

国立国語研究所学術情報リポジトリ

The method to identify the Japanese e-government system characters with the kanji dictionary characters and the missing kanji dictionary characters

| | |
|-------|--|
| メタデータ | 言語: jpn 出版者: 公開日: 2019-03-25 キーワード (Ja): キーワード (En): 作成者: 高田, 智和, TAKADA, Tomokazu メールアドレス: 所属: |
| URL | https://doi.org/10.15084/00002218 |

行政用文字の調査研究における文字同定

——辞書同定と辞書非掲載字に対する 文献資料・非文献資料による同定——

高田 智和
(国立国語研究所)

キーワード

行政用文字, 漢字字体, 文字同定

要 旨

汎用電子情報交換環境整備プログラムにおいて、行政用文字に対する文字同定では、7種の漢和辞典・漢字字典との照合による第1次同定、次いで、第1次同定で同定できなかった文字（辞書非掲載字）に対して、文献資料・非文献資料の実用例に基づく第2次同定という手順を設けた。行政用文字は、デザイン上の違いをもって「別の字」として扱うことがあるため、字体レベルではなく、字形レベルで辞書との照合を行った。また、辞書非掲載字の検討では、地名用例や「景観文字」用例を、文字同定のための資料とした。

1. はじめに

本稿は、国立国語研究所において進めている行政用文字に関する調査研究で行った文字の同定における経験をもとに、漢字字体研究の対象と方法について考えていく。具体的には次の2点について述べる。

- (1) 「同じ字」「別の字」とはどういうことなのか
- (2) 辞書にない文字を同定するためにはどうすればよいのか

2. 汎用電子情報交換環境整備プログラム

2.1. 行政用文字の調査研究

本稿で、「行政用文字」とは、今後、より高度の電子政府を実現していくにあたり、行政情報処理で必要とされる文字の集まりを指す。国立国語研究所では、経済産業省の委託を受け、2002年度から「汎用電子情報交換環境整備プログラム」を情報処理学会と日本規格協会と共同で進めている¹。これまでの7年間の調査研究において、住民基本台帳ネットワークシステムに搭載されている統一文字（以下、住基統一文字）21,039字（漢字19,432字）、戸籍の電算化に使うことができる戸籍統一文字約55,040字（漢字55,267字）、登記簿の電算化に使うことができる登記統一文字約68,067字（漢字65,597字）を集め、字典類との照合同定を行い、個々の文字につ

いて部首、画数、読み、文字コードなどの文字情報を整理・集積している（高田 2005, 2006a, 2006b；高田他 2008）。

2.2. 文字の同定

文字の同定とは、個々の文字が一体何をあらわす文字であるのかを確定する極めて基礎的な検討作業である。具体的には、現在市販されている中型規模以上の漢和辞典を用いて、それぞれの文字が辞書の見出し字にあるのかないのかわり分けていく。これが第1次同定である。続けて、辞書にない文字について、辞書以外の文献等を精査して、使用例の有無を確認し、第2次同定を行う。これは、JIS X 0208（通称「JIS 漢字」）の原典拠調査の方法（笹原 1996）に倣ったものである。文字同定の流れを図1に示す。

