Proposal to Encode the Sindhi Script in ISO/IEC 10646

Anshuman Pandey
Department of History
University of Michigan
Ann Arbor, Michigan, U.S.A.
pandey@umich.edu

August 2, 2010

1   Introduction

This is a proposal to encode the Sindhi script in the Universal Character Set (UCS). The Sindhi script described here is the Landa-based script that was standardized by the Government of Bombay in 1868. It is not related to the Arabic-based script used presently for writing the Sindhi language. The proposed script is intended to serve as the representative for the majority of the Landa-based scripts used in Sindh. With the exception of Khojki, most related Sindhi scripts may be unified with that proposed here.

The recommendations made here are based upon a discussion of Sindhi and other Landa-based scripts in the following documents:

- N3596 L2/09-101 "Proposal to Encode the Khojki Script in ISO/IEC 10646"
- N3766 L2/10-011 "A Roadmap for Scripts of the Landa Family"
- N3767 L2/10-012R "Preliminary Proposal to Encode the Sindhi Script in ISO/IEC 10646"
- N3768 L2/10-013R "Preliminary Proposal to Encode the Landa Script in ISO/IEC 10646"

A description of Landa and its Sindhi and Punjabi regional sub-families are given in the above documents. The justifications for encoding Khojki as an independent script are provided in N3596 L2/09-101.

The present document extends N3767 L2/10-012R. It provides additional details that will facilitate the encoding and implementation of Sindhi, including background information, a revision of the code chart and names list, and the inclusion of new specimens. A proposal summary form is also enclosed.

2   Overview

The Sindhi script proposed here for encoding was published as a standard by the Government of Bombay in 1868.[1] It was used for writing the Sindhi [snd] language. The script was developed by Narayan Jagannath

---

[1]   Grierson 1919: 18.

Mehta, the Deputy Educational Inspector in Sindh.[2] It is based primarily on the Khudawadi script used in Hyderabad, which was the most complete and well-regarded of the Landa-based scripts of Sindh; some glyphic elements of Shikarpuri were also adopted (see N3767 L2/10-012R for details).

The official script was known as 'Hindi Sindhi' or 'Hindu Sindhi'; the descriptors 'Hindi' and 'Hindu' refer to the derivation of the script from an 'Indian' (ie. Brahmi) model in order to distinguish it from an Arabic-based script for Sindhi that was developed by the Government of Bombay in 1852. The script is referred to here as 'Sindhi' and 'Standard Sindhi'.

The creation of a formal Sindhi script was driven by the need to standardize education and to develop a uniform medium for court records.[3] The Landa-based scripts of Sindh are unstandardized scripts whose character repertoire and glyph shapes differ from region to region. They are 'incomplete' scripts, as indicated by the name 'Landa' (*laṇḍā*), which means 'clipped'. The term refers to certain characteristics of the script, such as the absence of vowel signs, the use of a single letter for representing an aspirated and unaspirated consonant pair, the absence of word spacing and punctuation, the use of digits for consonant letters, and other shorthand practices. Standard Sindhi was developed to provide a 'complete' script that would fully represent the languages of Sindh.

In addition to its use in education and administration, the Standard Sindhi script was used for literary activities. It was adapted for print and at least two metal fonts were produced, specimens of which appear in Figure 5 and Figure 11. Sindhi epic tales such as *Dodo Chanesar* were published in the script (see Figure 4). Literary journals such as *Sukhiri* were published in the script beginning in 1899.[4] After the turn of the 20th century the British and Foreign Bible Society printed translations of Christian religious texts into the Sindhi language using 'Hindu Sindhi' (see Figure 5 and Figure 6).

Standard Sindhi was eventually replaced by the Arabic-based Sindhi script. This extended Arabic script is currently the official writing system for Sindhi in Pakistan and India. Sindhi is also written in Devanagari in India.

3   A Unified Encoding for the Scripts of Sindh

Numerous Landa-based scripts were used in Sindh (see charts in N3767 L2/10-012R). Several of these scripts are named after the regions, communities, and commercial activities with which they are associated: Arora, Baniya, Bhatia, Haidarabadi, Karadi, Khudawadi, Khwaja (Khojki), Hattai, Hatvanika, Landa, Larai, Lohanaki (Lohana), Maimon, Rajai, Sakkar, Shikarpuri, Sewhani Bhabhira, Thatta (Thattai), Vaniya, Wangai, Waniko. The majority of these scripts exhibit features fitting the description of Landa as an 'incomplete' script. Apart from Khojki and Khudawadi, these scripts are too minor to be encoded in the UCS.

Therefore, if the need arises to represent any of these minor Sindhi scripts, they should be unified with the Standard Sindhi proposed here. A unified block for Sindhi based on Standard Sindhi provides an effective means for managing local Sindhi scripts and for supporting the structural and orthographic features that distinguish them from other scripts associated with Landa. The Standard Sindhi script is an appropriate representative for a unified Sindhi block. It possesses the complete character repertoire necessary for uniquely representing sounds of the Sindhi languages. Moreover, its glyphs are based upon the well-regarded Khudawadi script and are attested in several other local varieties. The exception to this unification of Sindhi scripts is Khojki, which has been proposed for independent encoding on account of its use as a literary and ecclesiastical script (see N3596 L2/09-101).

---

[2] Dow 1976: 55.   [3] Government of the Bombay Presidency 1869: 213.   [4] Datta 1987: 1892.

