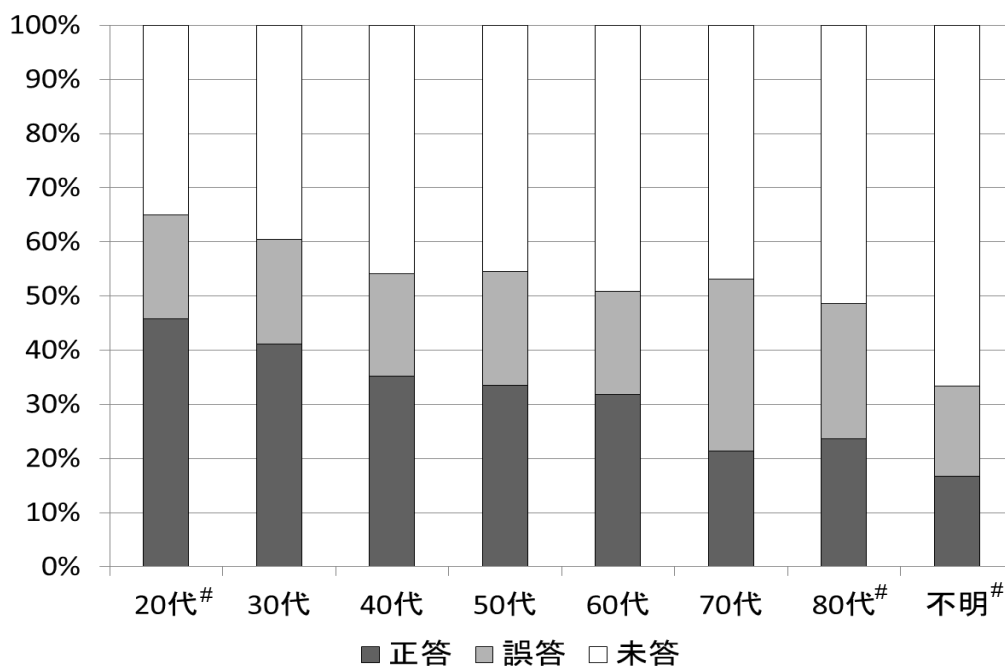
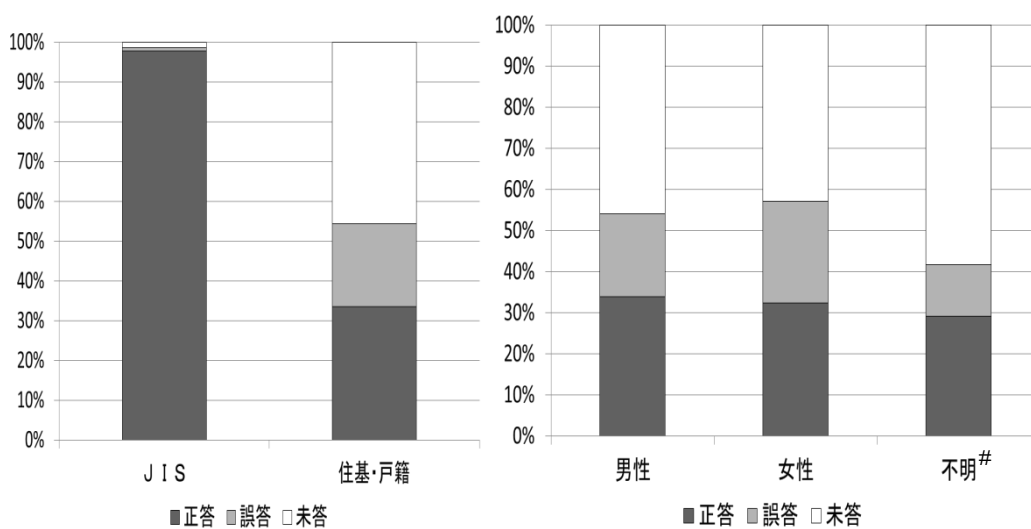
次の図表は、3 グループを合計して作成したグラフである。男女差についてはあまり明確な差はでなかったものの、女性は男性に比べ、認識速度が速い代わりに、誤認識が多いという傾向を持っている。

