

Title: Proposal for encoding the Yezidi script in the SMP of the UCS

Authors: Andrij Rovenchak, Dimitri Pirbari, and Erdal Karaca

Date: 11 March 2019

The preliminary version of the proposal was given in the document [L2/18-238](#)

1. Introduction.

The Yezidi script has a long history. According to some data, it dates back to the 13–14th centuries, while some scientists place the creation of this writing system in the 17–18th centuries (Amoêv 1999). The author of the script is unknown. The Yezidi script was used to write two manuscripts, *Maşḥaf Raş* and *Ktêbî Ĵalweh*, first published by Anastase Marie (1911), see samples in Figs. 1–2, while earlier manuscripts of the same titles were known in Arabic (cf. Joseph 1909).

The manuscripts are written on sheets of thin parchment made from processed gazelle leather. The pages of *Ktêbî Ĵalweh* contain 16 lines of text compared to 11 lines in *Maşḥaf Raş*. The first book contains a monologue of Tawûsê Melek, the principal angel in the Yezidi religion. In *Maşḥaf Raş*, cosmogony is described, which reaches beyond traditional Yezidi views (Amoêv 1999; Pirbari & Amoêv 2013). The language of *Maşḥaf Raş* is simpler than in *Ktêbî Ĵalweh*. The time of writing of those manuscripts as well as their authorship remains unclear. It is likely that they originate from 12th–13th centuries (Marie 1911; Omarkhali 2012; Avdoev 2016). It is believed that historically there existed Yezidi sacred manuscripts known as *Maşḥaf Raş* and *Ktêbî Ĵalweh*, but the originals were lost. Later copies of these manuscripts were found, which were written in a special Yezidi alphabet. But, unfortunately, their content was distorted. Therefore, the Yezidi clergy recognize the Yezidi writing, but do not consider the content of these manuscripts as a source of the Yezidi faith.

Texts in the Yezidi script are in the Kurmanji language known also as Northern Kurdish (https://en.wikipedia.org/wiki/Northern_Kurdish; <https://www.ethnologue.com/language/kmr>)

A decade ago, an attempt was made in Iraq to revive the script, even a Yezidi-Arabic dictionary was published. Unfortunately, in view of complex political and social situation the Yezidis live in, this idea has not been developed further.

In 2013, the Spiritual Council of Yezidis in Georgia decided to revive the Yezidi script and use it to write prayers, sacred books, on the organization letterhead (see Fig. 3), in the Yezidi heraldry, etc. For this purpose, two orientalists and specialists, Kêrîm Amoêv and Dimitri Pirbari, modernized the Yezidi alphabet and adapted it to the phonetic features of the modern Yezidi language. The book *Yezidi script* was published (Pirbari & Amoêv 2013), where each letter and its phonetic value are described in detail. Today the Yezidi script is used by clergymen in the Yezidi temple in Tbilisi. On the walls of the temple, the names of saints are also written in this alphabet (see Figs. 4–5). A book of prayers *Dua'yêd Êzdiyan* in the Yezidi script has been published recently (see sample pages in Figs. 6–7).

2. Structure. Yezidi is an alphabetic script written right to left in horizontal lines. Modern version of the alphabet includes letters both for consonants and vowels. Letters are written apart, unlike the Arabic script. In historical texts, some ligatures occur: 𐌒 (l-t), 𐌓 (l-c), 𐌔 (l-q), 𐌕 (l-k), 𐌖 (l-y), 𐌗 (l-v); see Fig. 1. Such ligatures are not used in the modern version of the Yezidi alphabet, but the joining behavior can be specified as an optional feature of the font. The following letter combinations are used instead of the abovementioned ligatures: 𐌒𐌐 (l-t), 𐌓𐌐 (l-c), 𐌔𐌐 (l-q), 𐌕𐌐 (l-k), 𐌖𐌐 (l-y), 𐌗𐌐 or 𐌗𐌐 (l-v).

In Yezidi, some diacritics inherited from the Arabic scripts are used in words of the Arabic origin. These are HAMZA < ْ > (denoting a glottal stop) and MADDA appearing over ELIF < َ > to denote long /a/ at the beginning of some words.

Comparison of different script to write Yezidi is shown in Table 1.

