

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

(Please read Principles and Procedures Document for guidelines and details before filling this form.)

See <http://www.dkuug.dk/JTC1/WG2/docs/summaryform.html> for latest Form.

See <http://www.dkuug.dk/JTC1/WG2/docs/principles.html> for latest Principles and Procedures document.

See <http://www.dkuug.dk/JTC1/WG2/docs/roadmaps.html> for latest roadmaps.

A. Administrative

1. Title:	<u>Proposal to encode Marwari LAM WITH BAR character</u>
2. Requester's name:	<u>Jonathan Kew, SIL International</u>
3. Requester type (Member body/Liaison/Individual contribution):	<u>Expert contribution</u>
4. Submission date:	<u>2003- -</u>
5. Requester's reference (if applicable):	_____
6. (Choose one of the following:)	_____
This is a complete proposal:	<u>Yes</u>
or, More information will be provided later:	_____

B. Technical - General

1. (Choose one of the following:)	_____
a. This proposal is for a new script (set of characters):	<u>No</u>
Proposed name of script:	_____
b. The proposal is for addition of character(s) to an existing block:	<u>Yes</u>
Name of the existing block:	<u>Arabic Supplementary</u>
2. Number of characters in proposal:	<u>1</u>
3. Proposed category (see section II, Character Categories):	<u>A</u>
4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000):	<u>1</u>
Is a rationale provided for the choice?	<u>Yes</u>
If Yes, reference: <u>No combining marks</u>	_____
5. Is a repertoire including character names provided?	<u>Yes</u>
a. If YES, are the names in accordance with the 'character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?	<u>Yes</u>
b. Are the character shapes attached in a legible form suitable for review?	<u>Yes</u>
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? <u>Jonathan Kew, SIL International</u>	_____
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	_____
<u>jonathan_kew@sil.org</u>	_____
<u>TrueType font generated with FontLab 4.5</u>	_____
7. References:	_____
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<u>No</u>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<u>Yes, see §3 below</u>
8. Special encoding issues:	_____
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)?	_____
<u>Yes: suggested Unicode character properties are included</u>	_____
9. 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/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.	

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain _____	<u>No</u>
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.)? If YES, with whom? <u>Local communities, linguists, and NGOs</u> If YES, available relevant documents: <u>See §4 below</u>	<u>Yes</u>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: <u>See §2 below</u>	<u>Yes</u>
4. The context of use for the proposed characters (type of use; common or rare) Reference: <u>See §2 below</u>	<u>Current publ'ns</u>
5. Are the proposed characters in current use by the user community? If YES, where? Reference: <u>Marwari communities in southern Pakistan</u>	<u>Yes</u>
6. After giving due considerations to the principles in <i>Principles and Procedures document</i> (a WG 2 standing document) must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: <u>Extension to BMP Arabic repertoire, see §1 below</u>	<u>Yes</u> <u>Yes</u>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<u>N/A</u>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference: _____	<u>No</u>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference: _____	<u>No</u>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference: _____	<u>No</u>
11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? If YES, is a rationale for such use provided? If YES, reference: _____ Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference: _____	<u>No</u>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<u>No</u>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference: _____	<u>No</u>

¹Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09)

Submitter's Responsibilities

The national body or liaison organization (or any other organization or an individual) proposing new character(s) or a new script shall provide:

1. Proposed category for the script or character(s), character name(s), and description of usage.
2. Justification for the category and name(s).
3. A representative glyph(s) image on paper:
If the proposed glyph image is similar to a glyph image of a previously encoded ISO/IEC 10646 character, then additional justification for encoding the new character shall be provided.
Note: Any proposal that suggests that one or more of such variant forms is actually a distinct character requiring separate encoding, should provide detailed, printed evidence that there is actual, contrastive use of the variant form(s). It is insufficient for a proposal to claim a requirement to encode as characters in the Standard, glyphic forms which happen to occur in another character encoding that did not follow the Character-Glyph Model that guides the choice of appropriate characters for encoding in ISO/IEC 10646.
Note: WG 2 has resolved in Resolution M38.12 not to add any more Arabic presentation forms to the standard and suggests users to employ appropriate input methods, rendering and font technologies to meet the user requirements.
4. Mappings to accepted sources, for example, other standards, dictionaries, accessible published materials.
5. Computerized/camera-ready font:
Prior to the preparation of the final text of the next amendment or version of the standard a suitable computerized font (camera-ready font) will be needed. Camera-ready copy is mandatory for final text of any pDAMs before the next revision. Ordered preference of the fonts is True Type or PostScript format. The minimum design resolution for the font is 96 by 96 dots matrix, for presentation at or near 22 points in print size.
6. List of all the parties consulted.
7. Equivalent glyph images:
If the submission intends using composite sequences of proposed or existing combining and non-combining characters, a list consisting of each composite sequence and its corresponding glyph image shall be provided to better understand the intended use.
8. Compatibility equivalents:
If the submission includes compatibility ideographic characters, identify the equivalent unified CJK Ideograph character(s).
9. Any additional information that will assist in correct understanding of the different characteristics and linguistic processing of the proposed character(s) or script.