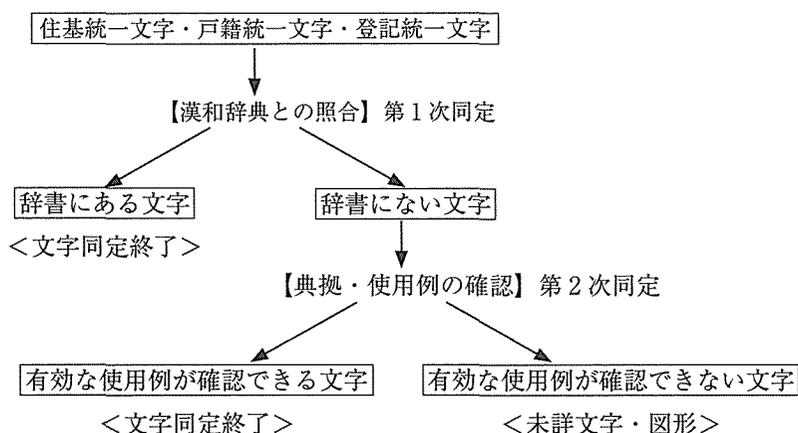


図1 文字同定の流れ

以下、第1次同定に関して、辞書掲出字と検討対象文字との間で、文字の形に違いがある場合、その違いをどのように扱うかという問題、また、第2次同定に関して、辞書にない文字の同定のために、辞書に代わる資料としてどのようなものがあるのかという問題について、順に述べていく。

3. 「同じ字」「別の字」とはどういうことなのか

3.1. 文字の形の単位

例えば、住基統一文字には (A) (B) のよく似ている2種類の文字が採録されている。第1画目が縦画か点かの違いがある。

(A) 𠄎 [住基 5E30] (B) 𠄎 [住基 B0AA]

『大漢和辞典』の掲出字は(C)である。(B)の住基統一文字とよく似ている。(B)は『大漢和辞典』にある文字としてよいだろうが、(A)は『大漢和辞典』にある文字としてよいのだろうか。そ

もそも、(A) と (B) (C) との形の違いを、どのように考えるべきなのだろうか。

(C) 歸 [大漢和 8930]

一つの解答として、常用漢字表の「(付) 字体についての解説 第1 明朝体活字のデザインについて」が参考になる。ここでは、(A) と (B) (C) との形の違いを活字設計上のデザインの違いととらえている (図2)。

(1) 点か、棒(画)かに関する例



図2 明朝体活字のデザインについて (一部)

漢字の形については、「字形」「字体」という枠組みで説明されることがある(樺島 1975; 石塚 1984; 野村 1984; 林 1984; 屋名池 2008 など)。「字形」「字体」は、音声学・音韻論で言う「音声」「音韻」の考え方に相当する。具体的・個別的に、目に見える形で実現された文字の形を「字形」、それに対して、人間の脳裏にあって、社会的な約束事として存在する抽象字形が、概ね「字体」とされる²。「歸」「歸」の例に適用すると、この二つの「字形」は、同一の「字体」である〔歸〕(「字体」は図形化できないので〔 〕で示す)の実現形となる。

また、住基統一文字には (D) (E) がある。(D) は旧字体 (康熙字典体), (E) は籀文・古文などとされ、〔歸〕とはそれぞれ「字体」が異なると考えられる。

(D) 歸 [住基 6B78] (E) 歸 [住基 3C55]

異なるそれぞれの「字体」の組を何と呼び表すか、定まった名称はないようであるが、「常用漢字表」の「字種」という用語がこれに相当すると考えられる。そこで、「字体」の上位に「字種」を置くと、漢字の形について、「字種」「字体」「字形」の階層を設定することができる。《歸》(「字種」は図形化できないので《 》で示す)を例に図示する (図3)。

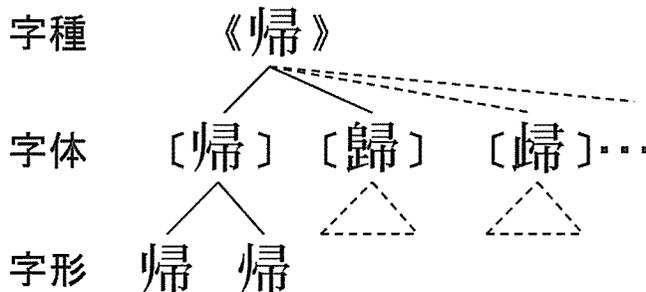


図3 「字種」「字体」「字形」の階層関係

したがって、「(A) 婦」と「(B) 婦」は字形レベル（デザインレベル）では「別の字」であるが、字体レベル・字種レベルでは「同じ字」, 「(A) 婦・(B) 婦」と「(D) 歸」と「(E) 婦」は字体レベルでは「別の字」, 字種レベルでは「同じ字」となる。

