

Unicode request for legacy Malayalam

Kirk Miller, kirkmiller@gmail.com

Neil Rees, neil.rees@biblesociety.org.uk

2021 July 16

Unicode support is needed by the British and Foreign Bible Society (BFBS) for an enormous back catalogue of published translations of Scripture that require digitization. One of these is in Romanized Malayalam, using unsupported <d l n r s t> for the retroflex consonants. (Capitals are not used, as an asterisk marks proper names, so only lower-case forms are requested.)

The letters all have a left hook, which is the reversed arm of an *r*. They were used in the early 20th century and for a time were used in the IPA as letters *qui sont déjà couramment employés dans des ouvrages spéciaux* (which are already widely used in specialist works – Passy & Jones 1921: 8).

Besides the ongoing digitization needs of the BFBS, Neil Rees notes current usefulness among the large Malayalam diaspora in the Commonwealth: *I know Malayalam speakers in my town who know the language but are not literate in it, because they were schooled in England and only know the Latin alphabet. So in such cases these old Romanised texts remain useful and helpful.* Indeed, a Romanized Tamil gospel, one that is fully supported by Unicode, has been digitized and made available online at: [*mārku eḷudhina suvisēsham 1952 \(rōman thamiḷ\)*](#), and the Romanized Malayalam material will be made available once it has Unicode support.

The expected letter <z> is not needed for Malayalam. <z> would be needed for full coverage of this convention, as noted by Marija Tabain, Editor of JIPA (p.c. 2020 Dec 17) and by Michael Ashby, President of the IPA (p.c. 2021 Jan 16). However, as the letters are no longer needed by the IPA and <z> is not yet attested in print, <z> is not requested.

Figure 1 shows the full set of letters used in this Romanized Malayalam. In contrast to the retroflex consonants, which were a regional standard, the unsupported vowel letters were adopted from the English Phonotypic Alphabet. These vowel letters are not all straightforward to encode, due to complications with their capital forms, so they will be requested in a separate proposal for the Phonotypic Alphabet.

Thanks to Deborah Anderson of the Universal Scripts Project for her assistance.

Characters

ɖ	1DF25 LATIN SMALL LETTER D WITH LEFT HOOK.
ɽ	1DF26 LATIN SMALL LETTER L WITH LEFT HOOK.
ɱ	1DF27 LATIN SMALL LETTER N WITH LEFT HOOK.
ɽ	1DF28 LATIN SMALL LETTER R WITH LEFT HOOK.
ɿ	1DF29 LATIN SMALL LETTER S WITH LEFT HOOK.
ɿ	1DF2A LATIN SMALL LETTER T WITH LEFT HOOK.

Properties

1DF25; LATIN SMALL LETTER D WITH LEFT HOOK; Ll; 0; L; ; ; ; ; N; ; ; ; ;
 1DF26; LATIN SMALL LETTER L WITH LEFT HOOK; Ll; 0; L; ; ; ; ; N; ; ; ; ;
 1DF27; LATIN SMALL LETTER N WITH LEFT HOOK; Ll; 0; L; ; ; ; ; N; ; ; ; ;
 1DF28; LATIN SMALL LETTER R WITH LEFT HOOK; Ll; 0; L; ; ; ; ; N; ; ; ; ;
 1DF29; LATIN SMALL LETTER S WITH LEFT HOOK; Ll; 0; L; ; ; ; ; N; ; ; ; ;
 1DF2A; LATIN SMALL LETTER T WITH LEFT HOOK; Ll; 0; L; ; ; ; ; N; ; ; ; ;

Chart

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Latin Extended-G																
U+1DF2x						ɖ	ɗ	ɛ	ɛ̃	ɛ̄						

References

*markkøṣ erutiya suviṣṣam (Gospel of Matthew) in Romanized Malayalam (1909). W.H. Christian, Eastbourne.
 R. Jorde (1924) nəˈwi:dzən. *Le Maître Phonétique*, vol. 2 (39), no. 7.
 G. Marin (1924) so:ma:li. *Le Maître Phonétique*, vol. 2 (39), no. 6.
 Paul Passy & Daniel Jones (1921) *L'écriture phonétique internationale : exposé populaire avec application au français et à plusieurs autres langues*, 2nd ed. Association phonétique internationale.

