

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

September 10, 2010

1 Introduction

This is a proposal to encode the Sindhi script in the Universal Character Set (UCS). The script described here is the Landa-based writing system 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 (see Section 3 for details on the proposed unification).

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

- 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”

The present document is a supplement to 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. An ISO/IEC 10646 proposal summary form is also enclosed.

2 Overview

The Sindhi script proposed for encoding was published as a standard by the Government of Bombay in 1868.¹ It was used for writing and printing the Sindhi [snd] language, which is spoken in Pakistan and India. This official script was known as ‘Hindu Sindhi’ and ‘Hindi Sindhi’; the adjectives ‘Hindu’ and ‘Hindi’ 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 ‘Standard Sindhi’ and ‘Sindhi’.

¹ Grierson 1919: 18.

The creation of a formal Brahmi-based script for Sindhi was driven by the need for a standard written medium for education and administration.² The intention was to use a script that would be familiar to merchant communities, who already employed various local Landa scripts, such as Khudawadi and Shikarpuri (see N3767 L2/10-012R for details). These scripts were used for routine writing and commercial purposes and were known as *haṭāī* ‘shopkeeper’, *vāṇiko* ‘merchant’, etc. They were unstandardized scripts whose character repertoires and glyph shapes differed widely across Sindh, from district to district. Moreover, they were considered ‘incomplete’ scripts, as indicated by the name *lanḍā* (Punjabi: ਲੰਡਾ *lanḍā*; Sindhi: لنڊو / لُنڊو *luṇḍo*; ‘tailless, clipped’). The term ‘clipped’ refers to general characteristics of the writing system, such as the absence of vowel signs, the use of one letter for writing all independent vowels, representing various consonants within an articulation class using a single letter, inattention to word spacing and phrase boundaries, and other shorthand practices. Such inconsistencies prevented the use of these scripts in administration, education, and publishing.

Standard Sindhi was established as a ‘complete’ script that could be used for fully representing the languages of Sindh. It was developed by Narayan Jagannath Vaidya, the Deputy Educational Inspector in Sindh.³ The standard is based upon the Khudawadi script used in Hyderabad, which was the most complete and well-regarded of the Landa scripts of Sindh. To the Khudawadi template were added some glyphic elements from the scripts used in Shikarpuri and Sakkar, which were also among the more developed Sindhi scripts. The main improvement introduced by Vaidya was the assignment of unique glyphs to each vowel and consonant sound of the Sindhi language, and the addition of vowel signs. A comparison of the original Khudawadi script and its Standard Sindhi form is given in Figure 7.

The script was first used for producing books for primary education and for maintaining court records. It was then extended to literary activities. The first books published in Standard Sindhi were lithographs, such as the Sindhi ballad *Dodo Chanesar* (see Figure 4). By 1870, books were being printed in Standard Sindhi using metal types produced in Bombay.⁴ At least two metal fonts were prepared, specimens of which appear in Figure 3 and Figure 11; the fonts used in the latter specimen appear to be produced in Vienna. The first literary journal, titled *Sukhiri*, was printed in the script in 1899.⁵ 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).

After the turn of the 20th century, Standard Sindhi was 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 the Devanagari script in India. The use of Landa-based scripts in modern-day Sindh and in Sindhi communities in India and elsewhere has all but disappeared.

3 A Unified Encoding for the Scripts of Sindh

Numerous Landa-based scripts were used in Sindh (see N3767 L2/10-012R). Several of these are named after the region, community, and commercial activities with which they were associated: Arora, Baniya, Bhatia, Hyderabad (Haidarabadi), Karadi, Khudawadi (Khudabadi), Khwaja (Khojki), Hatai, Hatvanika, Larai, Lohanaki (Lohana), Maimon, Rajai, Sakkar, Shikarpuri, Sewhani Bhabhira, Thatta (Thattai), Vaniya, Wangai, Waniko, etc. The majority of these scripts exhibit features fitting the description of Landa as an ‘incomplete’ script. Apart from Khojki and Khudawadi, these scripts are not sufficiently developed to be encoded independently in the UCS.

² Government of Bombay 1869: 213. ³ Hughes 1876: 373. ⁴ Government of Bombay 1874: 262. ⁵ Datta 1987: 1892.

