Title: Proposal to include Shughni/Roshani Alphabet

Action: For consideration by UTC and ISO/IEC JTC1/SC2/WG2

Author: Habib Zohoori

Date: Dec 26, 2023

Introduction

Shughni and Roshani, languages spoken in the Badakhshan province of Afghanistan and Tajikistan, face a unique challenge in their written form on the Afghanistan side. Currently, these languages are transcribed using Arabic characters. However, a linguistic nuance arises with the Arabic character denoting "th," which, despite its appearance with three dots, is pronounced as "S" in Shughni. This misalignment can lead to confusion, particularly when attempting to replace "Th/ θ".

In my role as a Shughni language Software Architect, I am dedicated to addressing this issue by providing accessible means of writing Shughni and Roshani on phones or computers, all offered freely to users. Part of this initiative involves introducing a new character that accurately represents the unique linguistic sound in Shughni and Roshani, currently unsupported by Unicode.

As part of this proposal, I seek to present this new character to the Unicode teams for inclusion. Beyond rectifying the confusion caused by the current use of the Arabic character for "th/ θ," the incorporation of this new character holds the potential to enhance the writing experience for language speakers, ensuring a more accurate and user-friendly representation of their linguistic heritage.

Requests:

The author requests the encoding for a character used in Shughni/Roshani language.

Character Name and Shape:

<table>
<thead>
<tr>
<th>Shape</th>
<th>Code</th>
<th>Phonetic</th>
<th>IPA</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>🳙</td>
<td>U+0892</td>
<td>th</td>
<td>θ</td>
<td>ARABIC THEH WITH RING BELOW</td>
</tr>
</tbody>
</table>

Proposal to include Shughni/Roshani Alphabet
Proposal to include Shughni/Roshani Alphabet

Letter forms:

<table>
<thead>
<tr>
<th>بِل</th>
<th>بِنَج</th>
<th>ثِهٍ</th>
<th>بٍ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final</td>
<td>Middle</td>
<td>Initial</td>
<td>Isolated</td>
</tr>
</tbody>
</table>

This character is commonly used when writing Shughni language using the Arabic script to represent the sound θ/th, ARABIC THEH WITH RING BELOW. It is pronounced as “Theh” or θeh

Figures used in writing, circled in red:

Proposal to include Shughni/Roshani Alphabet

Proposal to include Shughni/Roshani Alphabet
Proposal to include Shughni/Roshani Alphabet

Transliteration:
Yew ðod qandred char
ștor pydand sat ya khawr
Wam yewe yad hazra
Sat kanedž ya bechora
Khoďoy mahďyd yikdes dad
Har cha petawd kho awlad
Dahů karenen nest ye chind
Focaθen daď afyonennd
Tar mo ches ay taryoke
Kho pets taredaj make
Mev mo gapen khech neghauj
Asonandaθ weď har kho ghauj

Translation:
One was given to foreigners
Who got busy as Sheppard

Proposal to include Shughni/Roshani Alphabet
The rest was sent to thousandards
And turned into a slave
God dumb a father like this
Who throws his kids everywhere
No one else would do such a thing
All are an addict’s doings
Look at me, oh addict
Don’t turn your face from me
Listen to me carefully
Put them all in your ears
Example 2:

Shughni lesson on letter Theh (θ)

ث

Transliteration:
Soraya nor chakar tets
Kho khaima khezande yats thawj

Translation:
Sorya has gone out today
She has burned fire in front of her tent

Proposal to include Shughni/Roshani Alphabet
<table>
<thead>
<tr>
<th>Wooden plate</th>
<th>تاشج</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ash</td>
<td>ثير</td>
</tr>
<tr>
<td>Walking Stick</td>
<td>ماث</td>
</tr>
</tbody>
</table>

Transliteration:
Tathch/taθch
Theer/θeer
Math/Maθ

Proposal to include Shughni/Roshani Alphabet
Example 3:
Because there is no letter to represent the Shughni sound for TH/θ the “Arabic Theh” is used to write the language, which confuses readers whether to pronounce the letter as “S” or TH/θ.

Transliteration:
Shughni: Thaw, Thaw nedzdaw  
**Translation:** Marking or Tattoo, Tattooing  
Farsi: Khal az sokhtagi, khal zadan  
**Translation:** Mark from burning or Tattoo  
Shughni: Sobot, Sobot mo te tsa sodzjat banden mo chodjat  
**Translation:** evidence, ex: if they had evidence, they would’ve put me in Jail  
Farsi: agar balayam sobot mekardan, bandi me shodam  
**Translation:** if they had evidence, they would’ve put me in Jail  
Shughni: Thapast, Yo kho jay teer thapast nost, yats thapast wezod  
**Translation:** quickly, ex: He quickly sat down, the fire turned off all of a sudden  
Farsi: Tez, oh sareh dar jaye khod nishast, atash dafatan khamosh shod  
**Translation:** quickly, ex: He quickly sat down, the fire turned off all of a sudden  
Shughni/Shughni: Soraya  
**Translation:** Soraya (Name)  

Proposal to include Shughni/Roshani Alphabet
Farsi: khoshaye parween ba englisi pleiades) ya soraya nam khoshaye dar sorat falaki gaw ke ba haft khooharan ham mashoor hast, dar list masya ham nam gozari shoda hast. Khoshaye parween ba nam arabi An Soraya ham haft khoahar me goftand, chira ke ba tawr mahmool haft khoahar az besh az diwest sitara shonakhta shoda An ba chashm ghair masleh qabil deedan stand.