Figures

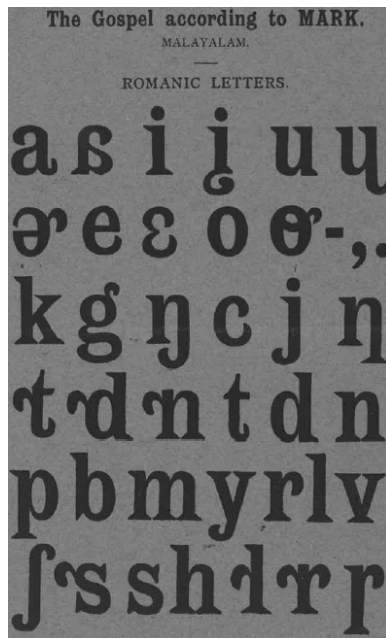


Figure 1. Back cover of Gospel (1909) with the romanized alphabet. The unsupported vowels will be requested in a separate proposal.

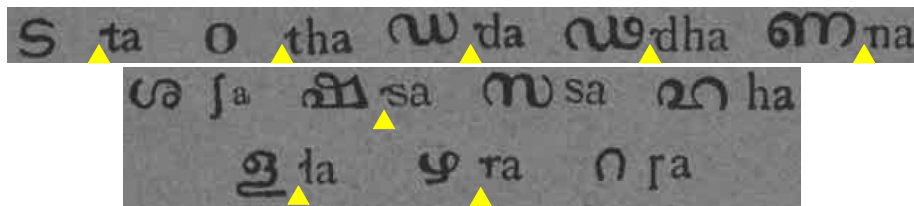


Figure 2. Front cover of the Gospel, identifying the left-hook letters with the retroflex consonants of Malayalam script.

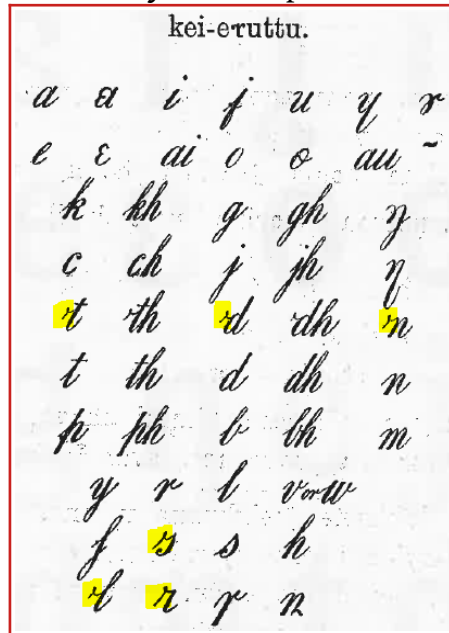


Figure 3. Chart of cursive forms, inside back cover, with the left hooks highlighted.

mukham dukkham leghu megham chsya psthām
 mudhan katha sthālam prsr̥tthana dhsnam buddhi
 phalam bhayam adhikam bhūmi bhāsa. hṛdaya-
 juddhi-ullavar bhāgyavanmṛ: avar *deivate kṣnum.

43 ran̄tu kṣju it̄tu. appṛlavan tan̄re j̄i-syanmṛe a-tukkal
 vidiccu: bhān̄d̄sr̄att̄il it̄ta e-l̄svarekks̄lum j̄ daridra-
 sya vidhava adhikam it̄t̄-irikkunnu, ennu satyam-syit̄tu

Figure 7. Gospel (1909) §10 and p. 37, showing the non-native <ṛ> that is found in Sanskrit loans.

Les linguales cacuminales ou rétroflexes des langues scandinaves
 (Suéd. *barn*, *fors*) et indiennes se marquent par un point derrière la
 lettre: ṛ, ṝ, etc. Le “r rétroflexe” del’ Hindoustani est représenté par ṝ.