Therefore, if the need arises to represent any of these ‘minor’ Sindhi scripts, they should be unified with the script proposed here. A unified block for Sindhi based upon Standard Sindhi provides an effective means for managing the various local Sindhi scripts and for supporting the structural and orthographic features that distinguish them from other scripts of the Landa family, such as the forms used for writing languages of the Lahnda [lah] group, eg. the Multani or Kirakki script for Seraiki [skr].

The Standard Sindhi script is an appropriate representative for a unified block for all Landa-based Sindhi scripts. It is modeled upon the well-regarded Khudawadi form and may be considered an extension of that script. It possesses the complete character repertoire necessary for uniquely representing sounds of the Sindhi languages. Moreover, its glyph shapes are attested in several other local Sindhi script varieties.

The exception to the unification of Sindhi scripts is Khojki, which has been proposed for independent encoding on the basis of its use as a literary and ecclesiastical script for the Nizari Ismaili community of South Asia. Information about Khojki and the justification for its independent encoding is provided in “Proposal to Encode the Khojki Script in ISO/IEC 10646” (N3596 L2/09-101).

4 Implementation Details

4.1 Script Name

The name ‘Sindhi’ is assigned to the script block. It is more appropriate than ‘Standard Sindhi’, ‘Hindu Sindhi’, or ‘Hindi Sindhi’. Although ‘Sindhi’ is a generic term, there is no other script in the UCS with this name. Also, the proposed Sindhi script has no relationship to the Arabic-based Sindhi script and there is unlikely to be confusion between the two scripts.

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 to the Supplementary Multilingual Plane (SMP) at the range U+11600..1164F. This allocation provides sufficient code points for completely encoding Sindhi and offers extra space for future additions.

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 code chart and names list are provided in Table 1 and Table 2.

4.5 Encoding Order

Different ordering schemes are attested for Sindhi consonant letters. In some sequences the implosive consonant letters (GGA, JJA, DDDA, BBA) appear before their unaspirated counterparts (GA, JA, DDA, BA), in others they are placed after. In some orders RRA appears before DDDA and DDA, and in others it is placed after them and before DDHA; also, some show SA before SHA. In the encoding proposed here, the letters for the implosive consonants are placed after their unaspirated counterparts, RRA is placed before DDHA, and SHA before SA.

4.6 Font

The font for Standard Sindhi is based upon the glyphs given in Figure 3. These glyphs are specimens of the Standard Sindhi published by the Government of Bombay. The glyph shapes agree with those given by Grierson in Figure 7 and those used for printing the text whose specimen appears in Figure 5.

4.7 Vowel Signs

Vowel signs function as in Devanagari. The $\text{f} \circ$ VOWEL SIGN I is written to the left of the consonant with which it combines.

4.8 Representation of Vowel Letters

Some independent vowel letters may be represented using a combination of a base vowel letter and a dependent vowel sign. This practice is not recommended. The atomic character for the independent vowel letter should always be used, as shown below:

RECOMMENDED	NOT RECOMMENDED
wI VOWEL LETTER AA	$\text{w} + \text{I}$ VOWEL LETTER A + O VOWEL SIGN AA
wE VOWEL LETTER E	$\text{w} + \text{E}$ VOWEL LETTER A + O VOWEL SIGN E
wAI VOWEL LETTER AI	$\text{w} + \text{AI}$ VOWEL LETTER A + O VOWEL SIGN AI
wO VOWEL LETTER O	$\text{w} + \text{O}$ VOWEL LETTER A + O VOWEL SIGN O
wAU VOWEL LETTER AU	$\text{w} + \text{AU}$ VOWEL LETTER A + O VOWEL SIGN AU

4.9 Consonant Conjuncts

Sindhi consonant clusters generally consist of two consonants. These are written using a visible Q VIRAMA, represented as $C_1 + \text{VIRAMA} + \text{ZWNJ} + C_2$. Half-forms and ligatures are not attested.

4.10 Nasalization

The O ANUSVARA is used for indicating nasalization. It is written above the letter with which it combines.