また、漢字の形については、「異体字」という概念で説明されることが多い（山田 1968；杉本 1978；佐藤 1987；笹原他 2003 など）。これは、規範的に「正しい」漢字（「正字」）に対して、意味や読みを共有するが形が異なるものを「異体字」として整理する方法であり、辞書や漢字研究において一般的に行われている。「(A) 婦・(D) 歸・(E) 婦」を例にすると、現代日本では、常用漢字表に掲出され、学校教育で行われている「(A) 婦」が「正字」、これ以外の「(D) 歸・(E) 婦」が「異体字」となる。「(D) 歸」は「旧字体」, 「(E) 婦」は「籀文」「古文」などと呼ばれる。「正字」「旧字体」「籀文」「古文」は、「(A) 婦」を「正字」とする規範のもとで、それぞれの字体に与えられた価値である。「正字」に対するものとしての「異体字」の考え方は、価値観に基づくものであるから、時代や地域が変わると規範も変わり、同じ字体であっても、与えられる価値が変わる。戦前の日本では、『康熙字典』を規範とするため、「(D) 歸」が「正字」であり、「(A) 婦・(E) 婦」は「異体字」となる。現代の中国では、簡化字「归」が「正字」であるから、「(A) 婦・(D) 歸・(E) 婦」はいずれも「異体字」となろう。「正・俗・新・旧・古・通・略・同」など、字体に与えられた価値は、いわばラベルのようなものであって、価値観によって張り替えが行われ、不変のものではない。

3.2. 第1次同定

「字種」「字体」「字形」と文字の形の単位を設定したところで、漢和辞典掲出字との対応を組んでみる。漢和辞典の見出し字では、「婦」と「婦」のような字形の差（デザインの差）を、別々に立項することは行われていないようなので、字体レベルで整理されていると考えられる。住基統一文字 (A) (B) (D) (E) を、『大漢和辞典』『角川大字典』『新大字典』の見出し字に対応させると表1のようになる。字形画像に文字コード、文字番号を添えて示す。(A) (B) (D) (E) は、字体レベルにおいて、『大漢和辞典』『角川大字典』『新大字典』にある文字と判断される。

表1 漢和辞典の見出し字との対応

| 住基統一文字 | 大漢和辞典 | 角川大字典 | 新大字典 |
|--------------|-----------|----------|----------|
| (A) 婦 [5E30] | | 婦 [4675] | |
| (B) 婦 [B0AA] | 婦 [8930] | | 婦 [4198] |
| (D) 歸 [6B78] | 歸 [16349] | 歸 [4674] | 歸 [7868] |
| (E) 婦 [3C55] | 婦 [16308] | 婦 [4667] | 婦 [7847] |

第1次同定では、同定辞書として、下記7種の漢和辞典・漢字字典を用いた。

- (1) 『大漢和辞典』(諸橋轍次, 修訂第二版第六刷, 大修館書店, 2001年)
- (2) 『角川大字源』(尾崎雄二郎・都留春雄・西岡弘・山田勝美・山田俊雄, 角川書店, 1992年)
- (3) 『講談社新大字典』(上田万年・岡田正之・飯島忠夫・栄田猛猪・飯田伝一, 講談社, 1993年)
- (4) 『増補改訂 JIS 漢字字典』(芝野耕司, 日本規格協会, 2002年)
- (5) 『大漢和辞典補巻』(鎌田正・米山寅太郎, 大修館書店, 2000年)
- (6) 『大漢語林』(鎌田正・米山寅太郎, 大修館書店, 1992年)
- (7) 『広漢和辞典』(諸橋轍次・鎌田正・米山寅太郎, 大修館書店, 1981年)