Figure 8. Passy & Jones (1921: 8). <ṛ ṝ> are illustrated; the Swedish examples *barn* and *fors* would be transcribed with <n̄ s̄>. Hindustani <ṝ> is perhaps best handled as U+AB4B LATIN SMALL LETTER SCRIPT R.

int̄dentl; ṝ iz retr̄fleks;
 ḍə tounz hæv n̄t jet bi'n
 m̄kt̄ in ḍis tekst̄.
 d̄ə'baɟʃi ḍəah 'ijə qər'ra
 wejn s̄əd'ʃa:l ṝdar d̄ügsü:n 'h
 'kenni s̄əd'ʃa:lka ṝdarka

Figure 9. Marin (1924: 15). <ṝ> in Somali.

t, d, l, n a dentl; l iz hævev̄ ælviolə a'ft̄ə ṝ. ṝ, ṝ, ṝ, ṝ a
 f̄ə:md̄ ɒn ḍə hæ:d̄ p̄ælit, ḍəast̄ bihamd̄ ḍə ti-θ-r̄id̄ɟ. ṝ iz ɔ:l̄sou
 ə saund̄ əv ḍə længwid̄ɟ, b̄ət̄ it̄ d̄aznt̄ ək̄s̄:r̄ in ḍis p̄æs̄id̄ɟ. ṝ hæz
 ən̄ əkaustik̄ kw̄ɔliti bitwi:n̄ s̄ ənd̄ j̄. r̄ iz liŋgwəl̄ ənd̄ rould̄.
 o, o:, ə, ə: st̄ænd̄ f̄ə o, o:, ə, ə: rispektivli. ʃ̄ɔ:t̄ ɛ, æ hæv
 ḍə væljuz̄ è, ø, iksept̄ b̄if̄ə:r̄ r̄, ṝ, ṝ, ṝ, ṝ. ḍə sek̄ənd̄ elim̄ənt̄

v̄no:n̄'vin:n̄ va ṝd̄n̄ 'f̄æst̄ 's̄l̄ut̄: v̄va:t̄ə va 'so:l̄n̄

Figure 10. Jorde (1924: 19). Retroflex consonants in Norwegian.

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 (P & P) from std.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 std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

A. Administrative

1. Title:	<input type="text"/>	<i>Additional phonetic click letters</i>
2. Requester's name:	<input type="text"/>	<i>Kirk Miller, Neil Rees</i>
3. Requester type (Member body/Liaison/Individual contribution):	<input type="text"/>	<i>individual</i>
4. Submission date:	<input type="text"/>	<i>2021 July 16</i>
5. Requester's reference (if applicable):	<input type="text"/>	
6. Choose one of the following:		
This is a complete proposal:	<input type="checkbox"/>	<i>yes</i>
(or) More information will be provided later:	<input type="checkbox"/>	

B. Technical - General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input type="checkbox"/>	<i>no</i>
Proposed name of script:	<input type="text"/>	
b. The proposal is for addition of character(s) to an existing block:	<input type="checkbox"/>	<i>yes</i>
Name of the existing block:	<input type="text"/>	<i>Latin Extended-G</i>
2. Number of characters in proposal:	<input type="text"/>	<i>6</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input type="checkbox"/>	<i>yes</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input type="checkbox"/>	<i>yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<input type="checkbox"/>	<i>yes</i>
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<input type="text"/>	<i>Kirk Miller</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<input type="text"/>	<i>SIL (Gentium Release)</i>
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input type="checkbox"/>	<i>yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input type="checkbox"/>	<i>yes</i>
7. 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)?	<input type="checkbox"/>	<i>no</i>

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 www.unicode.org for such information on other scripts. Also see Unicode Character Database (www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-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, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>no</i>
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? If YES, available relevant documents:	<i>yes</i> <i>Rees is involved in the digitization project.</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>republication</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>no</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>no</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>yes</i>
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:	<i>no</i>
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:	<i>no</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>no</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:	<i>no</i> <i>no</i>
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)	<i>no</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>no</i>