4.11 Usage of NUKTA

The ◌ NUKTA is used to represent sounds not native to Sindhi, such those that occur in Persian and Arabic loanwords. The NUKTA letters that appear in Sindhi records are given below along with the Arabic letters they represent:

	SINDHI		ARABIC	
<i>kha</i>	ڪ	KHA + NUKTA	خ	U+062E ARABIC LETTER KHAH
<i>ga</i>	ڳ	GA + NUKTA	غ	U+063A ARABIC LETTER GHAIN
<i>za</i>	ڙ	JA + NUKTA	ذ	U+0630 ARABIC LETTER THAL
<i>za</i>	ڙ	JA + NUKTA	ز	U+0632 ARABIC LETTER ZAIN
<i>za</i>	ڙ	JA + NUKTA	ض	U+0636 ARABIC LETTER DAD
<i>za</i>	ڙ	JA + NUKTA	ظ	U+0638 ARABIC LETTER ZAH
<i>fa</i>	ڦ	PHA + NUKTA	ف	U+0641 ARABIC LETTER FEH

The NUKTA may be written with any vowel or consonant letter. If other combining marks, such as a vowel sign or ANUSVARA, also appear with the base character, then the NUKTA should be written first.

4.12 Punctuation

Various punctuation is attested in Sindhi records, such as *danḍās* and Latin marks, such as periods, dashes, colons, and semi-colons. Sindhi *danḍās* may be unified with those of Devanagari.

4.13 Digits

Sindhi has a full set of decimal digits: ٠ ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩.

4.14 Number Forms

Fraction signs and currency marks are attested in Sindhi records. 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.

4.15 Collation

The collating order for Sindhi is as follows:

ٺ A > ٺI AA > ٺ I > ٺ II > ٺ U > ٺ UU > ٺ E > ٺ AI > ٺ O > ٺ AU > ٺ KA > ڪ KHA >
 ڳ GA > ڳI GGA > ڳ GH A > ڳ NG A > ڳ CA > ڳ CH A > ڳ JA > ڳ JJA > ڳ JHA > ڳ NYA >
 ڙ TTA > ڙ TTHA > ڙ DDA > ڙ DDDA > ڙ RRA > ڙ DDHA > ڙI NNA > ڙ TA > ڙ THA > ڙ DA >
 ڙ DHA > ڙ NA > ڙ PA > ڙ PHA > ڙ BA > ڙ BBA > ڙ BHA > ڙ MA > ڙ YA > ڙ RA > ڙ LA >
 ڙ VA > ڙ SHA > ڙ SA > ڙ HA > ڙ ANUSVARA > ڙ SIGN AA > ڙ SIGN I > ڙ SIGN II > ڙ SIGN U >
 ڙ SIGN UU > ڙ SIGN E > ڙ SIGN AI > ڙ SIGN O > ڙ SIGN AU > ڙ VIRAMA

A combination of a letter + ◌̣ NUKTA is sorted with the base letter. The collating rules for Sindhi NUKTA are identical to those for Devanagari.

4.16 Linebreaking

Linebreaking for Sindhi letters, vowel signs, ANUSVARA, NUKTA, VIRAMA, and digits follow rules for Devanagari.

4.17 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

- 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 of *Dodo Chanesar*: “Le héros légendaire Dodo représenté dans un livret 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.
- 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 Bombay (Presidency). Educational Department. 1869. *Report of the Director of Public Instruction in the Bombay Presidency, for the Year 1868–69*. Bombay: Education Society’s Press, Byculla.
- . 1874. *Report of the Director of Public Instruction in the Bombay Presidency, for the Year 1872–73*. Bombay: Government Central Press.
- Grierson, George A. 1919. *The Linguistic Survey of India*. Vol. VIII. Indo-Aryan Family. North-Western Group. Part III. Sindhī and Lahndā. Calcutta: Office of the Superintendent of Government Printing, India.
- Hughes, A. W. [compiler]. 1876. *A Gazetteer of the Province of Sindh*. 2nd ed. London: George Bell and Sons.

Jetley, Murlidhar Kishinchand. 1999. (سِنڌِي ٻوليءَ جون لپيون) [= *boliya jo sirishhto ain likhawata (sindhi boliya jawan lipiyon)*]. Delhi: 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>

———. 2010a. “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>

———. 2010b. “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

I would like to thank Michel Boivin (L'École des hautes études en sciences sociales (EHESS), Paris) for providing photographs of Sindhi records.