また、年代別のグラフを作成してみると、年代が上がるにつれ、認識速度が低下し、文字の誤認識も多くなっていることがわかる。つまり、文字数の多さは高齢者にとって負担となり、JIS 第 1 水準と第 2 水準の範囲でさえ 70 代や 80 代には負担のようである。高齢社会において、これ以上文字を増やすことは、一般国民の負担を増やすだけにしかならな

いと思われる。

図表 20 3 グループ合計の実験結果 (#参考値)

JIS(実験票 A)と住基・戸籍(実験票 B)の結果比較[左上]、実験票 B の男女別結果[右上]、実験票 B の年代別結果[下]



(出所：筆者作成)

なお、実験票 B の最高得点は自治体の 50 代男性で、正答が 20、誤答が 4 という結果であった。市民課などの経験があれば、このような外字の同定に慣れていると思われるが、

それでも4つの文字を取り違えている。

また、今回の実験では外国人被験者が1人だったため、外国人の分析はできなかった。参考まで結果を見ると、外国人にとって漢字はかなり大きなハンディキャップであると想定され、実験票Aと実験票Bの結果があまり変わらなかった。両方とも図形として認識しようとしていたようだ。グローバル化する日本社会において、漢字の増大が外国人を排斥するようなことになってはならないと考える。

補足しておくが、国民の誤認識だけが問題なのではなく、政府もまた混乱を引き起こしている。戸籍担当者の間では有名な話だそうだが、「龍」という字の「立」の上部の棒が立っているのか横になっているのかという違いで、戸籍の現場を混乱させているという。常用漢字表に掲載されているのは龍(棒が縦)という字だが、人名用漢字(戸籍法施行規則別表第二に掲載)として子どもの名付けに使えるのは龍(棒が横)という字なのである。法務省の戸籍統一文字情報¹⁸を検索してみればわかるが、「子の名に使える漢字」という条件で検索すると龍(棒が横)しか表示されない。

常用漢字表をよく見ると「竜(龍)」と表記されており、戸籍法施行規則第六十条では常用漢字表に掲載されている漢字を子の名に使えるとしているが、「括弧書きが添えられているものについては、括弧の外のものに限る」という条件が付いている。すなわち龍という字を子の名に使うことはできないのである。「パソコンで表示できる字を子どもの名付けにせず、パソコンで表示できない字を使えとはどういうことか」という議論はネット上でも問題になっている。なぜこのような細部にこだわり、ここまで厳密に区別する必要があるのだろうか。

6. 3 人間と機械のインタフェース

本章の冒頭で、人間対人間では問題にならなかった事象が、機械対機械や人間対機械という関係では大きな問題となって立ちはだかると述べた。これについて考察したい。

漢字という図形が人間対人間の間で使われる場合、厳密なパターンマッチングが行われるわけではなく、一定の揺らぎを許容した認識が行われる。例えば、元首相の菅直人氏の「直」という字では、上部が「十」となっているのが正字であり、「ナ」となっているのが外字だとされる。しかし、下記のように字の書体によって「十」となったり「ナ」となったりしているものの、人間の認識としては「直」であり、ことさら異なる字としては認識しない。

直(明朝体)、直(教科書体)、直(行書体)、直(楷書体)

付録図表の実験票Bを見てもらいたいが、文字を同定できたという回答のうち半数近くが誤認識をしている。厳密なパターンマッチングでは確かに誤認識であるが、人間の目は

¹⁸ <http://kosekimoji.moj.go.jp/kosekimojidb/mjko/PeopleSearch/EXECUTE>

「書体、書き癖、筆使い、書き間違い」などを許容しながら認識するため、本人に誤認識という自覚は無い。自治体職員は実務上の経験から民間人よりも誤認識率が低い、やはり間違えた本人に誤認識という自覚は無いだろう。漢字という図形が人間対人間の場面で使用される場合、一定の揺らぎを許容した認識が行われるからである。もし自治体職員に「厳密なパターンマッチング」を要求するなら、彼らは市民からの申請書をことごとくつき返すことになるだろう。

それでは漢字という図形が機械と機械の間で使用される場合はどうか。漢字は図形であるが、文字としてのコードを持っており、図形のパターンマッチングは行わない。つまり、図形とは関係なく、コードのマッチングを行うのみである。実験票 A で示された漢字も実験票 B で示された漢字も、コンピュータは同じスピードで 100%完璧にマッチングできる。換言すれば、人間の目にはほとんど同じに見える漢字であっても、コードが 1 ビットでも異なればまったく異なる文字として扱うことになる。