4   Implementation Details

4.1   Script Name

The name 'Sindhi' is assigned to the script block. It is more appropriate than 'Standard Sindhi', 'Hindi Sindhi', or 'Hindu Sindhi'. Although 'Sindhi' is generic, there is no other script with the name and the characters are sufficiently distinct from the Arabic-based script used for Sindhi so as to prevent confusion.

4.2   Encoding Model

The general structure (phonetic order, *mātrā* reordering, use of *virāma*, etc.) of Sindhi is similar to that of other Indic scripts based upon the Brahmi model. It is written from left-to-right. Sindhi should be implemented according to the *virāma* model, as used for Devanagari.

4.3   Allocation

Sindhi is currently allocated in the Supplementary Multilingual Plane (SMP) at the range U+11600..1165F. Only four columns are needed to fully represent Sindhi; the fifth column may be released.

4.4   Character Repertoire

A total of 69 characters is required to encode a basic character set for Sindhi. This repertoire consists of 37 consonant letters, 10 vowel letters, 9 vowel signs, 2 various signs, 1 VIRAMA, and 10 digits. Character names follow the UCS convention for Brahmi-based scripts. A preliminary code chart and names list are provided in Table 1 and Table 2.

4.5   Encoding Order

Charts show different ordering schemes for Sindhi. Some sequences place the implosive consonant letters before their unaspirated counterparts, others place them between the unaspirated and aspirated counterparts. The encoding order selected for Sindhi is based upon the general Brahmi order, which is attested in several charts.

4.6   Vowel Signs

Vowel signs function as in Devanagari. The ि◌ VOWEL SIGN I is written to the left of the consonant with which it combines.

4.7   Representation of Vowel Letters

Some atomic vowel letters may be represented using a sequence of a base vowel letter and a vowel sign. This practice is not recommended. The atomic character should always be used. The characters in question are specified below:

RECOMMENDED	NOT RECOMMENDED
ꟺꟷ VOWEL LETTER AA	ꟺ VOWEL LETTER A + ◌ꟷ VOWEL SIGN AA
ꟺ̃ VOWEL LETTER E	ꟺ VOWEL LETTER A + ◌̂ VOWEL SIGN E
ꟺ̃ VOWEL LETTER AI	ꟺ VOWEL LETTER A + ◌̃ VOWEL SIGN AI
ꟺ̃ VOWEL LETTER O	ꟺ VOWEL LETTER A + ◌̃ VOWEL SIGN O
ꟺ̃ VOWEL LETTER AU	ꟺ VOWEL LETTER A + ◌̃ VOWEL SIGN AU

4.8   Consonant Conjuncts

Consonant clusters are generally using with an explicit ◌ VIRAMA. Half-forms and ligatures are not used for representing consonant conjuncts.

4.9   Nasalization

Sindhi uses only ◌̇ ANUSVARA to indicate nasalization. There is no attested character corresponding to the CANDRABINDU found in other northern Indic scripts.

4.10   Usage of NUKTA

The ◌̣ NUKTA is used to represent sounds not native to Sindhi, such as those analogous to Devanagari KHHA, GHHA, FA, QA, and ZA (see bottom row in Figure 7).

4.11   Punctuation

Sindhi uses *daṇḍā*s and Latin marks for punctuation. Sindhi *daṇḍā*s may be unified with those of Devanagari.

4.12   Digits

Sindhi has a full set of digits.

4.13   Number Forms

Fraction signs and currency marks are found in Sindhi documents. They are identical to the characters encoded in the Common Indic Number Forms block (U+A830..U+A83F) and should be unified with them. See Pandey (2007) for more information on these characters.

4.14   Collation

The collating order for Sindhi is as follows:

ꟺ A > ꟺꟷ AA > ơ I > ơơ II > Ə U > Ə UU > ꟺ̃ E > ꟺ̃ AI > ꟺ̃ O > ꟺ̃ AU > ꟺ KA > ꗙ KHA > ꗨ GA > ꙍ GGA > ꙝ GHA > ꙟ NGA > ꙡ CA > ꙣ CHA > ꙥ JA > ꙧ JJA > ꙩ JHA > ꙫ NYA > ꖞ TTA > ꖟ TTHA > ꖠ DDA > ꖡ DDDA > ꖢ RRA > ꖣ DDHA > ꖤ NNA > ꖥ TA > ꖦ THA > ꖧ DA > ꖨ DHA > ꖩ NA > ꖪ PA > ꖫ PHA > ꖬ BA > ꖭ BBA > ꖮ BHA > ꖯ MA > ꖰ YA > ꖱ RA > ꖲ LA >

O VA > **ಠ** SHA > **ಡ** SA > **3** HA > **◌ा** SIGN AA > **ि** SIGN I > **◌ी** SIGN II > **◌ु** SIGN U > **◌ू** SIGN UU >
◌ॆ SIGN E > **◌ॅ** SIGN AI > **◌ॊ** SIGN O > **◌ॉ** SIGN AU > **◌ं** ANUSVARA > **◌्** VIRAMA

Combinations of a letter + ◌़ NUKTA are sorted with the base letter.

4.15   Linebreaking

Letters, vowel signs, and digits behave as in Devanagari.