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

	1160	1161	1162	1163	1164
0	𑍀 11600	𑍁 11610	𑍂 11620	𑍃 11630	𑍄 11640
1	𑍅 11601	𑍆 11611	𑍇 11621	𑍈 11631	𑍉 11641
2	𑍊 11602	𑍋 11612	𑍌 11622	𑍍 11632	𑍎 11642
3	𑍏 11603	𑍐 11613	𑍑 11623	𑍒 11633	𑍓 11643
4	𑍔 11604	𑍕 11614	𑍖 11624	𑍗 11634	𑍘 11644
5	𑍙 11605	𑍚 11615	𑍛 11625	𑍜 11635	𑍝 11645
6	𑍞 11606	𑍟 11616	𑍠 11626	𑍡 11636	𑍢 11646
7	𑍣 11607	𑍤 11617	𑍥 11627	𑍦 11637	𑍧 11647
8	𑍨 11608	𑍩 11618	𑍪 11628	𑍫 11638	𑍬 11648
9	𑍭 11609	𑍮 11619	𑍯 11629	𑍰 11639	𑍱 11649
A	𑍲 1160A	𑍳 1161A	𑍴 1162A	𑍵 1163A	
B	𑍶 1160B	𑍷 1161B	𑍸 1162B		
C	𑍹 1160C	𑍺 1161C	𑍻 1162C		
D	𑍼 1160D	𑍽 1161D	𑍾 1162D		
E	𑍿 1160E	𑎀 1161E	𑎁 1162E		
F	𑎂 1160F	𑎃 1161F	𑎄 1162F		

Figure 1: Proposed code chart for Sindhi.

Independent vowels

11600	𑍀	SINDHI LETTER A
11601	𑍁	SINDHI LETTER AA
11602	𑍂	SINDHI LETTER I
11603	𑍃	SINDHI LETTER II
11604	𑍄	SINDHI LETTER U
11605	𑍅	SINDHI LETTER UU
11606	𑍆	SINDHI LETTER E
11607	𑍇	SINDHI LETTER AI
11608	𑍈	SINDHI LETTER O
11609	𑍉	SINDHI LETTER AU

Digits

11640	𑍀	SINDHI DIGIT ZERO
11641	𑍁	SINDHI DIGIT ONE
11642	𑍂	SINDHI DIGIT TWO
11643	𑍃	SINDHI DIGIT THREE
11644	𑍄	SINDHI DIGIT FOUR
11645	𑍅	SINDHI DIGIT FIVE
11646	𑍆	SINDHI DIGIT SIX
11647	𑍇	SINDHI DIGIT SEVEN
11648	𑍈	SINDHI DIGIT EIGHT
11649	𑍉	SINDHI DIGIT NINE

Consonants

1160A	𑍀	SINDHI LETTER KA
1160B	𑍁	SINDHI LETTER KHA
1160C	𑍂	SINDHI LETTER GA
1160D	𑍃	SINDHI LETTER GGA
1160E	𑍄	SINDHI LETTER GHA
1160F	𑍅	SINDHI LETTER NGA
11610	𑍆	SINDHI LETTER CA
11611	𑍇	SINDHI LETTER CHA
11612	𑍈	SINDHI LETTER JA
11613	𑍉	SINDHI LETTER JJA
11614	𑍊	SINDHI LETTER JHA
11615	𑍋	SINDHI LETTER NYA
11616	𑍌	SINDHI LETTER TTA
11617	𑍍	SINDHI LETTER TTHA
11618	𑍎	SINDHI LETTER DDA
11619	𑍏	SINDHI LETTER DDDA
1161A	𑍐	SINDHI LETTER RRA
1161B	𑍑	SINDHI LETTER DDHA
1161C	𑍒	SINDHI LETTER NNA
1161D	𑍓	SINDHI LETTER TA
1161E	𑍔	SINDHI LETTER THA
1161F	𑍕	SINDHI LETTER DA
11620	𑍖	SINDHI LETTER DHA
11621	𑍗	SINDHI LETTER NA
11622	𑍘	SINDHI LETTER PA
11623	𑍙	SINDHI LETTER PHA
11624	𑍚	SINDHI LETTER BA
11625	𑍛	SINDHI LETTER BBA
11626	𑍜	SINDHI LETTER BHA
11627	𑍝	SINDHI LETTER MA
11628	𑍞	SINDHI LETTER YA
11629	𑍟	SINDHI LETTER RA
1162A	𑍠	SINDHI LETTER LA
1162B	𑍡	SINDHI LETTER VA
1162C	𑍢	SINDHI LETTER SHA
1162D	𑍣	SINDHI LETTER SA
1162E	𑍤	SINDHI LETTER HA

Various signs

1162F	𑍥	SINDHI SIGN ANUSVARA
-------	---	----------------------

Dependent vowel signs

11630	𑍦	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

Figure 2: Proposed names list for Sindhi.

