

## 悉曇文字の国際標準化の動向

川幡太一<sup>†1</sup> 鈴木俊哉<sup>†2</sup> 永崎研宣<sup>†3</sup> 下田正弘<sup>†4</sup>

悉曇文字は日本において、仏典の研究や菩薩の種字等に用いられるインド系文字の一種である。本報告では、日本の悉曇文字の国際符号化文字集合 (UCS) への提案活動に関して、その概要・標準化の経緯・および標準化にあたっての技術的課題および今後の予定について述べる。

### Recent Standardization Activities on Siddham Script

TAICHI KAWABATA<sup>†1</sup> TOSHIYA SUZUKI<sup>†2</sup> KIYONORI NAGASAKI<sup>†3</sup>  
MASAHIRO SHIMODA<sup>†4</sup>

Siddham script is used to describe the Buddhist text written in Sanskrit or related languages, or the symbols of Bodhisattvas in Japan. Recently, the encoding of Siddham text to Universal coded Character Set (UCS) has been proposed. This document reports architecture of Siddham script and recent activities on encoding it.

#### 1. はじめに

悉曇文字 (Siddham Script) は、4～5世紀頃の南インドで使われた文字である。ブラフミー文字を祖とするインド系文字の一種であり、インドでは現在はいれられなくなったが、8世紀頃に仏教とともに中国から日本に伝わった。日本では仏典のサンスクリット語の記述に広く用いられており、日本に現存する古い資料としては、法隆寺で所蔵されていた般若心経の法隆寺貝葉本が有名である。

さらに、日本における悉曇文字の用法としては、単にサンスクリット語の仏典を記述するというだけでなく、その字形に依拠し、仏尊の種字等としてのシンボリックな用法も確立した。これは、曼荼羅や守護札、石塔、卒塔婆等、様々なところに描かれ、世間にも広く見られるものとなった。

これまで悉曇文字には標準化された文字符号がなく、パソコン等での一般的な利用は困難であったが、2012年までに国際文字符号規格、具体的には ISO/IEC 10646 Information technology -- Universal Coded Character Set (UCS) 規格[1] への追加提案がなされ、2014年には国際規格として出版される見込みとなった。

本稿では、悉曇文字の符号化における、文字符号の体系と、2013年6月時点までにおける経緯について述べる。

#### 2. インド系文字の符号

世界の文字は、大別して表意文字と表音文字に分類できる。表意文字の代表的なものとしては漢字があり、表音文字にはラテン文字やギリシャ文字のようなアルファベット文字や、日本の仮名文字、韓国のハングルの他に、子音文字の周囲に母音修飾文字を付与することを特徴とするアブギダ文字がある。アブギダ文字の中でも、特にブラフミー文字を先祖とするデーヴァナーガリー等のインド系文字は、インド・ミャンマー等、南アジア・東南アジアの多くの国で現在でも使われている。

インド系文字では、母音修飾文字は子音文字の上下左右に付加される。母音修飾文字がない場合、子音文字は a 随伴母音を持つ。子音に母音がつかない場合は母音を打ち消すヴィラーム修飾文字を子音に修飾する。母音を挟まずに子音が連続する場合、子音文字が結合してリガチャを形成する場合が多い (図1)。この構造は、デーヴァナーガリー・タミール・ベンガル・カンナダ等の多数のインド系文字で一貫している。

क + ् + र → क्र  
क + ् + ष → क्ष  
ड + ् + क → ड़

図1 ヴィラーム符号とリガチャの例 [10]

インド系文字は、インド規格 IS 13194:1991[2] によって最初に系統的に符号化された。IS 13194:1991 は、子音・

<sup>†1</sup> 日本電信電話株式会社 NTT 未来ねっと研究所  
NTT Corporation, Network Innovation Laboratories

<sup>†2</sup> 広島大学  
Hiroshima University

<sup>†3</sup> 人文情報学研究所  
International Institute for Digital Humanities

<sup>†4</sup> 東京大学大学院人文社会系研究科  
University of Tokyo

独立母音・修飾母音・ヴィラーマ・鼻音化文字等で構成される。リガチャは独立して符号化されず、連続する子音文字の間にヴィラーマ符号を挿入し、表示系がこれをリガチャに置き換える。IS 13194:1991 はデーヴァナーガリーやタミール等 8 種類の文字を、共通の母音・子音文字符号位置で符号化し、文字切り替え符号によって別の文字に切り替えられる。IS 13194:1991 のアーキテクチャは、国際文字符号規格である UCS にも引き継がれ、そこでもヴィラーマ符号が重要な役割を果たしている。ヴィラーマ符号によってリガチャの合成を制御するには、「どの文字がヴィラーマ符号で、どの文字とどの文字を結合しようとしているか」を表示系が把握できなければならない。従って、漢字における外字の追加のような形で運用することは難しい。