4.16   Character Properties

The properties for Sindhi characters in the Unicode Character Database format are:

```
11600;SINDHI LETTER A;Lo;0;L;;;;;N;;;;;
11601;SINDHI LETTER AA;Lo;0;L;;;;;N;;;;;
11602;SINDHI LETTER I;Lo;0;L;;;;;N;;;;;
11603;SINDHI LETTER II;Lo;0;L;;;;;N;;;;;
11604;SINDHI LETTER U;Lo;0;L;;;;;N;;;;;
11605;SINDHI LETTER UU;Lo;0;L;;;;;N;;;;;
11606;SINDHI LETTER E;Lo;0;L;;;;;N;;;;;
11607;SINDHI LETTER AI;Lo;0;L;;;;;N;;;;;
11608;SINDHI LETTER O;Lo;0;L;;;;;N;;;;;
11609;SINDHI LETTER AU;Lo;0;L;;;;;N;;;;;
1160A;SINDHI LETTER KA;Lo;0;L;;;;;N;;;;;
1160B;SINDHI LETTER KHA;Lo;0;L;;;;;N;;;;;
1160C;SINDHI LETTER GA;Lo;0;L;;;;;N;;;;;
1160D;SINDHI LETTER GGA;Lo;0;L;;;;;N;;;;;
1160E;SINDHI LETTER GHA;Lo;0;L;;;;;N;;;;;
1160F;SINDHI LETTER NGA;Lo;0;L;;;;;N;;;;;
11610;SINDHI LETTER CA;Lo;0;L;;;;;N;;;;;
11611;SINDHI LETTER CHA;Lo;0;L;;;;;N;;;;;
11612;SINDHI LETTER JA;Lo;0;L;;;;;N;;;;;
11613;SINDHI LETTER JJA;Lo;0;L;;;;;N;;;;;
11614;SINDHI LETTER JHA;Lo;0;L;;;;;N;;;;;
11615;SINDHI LETTER NYA;Lo;0;L;;;;;N;;;;;
11616;SINDHI LETTER TTA;Lo;0;L;;;;;N;;;;;
11617;SINDHI LETTER TTHA;Lo;0;L;;;;;N;;;;;
11618;SINDHI LETTER DDA;Lo;0;L;;;;;N;;;;;
11619;SINDHI LETTER DDDA;Lo;0;L;;;;;N;;;;;
1161A;SINDHI LETTER RRA;Lo;0;L;;;;;N;;;;;
1161B;SINDHI LETTER DDHA;Lo;0;L;;;;;N;;;;;
1161C;SINDHI LETTER NNA;Lo;0;L;;;;;N;;;;;
1161D;SINDHI LETTER TA;Lo;0;L;;;;;N;;;;;
1161E;SINDHI LETTER THA;Lo;0;L;;;;;N;;;;;
1161F;SINDHI LETTER DA;Lo;0;L;;;;;N;;;;;
11620;SINDHI LETTER DHA;Lo;0;L;;;;;N;;;;;
11621;SINDHI LETTER NA;Lo;0;L;;;;;N;;;;;
11622;SINDHI LETTER PA;Lo;0;L;;;;;N;;;;;
11623;SINDHI LETTER PHA;Lo;0;L;;;;;N;;;;;
11624;SINDHI LETTER BA;Lo;0;L;;;;;N;;;;;
11625;SINDHI LETTER BBA;Lo;0;L;;;;;N;;;;;
11626;SINDHI LETTER BHA;Lo;0;L;;;;;N;;;;;
11627;SINDHI LETTER MA;Lo;0;L;;;;;N;;;;;
11628;SINDHI LETTER YA;Lo;0;L;;;;;N;;;;;
11629;SINDHI LETTER RA;Lo;0;L;;;;;N;;;;;
1162A;SINDHI LETTER LA;Lo;0;L;;;;;N;;;;;
1162B;SINDHI LETTER VA;Lo;0;L;;;;;N;;;;;
1162C;SINDHI LETTER SHA;Lo;0;L;;;;;N;;;;;
```

```
1162D;SINDHI LETTER SA;Lo;0;L;;;;;N;;;;;
1162E;SINDHI LETTER HA;Lo;0;L;;;;;N;;;;;
1162F;SINDHI SIGN ANUSVARA;Mn;0;NSM;;;;;N;;;;;
11630;SINDHI VOWEL SIGN AA;Mc;0;L;;;;;N;;;;;
11631;SINDHI VOWEL SIGN I;Mc;0;L;;;;;N;;;;;
11632;SINDHI VOWEL SIGN II;Mc;0;L;;;;;N;;;;;
11633;SINDHI VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
11634;SINDHI VOWEL SIGN UU;Mn;0;NSM;;;;;N;;;;;
11635;SINDHI VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
11636;SINDHI VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;
11637;SINDHI VOWEL SIGN O;Mn;0;NSM;;;;;N;;;;;
11638;SINDHI VOWEL SIGN AU;Mn;0;NSM;;;;;N;;;;;
11639;SINDHI SIGN NUKTA;Mn;7;NSM;;;;;N;;;;;
1163A;SINDHI SIGN VIRAMA;Mn;9;NSM;;;;;N;;;;;
11640;SINDHI DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
11641;SINDHI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
11642;SINDHI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
11643;SINDHI DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
11644;SINDHI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
11645;SINDHI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
11646;SINDHI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
11647;SINDHI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
11648;SINDHI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
11649;SINDHI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
```

5   References

The American Bible Society. 1938. *The Book of a Thousand Tongues: Being Some Account of the Translation and Publication of All or Part of The Holy Scriptures Into More Than a Thousand Languages and Dialects With Over 1100 Examples from the Text.* Edited by Eric M. North. New York and London: Harper & Brothers.

Boivin, Michel. Photograph of the cover of a book containing the tale of *Dodo Chanesar*: "Le héros légendaire Dodo représenté dans un livet composé en écriture khudabadi." http://www.mboivin.fr/photos.html.