このような事実から、実験票 B で示された漢字を扱う場合、人間対機械、人間と機械のインタフェースではどのような問題が生じてくるだろうか。まず、人間が「斎」や「邊」という漢字を入力する場合、膨大な候補リストから一つを選び出さなければならない。実験結果が示すように、半数は誤った漢字を選択するだろう。つまり、同一人物でありながら機関 A で入力された氏名と機関 B で入力された氏名のコードが異なり、コンピュータはまったく異なるものという認識をすることになる。人間どうしの間では一定の揺らぎを許容して「同一だ」と認識されるものが、コンピュータは「同一ではない」と判断する。人間とコンピュータの認識の相違が如実に生じてくるのである。

人間とコンピュータの認識や能力には大きな相違があり、コンピュータが億単位の文字が扱えるようになったからといって、普通の人間が日常的に扱えるものではないことをはっきりと認識すべきだろう。外字の問題についても、文字を増やすという議論ではなく、人間とコンピュータの認識や能力の相違という観点から解決策を見出していかななくてはならない。

7. 結論

7. 1 問題の整理

これまで指摘した問題を以下に整理した。

- ① 住基ネット統一文字と戸籍統一文字を含む約 6 万字について、文字情報基盤を拠りどころに整理し、UTF-16 と IVS/IVD で対応していても、市町村における「不明文字」約 5 万字に対応できず、依然として多くの外字が解決できない。また、IVS/IVD についても、類字字形の判断基準や登録権限が各省庁で統一されていないばかりか、「類字」をどのような場面でどのように使うのかが示されていないため、結局は文字の数を増やすことと同義になってしまっている。

- ② 外字の存在によって、市町村だけでも年間約 20 億円の経済損失が発生しており、民間等も含めれば日本社会全体で 30 億円くらいの経済損失になると推定される。金額的にはそれほど大きくないという見方があるかもしれないが、デジタル情報の流通を阻害し、情報化社会の健全な発展を阻んで日本社会全体のパフォーマンスを低下させることを考えれば、その重大さについて真剣に受け止めるべきである。また、外字の使用を強硬に主張する人は、ほんのわずかであると考えられ、一部の人のために日本社会全体がこのコストを負担すべきかどうか議論する必要がある。
- ③ 現在のコンピュータに標準的に実装されている JIS 第 1 水準・第 2 水準の範囲の漢字(約 6,000 字)でさえ、日本国民は使いこなせない。漢字検定の結果からは、1 級(約 6,000 字の範囲)の合格者は常に全受験者の 0.05%未満という数字が明らかになっている。まして、日本国民が住基ネット統一文字と戸籍統一文字を含む約 6 万字を使うことは不可能である。
- ④ JIS 第 1 水準の範囲の漢字を完璧に認識できる日本人であっても、住基ネット統一文字と戸籍統一文字を含む約 6 万字の範囲では、認識速度が落ちるばかりか、漢字の誤認識が多数発生する。実験票では、通常の事務で使う文字の 2 倍以上の大きさの文字を使っているため、通常の事務では実験結果以上の認識速度の低下と誤認識の発生が想定される。

これまで漢字の問題は文化・精神の問題であり、経済性や効率性などの観点から議論されるべきものではないとされてきた。しかし、従来人間対人間において使われていた漢字が、IT が社会に広く浸透して人間対機械という場面で使われてくると新たな問題が生じてくる。

人間対人間の場合は、漢字の曖昧さを一定の揺らぎのなかで許容し、あるときは交渉によって漢字の処理が行われる。しかし、機械は 1 ビット異なるだけでも曖昧さを許容せず、コードが異なればまったく異なった文字として処理される。IVS で類似した漢字を処理するとしても、異体字登録の判断や同一と見なす文字デザインの判断は人間が与えなければ処理できない。この問題は、人間対機械の場面において、相互の認識の相違として顕著に出現してくる。機械が同一でないと判断したものを人間は同一だと判断し、人間が同一だと判断したものを機械は同一ではないと判断する。

