

L2/24-108

Version:2

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:

Shape	Code	Phonetic	IPA	Name
	U+0892	th	θ	ARABIC THEH WITH RING BELOW

Proposal to include Shughni/Roshani Alphabet

Letter forms:

			
Final	Middle	Initial	Isolated

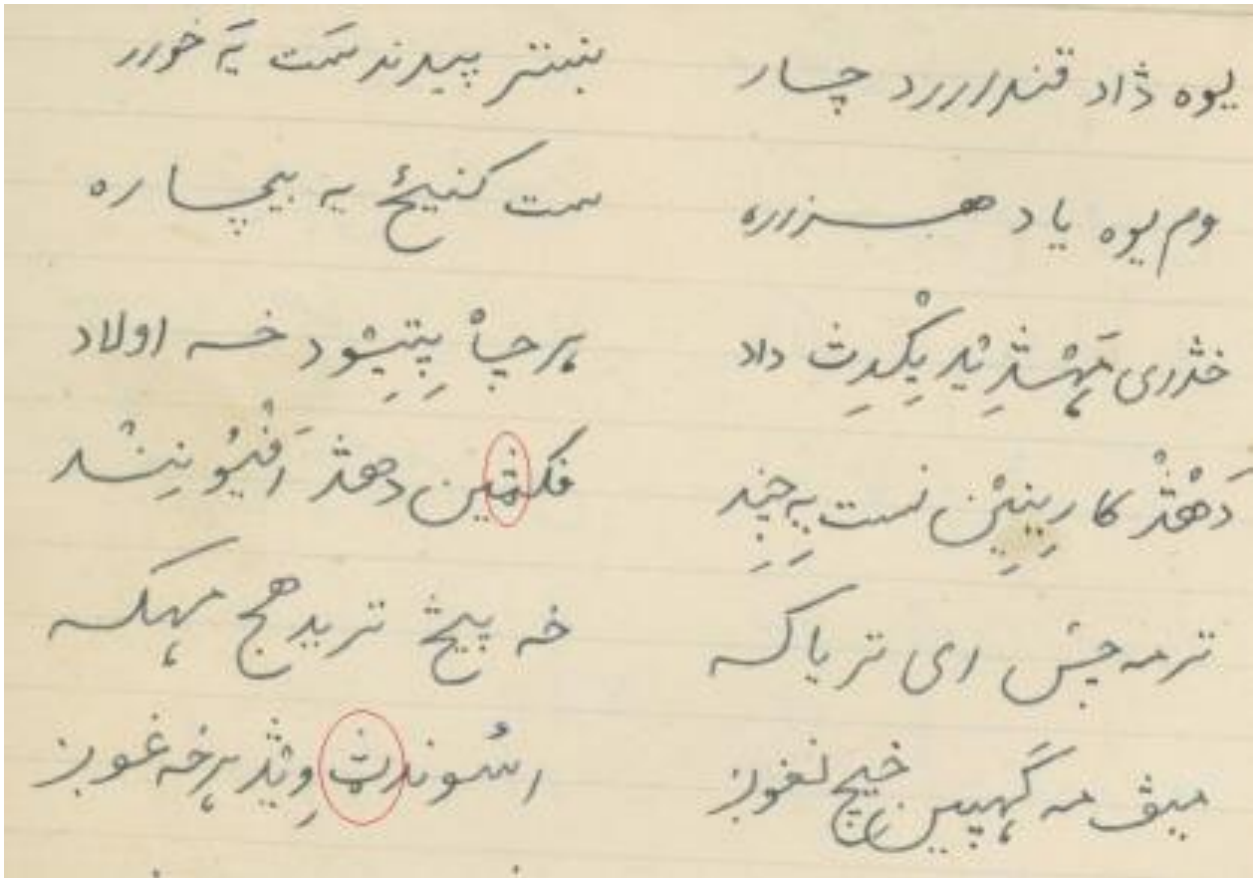
This character is commonly used when writing Shughni language using the Arabic script to represent the sound θ /th, ARABIC THEH WITH RING BELOW. It is pronounced as "Theh" or θ eh

Figures used in writing, circled in red:



Proposal to include Shughni/Roshani Alphabet

Proposal to include Shughni/Roshani Alphabet



Transliteration:

Yew ðod qandrerd char
 ştor pydand sat ya khawr
 Wam yewe yad hazra
 Sat kaneðz ya bechora
 Khoðoy mahðyd yikdes dad
 Har cha petawd kho awlad
 Dahð karenen nest ye chind
 Focaθen dað afyonennd
 Tar mo ches ay taryoke
 Kho pets taredaj make
 Mev mo gapen khech neghauj
 Asonandaθ weð 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



ثريا نر چكر تيخ
خو خيمه خيزنده ياش ثوج

Transliteration:

Soraya nor chakar tets

Kho khaima khezande yats **thawj**



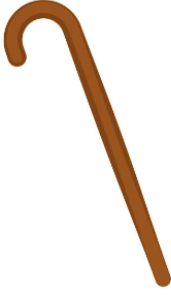
Translation:

Sorya has gone out today

She has burned fire in front of her tent

Proposal to include Shughni/Roshani Alphabet

ث

	Wooden plate	تاتچ
	Ash	تیر
	Walking Stick	مات

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 Teh” is used to write the language, which confuses readers whether to pronounce the letter as “S” or TH/ θ

