

ISO/IEC 10646 への変体仮名収録提案 —レパートリと符号化アーキテクチャ—

高田智和^{†1} 小林龍生^{†2} 田代秀一^{†2} 矢田勉^{†3}

概要: 仮名の異体字である変体仮名は、学術用途あるいは行政用途において、文字コード標準化のニーズがあるが、これまで文字コード標準に収録されることはなかった。これらのニーズを受けて、2015 年 10 月に国際規格 ISO/IEC 10646 への変体仮名収録提案を行った。本報告では、標準化提案のレパートリと符号化アーキテクチャについて述べる。

キーワード: 変体仮名, 平仮名, 文字コード, ISO/IEC 10646, Unicode

Proposal regarding Japanese *Hentaigana* for ISO/IEC 10646

TAKADA Tomokazu^{†1} KOBAYASHI Tatsuo^{†2}
TASHIRO Shuichi^{†2} YADA Tsutomu^{†3}

Abstract: *Hentaigana* are variants of *hiragana*, and are useful for administrative purpose such as the basic resident register, and academic purposes such as those relating to Japanology. However, *Hentaigana* have not been added to the standardized coded character sets of JIS X 0208 and ISO/IEC 10646. In October 2015, we proposed *Hentaigana* to ISO/IEC 10646. In this paper, we discuss the character set of *Hentaigana* and its encoded architecture.

Keywords: *Hentaigana*, *Hiragana*, coded character set, ISO/IEC 10646, Unicode

1. はじめに

国際文字コード規格である ISO/IEC 10646 は、現状において収録字数 10 万を超える巨大文字集合である。現在の世界諸言語の文字や記号類はほぼ収録され、新規追加収録の対象は、少数言語の文字や歴史的文献の文字に移行している。すなわち、人文学が対象とする歴史的文献を電子化・アーカイブ化において必要となる文字が、文字コード拡張の中心となりつつある。

日本語表記用の文字・記号類では、近年の ISO/IEC 10646 の拡張において、「国字」(統合漢字拡張 C, E), 戸籍や住民基本台帳など行政処理用途の固有名に使うとされる漢字群(統合漢字拡張 D, 審議中の統合漢字拡張 F), 絵文字が収録または収録審議中である。特定の古典籍・文書に使用が見られる漢字(異体漢字を含む)を除けば、Unicode による日本語文字の情報交換の環境は整いつつあると言える。しかし、特定用途であれば、このほかにもまだ文字コード標準化がされていない文字群が残っている。現代では平仮名の異体と認識されている変体仮名である。

平仮名は、平安時代中期には成立したとされ、そのころから 1 音に対して複数字体を持つ文字体系であった。教育

上の要請から、1 音に対して 1 字体と定めたのは明治 33 年(1900 年)の小学校令施行規則である。これは、社会全体に及ぼされる規則ではないものの、平仮名として定められなかった仮名字体は変体仮名と呼ばれるようになり、文章を綴る場からは急速に衰退していくのである。

現代の変体仮名と言えば、商店の看板や暖簾での使用がすぐに思い浮かぶであろうが、行政の場では人名の文字情報処理、あるいは、学術の場では、日本語文字・表記史研究における仮名字体の記述、古文書・古記録の翻刻などで、コンピュータを介した文字情報処理・文字情報交換の需要がある。特に行政においては、昭和 23 年(1948 年)の戸籍法施行規則において命名時の変体仮名使用が制限されるまで、名づけに変体仮名が使われてきた経緯があり、戸籍や住民基本台帳などの行政処理では、現在日常的に変体仮名が日常的に使われている。さらに付言すれば、現行法制下では、除籍簿の保存期間は 150 年と定められているため、向こう 200 年程度の期間、変体仮名は行政処理で必要とされるのである。

このような行政面と学術面での需要を受けて、2015 年 10 月に、国際文字コード規格 ISO/IEC 10646 への変体仮名収録提案を行った[a]。本発表では、提案した変体仮名のレパートリと符号化アーキテクチャについて述べる。