さらに、一般の日本人は JIS 第 1 水準・第 2 水準の漢字でさえ使いこなすことができず、住基ネット統一文字と戸籍統一文字では誤認識が多数発生するという事実は重要である。行政手続きは、日常的に迅速かつ正確に行われるべきであり、住基ネット統一文字や戸籍統一文字を使う合理性がない。漢字の認識速度の低下や誤認識は、行政手続きに混乱をもたらし、日本社会全体のパフォーマンスを著しく低下させることになるだろう。政府の報告書が示唆するようなコンピュータで対応できる文字数を増やす方法が、解決の手段とならないことは明白である。

そこで、次の2つの解決案を提示したい。

7. 2 問題解決のための提案1

(案1)行政手続きで使用する漢字（氏名や地名など）をJIS第1水準とJIS第2水準に制限し、それ以外の漢字の使用を法律で禁止する。現状でそれ以外の漢字を使っている場合は、JIS第1水準とJIS第2水準の範囲内の類似した漢字に置き換え、置き換え不可能な漢字については「ひらがな」または「カタカナ」に置き換えることとする。なお、氏名漢字の置き換えについては、本人の同意を得ることが望ましいが、不可能である場合は職権によって行うこととする。

漢字検定のデータから判断すれば、JIS第1水準あるいはもう少し狭い範囲が望ましいと思われるが、すでにPCなどでJIS第2水準が広く普及していること、JIS第3水準とJIS第4水準はまだ普及していないことを鑑みて、JIS第1水準とJIS第2水準の範囲とした。

また、強硬に外字の使用を主張する人に対しては、職権で置き換えができるよう法律に明記すべきである。総務省報告書¹⁹では、「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」との記載があるが、2000年の地方分権一括法の施行によって、通達の法的拘束力が無くなっているからである。

自治体の現場では、誤字であっても外字の使用を主張する市民と対峙せざるを得ず、それを退けるための法的な根拠が無いため、外字を作成せざるを得ない。自治体のヒアリング事例で見たように、福祉業務ではさほど問題にならない外字が、税業務になると極端に神経を使っている。その背景には、福祉のように給付を受ける手続きでは文句を言わないが、税金を納付するという場面では、文字の違いを理由に自治体の職員に不満をぶつけるという住民のエゴがある。このような住民を納得させるためにも、法的な強制力が是非とも求められる。

7. 3 問題解決のための提案2

(案2)行政手続き(戸籍業務を除く)で使用する漢字(氏名や地名など)をJIS第1水準とJIS第2水準に制限し、それ以外の漢字の使用を法律で禁止する。現状でそれ以外の漢字を使っている場合は、JIS第1水準とJIS第2水準の範囲内の類似した漢字に置き換え、置き換え不可能な漢字については「ひらがな」または「カタカナ」に置き換えることとする。なお、氏名漢字の置き換えについては、本人の同意を得ることが望ましいが、不可能であ

¹⁹ 富士ゼロックス株式会社(2012)

る場合は職権によって行うこととする。

案1に対して、戸籍業務のみ除外するという案である。定量的な調査結果が無いため要求の実態は良くわからないが、国民のなかには先祖代々の文字で自分のアイデンティティを保持したいという要求があるようだ。もしこのような要求が正当であるとして、国が認めるならば、その要求に対しては戸籍業務でのみ対応することとする。そして戸籍以外の業務については、住民基本台帳に記載された文字を使うこととする。

なお戸籍と住民基本台帳で氏名漢字が異なることになるが、個人の同一性については住民票コード（またはマイナンバー）を戸籍に記載することで確保する。

改めて付け加えておくが、より多くの漢字をコンピュータで表現しようという努力が間違っていると言いたいのではない。古文書に記載された現在では使われていない漢字、未解明の文字など、文化的な研究のためには何十万字、何百万字という文字を扱うことも専門的な分野では必要であろう。筆者が主張したいのは、日常的に迅速かつ正確に事務手続きを行う行政手続きおよびそれに関連する民間企業の手続きにおいて、そのような専門分野の文字コードを持ち込むべきではない、そのような技術の使い方は公益に反するということなのである。