ۙ	ۚ	ۛ	ۜ	۝	۞	۟	۠	ۡ	ۢ
ونڻ									
ۙ	ۚ	ۛ	ۜ	۝	۞	۟	۠	ۡ	ۢ
ۣ	ۤ	ۥ	ۦ	ۧ	ۨ	۩	۪	۫	۬
ۭ	ۮ	ۯ	۰	۱	۲	۳	۴	۵	۶
۷	۸	۹	۰	۱	۲	۳	۴	۵	۶
۷	۸	۹	۰	۱	۲	۳	۴	۵	۶
۷	۸	۹	۰	۱	۲	۳	۴	۵	۶

۱. جيڪي هند ۽ سنڌي سڙن مٿي ڏنل آهن ۽ سي سڀ ساڳيئي روپ رکندا آهن ۽ جڏهن لفظن جي منڍ ۾ ايندا آهن ۽ پر جڏهن وٽ يا پٺاڙي ۾ ايندا آهن ۽ ته ۙ کان سواءِ ٻيا سڀ روپ مٿانيندا آهن ۽ اهي هيٺ لکيل آهن :

اصل روپ :- ۙ, ۛ, ۜ, ۝, ۞, ۟, ۠, ۡ, ۢ, ۣ, ۤ, ۥ, ۦ, ۧ, ۨ, ۩, ۪, ۫, ۬, ۭ, ۮ, ۯ, ۰, ۱, ۲, ۳, ۴, ۵, ۶, ۷, ۸, ۹

مٿيل روپ :- ۙ, ۛ, ۜ, ۝, ۞, ۟, ۠, ۡ, ۢ, ۣ, ۤ, ۥ, ۦ, ۧ, ۨ, ۩, ۪, ۫, ۬, ۭ, ۮ, ۯ, ۰, ۱, ۲, ۳, ۴, ۵, ۶, ۷, ۸, ۹

۲. ونڻ سان ڪڏهن لکيل اهي سڙهين ٿا جن :

ۙ, ۛ, ۜ, ۝, ۞, ۟, ۠, ۡ, ۢ, ۣ, ۤ, ۥ, ۦ, ۧ, ۨ, ۩, ۪, ۫, ۬, ۭ, ۮ, ۯ, ۰, ۱, ۲, ۳, ۴, ۵, ۶, ۷, ۸, ۹

۳. اهڙي قطار کي ڪڪ يا ڏهه اکرِي چوندا آهن يعني هجي ڪرڻ جي حد ڪ ٿهه اکرِي قطار يا ڏهه اکرِي.

۴. سڀني هند ۽ سنڌي سڙن کي اصل يا مٿل روپ ۾ هميشه ساڳيا آواز آهن ۽ ٻين نه آهن جو هڪڙي سڙ کي ٻه يا وڌيڪ آواز آهن ۽ انهي ڪالهه کان لکڻ ۾ پنڪل ڪهڙو ناهي ۽ اهو هجي ڪرڻ ۾ ڇوڪ يا شڪ يا هڪ ڪڏهن نه ٿيندي ۽ مطلب ته جيتري قدر سنڌي ۾ آواز آهن ۽ اوتري قدر هند ۽ سنڌي سڙن جو بندوبست بلڪل پورو ۽ بي هيٺ آهي .

سنڌي لپي ماترائن سان گڏ (جهتمل وسيلي، ۱۸۹۲)

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).