^{†1} 国立国語研究所
National Institute for Japanese Language and Linguistics

^{†2} 情報処理推進機構
Information-technology Promotion Agency, Japan

^{†3} 大阪大学
Osaka University

a) ISO/IEC JTC1/SC2/WG2 第 64 回松江会議にて提案文書を提出・審議。ISO/IEC 10646 への追加提案を行ったのは、ISO/IEC 10646 規格対応の日本の窓口である情報処理学会情報規格調査会 SC2 専門委員会である。また、提案した変体仮名のレパートリと符号化アーキテクチャの原案を検討したのは情報処理推進機構にて進めている文字情報基盤整備事業である。

2. レパートリ

2015 年 10 月に提案した変体仮名のレパートリは、後述する行政用変体仮名 168 字と学術用変体仮名 264 字の和集合である[b]。本稿の末尾に一覧を示す。

2.1 行政用変体仮名

戸籍行政では、戸籍簿に記載することのできる文字の網羅を志向した戸籍統一文字[c]を定めている(平成 16 (2004) 年 4 月 1 日付け法務省民一第 928 号民事局長通達)。戸籍統一文字には 168 字の変体仮名が収録されている。この 168 字は、住民基本台帳ネットワーク統一文字にも収録されている。戸籍簿・住民基本台帳は住民記録の根本台帳であるから、戸籍統一文字の変体仮名を代表させて行政用変体仮名として差支えないであろう。

変体仮名は、国際文字コード規格 ISO/IEC 10646 はもとより、国内文字コード規格 JIS X 0208 や JIS X 0213 にこれまで収録されたことがないが、メインフレームを扱うベンダー各社は固有文字(いわゆる外字)として、独自の拡張文字セットに変体仮名を収録していた。ベンダー各社のコードブックを参照すると、富士通は 168 字(『FUJITSU JEF 文字コード索引辞書』、1980 年による)、日本電気は 152 字(『日本電気標準文字セット辞書<拡張>』、1983 年による)、日立製作所は 152 字(『HITAC 文字パターン辞書/コードブック』、1987 年)を収録していることがわかる。変体仮名の拡張収録は、人名処理の実務用途を満たすためのものと推測される。そして、戸籍統一文字の変体仮名と、各社の変体仮名を比較すると、レパートリが極めて近似している。メインフレーム以降、行政処理実務で使われていた変体仮名が、おおむねそのまま戸籍統一文字に収録されたと考えるのが妥当である。

2.2 学術用変体仮名

ここで学術用変体仮名と呼んでいるのは、国立国語研究所共同研究プロジェクト「文字環境のモデル化と社会言語科学への応用」にて選定した「学術情報交換用変体仮名」[d]のことである。日本語文字・表記史研究での仮名字体の記述と、日本史学における一部の古文書翻刻[e]における需要を満たすことを目的に選定された変体仮名セットである。

変体仮名セットの選定では、活版印刷以降の活字及びデジタルフォントの変体仮名を集め、仮名字典類での採録状況を勘案して、基礎セットを作成することにした[f]。変体

仮名と言うと、書道手本となるような古筆の仮名を思い浮かべるであろうが、字形変化が大きい筆写文字よりも、パターン化されてゆれの幅が小さい活字やデジタルフォントの方が、字形がより安定しているため、主たる選字の対象とした。以下の 6 種である。

- (1) 精興社変体仮名 134 字
- (2) 中西印刷変体仮名 154 字
- (3) 築地活版五号明朝体活字変体仮名(明治 27 年)〔府川充男『聚珍録 第三編仮名』(三省堂、2005 年)による〕 115 字
- (4) 教科書活字(明治 25 年・明治 27 年)〔板倉雅宣『教科書体変遷史』(朗文堂、2003 年)による〕 明治 25 年: 49 字 明治 27 年: 50 字
- (5) 今昔文字鏡変体仮名 380 字
- (6) Koin 変体仮名[g] 335 字