8. おわりに

長年、自治体の現場でシステムを開発してきた筆者は、「本人でさえ日常生活で使わないような字を使うことは不合理だ」、「何万もある漢字を正しく使うことなど不可能だ」という自治体職員の悲痛な声をずっと聞いてきた。そして民間企業でも、社員の氏名を二重管理するという、信じられないようなデータベース管理を強要されている。漢字は日本の文化・精神だと主張する人たちは、このような自治体や民間企業の現場の実態を知らないのではないだろうか。

筆者は決して日本の文化や精神を破壊しろと主張しているわけではない。代々家に伝わる文字を大切にすることは大事であろうし、プライベートで使うことを禁じようとしているわけではない。行政の手続きなど、日常的な場面で使うことを制限したいと主張しているだけなのである。

年間 30 億円のコスト負担など大した金額ではないと主張する人がいるかもしれない。しかし、このような棘(トゲ)がデジタル情報の流通を妨げ、あちこちで引っかかるような事態となれば、健全な情報化社会の発展を阻害し、日本社会全体のパフォーマンスを低下させることになる。「PC で外字が出てこないから、多くの漢字のなかから外字を探すのが大変だから、電子申請より紙のほうが楽だ」ということになり、日本の負の循環は始まっていく。実際にそのような事態になっているからこそ、日本の電子政府ランキングが低下し

続けているとも言える。

外字とは、デジタル情報の流れに引っかかる棘(トゲ)のようなもの。情報化社会の棘(トゲ)と言っても良い。大した痛みがないからと言って、今ここで棘(トゲ)を抜かないで放っておくと、やがて体内をめぐって心臓に達し、心臓を突き刺すことになりかねない。今のうちに早く棘(トゲ)を抜いて処置すべきである。民間企業は秒単位の時間や銭単位のコストを削減しながら、これまで生き延びてきた。行政もそのような考え方を取り入れなければ、生き残ることはできないことを肝に銘じるべきではないだろうか。

【追補 1】

「6.2 行政手続きで使われる漢字の認識能力について」で触れた「龍」の字に関する問題について、補足しておきたい。

右は小池和夫(2007) P96 の「龍」の字に関する記載部分である。この文章には、下記 2 点の間違ひがある。

・『常用漢字表』に記載されている「竜」の旧字は龍ではなく龍である。

・「戸籍統一文字」で子どもの名付けに使えているのは龍ではなく龍である。

しかし、この著者は、日本規格協会符号化文字集合調査研究委員会の委員として JIS X0213 の制定に携わった専門家であり、右に添付したものは第 2 刷である。つまり、単純な間違いや誤植であるとも思えない。

この著者の文章が本来正しかったとすると、下記のようなことが考えられる。

- ・政府が『常用漢字表』の内容を後にすり替えた。
- ・政府が「戸籍統一文字」のミスを後に修正した。

筆者がここで指摘したいのは、「誰が間違えたのか」ということではない。文字の専門家および政府でさえ間違えるような漢字の区別を、なぜ一般国民に強いる必要があるのかということなのである。行政や民間企業の実務の現場では、「謎がまた一つ増えました」ということでは済まされない問題なのである。

法務省によれば、子どもの名付けに使える「竜」の旧字は、龍です。これは『常用漢字表』に括弧付きで載っている字と同じ形です。一方、JISに示され、一般のフォントとして出回っているのは龍です。「戸籍統一文字」ではこの二つを区別しています。いや、不思議なことに、官報の『戸籍法施行規則別表第二』（人名用漢字）に載っているのは龍ですが、「戸籍統一文字」で子どもの名付けに使えているのは龍なのです。謎がまた一つ増えました。

【追補2】

もう一人の専門家の著書『漢字百珍 日本 の異体字入門』²⁰の一部を下記に示した。著者の肩書きは早稲田大学名誉教授（文学博士）であり、他に『あて字用例辞典』や『現代語源小辞典』などの専門書を著している。

『漢字百珍』P178-179 では下記のような記述があり、国語審議会に対する辛らつな意見が述べられている。時代に即して合理的に物事を考え、外字問題を解決していくのは、本来国語審議会の役割ではないだろうか。国語審議会の存在意義が問われていると言っても良いだろう。