同定辞書との照合結果を表2に示す。字体レベルで、住基統一文字は7.7%, 戸籍統一文字は1.9%が辞書にない文字である³。

表2 同定辞書との照合結果

| | 住基統一文字 | 戸籍統一文字 |
|---------|--------|--------|
| 同定辞書にあり | 17,933 | 54,194 |
| 同定辞書になし | 1,499 | 1,073 |
| 計 | 19,432 | 55,267 |

4. 辞書にない文字を同定するためにはどうすればよいのか

4.1. 第2次同定

第2次同定では、辞書にない文字を同定するために、主に漢和辞典以外の文献資料・非文献資料を対象に、使用例を求めていくことになる。もっとも、第1次同定に用いた辞書は7点に過ぎないので、ほかの辞書を調べると同定できるものもあるだろう。

漢和辞典は元来漢籍を読むことを第一の目的に編纂されているので、典拠となる漢籍に出現しない字種や字体は、収集範囲の外にあると考えられる。したがって、第2次同定において同定資料とすべきものは、日本国内のものに照準を合わせていく方針が立てられる。

4.2. 「国字」

元来中国語を書き表すために中国で作られた文字を漢字というのに対して、漢字に似せて日本で作られた文字は「国字」と呼ばれる。漢和辞典の見出し字に見えない文字は、「国字」である可能性が高い。そこで、『角川大字源』の付録「国字一覧」を検してみると、住基統一文字の辞書非掲載字1,499字のうち30字、戸籍統一文字の辞書非掲載字1,073字のうち958字を見出すことができる。

「国字一覧」の凡例には、一覧作成のための依拠資料に関して、以下のように記されている。

日本で造った漢字まがいの文字を、ひろく国字（倭字〈和字〉）というが、ここに挙げるものは、従来もしくは最近、それとして諸書に取り上げたものを、広く、多く示すこととした。この際、最近の菅原義三氏、エツコ＝オバタ＝ライマン氏の調査の恩恵を受けたことを記して敬意を表する。

ここにある「最近の菅原義三氏、エツコ＝オバタ＝ライマン氏の調査」は、次の2書である。

- (1)『国字の字典』（飛田良文監修，菅原義三編，東京堂出版，1990年）
- (2)『叢書・ことばの世界 日本人の作った漢字 国字の諸問題』（エツコ＝オバタ＝ライマン，南雲堂，1990年）

『国字の字典』は、『異體字辨』や『國字考』などの江戸時代の文字研究書、『新撰字鏡』（小学篇字）や古本節用集などの日本古辞書、『歌舞伎評判記集成』（歌舞伎評判記研究会，岩波書店，1972-77年）などの近世文学関連書、『日本姓氏大辞典』（丹羽基二，角川書店，1985年）などの固有名（姓氏）関連書を中心に，辞書にない「国字」を収集している。一方、『日本人の作った漢字』は，江戸時代の文字研究書をはじめ，国土地理院発行5万分の1地形図の悉皆調査による地名，全国の電話帳の悉皆調査による苗字から，辞書にない「国字」を集め考察している。

戸籍統一文字の辞書非掲載字1,073字のうち，「国字一覧」に見出された958字については，『国字の字典』『日本人の作った漢字』の典拠資料に遡って追跡調査を行い，調査結果を国立国語研究所編（2009）にまとめている。なお，「国字一覧」に見出された958字中，『国字の字典』に採録されているもの930字，『日本人の作った漢字』に採録されているもの293字である。

戸籍統一文字の辞書非掲載字の大部分は，『角川大宇源』の「国字一覧」に見出されたが，住基統一文字の辞書非掲載字の多くは，これまでに集められた「国字」には該当しないものである。以下に，住基統一文字の辞書非掲載字を例に，文献資料・非文献資料によって文字を同定するに至った事例を紹介する。