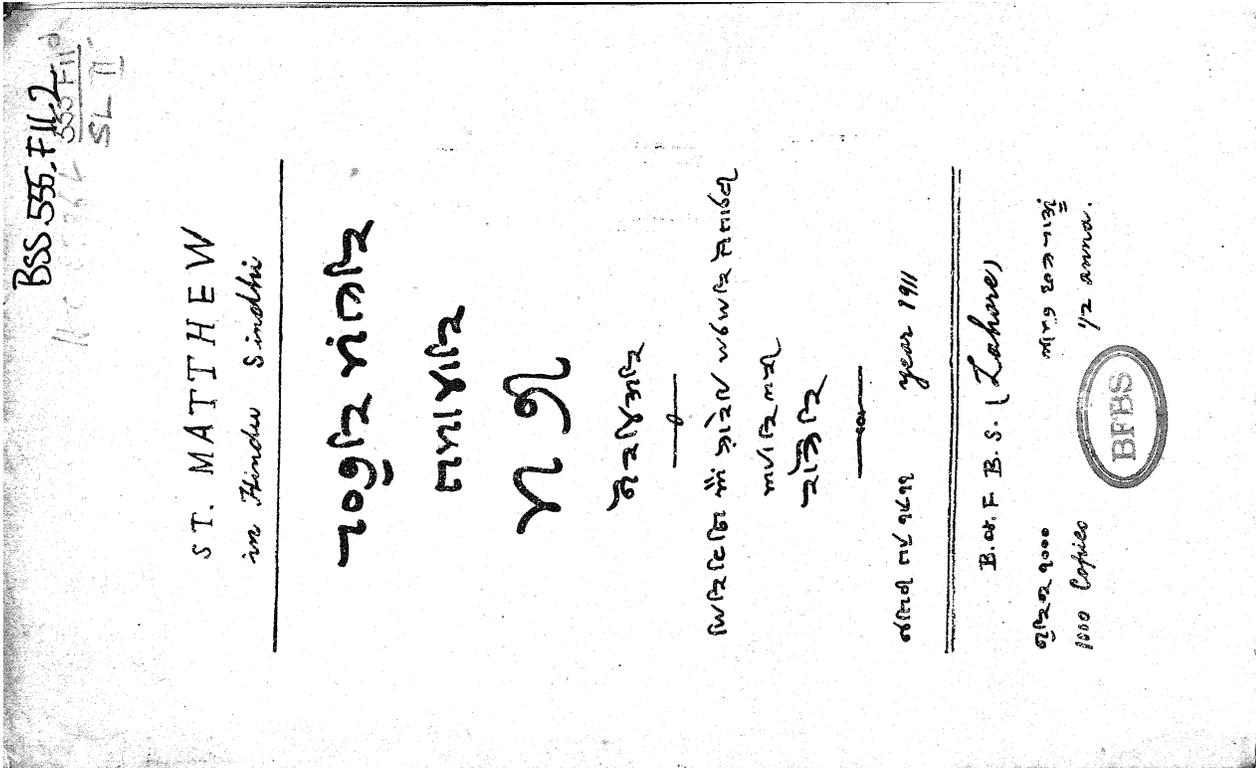


Figure 5: Cover and first page of St. Matthew in Standard Sindhi (from British and Foreign Bible Society 1911).

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	ڪ	ڪ	15	ra	No. 32 used instead.	ڙ	29	bha	No. 28 used instead.	ڳ
2	kha	ڪھ	ڪھ	16	ḍḍa	ڙھ	ڙھ	30	ma	م	م
3	gga	گگ	گگ	17	ḍa	ڙ	ڙ	31	ya	No. 9 used instead.	ڙ
4	ga	گ	گ	18	ḍha	ڙھ	ڙھ	32	ra	ر	ر
5	gha	No. 4 used instead.	گھ	19	ṇa	ڙڙڙ	ڙڙڙ	33	la	ل	ل
6	ṇa	ڙڙڙ	ڙڙڙ	20	ta	ٽ	ٽ	34	va, wa	و	و
7	cha	ڇ	ڇ	21	tka	ٽڪ	ٽڪ	35	sa	س	س
8	chha	ڇھ	ڇھ	22	da	ڌ	ڌ	36	sha	No. 35 used instead.	ڌھ
9	ja	ج	ج	23	dha	No. 22 used instead.	ڌ	37	ha	ھ	ھ
10	jja	جج	جج	24	na	ڻ	ڻ	38	khē	No. 2 used instead.	ڪھ
11	jha	جھ	جھ	25	pa	پ	پ	39	zē	No. 9 used instead.	ڙھ
12	ṇa	ڙڙڙ	ڙڙڙ	26	pha	ڦ	ڦ	40	ghain	No. 4 used instead.	ڳھ
13	ṭa	ڙ	ڙ	27	bba	ڙڙ	ڙڙ	41	fē	No. 26 used instead.	ڙھ
14	ṭha	ڙھ	ڙھ	28	ba	ڙ	ڙ				

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

VOWELS.

In improved Hindī Sindhī character.

a	ā	i	ī	u	ū
𑂀	𑂁	𑂂	𑂃	𑂄	𑂅
e	ai	ō	au	am̄	
𑂆	𑂇	𑂈	𑂉	𑂊	
ka	kā	ki	kī	ku	kū
𑂋	𑂌	𑂍	𑂎	𑂏	𑂐
kō	kai	kō	kau	kam̄	
𑂑	𑂒	𑂓	𑂔	𑂕	

The only vowel character in Khudāwāḍī is 𑂎

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

[No. 4.]