しかし、現代中国の簡体字は、いきあたりばったりの権威主義をふりかざす日本の常用漢字とは違って、システムに一貫性があります。同じ構造のものは繁→簡と同じように簡略化されていて、中国人のほうが日本人より合理性に富んでいるようです。たとえば、〈獨→独〉として、〈旁の（蜀）を（虫）と置換するのを認めるならば、〈焔（燭）〉も〈焔（濁）〉も正式な字として認めるわけです。日本のように、個人的主観などから〈独〉はよくても〈焔〉や〈焔〉はいけないといった不合理な取りきめなどする必要はありません。〈区（區）〉を認めるならば〈欧〉も〈謳〉も〈呕〉もいいはずだし、〈鷗（鷗）〉も認めるわけで、この方針こそ当然ではありませんか。〈鷗〉は非という馬鹿げた規制は、はじめから必要ないわけです。漢字は本来的に、こうした合理性をもって創作されているのです。そうでなければ、この地球上は異なる漢字で埋めつくされてしまうでしょう。〈鷗〉なども、国語審議会の委員の無知ゆえに長年認められず、私生児のように新聞やコンピューター上に出没しているわけです。こうした不条理な扱いこそ、漢字への無知を自ら告白しているようなものではありませんか。審議の資料や経緯の情報を是非とも公開してほしいと思います。

²⁰ 杉本つとむ(2001)

【参考文献一覧】

- ・ EGDI (2012), United Nations E Government Survey 2012,
<http://unpan1.un.org/intradoc/groups/public/documents/un/unpan048065.pdf>
- ・ 榎並利博(2000)、『自治体の IT 革命』東洋経済新報社
- ・ 円満字二郎(2005)、『人名用漢字の戦後史』岩波新書
- ・ 大分県商工労働部情報政策課(2010)、「大分県の自治体クラウドへの取組」2010年2月18日発表資料
- ・ 経済産業省・独立行政法人情報処理推進機構(IPA)(2011)、「文字情報基盤構築に関する研究開発事業調査報告書」2011年3月25日
- ・ 小池和夫(2007)、『異体字の世界 旧字・俗字・略字の漢字百科』河出文庫
- ・ 島田達巳・榎並利博(2010)、「韓国電子政府・電子自治体事情(後編)」『月刊 LASDEC』2010年10月号
- ・ 杉本つとむ(2001)、『漢字百珍 日本の異体字入門』八坂書房
- ・ 奈良県河合町総務課(2012)、奈良県基幹システム共同化検討会「自治体クラウド導入の効果と期待～奈良県における取り組みから～」2012年2月9日発表資料
- ・ 深沢千尋(2011)、『文字コードの超研究 改訂第2版』株式会社ラトルズ
- ・ 富士ゼロックス株式会社(2012)、「市区町村が使用する外字の実態調査報告書(平成23年度総務省請負調査)」2012年3月
- ・ 矢野啓介(2010)、『プログラマのための文字コード技術入門』技術評論社

【参考サイト一覧】

- ・ 財団法人日本漢字能力検定協会
<http://www.kanken.or.jp/index.php>
- ・ 文字情報基盤構築に関する研究開発事業 調査報告書付属資料
<http://ossipedia.ipa.go.jp/article/31/>
- ・ 平成22年度電子経済産業省推進費(文字情報基盤構築に関する研究開発事業) 委員会議事次第・配布資料・議事録一覧
<http://ossipedia.ipa.go.jp/article/9/>
- ・ IPA MJ 文字情報検索システム(簡易版)
http://ossipedia.ipa.go.jp/ipamjfont_glyph/html/GldVisitor/GldVisitorGlyphSearchBasic.html?clear=true

巖	<input type="radio"/>	議	①	顧	<input type="radio"/>	釀	①
懸	<input type="radio"/>	耀	②	轟	<input type="radio"/>	馨	②
欄	<input type="radio"/>	襲	③	露	<input type="radio"/>	麟	③
灘	<input type="radio"/>	讓	④	翳	<input type="radio"/>	鐘	④
艦	<input type="radio"/>	纏	⑤	馨	<input type="radio"/>	魔	⑤
纏	<input type="radio"/>	欄	⑥	鐘	<input type="radio"/>	驚	⑥
纂	<input type="radio"/>	巖	⑦	響	<input type="radio"/>	露	⑦
議	<input type="radio"/>	艦	⑧	釀	<input type="radio"/>	騰	⑧
競	<input type="radio"/>	懸	⑨	麟	<input type="radio"/>	轟	⑨
襲	<input type="radio"/>	灘	⑩	魔	<input type="radio"/>	響	⑩
讓	<input type="radio"/>	競	⑪	騰	<input type="radio"/>	顧	⑪
耀	<input type="radio"/>	纂	⑫	驚	<input type="radio"/>	翳	⑫