4.3. 地名文字

地名を表す文字には，ある地点，ある地域に特有の文字が存在することが知られている（読売新聞社会部 1975；笹原 2007 など）。

撞

[住基 B23A・戸籍 145300]

この文字は，福島県郡山市の地名「鐘撞田（かねつきた）」に用いられる。現地郡山市役所の「コード表」（図4），2007年1月1日現在の課税台帳（図5）で，同じ字体を確認することができる。

また、福島地方法務局郡山支局の公図（図6）においても同様である。行政上の「正式な」表記は「鐘撞田」である。

89-74 鐘撞田 カネツキタ

図4 コード表

郡山市 西田町三町目字鐘撞田

図5 課税台帳

郡山市西田町三町目字鐘撞田

図6 公図

ところが、福島県歴史資料館で保管されている明治期の土地関係資料には、現在の「正式な」表記に用いる「撞」のほかに、「撞・衝」が見える（図7、図8、図9）。「衝」と「衝」とは異体関係にある。

田 撞 鐘 田 撞 鐘

図7 地籍帳「撞」

図8 地籍帳「撞」

鐘
衝
田

図9 地籍図「衝(衝)」

以上の事例から、「撞」は「つく」と訓む「撞」と「衝（衝）」とが混同された結果生じた文字であると推測される。規範的には「誤字」と呼ばれるものであろうが、地名表記として固定されたものであるので、「新しい文字」と言うべきであらうか。「撞」の発生についてまとめると図10のようになる。

なお、地名に由来する住基統一文字（「地名外字」）については、国立国語研究所編（2008）で報告している。

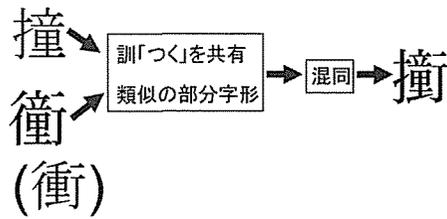


図 10 「撞」の発生

4.4. 人名文字

行政文書には人名表記で異体が使われていることが知られているが、その実態は定かではない。人名は個人情報に属するものであり、人名文字の実態把握は今後も難しいであろう。

𠄎 [住基 B263]

この文字は「政」の異体である。図 11 は東京都墨田区の本所松坂町公園内の顕彰碑（文面は昭和 8 [1933] 年揮毫）から採取した人名用例である。

「𠄎」に類似する字体に「𠄎 [登記 01036550]」がある。字体の変形の類型に、構成要素の左右が上下に（上下が左右に）配置を変える「動用」と呼ばれる現象がある（「峰—峯」「鶯—鶯」など）。「政—𠄎」は「動用」にあたる。さらに、構成要素としての「女（ほくによろ・のぶん）」と「文」は交替することがあり、「𠄎—𠄎」は「女」と「文」とが交替した字体である。「政」を定点とするならば変形の過程は図 12 のように推定される。



図 11 「𠄎助」



図 12 「政」の変形

4.5. 「景観文字」

街路の看板などの表示に使われている文字を、ここでは「景観文字」と呼ぶ（高田他 2008；田島他 2008）。

𠄎 [住基 B2F2]

この文字は、「石𠄎」「木𠄎」などの語表記に使われ、「材」の異体である。愛知県岡崎市花崗町には 10 数軒の石材店があり、3 種の字体を確認できる（図 13、図 14、図 15）。

石 材 店

図13 「材」

石 杖 店

図14 「杖」

石 棧 店

図15 「棧」

「杖」は観智院本『類聚名義抄』（佛下本 51 ウ 3）などの古辞書に確認できる。筆写字体として古くから使われてきた字体である⁴。「才」の異体に「戈」があることを参考にすると、「材」から「杖」が派生したと推測される。楷書体以前の隷書体の字体が、「材」に近い形状であることがこの推測を支持する。また、「杖」は、「丈」に対して「丈」,「土」に対して「土」のように、書法上の安定性を持たせるための点が付加されたものか、あるいは、「戈」や「伐」からの類推によって発生した字体と考えられる⁵。「材」から直接「杖」が派生したのではなく、「杖」を経由して「棧」が派生したと考えるのが妥当であろう（図16）。