British and Foreign Bible Society. 1911. *St. Matthew in Hindu Sindhi.* Lahore.

Datta, Amaresh. 1987. *Encyclopaedia of Indian literature*, vol. 2 (Devraj to Jyoti). New Delhi: Sahitya Akademi.

Dow, Hugh. 1976. *A Note on the Sindhi Alphabet. Asian Affairs*, vol. 7, no. 1, February 1976, pp. 54–56.

Faulmann, Carl. 1880. *Das Buch der Schrift: Enthaltend die Schriftzeichen und Alphabete aller Zeiten und aller Völker der Erdkreises*. Zweite Vermehrte und verbesserte Auflage. Wein: Der Kaiserlich-Königlichen Hof- und Staatsdruckerei.

Government of the Bombay Presidency. 1869. *Report of the Department of Public Instruction in the Bombay Presidency, for the Year 1868–69.* Bombay: Education Society's Press, Byculla.

Jetley, Murlidhar Kishinchand. 1999. بوليءَ جو سِرشتو ۽ لِكَاوَتَ (سِنڌي بوليءَ جون لِپِيون) [= *boliya jo sirishto ain likhawata (sindhi boliya jawan lipiyon)*. Pune: Akhil Bharatiya Sindhi Sahitya Vidvat Parishad.

Pandey, Anshuman. 2007. "Proposal to Encode North Indic Number Forms in ISO/IEC 10646." ISO/IEC JTC1/SC2/WG2 N3367 L2/07-354. October 7, 2007. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3367.pdf

———. 2009. "Proposal to Encode the Khojki Script in ISO/IEC 10646." N3596 L2/09-101. March 25, 2009. `http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3596.pdf`

———. 2010. "A Roadmap for Scripts of the Landa Family." N3766 L2/10-011R. February 9, 2010. `http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3766.pdf`

———. 2010. "Preliminary Proposal to Encode the Sindhi Script in ISO/IEC 10646." N3767 L2/10-012R. February 9, 2010. `http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3767.pdf`

Stack, George. 1849. *A Grammar of the Sindhi Language*. Bombay: American Mission Press.

6   Acknowledgments

	1160	1161	1162	1163	1164
0	ᴍ 11600	४ 11610	ᴚ 11620	◌। 11630	๐ 11640
1	ᴍᴵ 11601	੬ɜ 11611	∨ 11621	୮◌ 11631	੧ 11641
2	ᢐ 11602	ഖ 11612	⤳ 11622	◌⸣ 11632	੨ 11642
3	ഄ 11603	ᴍᢐ 11613	⊅ 11623	◌ 11633	Ɜ 11643
4	Ꮹ 11604	ഖ 11614	Ꮤ 11624	◌ 11634	४ 11644
5	Ꮹ 11605	੩· 11615	ഄ 11625	◌ 11635	੫ 11645
6	ᴟ̃ 11606	ᴄ 11616	Ꮤ 11626	◌ 11636	੬ 11646
7	ᴟ̃ 11607	ᴇ 11617	ᴫ 11627	◌ 11637	੭ 11647
8	ᴟ̃ 11608	Ɜ 11618	ᴣᴵ 11628	◌ 11638	८ 11648
9	ᴟ̃ 11609	ੲ 11619	੨ 11629	◌ 11639	౺ 11649
A	ᴫ 1160A	੭ 1161A	ᴤ 1162A	◌ 1163A	
B	੫ 1160B	੬ 1161B	0 1162B		
C	ᴆ 1160C	ⅲ 1161C	ᴄ 1162C		
D	ᴅᴵ 1160D	੭ 1161D	ᴨ 1162D		
E	ᴕ 1160E	ᴍ 1161E	⊃ 1162E		
F	੨ 1160F	ᴕ 1161F	˙◌ 1162F		

Figure 1: Proposed code chart for Sindhi

8

Independent vowels

11600	ℳ	SINDHI LETTER A
11601	ℳІ	SINDHI LETTER AA
11602	ᚋ	SINDHI LETTER I
11603	ᚋᚋ	SINDHI LETTER II
11604	ᘓ	SINDHI LETTER U
11605	ᘒ	SINDHI LETTER UU
11606	ᚌ	SINDHI LETTER E
11607	ᚌ	SINDHI LETTER AI
11608	ᚌ	SINDHI LETTER O
11609	ᚌ	SINDHI LETTER AU

Consonants

1160A	ℳ	SINDHI LETTER KA
1160B	�9	SINDHI LETTER KHA
1160C	ℬ	SINDHI LETTER GA
1160D	ᶁ	SINDHI LETTER GGA
1160E	ᶀ	SINDHI LETTER GHA
1160F	ꝛ·	SINDHI LETTER NGA
11610	४	SINDHI LETTER CA
11611	ᴇ	SINDHI LETTER CHA
11612	ᴂ	SINDHI LETTER JA
11613	ℋᚋ	SINDHI LETTER JJA
11614	ᴂ	SINDHI LETTER JHA
11615	ɜ·	SINDHI LETTER NYA
11616	ᴄ	SINDHI LETTER TTA
11617	ᴇ	SINDHI LETTER TTHA
11618	₃	SINDHI LETTER DDA
11619	ℰ	SINDHI LETTER DDDA
1161A	ℨ	SINDHI LETTER RRA
1161B	ᴜ	SINDHI LETTER DDHA
1161C	ᴽ	SINDHI LETTER NNA
1161D	ᴣ	SINDHI LETTER TA
1161E	ℳ	SINDHI LETTER THA
1161F	ᴀ	SINDHI LETTER DA
11620	ᴀ	SINDHI LETTER DHA
11621	ᴗ	SINDHI LETTER NA
11622	ᴛ	SINDHI LETTER PA
11623	ᴢ	SINDHI LETTER PHA
11624	ᴡ	SINDHI LETTER BA
11625	ᴈ	SINDHI LETTER BBA
11626	ᴫ	SINDHI LETTER BHA
11627	ᴨ	SINDHI LETTER MA
11628	ℳ	SINDHI LETTER YA
11629	ꝛ	SINDHI LETTER RA
1162A	ᴥ	SINDHI LETTER LA
1162B	ᴏ	SINDHI LETTER VA
1162C	ᴕ	SINDHI LETTER SHA
1162D	ᴛ	SINDHI LETTER SA
1162E	ᴢ	SINDHI LETTER HA