(1)と(2)は東京大学史料編纂所の『大日本古文書』『大日本古記録』の印刷会社が保有する変体仮名セットであり、日本史学での用途をおおよそ満たすものと推測される。(3)は近代活版印刷の活字デザインが安定したとされる時期の活字変体仮名である。(4)は明治期読本に用いられた変体仮名で、初学者が利用する教科書で使われたことから、基本的な変体仮名であると考えられる。(5)と(6)は研究者の利用が多い現代の変体仮名フォントである。なお、これらの変体仮名セットには、濁点や半濁点付きの変体仮名も含まれるが、学術用途での優先順位は低いものと判断し、今回は集字の対象から除外することにした。

さらに、活字やデジタルフォントの変体仮名に加えて、仮名字体研究論考 12 種の字体表に出現する変体仮名と、『大日本古文書』『大日本古記録』の編纂で必要な変体仮名(上記の(1)(2)に含まれていない仮名)を追加した。なお、仮名字体研究論考 12 種は次のとおりである。

- (1) 池田亀鑑『古典の批判的処置に関する研究』(岩波書店、1941 年)
- (2) 山田俊雄『日本語の歴史 別巻 言語史研究入門』(平凡社、1976 年)
- (3) 浜田啓介「板行の仮名字体—その収斂的傾向について—」(『国語学』118、1979 年)
- (4) 築島裕『日本語の世界 5 仮名』(中央公論社、1981 年)
- (5) 玉村禎郎『春色梅兒誉美』における仮名の用字法(『国語文字史の研究 2』、和泉書院、1994 年)
- (6) 松下なるみ「平仮名の字体と字源」(『漢字百科大事

b) 文字情報基盤整備事業の Web ページにも提案文字の一覧がある。
<http://mojikiban.ipa.go.jp/4436.html>

c) <http://kosekimoji.moj.go.jp/kosekimojidb/mjko/PeopleTop>

d) <http://kana.ninjal.ac.jp/>

e) 東京大学史料編纂所によれば、和歌や仮名消息の翻刻で変体仮名を使うとのことである。

f) 変体仮名検討グループは、銭谷真人(国語研)、小助川貞次(富山大)、當山日出夫(立命館大)、斎藤達哉(専修大)、矢田、高田の 6 名。第 30 回表記研究会(2014 年 9 月 28 日)、日本語学会 2015 年度春季大会(2015 年 5

月 24 日)の研究発表や、日本史学、書学書道史、印刷史、書誌学などの専門家への個別照会も随時行い、選定方法とレパートリの検討を行った。

g) <http://www10.plala.or.jp/koin/koinhentaigana.html>

典』, 明治書院, 1996 年)

- (7) 久保田篤『『浮世風呂』の平仮名の用字法』(『成蹊国文』30, 1997 年)
- (8) 内田宗一「黄表紙・洒落本の仮名字体」(『国語文字史の研究 4』, 和泉書院, 1998 年)
- (9) 内田宗一『『修紫田舎源氏』の仮名字体一作者自筆稿本と板本の比較考察』(『待兼山論叢文学篇』32, 1998 年)
- (10) 内田宗一「馬琴作合巻『金比羅船利生纜』の仮名字体」(『国語文字史の研究 5』, 和泉書院, 2000 年)
- (11) 久保田篤「江戸板本の表記の多様性一洒落本『傾城買二筋道』の場合」(『成蹊国文』42, 2009 年)
- (12) 矢田勉『国語文字・表記史の研究』(汲古書院, 2012)

このようにして集字した個々の変体仮名を字形の類似でまとめ、字形群の文字と仮名字典類との照合を行った。参照した仮名字典類は次の 5 点である。いずれも用例典を明示している仮名辞典類であり、典拠のある変体仮名を選定するための照合作業である。