UCS とこれを基礎とする Unicode は、世界の全ての文字を対象とする文字符号やその符号化方式を規定する国際標準規格であり、現在ほとんどすべてのパソコンやスマートフォン等の基盤として採用されている。現在、UCS は、国際標準化機構 (ISO) と国際電気標準会議 (IEC) の合同委員会 1・副委員会 2・第 2 ワークグループ (ISO/IEC JTC 1/SC 2/WG 2、以下 WG2) にて標準規格の開発作業が進められている。WG2 は日本を含む世界の約 30 カ国が参加メンバーとして活動し、1 年に 1～2 回程度の会議を行い、そこで規格の開発・投票を行う。

### 3. 悉曇文字の符号化の経緯

#### 3.1 最初の国際符号化提案

UCS への悉曇文字の追加提案は、2012 年に米国ミシガン大学の Anshuman Pandey 氏によって WG2 へ提案された [3,4]。初期提案が 5 月 [3]、本提案が 8 月 [4] である。この提案では、悉曇文字はインド系文字の一種として提案され、独立母音文字 (U+11580～U+1158D)・子音文字 (U+1158E～U+115AE)・修飾母音文字 (U+115AF～U+115C0) およびその他の記号類 (U+115C1～U+115C6) から構成される。

a

UCS における既存のインド系文字と共通のアーキテクチャに基づく符号化であるため、排列に関しては若干の差があるものの、各文字はデーヴァナーガリー等の文字と原則として 1:1 に対応し、悉曇十八章等に記述される悉曇の全ての音節組み合わせを表現可能である (図 2)。

a 符号位置は本稿執筆中で予定されている位置である。正式な規格票出版までは確定していないことに留意されたい。

|   | 1158    | 1159    | 115A    | 115B    | 115C    | 115D | 115E | 115F |
|---|---------|---------|---------|---------|---------|------|------|------|
| 0 | 𑀀 11580 | 𑀁 11590 | 𑀂 115A0 | 𑀃 115B0 | 𑀄 115C0 |      |      |      |
| 1 | 𑀅 11581 | 𑀆 11591 | 𑀇 115A1 | 𑀈 115B1 | 𑀉 115C1 |      |      |      |
| 2 | 𑀊 11582 | 𑀋 11592 | 𑀌 115A2 | 𑀍 115B2 | 𑀎 115C2 |      |      |      |
| 3 | 𑀏 11583 | 𑀐 11593 | 𑀑 115A3 | 𑀒 115B3 | 𑀓 115C3 |      |      |      |
| 4 | 𑀔 11584 | 𑀕 11594 | 𑀖 115A4 | 𑀗 115B4 | 𑀘 115C4 |      |      |      |
| 5 | 𑀙 11585 | 𑀚 11595 | 𑀛 115A5 | 𑀜 115B5 | 𑀝 115C5 |      |      |      |
| 6 | 𑀞 11586 | 𑀟 11596 | 𑀠 115A6 |         | 𑀡 115C6 |      |      |      |
| 7 | 𑀢 11587 | 𑀣 11597 | 𑀤 115A7 |         | 𑀥 115C7 |      |      |      |
| 8 | 𑀦 11588 | 𑀧 11598 | 𑀨 115A8 | 𑀩 115B8 | 𑀪 115C8 |      |      |      |
| 9 | 𑀫 11589 | 𑀬 11599 | 𑀭 115A9 | 𑀮 115B9 | 𑀯 115C9 |      |      |      |
| A | 𑀰 1158A | 𑀱 1159A | 𑀲 115AA | 𑀳 115BA |         |      |      |      |
| B | 𑀴 1158B | 𑀵 1159B | 𑀶 115AB | 𑀷 115BB |         |      |      |      |
| C | 𑀸 1158C | 𑀹 1159C | 𑀺 115AC | 𑀻 115BC |         |      |      |      |
| D | 𑀼 1158D | 𑀽 1159D | 𑀾 115AD | 𑀿 115BD |         |      |      |      |
| E | 𑁀 1158E | 𑁁 1159E | 𑁂 115AE | 𑁃 115BE |         |      |      |      |
| F | 𑁄 1158F | 𑁅 1159F | 𑁆 115AF | 𑁇 115BF |         |      |      |      |

