TO: UTC
FROM: Debbie Anderson, SEI, UC Berkeley
DATE: April 17 2023 -- Revised July 2, 2023
SUBJECT: Draft Template for character additions to an existing script in Unicode

Attached is a draft template for proposal authors who wish to propose character additions to an existing script in Unicode.

Coversheet for New Proposal Submissions

This coversheet is to provide information used by the review committees to help in their evaluation of proposals. Information provided in the coversheet is treated as confidential by the review committees and will not be published as a part of the proposal document, in the event that the proposal is deemed appropriate for posting.
Required field

*Your name (and any other proposal authors):

*Email (for main proposal author):

Your address:

*Date:

*Describe your relationship to the user community for the script/characters. *Examples:*
- *I have been contracted to prepare the proposal by the user community by the Language Board of [ ].*
- *I have been in email contact with [names of users] on [dates].*
- *I am the creator of the script/characters.*
- *I am a member of the user community, as speaker of the [ ] language using the [ ] script.*
- *I discovered the script or characters during research as a student of [field] at [institution].*

*How do you define the user community (for this script or character[s])?
Examples:
- *Modern users of the Bamum script who live in Fumban, Cameroon, and want to be able write the script on their mobile devices.*
- *Researchers in the Middle Eastern Languages and Cultures Department at UC Berkeley who are studying cuneiform tablets from Iraq and want to put their text online.*

TEMPLATE FOR CHARACTER ADDITIONS TO AN EXISTING SCRIPT

To: Unicode Technical Committee
From: [your name and affiliation]
Subject: [Additions to {Unicode script name}]
Date:

Include a short paragraph summarizing what is being asked the UTC to do (such as, "We are requesting the addition of one Arabic character").

I Background

Provide background on the proposed character additions, addressing questions listed below. This information is needed in order to understand the use of the characters today. For modern character additions, widespread usage is looked upon favorably.

Information to include:

- Who is the user community (for example, scholars, government, school children, or elders of a community) for whom these characters are needed?

- What is the approximate size of the user community?

- What is the time period and geographical area (i.e., countries) when/where the characters have been used?

   For modern-use characters
   - o Are they being taught (in public schools or private schools or informally)?
   - o Are they included in primers?
   - o Do they appear in newspapers or other publications and/or signage? (Provide examples in section VIII. Note that examples besides a chart are needed.)
   - o Is there government support for the characters?
   - o Is there a font currently being used to print the characters?
   - o Do the characters appear on any keyboards today? If so, provide information on the keyboard(s) that use the characters

   Tip: A criterion for adding a new, modern-use character is that the character has spread across a community (beyond the creator).

   For historical characters
   - o Were the characters used in certain contexts (such as, in religious materials or to document commercial transactions)?
   - o What is the importance of getting these historical characters into Unicode?

- Is there an ongoing digitization project that requires the characters?

- What language(s) are they used in? Give the ISO 639 code for the language(s) using the character.

- If the characters were proposed earlier, give proposal document numbers. [*Tip*: Do a search on the Unicode website to see if any reference is made to the additions.]

II Proposed characters

List of proposed characters
List the proposed characters. For each character, provide a glyph (image) and a proposed name for each and refer to at least one example in the figures.

Example: (from L2/20-115):

ʄ   U+A7F0 LATIN SMALL LETTER ESH WITH DOUBLE BAR. Figures 1–9.

Glyph: Pick a representative glyph that most users would recognize.

Code point: You can suggest proposed code points for the characters, but this is not necessary, since the standards committees will assign code points. If you do suggest code points, verify they are available by comparing the latest script code chart and the Pipeline of proposed characters. If you don't suggest code points, use XXXX0…XXXX9, XXXXA, …XXXXF, XXX10, XXX11, etc.

Name: It is best to suggest a name that fits with the same pattern as others in the same script and similar characters in other scripts. Check the names list or the code charts for the naming patterns.

Figure: At least one example of each character is required. Examples help to establish that the characters need to be interchanged digitally. In many cases, this can be demonstrated by providing two examples from different authors and different publishers. In some cases, one figure may suffice. The character should be circled in the figure with a caption that identifies the character and provides a reference to the source. (See further guidelines in section VII.)