- (1)『変体仮名の手引 (改訂増補 25 版)』(中野幸一編, 武蔵野書院, 2009 年)
- (2)『字典かな (新装版)』(笠間影印叢刊刊行会編, 笠間書院, 2003 年)
- (3)『かな名跡大字典』(筒井茂徳編, 角川書店, 1981 年)
- (4)『かな連綿字典』(竹田悦堂監修・佐野榮輝編, 雄山閣出版, 1990-1992 年)
- (5)『用例かな大字典』(中田易直他編, 柏書房, 1978 年)

以上の過程を経て選定した変体仮名は 264 字である。この中には、特に平安時代の用例で、一つの字母が複数の音で使われる場合がある。以下の 10 字母である。

悪: あ, を 家: か, け 期: こ, き
徒: つ, と 等: と, ら 而: に, て
子: ね, こ 馬: め, ま 夜: や, よ
无: む, も, ん

2.3 日本提案変体仮名

本節の冒頭に述べたように、日本提案変体仮名は、行政用変体仮名 168 字と学術用変体仮名 264 字の和集合である。先に述べた一字母が複数音に対応する仮名については、音ごとに別字とみなしたため、提案したレパートリは 299 字となった。なお、末尾の一覧では、音ごとに別字とみなしたのものには*を付している。

また、日本提案変体仮名は、3 層で整理している。第 1 層は音価、第 2 層は字母、第 3 層は字体である。仮名字典類で一般的な整理方法である。

3. 符号化アーキテクチャ

3.1 変体仮名の符号化の考え方

符号化モデルを考える場合、まず行わなければならないことは、何をどのレベルで符号化するか符号化対象を明確にすることである。ただし、既存の ISO/IEC 10646 の図形文字の全てについて、符号化対象が明確であるわけではない。現に、既存の平仮名ブロックは、音を符号化したものか、形を符号化したものか、五十音図の位置を符号化したものか、じつは判然としていない。それぞれの関係が、日常的な日本語のレベルでは安定しているので、不都合が生じないだけである。

変体仮名について、符号化対象は次の 2 点になると考えられる。

- (1) 変体仮名の字母 (字源の漢字) の情報
- (2) 変体仮名が表す音価 (phonetic value) の情報

このほかに、変体仮名の字形の情報が符号化対象の候補となろうが、具体字形としては草書体の漢字との区別が難しいものもあり、また草体化によって実現された字形は、ほぼ無限でかつ連続的だと考えられるため、同一字母の変体仮名を具体字形に依拠して区別することは、符号化という観点からは意味をなさないものと考えられる。

字母と音価のどちらを符号化対象としても、変体仮名の実利用の面からは、二つの問題への解決策が求められる。一つ目は同一字母複数字体の問題である。

か (可) かう

な (奈) 奈ふあ

これまでの日本語文字・表記史研究において、複数の仮名字体の機能的使い分けが報告されている。特定の語を特定の仮名字体で書く事例、語頭・語中語尾の位置によって仮名字体を使い分ける事例、同音節が隣り合う場合に仮名字体を使い分ける事例などである。これらの機能的使い分けが、同じ字母を持つ複数の仮名字体によって発生する場合もある。

二つ目は同一字母複数音価の問題である。

や (夜) 夜 よ (夜) 夜

む (无) ええ も (无) ええ ん (无) ええ

この問題は、音価情報を符号化すると考えれば、異なる符号位置が同一字体を持つ場合だとみなすことができる。一方、字母にのみ注目すれば、ある符号位置を持つ変体仮名が、文脈によって複数の音価を持つ場合だとみなすことができる。

3.2 SVS モデル

SVS モデルは、字母情報と音価情報を組として共に保存するモデルの一つである。現行の平仮名を、音価に対して符号位置を与えたものであると考え、個々の変体仮名を同一音価の、字母ごとに異なる視覚表現と捉えて、VS (Variation Sequence) を用いて符号化するものである。この方式では、同一字母複数字体の問題に対しても、同一字母複数音価の問題に対しても、異なる VS を付与することで解決している。