ٲٲٲر شٲٲ = کوزه از خاکستر.
ٲٲٲر = خاکستر.
ٲٲٲر گویٲش = آتشفشان.
ٲٲٲوداو = سوختاندن.
ٲٲٲداو = سوختن.
ٲٲٲر-چون = جای برای جمع آوری خاکستر.
ٲٲیک = لکنٲ به اختلال در روانی گفتار و الگوی زمانی آن که سخن گفتن را دشوار می‌سازد گفته می‌شود. لکنٲ عمدتاً و به‌طور کلی جز در موارد استثنایی در دوران خردسالی آغاز شده و یک پدیده خاص دوران کودکی است و معمولاً از سنین ۲ تا ۴ سالگی شروع می‌شود. لکنٲ از مهم ترین و متداول‌ترین اختلالات تکلمی است. در سنین ۶ تا ۷ سالگی نیز که مصادف با زمان آغاز مدرسه‌است به لحاظ ویژگی‌های عاطفی و سازگاری اجتماعی خاص این دوره موقعیت مناسبی است برای بروز لکنٲ در میان کودکانی که بیشتر مستعد چنین اختلالی هستند. گاهی از موارد در دوران بلوغ نیز افرادی که قبلاً به نحوی لکنٲ داشته‌اند دچار لکنٲ زبان می‌شوند. در واقع سن، شرایط و ویژگی‌های بلوغ نیز احتمالاً در تغییر شکل یا تشدید لکنٲ مؤثرند. لکنٲ زبان در میان پسران بیشتر از دختران است و در بررسی‌ها و تحقیقات مختلف نسبتی

ث

th/θ

ٲٲاو، ٲٲاو نندداو = خال از سوختگی، خال زدن.
ٲٲوت = ثبوت؛ ~ مو-ٲے ځه سډډجت، بند-ٲن مو چورجت = اگر بالاٲم ثبوت می‌کردند، بندی می‌شدم؛ ثبوت چ. = ثبوت ک. ٲ. دین، حسابٲ.
ٲٲست، ٲٲست، ٲٲست؛ ٲٲست، ٲٲست = تیز، سریع، دفعٲن؛ یو ځو جای تیر ٲٲست نوست = او سریع در جای خود نشسٲ؛ یاځ ٲٲست وږد = آتش دفعٲن خاموش شد؛ بارون ٲٲست دٲد = باران سریع میبارد؛ یو ٲات بساه ٲٲست ځو جای ٲے نوست = ٲاد شاه زود در جای خود نشسٲ (فلک).
ٲٲوت = ٲٲوت؛ ٲٲوتمند = ٲٲوتمند.

ٲٲریا، ٲٲریا = خوشهٲ پروین (به انگلیسی: Pleiades)، یا ٲٲریا (ام ۴۵) نام خوشه‌ای در صورت فلکی گاو که به «هفت خواهران» هم مشهور است. در لیست مسیه ام ۴۵ نام گذاری شده‌است. خوشه پروین با نام عربی آن، ٲٲریا، هم در فارسی نام برده شده‌است. در گذشته به آن هفت خواهر هم می‌گفتند چرا که به‌طور معمول فقط هفت ستاره از بیش از دویست ستاره شناخته شده آن، با چشم غیر مسلح قابل دیدن هستند و

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

Farsi/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 khoharan 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 khahar 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

Translate: Someone with speech disorder (articulation disorder)

Farsi: Lankat ba ikhtilal dar rawani goftar wa algoye zamani aan ke sokhan goftan ra doshwar me sazaad gofta meshawad. Lankat omdatan 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

Zahoori, M. S. (1970). Nedaye Dil Poetry. Balkh; Afghanistan.

Arkan, S., & Zahoori, M. S. (2011, January 23). يرظ هاش بهذم طست ینغش یابفلا ... - shughnan. Shughni_Language-Mazhab_Shah_Zahoori. http://www.shughnan.com/wp-content/uploads/2013/08/Shughni_Language-Mazhab_Shah_Zahoori.pdf

Arkan, S. (2022). . عتسماخ زا نروم = پيش ريث - shughnan.com. Shughni Vocabulary. <http://www.shughnan.com/wp-content/uploads/2021/05/shughnani-alefba-5-c.pdf>

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 <http://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 <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest *Roadmaps*.

A. Administrative

1. Title:	Proposal to include Shughni/Roshani Alphabet
2. Requester's name:	<i>Habib Zohoori</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual Contributor</i>
4. Submission date:	<i>Dec 26, 2023</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/>
(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):		
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/>	
Name of the existing block:	<i>Arabic Extended-B</i>	
2. Number of characters in proposal:	<i>1</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>

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)

Proposal to include Shughni/Roshani Alphabet

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"/>	<input type="checkbox"/>

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. 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

Please see section named character and shape

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:	<i>Badakhshan Afghanistan and Tajikistan, where the Shughni /Roshani Native Speakers Reside</i>
4. The context of use for the proposed characters (type of use; common or rare)	common
Reference:	<i>The proposed alphabet is very common in Shughni/Roshani Language.</i>
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

Proposal to include Shughni/Roshani Alphabet

If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	<i>No</i>
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?	<i>No</i>
If YES, reference:	
12. Does the proposal contain characters with any special properties such as	
control function or similar semantics?	<i>No</i>
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	<i>No</i>
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	