Translation: The branches of pleiades. Rest is explanation of the word “Soraya” and how it means different things

Shughni: Theer Shap
Translation: Ash container/jar

Farsi: Koza az khaistar
Translation: Ash container/jar

Shughni: Theer goshts
Translation: Volcano

Farsi: Atash fishan
Translation: Volcano

Shughni: thewdaw
Translation: To Burn

Farsi: shokhtandan
Translation: To Burn

Shughni: thew
Translation: Burn

Farsi: sokhtan
Translation: Burn

Shughni: theerjawn
Translation: Ash Yard

Farsi: Jaye baraye jamh awardan khakestar.
Translation: A place where ashes are thrown

Shughni: Theek
Translation: Someone with speech disorder (articulation disorder)

Farsi: Lankat ba ikhtilal dar rawani goftar wa algoye zamani aan ke sokhan goftan ra doshwar me sazad gofta meshawad. Lankat omdata ba tawr koli joz dar mawrod istisnayi dr dwran kwdky hast, w mahmolan az sineen 2 ya 4 salagi shoro me shawad. Lankat az mohim tareen wa motadawil tareen ikhtilal takmeely hast, dar sineen 6 ya 7 salagi nez as masadif ba zaman aghaz madrasa hast ba lihaz wizhagi haye atifi wa sazgary ijtimah khas een dawra moqhat monasibi hast baraye baroz lankat dar miyan kodakani ke beshtar mosthad chinen ikhtilal hastand. Gahi az mawarid dar daroon bologh nez afradi qablan ba nahwi lankat dashta and dochar lankat zaban meshawad. Dar waqeh sin, sharayet wa wizhagi haye bologh nez ehtimal dar tagheer shall ya tashdeedk lankat mosarand. Lankat zaban dar miyan pisaran bishtar az dokhtaran hast wa dar barrasi ha wa tahqeeqat mokhtalif nasbati..

Translation: articulation disorder is usually when someone cannot articulate his/her words better. This usually happens when kids start learning the language between the age of 2/4 and as they grow older, the disorder goes away. Some people continue with the disorder into their teenage and adult life.

Reference:

Proposal to include Shughni/Roshani Alphabet


Proposal to include Shughni/Roshani Alphabet
<table>
<thead>
<tr>
<th>A. Administrative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Title:</strong> Proposal to include Shughni/Roshani Alphabet</td>
</tr>
<tr>
<td><strong>2. Requester’s name:</strong> Habib Zohoori</td>
</tr>
<tr>
<td><strong>3. Requester type (Member body/Liaison/Individual contribution):</strong> Individual Contributor</td>
</tr>
<tr>
<td><strong>4. Submission date:</strong> Dec 26, 2023</td>
</tr>
<tr>
<td><strong>5. Requester’s reference (if applicable):</strong></td>
</tr>
<tr>
<td><strong>6. Choose one of the following:</strong></td>
</tr>
<tr>
<td>- This is a complete proposal: ✓</td>
</tr>
<tr>
<td>- (or) More information will be provided later:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Technical – General</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Choose one of the following:</strong></td>
</tr>
<tr>
<td>- This proposal is for a new script (set of characters):</td>
</tr>
<tr>
<td>- Proposed name of script:</td>
</tr>
<tr>
<td>- The proposal is for addition of character(s) to an existing block: ✓</td>
</tr>
<tr>
<td>- Name of the existing block: Arabic Extended-B</td>
</tr>
<tr>
<td><strong>2. Number of characters in proposal:</strong> 1</td>
</tr>
<tr>
<td><strong>3. Proposed category (select one from below - see section 2.2 of P&amp;P document):</strong></td>
</tr>
<tr>
<td>- A-Contemporary</td>
</tr>
<tr>
<td>- B.1-Specialized (small collection) ✓</td>
</tr>
<tr>
<td>- B.2-Specialized (large collection)</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>C-Major extinct</th>
<th>D-Attested extinct</th>
<th>E-Minor extinct</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-Archaic Hieroglyphic or Ideographic</td>
<td>G-Obscure or questionable usage symbols</td>
<td></td>
</tr>
</tbody>
</table>

4. Is a repertoire including character names provided?
   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. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? Habib Zohoori
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): Habib Zohoori – habibullah.zahoori@gmail.com

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:
   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

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 Unicode Character Database (http://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.
C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?  
   No
   If YES explain

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
   If YES, with whom?  
   If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?  
   Yes  
   Reference: Badakhshan Afghanistan and Tajikistan, where the Shughni /Roshani Native Speakers Reside

4. The context of use for the proposed characters (type of use; common or rare)  
   common  
   Reference: The proposed alphabet is very common in Shughni/Roshani Language.

5. Are the proposed characters in current use by the user community?  
   Yes  
   If YES, where? Reference: It is mostly used when handwriting, but since computer does not support it; they use random characters to identify it.

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?  
   Yes  
   If YES, is a rationale provided?  
   Yes  
   If YES, reference: Character could be kept with their counterparts in Arabic block

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
   If YES, is a rationale for its inclusion provided?
   If YES, reference:

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?  
   No
   If YES, is a rationale for its inclusion provided?
   If YES, reference:

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?  
    No
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Does the proposal include use of combining characters and/or use of composite sequences?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, is a rationale for such use provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>12. Does the proposal contain characters with any special properties such as control function or similar semantics?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, describe in detail (include attachment if necessary)</td>
<td></td>
</tr>
<tr>
<td>13. Does the proposal contain any Ideographic compatibility characters?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, are the equivalent corresponding unified ideographic characters identified?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
</tbody>
</table>