同一字母複数字体の例

お 《HIRAGANA LETTER KA, VS_m》

う 《HIRAGANA LETTER KA, VS_m》

同一字母複数音価の例

お 《HIRAGANA LETTER YO, VS_m》

お 《HIRAGANA LETTER YO, VS_m》

SVS モデルは、学術利用の際も戸籍実務等における行政利用の際も、変体仮名が利用できない環境では、現行の平仮名に置き換えることが日常的に行われており、その置き換えが VS のメカニズムで自動的に可能となるというメリットがある。

2015 年 10 月の提案に先立つ意見照会では、SVS モデルを採用したが、UTC から次に述べる独立符号方式の提案があり、本提案ではそちらに変更することとした。

3.3 独立符号モデル

独立符号モデルは、字母＋音価それぞれに対して、HIRAGANA SUPPLEMENT ブロックの独立した符号位置を与えようというものである。この場合、同一字母複数字体問題については、デフォルト表示問題を回避するため、VS を用いずに、それぞれ独立した符号位置を付与することが望ましい。

また、同一字母複数音価問題については、この符号化モデルを字母＋音価と考えると、同一字体に対して複数の符号位置を与える必要が生じ、運用に混乱が生じる恐れがある。事実、2015 年 10 月の審議時には、同一字母複数音価に対して UTC から意見があり、レパトリの修正を検討することとなった。

同一字母複数字体の例

お 《HENTAIGANA LETTER KA VARIANT1》

う 《HENTAIGANA LETTER KA VARIANT2》

同一字母複数音価の例

や 《HENTAIGANA LETTER YA VARIANT1》

や 《HENTAIGANA LETTER YA VARIANT2》

お 《HENTAIGANA LETTER YA VARIANT3》

代 《HENTAIGANA LETTER YO VARIANT1》

年 《HENTAIGANA LETTER YO VARIANT2》

お 《HENTAIGANA LETTER YO VARIANT3》

4. おわりに

2015 年 10 月に提案した変体仮名は受理され、符号標準化のプロセスにのった。しかし、レパトリや文字名については今後も審議が行われる。より良い形で変体仮名が ISO/IEC 10646 に収録されるよう活動を続けていく所存である。