3. Collating order. Collation order is that in the code chart. It reflects the alphabetical order preferred by the Yezidi community in Georgia. To our knowledge, this issue is not under active consideration in any other community. Note that the original order of letters was based on the Perso-Arabic script, cf. Fig. 8, where the classical version of the Yezidi alphabet is shown.

The newly created letters are: ʼ, ȳ, ɖ, ɗ, ʒ, ʝ, ɟ, ɣ, ɸ, ɹ.

In the classical version of the alphabet, the shape ʼ for PHE (often simplified to a straight line, more specifically, ʼ) was used to denote both [p] and [pʼ] alongside the shape ʒ, cf. Figs. 1 and 2. In the modern version of the alphabet, the two letters are used for different phonemes.

4. Character names. The usual UCS conventions are used. Glottalization is marked by H (PE versus PHE).

ɹ YEZIDI LETTER YOT serves to denote the consonant [j] or vowel [i] (in the Roman orthography, ⟨y⟩ and ⟨î⟩, respectively). This letter serves to denote a vowel if placed at the beginning of a word before a consonant or after the ELIF.

ɸ YEZIDI LETTER ET serves to denote the vowel [e], which is written as ⟨ê⟩ in the Roman orthography.

5. Linebreaking. Letters and digits behave as do the letters and numbers in Arabic. However, YEZIDI HYPHENATION MARK can be placed above the last letter in the line to denote word breaks. In old texts, this mark also appears at the beginning of the following line (right-most position, see Fig. 2).

6. Punctuation and digits. No script-specific punctuation is known in Yezidi, ⟨,⟩ ARABIC COMMA (U+060C), ⟨;⟩ ARABIC SEMICOLON (U+061B), and ⟨?⟩ ARABIC QUESTION MARK (U+061F) can be used additionally to ordinary FULL STOP and COLON. Arabic digits (U+0660...U+0669) were applied in older texts (cf. Anastase 1911) while Western digits (0...9) are preferred nowadays, see page numbers and edition year in Fig. 6. Unfortunately, there are no texts available so far to illustrate the use of digits inside texts.

7. Historical diacritical marks. Several diacritics occur in one of classical manuscript texts, *Maşhaf Raş* (see Fig. 9). The meaning of some diacritics is unclear. For example, the meaning of the dot below the letter RA in Fig. 9 is not known. For some other marks, no special meaning can be assigned. This includes the decorative horizontal parenthesis in Fig. 9, as well as a dot between letters in the lower right part of Fig. 1 (the latter could be an accidental scribal sign).

However, there are two letters with diacritics that occur in historical texts and change the pronunciation of the base letter:

LAM WITH DOT ABOVE denotes the “hard” [t] as opposed to the soft (palatalized) [lʲ].

YOT WITH CIRCUMFLEX ABOVE changes the pronunciation from [î] to [ê].

These two precomposed characters, LAM WITH DOT ABOVE and YOT WITH CIRCUMFLEX ABOVE, are separately proposed here.

Presently, we do not propose to encode other historical diacritical marks as they are attested neither in other manuscripts nor in modern texts. These can be encoded in future on the reserved codepoints if required.

On the other hand, we do not consider it possible to decompose the shape of VA ⟨ɹ⟩ as UM plus a combining dot below. This shape is the basic one used to represent the consonant [v].

Table 1: Yezidi in different scripts.

Yezidi	Arabic	Latin	Cyrillic	IPA	Proposed Unicode name
ا	ا	A a	А а	a	YEZIDI LETTER ELIF
ب	ب	B b	Б б	b	YEZIDI LETTER BE
پ	پ	P p	П п	p	YEZIDI LETTER PE
پ'	پ'	P' p'	П' п'	p'	YEZIDI LETTER PHE
ت	ت	T' t'	Т' т'	t'	YEZIDI LETTER THE
س	س	s		s	YEZIDI LETTER SE
ج	ج	C c	Ц ц	ɟ	YEZIDI LETTER CIM