INDO-ARYAN FAMILY. NORTH-WESTERN GROUP.

SINDHI.

STANDARD DIALECT.

DISTRICT KARACHI.

SPECIMEN I.

Heḳṛē māḅhñ-khé bba puṭ huā.

Ā: tñ-mā nandhē pāh-jē-piy-khé

chayō, 'ē bābā māḅ-mā jēkō bhānō

mukṭ-jō thīē, sō mī-khé ḳḳē.

Ā: hun' pāh-jō māḅ virāḅē (for virāḅē)

hunnar-khé ḳḳinō. Thōriṅṅ. q. 2

ḳḳihār-khā pō nandhō puṭ

sabb' gaḳḳ' karē heḳṛē

ḳḳāḅhē-ḳḳēḅā-ḳḳē haliyō

viyō, ā' pāh-jō māḅ

bē-hudīr (for bē-hudī) halat-mē viṅāē

chhaḳḳiyā. Ā' jaddhī sabb'

khapūē rahiyo tāh-khā pō

un'-ḳḳēh-mē ḳḳāḅhō ḳḳukār' achi

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

SINDH- UND MULTANSCHRIFT.

Sindh		Multan		Sindh		Multan		Sindh		Multan	
Zeichen	Wert	Zeichen	Wert	Zeichen	Wert	Zeichen	Wert	Zeichen	Wert	Zeichen	Wert
ᳵ	a	ᳶ	a	᳷	ña, ña			᳸	pha	᳹	pha
ᳺ	i	᳻	i	᳼	ṭa			᳽	ba	᳾	ba
᳾	u	᳿	u	᳽	ṭha			᳾	bha		
᳼	ka	᳽	ka	᳾	ḍa	᳾	ḍa	᳿	ma	᳽	ma
᳽	kha	᳾	kha	᳾	ṇa	᳾	ṇa	᳽	ya	᳾	ya
᳾	ga	᳽	ga	᳾	ṭa	᳾	ṭa	᳽	ra	᳾	ra
᳾	gha			᳽	ṭha	᳾	ṭha	᳽	la	᳾	la
᳾	tša	᳽	tša	᳾	ḍa	᳾	ḍa	᳽	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 ᳵ᳾᳽ *tuhūḍzo*, zweitens weil ihre Zeichen sich an die semitischen Schriften anlehnen, andererseits sich in der Devanagarschrift vorfinden, wenn diese von der Paliform abweicht; so hat ᳽ dieselbe Bedeutung auf Pehlewi-Münzen, ᳽ entspricht dem ᳽ im Aramäischen, ᳽ erinnert an die phönikische Form ᳽ , ᳽ an das phönikische ᳽ . Den von den Inschriftformen abweichenden Devanagarizeichen ᳽ *ka* entspricht Multan ᳽ , dem ᳽ *kha* (Inschrift ᳽) entspricht Sindh ᳽ *gha*, ᳽ kann ebensowohl von der Inschrift ᳽ , wie von Sindh ᳽ *ga* abstammen, dem ᳽ *tšha* entspricht Sindh ᳽ *tšha*, dagegen Inschrift ᳽ *tšha* dem Devanagari ᳽ *la*, Devanagari ᳽ hat seine Analogie nur in Sindh ᳽ (dagegen Inschrift ᳽ *ṇa*), ebenso ᳽ Sindh ᳽ (Inschrift ᳽), ᳽ *p*, Sindh ᳽ (Inschrift ᳽).