Additional information on characters
Phonetic value
Provide the phonetic value of the characters (preferably with IPA or a phonetic description), if applicable.

Example (from L2/19-118):

The first of these characters is commonly used when writing Wolof using the Arabic script to represent a few phonemes ñ, č, nǰ, nč. JEEM WITH THREE DOTS ABOVE typically functions to create nasalised sounds that are common in Wolof.

Joining information
Do any characters join or touch with one another? If so, describe how they join and provide examples in the figures. The more detail, the better. This is especially helpful for font designers.

Annotation (optional)
You can also suggest an annotation for the names list, if you wish. (The annotations appear after bullets in the names list, see example below.)

The editors will take the suggested annotations into consideration, though the names list is not intended to contain encyclopedic information. An annotation is particularly useful to guide users in determining which character to use and for those implementing Unicode characters in software and fonts. In the example below, adding the language can assist font developers so they know to include this character in a font that will be used for Punjabi, for example.

> Example (with a suggested annotation after the bullet, here added since the character is only used for the Punjabi language) (from L2/19-111):

> 08C5 ARABIC LETTER LAM WITH SMALL ARABIC LETTER TAH ABOVE
> • Punjabi language

If a proposed character is a mark of punctuation
Describe the character's behavior (for example, if a given character is used as a paragraph marker). Provide examples that demonstrate the character's use.

III Properties
General Category and other properties
Provide the General Category of the characters and other properties.

Tip: Find characters that you think are used in a similar way, and use the properties contained in https://www.unicode.org/Public/UCD/latest/ucd/UnicodeData.txt, modifying as needed.

> Example (from L2/19-111):

> 08C5;ARABIC LETTER LAM WITH SMALL ARABIC LETTER TAH
> ABOVE;Lo;0;AL;;;;;N;;;;;

The fields include the following (with links to descriptions and range of values):

> Code_point (provide your suggested code point or XXXX0, etc.)
> NAME (provide your suggested name)
> General_Category [range of gc values]
> Canonical_Combining_Class [range of CCC values]
> Bidi_Class [range of Bidi_Class values]
> Decomposition_Type/Decomposition_Mapping
> Numeric_Type
> Numeric_Value
> Bidi_Mirrored
> Unicode_1_Name (obsolete as of Unicode 6.2.0)
> ISO_Comment (obsolete as of Unicode 5.2.0)
> Simple_Uppercase_Mapping

Simple_Lowercase_Mapping
Simple_Titlecase_Mapping

Line breaking information

Provide information on line breaking behavior. Are there restrictions on line breaking before or after the proposed characters?

Script information

Identify which script(s) the character(s) are used in.

Properties for Arabic and Indic characters

Additional information is required for characters that join, such as Arabic, and for various Indic scripts. (See details in https://www.unicode.org/Public/UCD/latest/ucd/ArabicShaping.txt and, for Indic scripts, https://www.unicode.org/Public/UCD/latest/ucd/IndicPositionalCategory.txt and https://www.unicode.org/Public/UCD/latest/ucd/IndicSyllabicCategory.txt.)

IV Collation

What is the expected sorting order of the characters in the script? In other words, where would you expect them to occur in a dictionary or list of words?

Give the order characters are expected to occur. A prose description is acceptable or you can use the following format (which shows ALEPH occurs before GENERIC ALEPH-NUN, etc.):

ALEPH < GENERIC ALEPH-NUN < ALEPH WITH LEFT TAIL < BETH <

GENERIC BETH-YODH < GIMEL-HETH < WAW < ZAYIN < FINAL HETH <

YODH < KAPH < LAMEDH < MEM < NUN < SAMEKH < PE <

(Note: The characters in a code chart do not necessarily reflect the expected order in a dictionary.)

V References

Provide reference works that you used and standard reference works that can be consulted, with full bibliographic information.

> Example (from L2/19-111):
> Muhammad Yar. Circa 1792. *Afarinish Nama*. Courtesy British Library.

VI Acknowledgements

(optional)

VII Examples

Provide examples of each proposed character in a line of printed text, circling the characters. In the caption, name the circled character and provide the reference.