日本提案変体仮名一覧

あ (安) あ (悪) 愛 (愛) 雲 (阿) 何

い (以) 以 (伊) 伊 (意) 意 (移) 悔

う (宇) 宇 (憂) 有 (有) 有 (雲) 雲

え (江) 江 (盈) 盈 (縁) 縁 (衣) 衣 (要) 要

お (於) 於 (隠) 隠

か (佳) 佳 (加) 加 (可) 可 (嘉) 嘉 (家) 家 (我) 我

(賁) 賁 (賀) 賀 (閑) 閑 (香) 香 (駕) 駕

き (喜) 喜 (幾) 幾 (支) 支 (期) 期 (木) 木 (祈) 祈

(貴) 貴 (起) 起

く (久) 久 (九) 九 (供) 供 (俱) 俱 (具) 具 (求) 求

け (介) 介 (家) 家 (希) 希 (氣) 氣 (計) 計 (遣) 遣

こ (古) 𠂔 * (子) 子 (故) 𠂔 * (期) 𠂔 (許) 𠂔
 さ (乍) 𠂔 (佐) 𠂔 (左) 𠂔 (差) 𠂔 (散) 𠂔 (斜) 𠂔
 (沙) 𠂔
 し (之) 𠂔 (事) 𠂔 (四) 𠂔 (志) 𠂔 (新) 𠂔
 す (受) 𠂔 (壽) 𠂔 (數) 𠂔 (春) 𠂔 (須) 𠂔
 せ (世) 𠂔 (勢) 𠂔 (聲) 𠂔
 そ (所) 𠂔 (曾) 𠂔 (楚) 𠂔 (蘇) 𠂔 (處) 𠂔
 た (堂) 𠂔 (多) 𠂔 (當) 𠂔
 ち (千) 𠂔 (地) 𠂔 (智) 𠂔 (知) 𠂔 (致) 𠂔 (遲) 𠂔
 っ (川) 𠂔 * (徒) 𠂔 (津) 𠂔 (都) 𠂔
 て (亭) 𠂔 (低) 𠂔 (傳) 𠂔 (天) 𠂔 (帝) 𠂔
 (亘) 𠂔 * (而) 𠂔 (轉) 𠂔
 と (土) 𠂔 (度) 𠂔 * (徒) 𠂔 (東) 𠂔 (登) 𠂔 (砥) 𠂔
 * (等) 𠂔
 な (南) 𠂔 (名) 𠂔 (奈) 𠂔 (菜) 𠂔 (那) 𠂔
 (難) 𠂔
 に (丹) 𠂔 (二) 𠂔 (仁) 𠂔 (兒) 𠂔 (爾) 𠂔 * (而) 𠂔
 (耳) 𠂔
 ぬ (努) 𠂔 (奴) 𠂔 (怒) 𠂔
 ね * (子) 𠂔 (年) 𠂔 (根) 𠂔 (熱) 𠂔 (禰) 𠂔
 の (乃) 𠂔 (濃) 𠂔 (能) 𠂔 (農) 𠂔
 は (八) 𠂔 (半) 𠂔 (婆) 𠂔 (波) 𠂔 (盤) 𠂔 (破) 𠂔
 (者) 𠂔 (葉) 𠂔 (頗) 𠂔
 ひ (悲) 𠂔 (日) 𠂔 (比) 𠂔 (避) 𠂔 (非) 𠂔 (飛) 𠂔

ふ (不) 𠂔 (婦) 𠂔 (布) 𠂔
 へ (倍) 𠂔 (弊) 𠂔 (遍) 𠂔 (邊) 𠂔 (部) 𠂔
 ほ (保) 𠂔 (報) 𠂔 (奉) 𠂔 (寶) 𠂔 (本) 𠂔
 (豐) 𠂔
 ま (万) 𠂔 (末) 𠂔 (滿) 𠂔 (萬) 𠂔 * (馬) 𠂔
 (麻) 𠂔
 み (三) 𠂔 (微) 𠂔 (美) 𠂔 (見) 𠂔 (身) 𠂔
 む * (无) 𠂔 (武) 𠂔 (無) 𠂔 (牟) 𠂔 (舞) 𠂔
 め (免) 𠂔 (面) 𠂔 * (馬) 𠂔
 も * (无) 𠂔 (母) 𠂔 (毛) 𠂔 (茂) 𠂔 (裳) 𠂔
 や (也) 𠂔 * (夜) 𠂔 (屋) 𠂔 (耶) 𠂔
 ゆ (游) 𠂔 (由) 𠂔 (遊) 𠂔
 よ (代) 𠂔 (余) 𠂔 * (夜) 𠂔 (與) 𠂔 (餘) 𠂔
 ら * (等) 𠂔 (羅) 𠂔 (良) 𠂔
 り (利) 𠂔 (李) 𠂔 (梨) 𠂔 (理) 𠂔 (里) 𠂔 (離) 𠂔
 る (流) 𠂔 (留) 𠂔 (累) 𠂔 (類) 𠂔
 れ (禮) 𠂔 (連) 𠂔 (麗) 𠂔
 ろ (呂) 𠂔 (婁) 𠂔 (樓) 𠂔 (路) 𠂔 (露) 𠂔
 わ (倭) 𠂔 (和) 𠂔 (王) 𠂔
 ゐ (井) 𠂔 (居) 𠂔 (爲) 𠂔 (遣) 𠂔
 𠂔 (惠) 𠂔 (衛) 𠂔
 を (乎) 𠂔 (尾) 𠂔 * (惡) 𠂔 (緒) 𠂔 (越) 𠂔
 (遠) 𠂔
 ん * (无) 𠂔