Various signs

1162F	◌̇	SINDHI SIGN ANUSVARA

Dependent vowel signs

11630	◌ı	SINDHI VOWEL SIGN AA
11631	◌	SINDHI VOWEL SIGN I
11632	◌	SINDHI VOWEL SIGN II
11633	◌	SINDHI VOWEL SIGN U
11634	◌	SINDHI VOWEL SIGN UU
11635	◌	SINDHI VOWEL SIGN E
11636	◌	SINDHI VOWEL SIGN AI
11637	◌	SINDHI VOWEL SIGN O
11638	◌	SINDHI VOWEL SIGN AU

Various signs

11639	◌	SINDHI SIGN NUKTA
1163A	◌	SINDHI SIGN VIRAMA

Digits

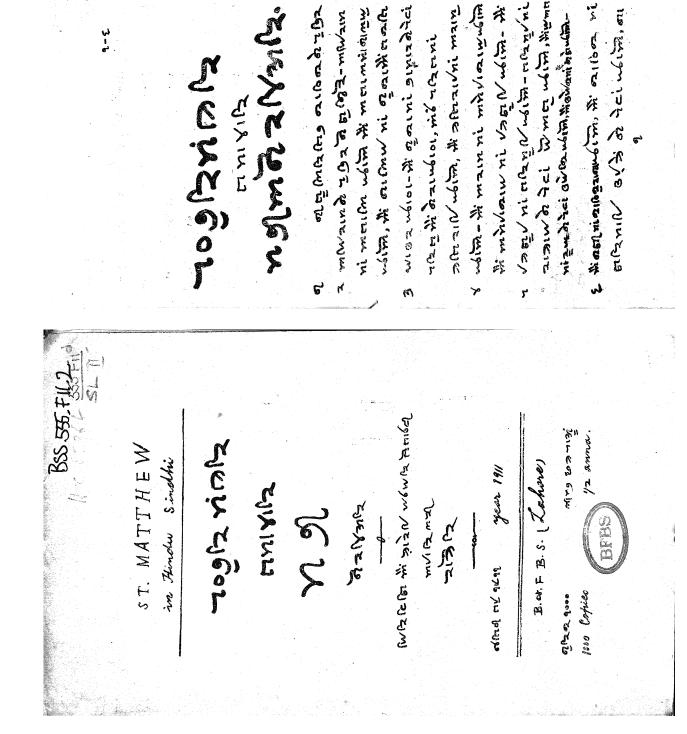|		
11640	ᴏ	SINDHI DIGIT ZERO
11641	ᴦ	SINDHI DIGIT ONE
11642	ᴩ	SINDHI DIGIT TWO
11643	ᴣ	SINDHI DIGIT THREE
11644	ᴘ	SINDHI DIGIT FOUR
11645	ᴚ	SINDHI DIGIT FIVE
11646	ℰ	SINDHI DIGIT SIX
11647	ᴜ	SINDHI DIGIT SEVEN
11648	ᴄ	SINDHI DIGIT EIGHT
11649	ᴜ	SINDHI DIGIT NINE

Figure 2: Proposed names list for Sindhi

9

Figure 3: Characters of the Standard Sindhi script (from M. K. Jetley 1999: 90).

Figure 4: Cover of a book containing the epic *Dodo Chanesar* written in Standard Sindhi (photo from Michel Boivin).

Figure 5: Cover and first page of St. Matthew in Standard Sindhi (from British and Foreign Bible Society 1911).

SIHONG　　　　　　　　　869

(No specimen available.)

Spoken in the Sihong district of southern Borneo, Netherlands East In-dies. First publication, Genesis, chapters 1–26:5, in 1858 at Band-ermassin, Borneo, by the Rhenish Missionary Society; tr. by E. Denninger.

SINDHI: Kachchhi dialect　　871

Devanagari characters

तूं पिंढ बे जुदा के विको तींओ जुदा आहे जम एं उन बीच मंदगी कर ।

Mt. 4.10

Spoken by about 500,000 people in Bombay and Baroda, western India. First publication, St. Matthew's Gospel in 1834 at Bombay by the BFBS; tr. by James Gray, chaplain to the East India Company.

(For No. 872 see next page.)