材 → 杖 → 棧

図16 「材」の変形

「杖」や「棧」は、岡崎市の石材店だけではなく、全国の石材店で見る事ができる。また、石材店だけでなく木材店の看板でも、これらの字体が使われている。現在では、「杖」や「棧」は、建築・加工のための資材を扱う業種における特有の文字になりつつある可能性がある。

棧 [住基 AE69]

この文字は「協」の異体である。図17は岡崎市美合町で採取した看板の用例である。「濫一洪」「攝一撰」「壘一壘」など、新旧字体の対応に類型が求められる。同一の構成要素が上の一つ、下に二つ配置されたとき、下二つの構成要素を二点（繰り返し符号）に置き換えることが行われる。生産性の高い略字体の作り方である。蜂矢（2005）は、「棧」を1965～75年度以前から一般的に用いられていたものとしている。



図17 「連絡棧議会議」

また、図18の「轟」の略字体は、青森県深浦町の地名「轟木（とどろき）」の表記において観察され、「洪一撰一壘」などの類型である。「轟」を地名に用いるのは全国で深浦町だけのようであり、多くの人には「轟」にはなじみがないであろうが、轟木地区の住民にとっては身近な文字であることは疑いようもない。限られた地域で流通している文字であるが、よく使われる文字ならば簡略化されるという筆記経済に適った事例である。同時に、同一構成要素を二点に置き換える簡略化の方法の一般性を裏付ける事例である。



図18 「轟木」

4. おわりに

本稿では、行政用文字に関する調査研究における文字の同定について述べた。第1次同定では辞書を利用し、従来の文献を用いる文字の調査研究の手法に拠った。第2次同定では、文献を用いる文字研究においてほとんど扱われてこなかった非文献資料にも手を広げていった。街路の看板などの「景観文字」や、社会的な規範から外れる字体である略字体などは、漢字字体研究において蓄積が乏しいものである。しかし、「景観文字」や手書きに見られる略字体は、日常生活に関わりの深いものであり、言語生活の「書く」「読む」（あるいは「見る」）の領域に含まれるものである。言語と言語を使う人間や周囲の環境に視点を置き、話し言葉を中心に展開してきた言語生活の研究手法は、文字研究に応用できる部分もあるだろう。文字と文字を使う人間や周囲の環境に視点を置くことで、文字研究に次の展開が訪れることを期待するものである。

註

- 1 汎用電子情報交換環境整備プログラムでは、国立国語研究所が文字情報の整理・体系化（文字同定と文字同定情報の集積）、情報処理学会が既存の文字コードに不足する文字の国際提案活動（ISO/IEC 10646 への追加提案）、日本規格協会が文字グリフのデザインの検討（平成明朝体グリフの制作）をそれぞれ分担し、2008年度末（2009年3月）をもって終了した。
- 2 「字体」が示す概念は論者によって異なり、必ずしも一致してはいない。しかし、「字形」に対するものとしての位置づけは一致している。
- 3 本稿執筆時点で登記統一文字の辞書同定は完了していないため、登記統一文字に対する第1次同定の結果を記述することができない。なお、登記統一文字には、「龙（龍）・华（華）・郑（鄭）」などの中国簡化字が登録されているため、適宜中国辞書も参照した。
- 4 「𣎵」は登記統一文字に採録されている（登記 01042820）。
- 5 「𣎵」は、観智院本『類聚名義抄』では「クヒ、クヒウツ、ヒクハシ」と和訓があり（佛下本 48ウ5）、「𣎵」との関係がうかがわれる。「景観文字」の「𣎵」とは、同形異字とみるべきであろう。