図 2 [4] における悉曇文字の提案

この提案による悉曇文字の符号化は、悉曇の本来の目的であるインド招来仏典の原語による表現には十分であるとみなされていた。しかし、悉曇文字は日本に渡来した際、このような本来の目的以外に、守護札や卒塔婆等において、仏尊の種字としてのシンボリックな利用が広く行われている。

日本の幾つかの宗派では、このような諸尊の種字としてシンボリックに悉曇を用いる場合、インド系文字の用法では重要視されていなかった異体字が重要な意味の違いを持つ。しかしながら、米国の悉曇文字提案においては、異体字は地域的な書体差であるため全てフォント切り替えによって

対応すべきものと扱われており、同一音価の字形群には一つの符号位置しか与えない符号構造であった。

この悉曇文字符号化の米国提案は、現在の悉曇文字の用法には適していないとして審議の延期を提案する文書が日本から提出された[5]。米国提案がフォント切り替えを主張する根拠であった、日本・中国と朝鮮の字形差は、筆記具または書写・版刻の差に由来する書体差の可能性があり、それらはフォント切り替えで対応するとしても、同一書体の中で異体字を使い分けている事例もあるため(図 3)、専門家の判断を仰ぐための審議の延期を提案するものであった。

しかし、早急な符号化を求める米国は、具体的な内容でなく延期するには十分な根拠でないとし、WG2 チェンマイ会議(2012 年 10 月 22 日から 26 日まで)においては、米国提案が、UCS 第三版・追補 2 として、2013 年末までに出版されることが合意された。



図 3 同一書体における異体字の利用例[11]

### 3.2 日本の提案と WG2 ヴィリニウス会議

悉曇文字提案 [4]がそのまま標準化された場合、日本では異体字を利用するための非互換な拡張が発生するなどして悉曇文字符号の利用に混乱が起こる可能性がある。そのため、日本は 2013 年 4 月までに、国内でよく利用されている異体字を収集して追加すること、また日本で追加する異体字は米国の提案の文字と同じタイミングで符号化することを要望する文書を提出し[6]、WG2 ヴィリニウス会議(2013 年 6 月 10 日から 14 日まで)において、提案の説明と議論を行った。

その結果、次節に述べるような課題があったものの、会議中に提案を修正・再提出することにより、2013 年 6 月 14 日までに、米国提案の悉曇文字に加えて、以下の文字を追

b 米国提案は日本の悉曇文書から例示字形を採取したが、中国、朝鮮の悉曇文書に関しても調査しており、特に朝鮮の悉曇文書の字形が大きく異なるため、それらにも対応するにはフォント切り替えが必須であり、日本の悉曇文書に見られる字形差も文字符号レベルでは統合すべきだという論理であった。初期提案(WG2 N4185)では統合を明確にしていなかったが、本提案(WG2 N4294)では明確に統合すべきだと主張している。

加する提案が WG2 において承認された。(図 4)

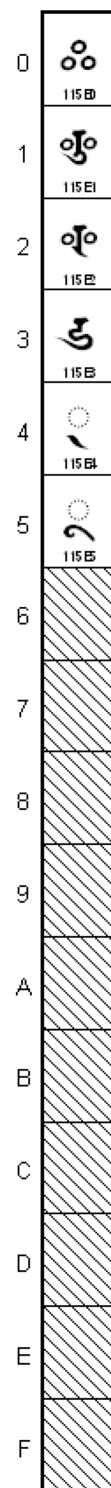


図 4 日本による異体字の追加提案

| 符号    | 文字名称                                 |
|-------|--------------------------------------|
| 115E0 | SIDDHAM LETTER I VARIANT FORM A      |
| 115E1 | SIDDHAM LETTER I VARIANT FORM B      |
| 115E2 | SIDDHAM LETTER II VARIANT FORM A     |
| 115E3 | SIDDHAM LETTER U VARIANT FORM A      |
| 115E4 | SIDDHAM VOWEL SIGN U VARIANT FORM A  |
| 115E5 | SIDDHAM VOWEL SIGN UU VARIANT FORM A |