SLAVE　　　　　　　　　873

Syllabic characters

15 ᑕᑐ ᐅᐊ ᓂᐸᐊᑕᓕ ᐅᐊᑐᓕ ᐊᑌ, ᐊᑦᑐ ᒡᑎ ᑊᐱᐅ ᐅᒧᒧ. 16 ᑊᑕ ᐁᑲᐅᕀ ᑕᐅᑲ ᓕᐳ, ᑌ ᐅᕐ ᒍᒧᑎ ᐱ ᒡᑌᓂᕐᑭ, ᑕᑐ ᐅᐊ ᕃᓂᐸᐊᑕᓕ. ᐅᐊᑐᓕ. ᑲ ᐊᑌ, ᐊᑦᑐ ᒡᑎ ᑊᐱᐅ ᐅᒧᒧ. 17 ᑊᑕ ᑊᑕᐅ ᑌ ᐅᕐ ᑖ ᑧ ᑕᐊ,

Roman characters

15 Eyitta tene oyi yego ye keniethet nite tsiedathe ka ile; kulu ithlasi onda ga agonte. 16 Tta Niotsi otie nun gonieto eyitta te yazi inthligi zon nakhe ganiti, tene oyi ye keniethet nite tsiedethet ga ile, kulu ithlasi oonda ga olili. 17 Tta Niotsi te Yazi nun tsun dea nun

1905

Spoken by Indians living along the Mackenzie River, northwestern Can-ada, North America. First publication, St. Mark's Gospel (syllabic characters) in 1868 by the BFBS; tr. by W. W. Kirkby of the CMS. St. John's Gospel, 1870; St. Matthew's and St. Mark's Gospels, 1886; tr. by W. C. Bompas, transliterated by William Day Reeve; St. Luke's and St. John's Gospels, 1890. New Testament, 1891. Edi-tions in roman characters: St. Mark's Gospel, 1874; tr. by William Carpenter Bompas of the CMS. The Gospels, 1883; The Acts, 1890; Romans–Revelation, 1891. The Gospels, 1905; revised by W. Spend-love. CP: BFBS.

SINDHI　　　　　　　　　870

Devanagari characters

सांईं, जो तुंझ जो खुदा, तैंईं बे पूविंओ, चैं रखो उमे जो ईं बंदगी किजं ।

Mt. 4.10; 1850

Arabic characters (for Moslems)

چالاءِ تہ حدا دنيا کي امرو پيار ڪيو آهي, جو پنهنجو هڪڙوئي پُٽ ڏنو اٿس, تہ جيڪوڪ مٿس وساهہ آڻي, سو چٽ نہ ٿئي, پر هميشہ جو جياپو پائي ♦

1917

Arabic characters (for Hindus)

چوتہ ايتر حڪت کي امرو پيار ڪيو آهي, جو پنهجو هڪڙوئي پٽو ڏنو اٿس تو جڪوڪ مٿس وسراش ڪوي سوملن نہ ٿُنجبا پرنت جبون پائي ♦

1908

Gurmukhi characters

ਭਾ ਖਾ ੩ ਈਸ਼੍ਮੂਤ ਜ਼ਰਾਤ ਥੇ ਇਹੜੇ ਪਿਆਰੇ ਰਖਗੇ ਜੋ ਪਹੰਜੇ ਹਿਕਿੜੇ ਈ ਜ਼ਲਯਲ ਪੁਟੂ ਡਿਨਾਇਟੀ ੩ ਜੋਰੇ ਬੇ ਤਹਿ ੩ ਵੈਮਾਤੂ ਆਲੇ ਮੇ ਨਾਮੁ ਢ ਧਿਠੇ ਵੇਤਰਿ ਸਰਾ ਜ਼ਿਮਲ ਲਹੇ ॥

1877

Banya characters

ꢪꢿꢀ�9 ꢬꢸꢺꢸꢪꢬ ꢥꢙꢰꢎꢙꢾ ꢪꢺꢄꢿ/ꢬꢑꢾꢬꢒꢸ ꢪꢺꢽ ꢤꢷ ꢭꢲꢪꢑꢸ ꢡꢸꢬ ꢡꢶꢸꢥ ꢪꢪꢭꢺ 9 ꢠꢺꢭꢷꢿ ꢘ ꢪꢭꢺ ꢆꢬꢭꢡ ꢁꢸꢷꢸ ꢥꢭꢺ ꢥꢶꢪꢷꢪ ꢬꢑꢷꢰꢺ ꢡꢭꢞ. ꢑꢷꢪꢷ.

1926

Spoken by some 3,000,000 people in Sindh and neighboring districts, western India. First publication, St. Matthew's Gospel in 1825 at Serampore by the Mission Press; tr. by the Serampore missionaries (See No. 87). St. Matthew's Gospel (Devanagari characters), BFBS, Bom-bay, 1850; tr. by G. Stack. St. John's Gospel (Arabic characters), BFBS, Karachi, 1858; tr. by A. Burn of the CMS; reprinted in Gur-mukhi characters, 1859. Genesis, 1860; St. Matthew's and St. Mark's Gospels, London, 1868; St. Luke's Gospel, 1869; the Acts, 1870. Mr. Burn's translations were issued both in Arabic and Gurmukhi characters. New Testament, 1890; tr. by C. W. Isenberg and G. Shirt. The Psalter, BFBS, Ludhiana, 1879; tr. by Mr. Shirt: Proverbs, 1885; Isaiah, 1890. Revised Version: St. Mark's and St. John's Gospels, BFBS, Lahore, 1908; tr. by a committee of missionaries and Sindhi schol-ars headed by R. Redman, issued in editions prepared for Moslems and for Hindus. St. Matthew's and St. Luke's Gospels, 1911. St. Matthew's Gospel in Banya characters, 1913.