参考文献

- 石塚晴通（1984）『圖書寮本日本書紀研究篇』、汲古書院
- エツコ＝オバタ＝ライマン（1990）『叢書・ことばの世界 日本人の作った漢字 国字の諸問題』、南雲堂
- 樺島忠夫（1975）「文字体系の構造」、『計量国語学』75、13-22、計量国語学会
- 国立国語研究所編（2008）『平成19年度汎用電子情報交換環境整備プログラム 文字対応作業委員会資料（「地名外字」資料）』、国立国語研究所
- 国立国語研究所編（2009）『平成20年度汎用電子情報交換環境整備プログラム 文字対応作業委員会資料（「辞書非掲載字」資料）』、国立国語研究所
- 笹原宏之（1996）「『JIS X 0208』における音義未詳字に対する原典による同定——『標準コード用漢字表（試案）』と『国土行政区画総覧』——」、『国語学研究与資料』20、1-17、国語学研究与資料の会

- 笹原宏之 (2007) 『国字の位相と展開』, 三省堂
- 笹原宏之・横山詔一・エリク = ロング (2003) 『現代日本の異体字——漢字環境学序説——』, 三省堂
- 佐藤稔 (1987) 「異体字」, 『漢字講座 3 漢字と日本語』, 183-210, 明治書院
- 杉本つとむ (1978) 『杉本つとむ日本語講座 1 異体字とは何か』, 桜楓社
- 高田智和 (2005) 「公共サービスと漢字」, 『日本語学』 24(13), 58-66, 明治書院
- 高田智和 (2006a) 「「新しい文字」の発生とその要因」, 『日本語学』 25(9), 28-36, 明治書院
- 高田智和 (2006b) 「漢字字体研究의 文字情報 데이터 베이스」, 『漢文讀法과 東아시아의 文字』, 621-630, 태학사
- 高田智和・井手順子・虎岩千賀子 (2008) 「行政用文字の調査研究——汎用電子情報交換環境整備プログラム——」, 『日本語科学』 23, 95-110, 国立国語研究所
- 高田智和・田島孝治・米田純子 (2008) 「文字生活研究における「景観文字調査」」, 『情報処理学会研究報告』 2008-CH-79, 1-8, 情報処理学会
- 田島孝治・米田純子・高田智和 (2008) 「「景観文字調査」のための調査結果分類・可視化ツールの試作」, 『人文科学とコンピュータシンポジウム論文集 サービス指向のデジタル技術へ——人文科学のポテンシャル——』, 291-296, 情報処理学会
- 野村雅昭 (1984) 「同字と別字のあいだ」, 『日本語学』 3(3), 23-31, 明治書院
- 蜂矢真郷 (2005) 「一九六五～七五年度頃の略字」, 『国語文字史の研究 8』, 197-214, 和泉書院
- 林大 (1984) 「字体・字形・書体をめぐって」, 『日本語学』 3(3), 10-15, 明治書院
- 飛田良文監修・菅原義三編 (1990) 『国字の字典』, 東京堂出版
- 前田富祺 (1992) 「国語文字史研究の課題」, 『国語文字史の研究 1』, 1-30, 和泉書院
- 屋名池誠 (2008) 「文字・表記 (理論・現代)」, 『日本語の研究』 4(4), 70-76, 日本語学会
- 山田俊雄 (1968) 「漢字字形の史的研究の問題とその一方向」, 『国語学』 72, 110-126, 国語学会
- 読売新聞社会部編 (1975) 『日本語の現場第一集』, 読売新聞社

付 記

本稿は、平成 20 年度国立国語研究所公開研究発表会（国立国語研究所、2008 年 12 月 19 日）において、「文字の研究法——漢字字体研究の対象と方法——」と題して口頭発表したものに基づいている。

高田 智和 (たかだ ともかず)

国立国語研究所

190-8561 東京都立川市緑町 10-2

ttakada@kokken.go.jp