また、これらの提案を、原提案 ([4]) と同時に国際規格として出版するため、[4]の提案が含まれる UCS 第3版追加2は、承認投票 (FDIS 投票) は行なうが、実際には出版せず、日本の[6]の提案が含まれる UCS 第4版に組み入れて、同時に2014年4月の承認投票以降に出版することが決定した。[7]

## 4. 提案にあたっての課題

### 4.1 字形選択子

日本は当初、悉曇文字の異体字を字形選択子を用いて符号化することを提案した。ただし、UCS では、結合文字に字形選択子を使うことが禁止されているため、結合クラスが0の結合文字に対してのみ、字形選択子を使えるよう、規格を修正することも合わせて提案した。

しかし、WG2 の会議における議論の結果、字形選択子は広義の結合文字であり、結合文字に字形選択子を利用できるようにする提案はアーキテクチャに反しているということが判明し、この提案は認められなかった。また議論の結果、既存の他のインド系文字においても、異体字を別符号化している例が幾つかあることが判明した。WG2 での議論の結果、日本が必要とする悉曇の異体字もこれに倣い、字形選択子を用いず、独立した符号位置で提案することとし、提案文書を修正した。[6]

### 4.2 文字名

悉曇の異体字を独立符号として提案する場合、UCS の規定に従い、異体字にも適切なアルファベットによる文字名を決める必要がある。しかしながら、これらの異体字は必ずしも定着した名称を持たない。そのため、異体字の文字名としては、すでに符号化されている楔型文字等の異体字名を参考に、"VARIATION FORM A, B" という非固有名詞的な名称を用いることにした。

### 4.3 検索・照合

悉曇の異体字は、シンボルとして検索される場合を除き、一般の検索では同じ字音で検索できることが望ましい。そのために、互換正規化 (NFKC/NFKD) [8]によって悉曇の異体字を同音の基底文字に変換できるようにする互換分解を提案に含めることを当初検討した。しかし、会議で関係者と議論した結果、互換正規化による検索・照合は、文字そのものの変更を伴う上、実際には限られた応用でしか用いられないなど、不適切であるとの指摘があり、Unicode 照合アルゴリズム[9]に基づく検索・照合ルールの整備を引き続き目指すこととした。UCS で符号化されたルーン文字等においても、異体字の照合には正規化ではなく本アルゴリズムを用いている。

## 5. まとめ

悉曇文字の UCS 符号化は、最初にアメリカのミシガン大学の Anshuman Pandey 氏によって2012年8月にインド系

文字の符号化アーキテクチャに基づく形で提案された([3, 4])。しかしこの提案は、日本における一般的な悉曇文字の用法には必ずしもそぐわなかった。そのため、日本は WG2 N4407 として、アメリカの提案した文字に対し、異体字選択子を用いて異体字を符号化する提案を行った[6]。2013年6月に WG2 ヴィリニウス会議では異体字選択子を用いた提案は認められなかったため、あらためて別の符号位置に異体字を符号化する提案 (WG2 N4407R) を行い、ISO/IEC 10646 第4版において、米国による提案と同時に符号化することが認められた[7]。

## 参考文献

- 1) ISO/IEC 10646: 2012: Information technology -- Universal Coded Character Set (UCS), International Standard Organization, Geneva (2012)
- 2) IS 13194: Indian Standard Indian Script Code For Information Interchange, Bureau of Indian Standards, New Delhi (1991).
- 3) Anshuman Pandey: "Preliminary Proposal to Encode the Siddham Script in ISO/IEC 10646", WG 2 N4185 (2012)
- 4) Anshuman Pandey: "Proposal to Encode the Siddham Script in ISO/IEC 10646", WG 2 N4294 (2012)
- 5) Suzuki Toshiya: "Feedback to Siddham Proposal", WG 2 N4361
- 6) Taichi Kawabata, *et. al.*: "Proposal to Encode Variant Forms of Siddham Script", WG 2 N4407 (N4407R) (2013)
- 7) WG2 Vilinius Meeting Resolution, WG 2 N4404 (2013)
- 8) Mark Davis, *et. al.*: UAX#15 Unicode Normalization Forms (2012)
- 9) Mark Davis, *et. al.*: UAX#10 Unicode Collation Algorithm (2012)
- 10) The Unicode Consortium, Unicode Standard ver. 6.0.0, The Unicode Consortium, Mountain View (2011).
- 11) 大正一切経刊行会: 大正新脩大蔵経, 大蔵出版, vol.84, p.407 (1924-1934)