Figure 6: A specimen of the New Testament printed in the Sindhi script (from The American Bible Society 1938: 297.) The script labeled 'Banya characters' is Standard Sindhi.
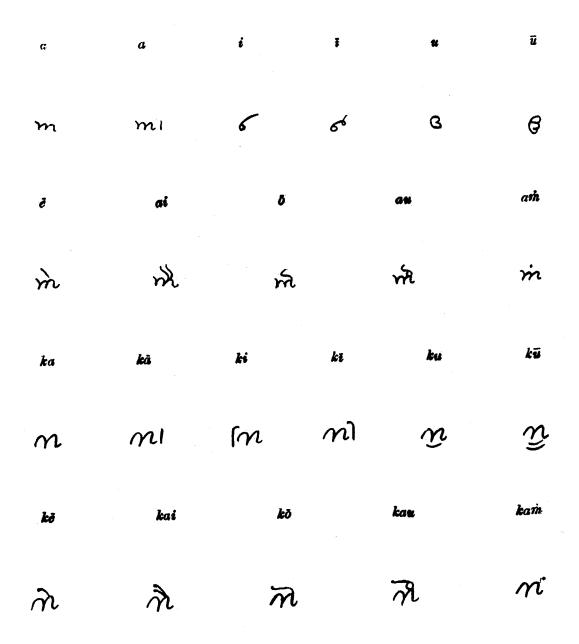
13

CONSONANTS.

No.	Roman Character.	Original Khudawadi Character.	Improved Hindi Sindhi Character.	No.	Roman Character.	Original Khudawadi Character.	Improved Hindi Sindhi Character.	No.	Roman Character.	Original Khudawadi Character.	Improved Hindi Sindhi Character.
1	*ka*			15	*ṛa*	No. 32 used instead.		29	*bha*	No. 28 used instead.	
2	*kha*			16	*ḍḍa*			30	*ma*		
3	*gga*			17	*ḍa*			31	*ya*	No. 9 used instead.	
4	*ga*			18	*ḍha*			32	*ra*		
5	*gha*	No. 4 used instead.		19	*ṇa*			33	*la*		
6	*ṅa*			20	*ta*			34	*va, wa*		
7	*cha*			21	*tha*			35	*sa*		
8	*chha*			22	*da*			36	*sha*	No. 35 used instead.	
9	*ja*			23	*dha*	No. 22 used instead.		37	*ha*		
10	*jja*			24	*na*			38	*khē*	No. 2 used instead.	
11	*jha*			25	*pa*			39	*zē*	No. 9 used instead.	
12	*ña*			26	*pha*			40	*ghain*	No. 4 used instead.	
13	*ṭa*			27	*bba*			41	*fē*	No. 26 used instead.	
14	*ṭha*			28	*ba*						

Figure 7: A comparison of consonant letters of Khudawadi and Standard Sindhi (from Grierson 1919: 20).

Figure 8: Vowel letters and consonant-vowel combinations in Standard Sindhi (from Grierson 1919: 19).

The first few lines of the preceding specimen in the Khudāwadi character.

The first few lines of the preceding specimen in the improved Hindu Sindhi character.

88

[No. I.]

INDO-ARYAN FAMILY.   NORTH-WESTERN GROUP.

SINDHI.

STANDARD DIALECT.   DISTRICT HYDEBABAD.

SPECIMEN I.

NĀGARĪ CHARACTER.

VOL. VIII, PART I.

02

Figure 9: Text in Devanagari, Khudawadi, and Standard Sindhi (from Grierson 1919: 99, 101).

[No. 4.]

INDO-ARYAN FAMILY.   NORTH-WESTERN GROUP.

SINDHI.

STANDARD DIALECT.   DISTRICT KARACHI.

SPECIMEN I.

Figure 10: Specimen of hand-written Standard Sindhi (from Grierson 1919: 115–116).

VOL. VIII, PART I.   Q 2

17

SINDH- UND MULTANSCHRIFT.

Sindh		Multan		Sindh		Multan		Sindh		Multan	
Zeichen	Wert	Zeichen	Wert	Zeichen	Wert	Zeichen	Wert	Zeichen	Wert	Zeichen	Wert
	a		a		ṅa, ńa				pha		pha
	i		i		ṭa				ba		ba
	u		u		ṭha				bha		
	ka		ka		ḍa		ḍa		ma		ma
	kha		kha		ṇa		ṇa		ya		ya
	ga		ga		ta		ta		ra		ra
	gha				tha		tha		la		la
	tša		tša		da		da		va		va
	tšha		tšha		dha		dha		sa		sa
	dža		dža		na		na		ha		ha
	džha				pa		pa		tra		tra

Die vorstehenden Schriften sind in doppelter Art beachtenswert, erstens weil sie in gleicher Weise wie die semitischen Schriften die Vokale in der Mitte der Wörter nicht schreiben, z. B. Sindhisch ᴣᴣᴣᴧ *tuhidžo*, zweitens weil ihre Zeichen sich an die semitischen Schriften anlehnen, anderseits sich in der Devanagarischrift vorfinden, wenn diese von der Paliform abweicht; so hat ᴣ dieselbe Bedeutung auf Pehlewi-Münzen, ᴣ entspricht dem ᴣ im Aramäischen, ᴣ erinnert an die phönikische Form ᴧ χ, ᴧ an das phönikische ᴧ *š*. Den von den Inschriftformen abweichenden Devanagarizeichen क *ka* entspricht Multan ᴣ, dem ख *kha* (Inschrift ᴣ) entspricht Sindh ᴣ *gha*, ᴨ kann ebensowohl von der Inschrift ᴨ, wie von Sindh ᴧ *ga* abstammen, dem ᴣ *tša* entspricht Sindh ᴣ *tšha*, dagegen Inschrift ᴧ *tša* dem Devanagari ळ *la*, Devanagari ण hat seine Analogie nur in Sindh ᴧ (dagegen Inschrift ᴣ *ṇa*), ebenso व Sindh ᴧ (Inschrift ᴧ), फ *p*, Sindh ᴧ (Inschrift ᴧ).