齋	○	齋	①
齋	○	齋	②
齋	○	齋	③
齋	○	齋	④
齋	○	齋	⑤
齋	○	齋	⑥
齋	○	齋	⑦
齋	○	齋	⑧
齋	○	齋	⑨
齋	○	齋	⑩
齋	○	齋	⑪
齋	○	齋	⑫

邊	○	邊	①
邊	○	邊	②
邊	○	邊	③
邊	○	邊	④
邊	○	邊	⑤
邊	○	邊	⑥
邊	○	邊	⑦
邊	○	邊	⑧
邊	○	邊	⑨
邊	○	邊	⑩
邊	○	邊	⑪
邊	○	邊	⑫

研究レポート一覧

No.400	電子行政における外字問題の解決に向けて －人間とコンピュータの関係から外字問題を考える－	榎並 利博 (2013年2月)
No.399	中国の国有企業改革と競争力	金 堅敏 (2013年1月)
No.398	チャイナリスクの再認識 －日本企業の対中投資戦略への提言－	柯 隆 (2012年12月)
No.397	インド進出企業の事例研究から得られる示唆	長島 直樹 (2012年10月)
No.396	再生可能エネルギー拡大の課題 －FITを中心とした日独比較分析－	梶山 恵司 (2012年9月)
No.395	Living Lab(リビングラボ) －ユーザー・市民との共創に向けて－	西尾 好司 (2012年9月)
No.394	ドイツから学ぶ、3.11後の日本の電力政策 ～脱原発、再生可能エネルギー、電力自由化～	高橋 洋 (2012年6月)
No.393	韓国企業の競争力と残された課題	金 堅敏 (2012年5月)
No.392	空き家率の将来展望と空き家対策	米山 秀隆 (2012年5月)
No.391	円高と競争力、空洞化の関係の再考	米山 秀隆 (2012年5月)
No.390	ソーシャルメディアに表明される声の偏り	長島 直樹 (2012年5月)
No.389	超高齢未来に向けたジェロンとロジー(老年学) ～「働く」に焦点をあてて～	河野 敏鑑 (2012年4月) 倉重佳代子
No.388	日本企業のグローバルITガバナンス	倉重佳代子 (2012年4月)
No.387	高まる中国のイノベーション能力と残された課題	金 堅敏 (2012年3月)
No.386	BOP市場開拓のための戦略的CSR	生田 孝史 (2012年3月)
No.385	地域経済を活性化させるための新たな地域情報化モデル －地域経済活性化5段階モデルと有効なIT活用に関する研究－	榎並 利博 (2012年2月)
No.384	組織間の共同研究活動における地理的近接性の意味 －特許データを用いた実証分析－	齊藤有希子 (2012年2月)
No.383	企業集積の効果 －マイクロ立地データを用いた実証分析－	齊藤有希子 (2012年2月)
No.382	BOPビジネスの戦略的展開	金 堅敏 (2012年1月)
No.381	日米におけるスマートフォンの利用実態とビジネスモデル	田中 辰雄 (2012年1月) 浜屋 敏
No.380	「エネルギー基本計画」見直しの論点 －日独エネルギー戦略の違い－	梶山 恵司 (2011年11月)
No.379	ロイヤルティとコミットメント －百貨店顧客の評価に基づく実証分析から－	長島 直樹 (2011年10月)
No.378	中国経済の行方とそのソブリンリスク	柯 隆 (2011年10月)
No.377	Startup Acceleratorの現状と展望 －変化する起業の形から考える今後のICTビジネス－	湯川 抗 (2011年9月)
No.376	生物多様性視点の地域成長戦略	生田 孝史 (2011年8月)

<http://jp.fujitsu.com/group/fri/report/research/>

研究レポートは上記URLからも検索できます



富士通総研 経済研究所

〒105-0022 東京都港区海岸1丁目16番1号 (ニューピア竹芝サウスタワー)
TEL.03-5401-8392 FAX.03-5401-8438
URL <http://jp.fujitsu.com/group/fri/>