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

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

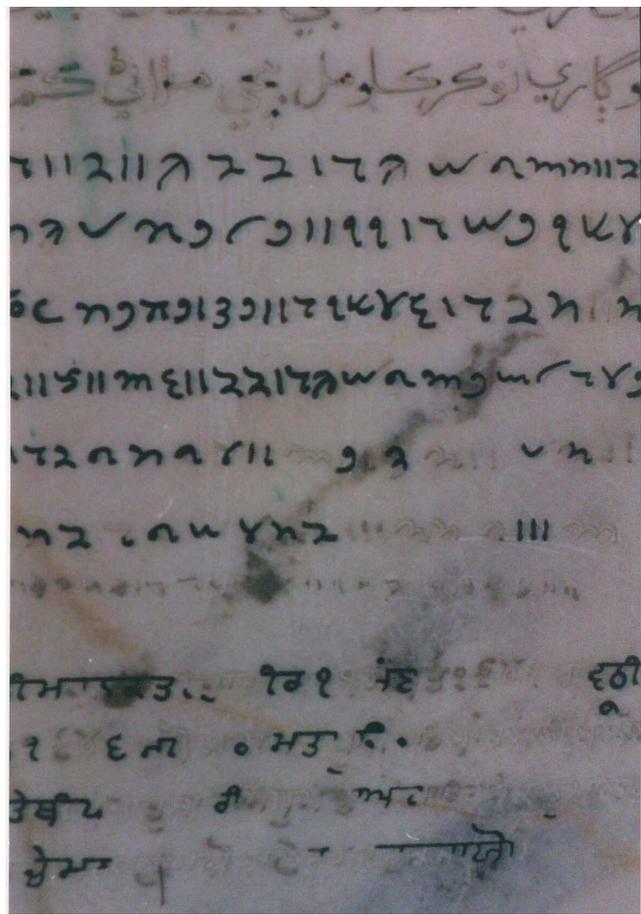
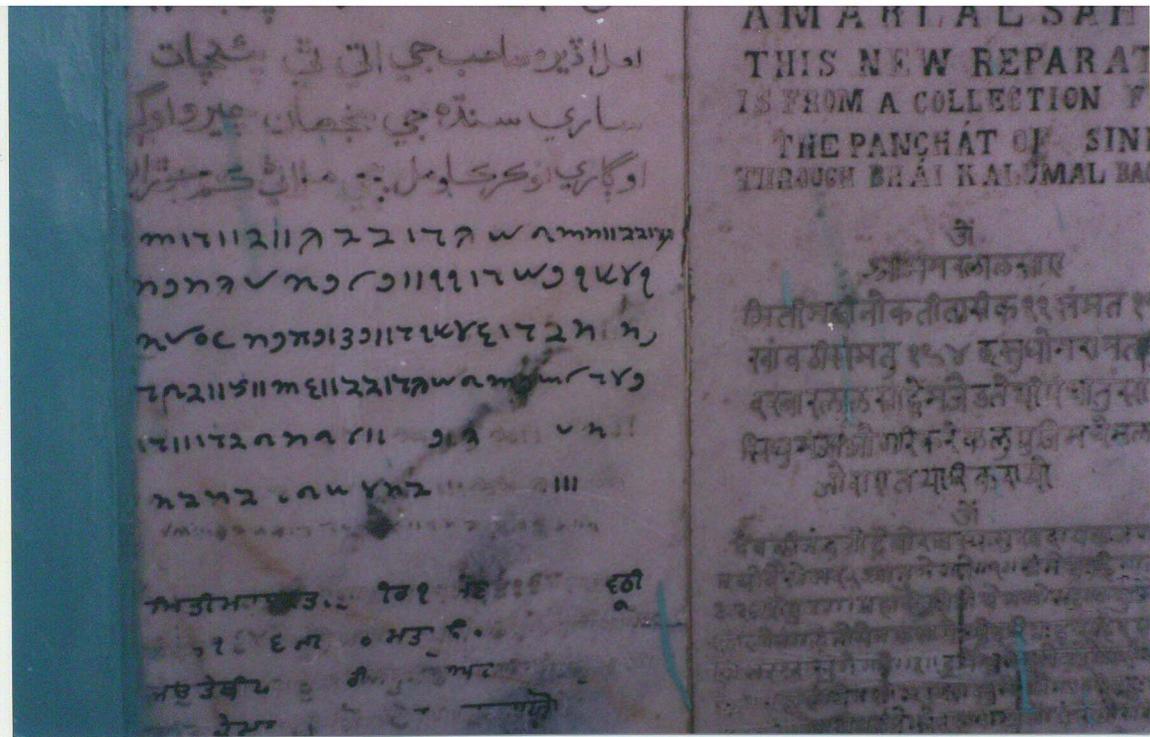


Figure 12: Top: An inscription in Khudawadi in an *ex-voto*, alongside the Latin, Devanagari, Arabic, and Gurmukhi scripts, at the shrine of Udero Lal in Hyderabad, Sindh. Bottom: detail of a portion of the Khudawadi text. Photos taken by Michel Boivin in 2000.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646⁶

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: **2010-09-10**
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 proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed 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 behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence 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 associated 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.**

⁶ 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**