Die auffallenden Abweichungen der beiden Schriften von Sindh und Multan beweisen, dass von einer einfachen Entlehnung keine Rede sein kann, denn wir finden Sindh ᴧ *tša* als Multan ᴧ *ra*, Sindh ᴧ *tha* als Multan ᴧ *pha*, dagegen befindet sich Multan ᴧ *tšha* in Übereinstimmung mit Sindh ᴧ *tša*.

Mit Rücksicht auf die Einwanderung der arischen Stämme von Norden liegt die Vermuthung nahe, dass die vorstehenden Schriften des nördlichen Indiens den Grundstock der vedischen Schrift abgaben. Durch die Herrschaft der magadhischen Schrift während der Zeit, wo in ganz Vorderindien der Buddhismus herrschte, wurde die heimische Schrift in den Schatten gestellt. Man zog die fremde Schrift wegen ihres Reichthums an Zeichen und wegen ihrer festen Schreibregeln vor. So kam es, dass die heimische Schrift, welche sich trotzdem im Volke erhielt, zu den gelehrten Werken nicht verwendet und diesfalls die Devanagarischrift vorgezogen wurde.

121

Figure 11: Comparison of metal-font glyphs for Standard Sindhi and Multani (from Faulmann 1880: 121).

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[5]

Please fill all the sections A, B and C below. Please read Principles and Procedures Document (P & P) from
http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.

A. Administrative

1. Title: **Proposal to Encode the Sindhi Script in ISO/IEC 10646**
2. Requester's name: **University of California, Berkeley Script Encoding Initiative (Universal Scripts Project);
author: Anshuman Pandey (pandey@umich.edu)**
3. Requester type (Member Body/Liaison/Individual contribution): **Liaison contribution**
4. Submission date: **2010-08-02**
5. Requester's reference (if applicable): **N/A**
6. Choose one of the following:
   (a) This is a complete proposal: **No**
   (b) or, More information will be provided later: **Yes**

B. Technical - General

1. Choose one of the following:
   (a) This proposal is for a new script (set of characters): **Yes**
       i. Proposed name of script: **Sindhi**
   (b) The proposal is for addition of character(s) to an existing block: **No**
       i. Name of the existing block: **N/A**
2. Number of characters in proposal: **69**
3. Proposed category: **C - Major extinct**
4. Is a repertoire including character names provided?: **Yes**
   (a) If Yes, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?:
       Yes
   (b) Are the character shapes attached in a legible form suitable for review?: **Yes**
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
publishing the standard?: **Anshuman Pandey**; **True Type format**
   (a) If available now, identify source(s) for the font and indicate the tools used: **The characters of the digitized
   Sindhi font are based on normalized forms of printed Standard Sindhi characters. The font was designed
   by Anshuman Pandey using FontForge.**
6. References:
   (a) Are references (to other character sets, dictionaries, descriptive texts etc.) provided?: **Yes**
   (b) Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed
       characters attached?: **Yes**
7. Special encoding issues:
   (a) Does the proposal address other aspects of character data processing (if applicable) such as input, presentation,
       sorting, searching, indexing, transliteration etc. (if yes please enclose information)? **Yes; see proposal for
       additional details.**
8. Additional Information: Submitters are invited to provide any additional information about Properties of the pro-
posed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the pro-
posed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency
information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing be-
haviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equiv-
alence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org
for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associ-
ated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for
inclusion in the Unicode Standard. **Character properties and numeric information are included.**

---

[5] Form number: N3102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-
11, 2005-01, 2005-09, 2005-10, 2007-03)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?: **No**
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? **No**
    (a) If Yes, with whom?: **N/A**
        i. If Yes, available relevant documents: **N/A**
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **Yes**
    (a) Reference: **See text of proposal**
4. The context of use for the proposed characters (type of use; common or rare): **Common**
    (a) Reference: **See text of proposal**
5. Are the proposed characters in current use by the user community?: **Yes**
    (a) If Yes, where? Reference: **In India and Pakistan**
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?: **No**
    (a) If Yes, is a rationale provided?: **N/A**
        i. If Yes, reference: **N/A**
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **Yes**
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No**
    (a) If Yes, is a rationale for its inclusion provided?: **N/A**
        i. If Yes, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**
    (a) If Yes, is a rationale provided?: **N/A**
        i. If Yes, reference: **N/A**
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? **Yes**
    (a) If Yes, is a rationale for its inclusion provided? **Yes**
        i. If Yes, reference: **See text of proposal**
11. Does the proposal include use of combining characters and/or use of composite sequences? **Yes**
    (a) If Yes, is a rationale for such use provided? **Yes**
        i. If Yes, reference: **See text of proposal**
    (b) Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **N/A**
        i. If Yes, reference: **N/A**
12. Does the proposal contain characters with any special properties such as control function or similar semantics? **Yes**
    (a) If Yes, describe in detail (include attachment if necessary): **Virama**
13. Does the proposal contain any Ideographic compatibility character(s)? **No**
    (a) If Yes, is the equivalent corresponding unified ideographic character(s) identified? **N/A**
        i. If Yes, reference: **N/A**