Printed materials are generally preferred, if possible.

It is highly advisable to include texts that show the character in running text, that is, not just a chart or a keyboard, but sentences that use the character. Such examples help demonstrate spacing and where line breaks can occur.

Provide a transliteration of the text of a few examples.

NOTE: If no examples are provided, the proposal will very likely not be reviewed, stalling the proposal. Only those characters with evidence will be considered.
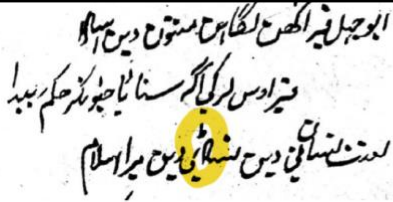
Example (from L2/19-111):

Figure 3. ARABIC LETTER LAM WITH SMALL ARABIC LETTER TAH (*Afarinish Nama*)
نسلاؠ (*noon seen rlam alef yeh*) [nasllai] {নমলাঈ} meaning "of forefather or from generation".

IX ISO Proposal Summary

Include the ISO proposal summary forms at the end. This is not necessary for the preliminary version of a proposal.

Next Steps

1. For submission to Unicode for review by Script Ad Hoc

- Number all the pages.
- Send a copy to Unicode (docsubmit@unicode.org) so it can be routed to the Script Ad Hoc to review.

  The Script Ad Hoc group meets monthly and generally will review those proposals that have included information described in this template and have shown the need for the characters to be carried in plain text. Typically proposals take several revisions until all the needed information is provided. Comments on proposals that are reviewed by the Script Ad Hoc appear in the quarterly Script Ad Hoc recommendations.

NOTE: Not all proposed scripts and characters are approved. Some newly created scripts may not yet be ready for encoding, as the script generally needs to be widely accepted, used in various publications, and taught in schools.

Note: If the characters are eventually accepted, a font will be required.

- If you do not have a font, you can create a font using Glyphs or other font-editing software. When creating glyphs that uses another font as its base, use a font that allows modification (such as Noto or a font with a SIL Open Font License).
- If a character includes combining marks (such as diacritics that appear above or below another character such as ◌̀), do not include the dotted circle.
- In the font, map the glyphs to the proposed code points.
- The glyphs for the proposed characters should try to match the style found in the current code charts.

2. After the proposal is considered mature by the Script Ad Hoc

Once the Script Ad Hoc has decided the proposal has provided full information on the proposed characters and a strong case has been made to encode them, the document will be posted in the Unicode document register and then will be reviewed by the Properties and Algorithms group, which may suggest changes. Once both the Script Ad Hoc and Properties and Algorithms have agreed the proposal is ready for encoding, the proposal goes to the Unicode Technical Committee (UTC).

The UTC will review the proposal and, if it approves the characters, will provisionally assign code points. The UTC will at a later point formally accept the characters and identify a Unicode version where the characters will be published. Before publication, the characters will go through an alpha and beta review phase.

Proposers should be aware that the entire approval process – if successful – can take at least two years.

Proposals that are considered mature and advance to the UTC are required to have a font with the glyphs.

3. After publication

Once the characters are published in Unicode, having a font and keyboard ready – with the characters at the approved Unicode code points -- can speed up the use of the characters. In addition, operating systems may need to be updated in order to support the new characters. Working with software, platform and font vendors may be useful in getting support for the new characters.

For support in locale data, see CLDR project.

Additional tips

Basically, the more information on the use and behavior of a character (with clear examples), the better. This information can be used by the committees to verify the character as proposed and to check if the specified properties are correct.

Consulting with linguists or specialists in the script is very helpful, since they can provide phonetic information and review the proposal. Also, if there are local or governmental language commissions or language institutes or societies, involving them can also be very beneficial.

Good examples to use as a guide:
(adding a letter to Limbu) L2/11-008
(adding a letter to Lisu) L2/18-338

The most successful proposal authors are those that work collaboratively with the standards committees. The committees review proposals to ensure the characters follow the Unicode Standard. Typically, proposal authors need to revise their proposals several times in order to present all the required information.