1. Proposed character

The following character is proposed for encoding in the UCS:

Glyph	Code	Character name	GC	CC	Bidi	Shaping
ﻝ̣	076A	ARABIC LETTER LAM WITH BAR THROUGH	Lo	0	AL	LAM

This character is used in writing the Marwari language in southern Pakistan, to represent a flapped retroflexed lateral. It has also been suggested for use in the Gawri language of northern Pakistan (see Baart & Sagar 2002) to represent a lateral fricative, but it is unclear how widely it has been adopted there. The character is proposed for encoding at this time on the basis of its established usage in Marwari, but naturally this does not preclude possible use in other languages and communities.

As this character is an extension to the standard Arabic script, as adapted for use in South Asian languages, and is in current use, it should be encoded on the BMP; a location in the new Arabic Supplementary block is suggested.

Contextual shaping for this character is similar to that of U+0644 ARABIC LETTER LAM, including the requirement to form ligatures with ALEF and related characters (examples can be seen in figures 4, 6, 7, and 9) even in simplified cursive styles of the script.

2. Background information

The Marwari language is spoken by over 200,000 people in the Sindh and southern Punjab provinces of Pakistan (see http://www.ethnologue.com/show_language.asp?code=MRI). While Marwari has not traditionally been written, with most available education being in Sindhi or Urdu, there has been some use of written Marwari within the past 20 years. The Ethnologue cites a mother-tongue literacy rate of below 1%; nevertheless, a number of works in Marwari have been published in recent years (see §4 References).

Two approaches have been taken to writing Marwari: in Sindh province, an orthography based on that of Sindhi has been used, while in the northern Marwari area (southern Punjab) the orthography has instead been based on Urdu. In both cases, however, the orthography has been extended by the use of a “barred LAM” character to represent a flapped retroflexed lateral, not found in either Sindhi or Urdu.

Most existing literature in Marwari comes from church-related organizations and other NGOs working in the fields of education and community development among the Marwari community. Examples using both southern (Sindhi-based) and northern (Urdu-based) orthographies are shown below.

It may be noted that the name “Marwari” is also used to refer to the standard form of the Rajasthani language spoken in India (see http://www.ethnologue.com/show_language.asp?code=MKD). It is unclear exactly how closely related the Pakistani “Marwari” language is to this. In any case, Marwari in India is written with the Devanagari script, and is therefore not relevant to the present character proposal.

3. Samples showing use of the proposed character

3.1 Southern Marwari (Sindhi-based orthography)

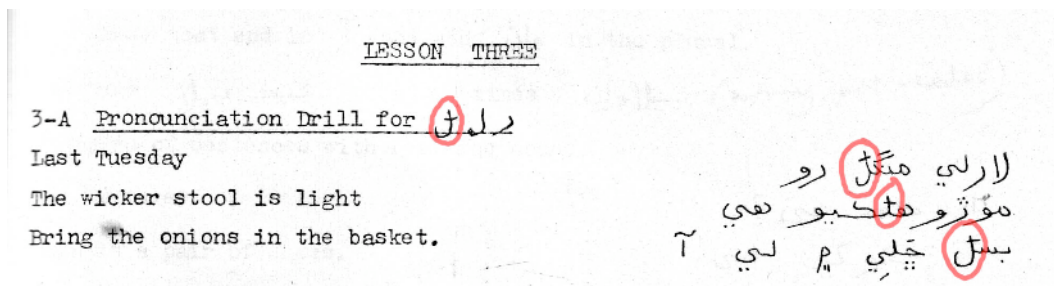
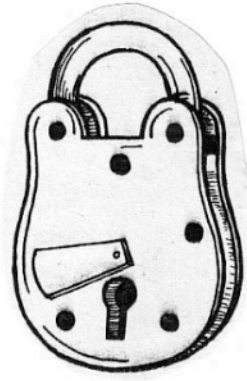


Figure 1: Saint (1986), page 12

30



ٺاڻو



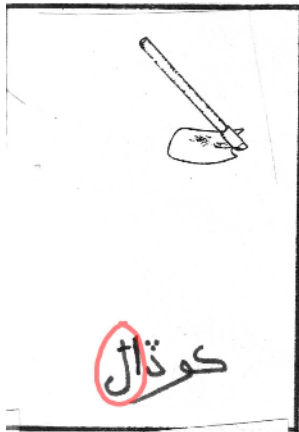
ڦڙ



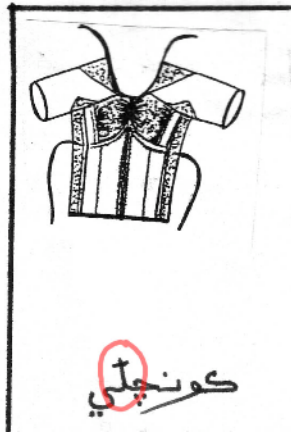
ٻاڻيا

ٻاڻيا ڦڙ ڪاوي وا .
ٻاڻيا ٺاڻو لڱاوي وا .

Figure 2: Saint (1988), page 30



ڪوٽال



ڪونجڻي

Figure 3: ICF (1987), no page #

آن پڇي سات برسان رو ڪال پڙشي.
 پڇي يوسف بادشاه نا صلاح ڏنهي ڪه ڪي ڪرڻو چهجي.
 بادشاه نا يوسف ري صلاح ڏاڍي ٺاهوڪي لاڳي.
 آن بادشاه يوسف نا مصر رو وزير ٺايو. (

جڏي يوسف وزير ٿيو
 تو اوئي ري عمر تريس برس هتي.
 پڇي سات برسان تائين گهڻي ٺاهوڪي پوک هئي.
 آن يوسف سڄي ملڪ ۾ ڦريو آن اي حڪم ڏنو ڪه
 اي ڏاڏا گاوان ري گدامان ۾ پيٽا ڪرو.
 اوئي پيٽا اتري ڪٽڪ هتي جڪو ٿوئن ڏکي هتي.

Figure 4: PBS (1990), page 12

ڏيغو ڇڪ هيٺا راکن

۱۶ ڪوئي به اهڙو منڪ ڪوئي جڪو ڏيغو پال رو ڇڪ سان ڇڪي
 جان اوئي نا مانڀي هيٺا راکي. پن اوئي نا جائگي ۾ راکي جان
 هر پيٽرا آون آئي نا آڃاڻو جڙي. ۱۷ ڪيان ڪه ڪوئي لڪيوڙي بست
 اهڙي ڪوئي جڪو چئي ٿي وي، آن ڪوئي اهڙو پيڏ ڪوئي جڪي ري ڪهر

Figure 5: PBS (1991), page 45

(۹۵)

جگت ۾ پاپ ڪيان آيو

۱: عدن ڏيش ۾ ايشور باغ لگاڻو جيئڻ ۾ نر ناري نان بسايو
 اس ٻڌ پاپ جگت ۾ آيا رے پائيو پاپ جگت ۾ آيا

۲: پيارے پريو اے وچن سٺايا پاپ پرکو رو ڦل متي ڪايا
 اوے ڦل رے ٿيڙا متے جايا اوئي ڦل ني ڪايا

۳: نانگ رے روپ ۾ شيطان آيو نر ناري نان ڦل ڪلايو
 پاپ پرک رو ڦل ڪلاے نر ناري نان پلايا

Figure 6: SCF/PCF (n.d. #1), page 98; compare figure 7 for Urdu-based orthography

3.2 Northern Marwari (Urdu-based orthography)

(۹۵)

جگت میں پاپ کیاں آیو

- ۱- عدن ڈیس میں ایشور باغ لگا یو میں نرناری ناں بسپا یو
- اس پڈپاپ جگت میں آیائے بھائیو پاپ جگت میں آیا
- ۲- پیارے پر بھولے وچن سنایو پاپ پرکھ رو پھل متی کھایا
- اوے پھل سے نیرا متے جایا اوئی پھل تی کھایا
- ۳- نانگ رے روپ میں شیطان آلو نرناری ناں پھل کھلایو
- پاپ پرکھ رو پھل کھلے نرناری ناں بھلایا

Figure 7: SCF/PCF (n.d. #2), page 98

(۵) ۱ پطرس ۳
پن مسع ناں ایشور جان کر آپرے منوں
میں یوینتر سمجھو آن جکو کوئی نقان ساں
منازہ امید ری وجد پوچھے اوئے ناں بلتی
ڈیوٹ ساروں ہر بندہ تیار رو
پن حلیہی آن لوچن رے بندہ -

Figure 8: ARP (n.d.), #75 of 100 portions

۲۱
تھیں اوئے ناں کہو کہ
منال ای پتی ڈرے
جیسو اوئے ناں بلتی ڈرے کہ
منال تھیں ماتھے نیا کر ٹ کر
جاں مال پلا وٹن آلو
کے کھٹا یو

Figure 9: PBS (n.d.), page 21

4. References

[ARP Mission?]. n.d. 100 Hawale [100 Scripture passages], Marwari boli [language].

Baart, Joan L. G. and Muhammad Zaman Sagar. 2002. *The Gawri language of Kalam and Dir Kohistan*.
http://www.geocities.com/kcs_kalam/gawri.pdf.

ICF [Indus Christian Fellowship]. 1987. Marwari Bhili Topical Vocabulary, Trial Edition.

Pakistan Bible Society. 1990. Genesis New Reader Portion Book 5 (Joseph): Marwari language.

———. 1991. Gospel of Luke: Marwari language.

———. n.d. Parables of Jesus (Part 1): Marwari language.

Saint, Helen. 1986. Marwari Language Learning Course.

———. 1988. Marwari Primer, Trial Edition. Pakistan: ICF [Indus Christian Fellowship].

Sindhi Christian Fellowship/Pakistani Christian Fellowship. n.d. (#1). Christian Songs (Sindhi Script).

———. n.d. (#2). Christian Songs (Urdu Script).