Yezidi	Arabic	Latin	Cyrillic	IPA	Proposed Unicode name
𐌆		Ç ç	Ч ч	tʃ	YEZIDI LETTER CHIM
𐌇	چ	Ç' ç'	Ч' ч'	tʃʰ	YEZIDI LETTER CHHIM
𐌈	ح	H' h'	Һ' һ'	h	YEZIDI LETTER HHA
𐌉	خ	X x	Х х	x	YEZIDI LETTER XA
𐌊	د	D d	Д д	d	YEZIDI LETTER DAL
𐌋	ذ	z		z ^ɕ	YEZIDI LETTER ZAL
𐌌	ر	R r	Р р	r	YEZIDI LETTER RA
𐌍	ړ	R' r'	Р' р'	rʰ	YEZIDI LETTER RHA
𐌎	ز	Z z	З з	z	YEZIDI LETTER ZA
𐌏	ژ	J j	Ж ж	ʒ	YEZIDI LETTER JA
𐌐	س	S s	С с	s	YEZIDI LETTER SIN
𐌑	ش	Ş ş	Ш ш	ʃ	YEZIDI LETTER SHIN
𐌒	ص	ş		s ^ɕ	YEZIDI LETTER SAD
𐌓	ض	ḍ		d ^ɕ	YEZIDI LETTER DAD
𐌔	ط	T t	Т т	t	YEZIDI LETTER TA
𐌕	ظ	z'		zʰ	YEZIDI LETTER ZE
𐌖	ع	' / E' e'	Ə' ə'	ʔ	YEZIDI LETTER EYN
𐌗	غ	X' x'	Г' г'	ɣ	YEZIDI LETTER XHEYN
𐌘	ف	F f	Ф ф	f	YEZIDI LETTER FA
𐌙	ڤ				YEZIDI LETTER VA
𐌚	ڤا	V v	В в	v	YEZIDI LETTER VA ALTERNATE FORM
𐌛	ق	Q q	Q q	q	YEZIDI LETTER QAF
𐌜	ك	K k	К к	k	YEZIDI LETTER KAF
𐌝	ك	K' k'	К' к'	kʰ	YEZIDI LETTER KHAF
𐌞	گ	G g	Г г	g	YEZIDI LETTER GAF
𐌟	ل	L l	Л л	l	YEZIDI LETTER LAM
𐌠	م	M m	М м	m	YEZIDI LETTER MIM
𐌡	ن	N n	Н н	n	YEZIDI LETTER NUN
𐌢	و	U u	Ö ö	u	YEZIDI LETTER UM
𐌣	وو	Û û	У у	uː	(see Note under the table)
𐌤	و	W w	В в	w	YEZIDI LETTER WAW

9. Bibliography

- Amoêv, K. [Амоев, К.] 1999. *Езидские священные книги* (Тбилиси).
- Bittner, M. 1911. Die beiden heiligen Bücher der Jeziden im Lichte der Textkritik. *Anthropos* 6(3): 628–639.
- Dua'yêd Êzdiyan* = *Дѡа'ѣд Эздиѣан* = *ܕܘܥܝܬܐ ܐܙܕܝܐܢ*. 2018 (T'bilîsî: Publishing House "UNIVERSAL").
- Joseph, I. 1909. Yezidi texts. *The American Journal of Semitic Languages and Literatures* 25(2): 111–156.
- Marie, A. 1911. La découverte récente des deux livres sacrés des Yézîdis. *Anthropos* 6(1): 1–39.
- Mingana, A. 1921. Sacred Books of the Yezidis. *Journal of the Royal Asiatic Society of Great Britain and Ireland*, No. 1: 117–119.
- Pirbari, D. & Amoêv, K. [Пирбари, Д. & Амоев, К.] 2013. *Езидская письменность* (Тбилиси: Некери).
- Avdoev, T. [Авдоев, Т.] 2016. *Историко-теософский аспект езидизма* (Москва: Изд-во «Эра»).
- Omarkhali, Kh. [Омархали, Х.] 2012. Йезидская письменная религиозная литература. Священное Писание, Джелва, Масхафа Раш и рукописи. In: *Лазаревские чтения. Вып. 1*. Отв. ред. О. И. Жигалина (Москва: Институт востоковедения РАН), pp. 159–184.

10. Acknowledgements

This project was partly supported by a grant from the United States National Endowment for the Humanities (PR-253360-17), which funds the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.

11. Figures

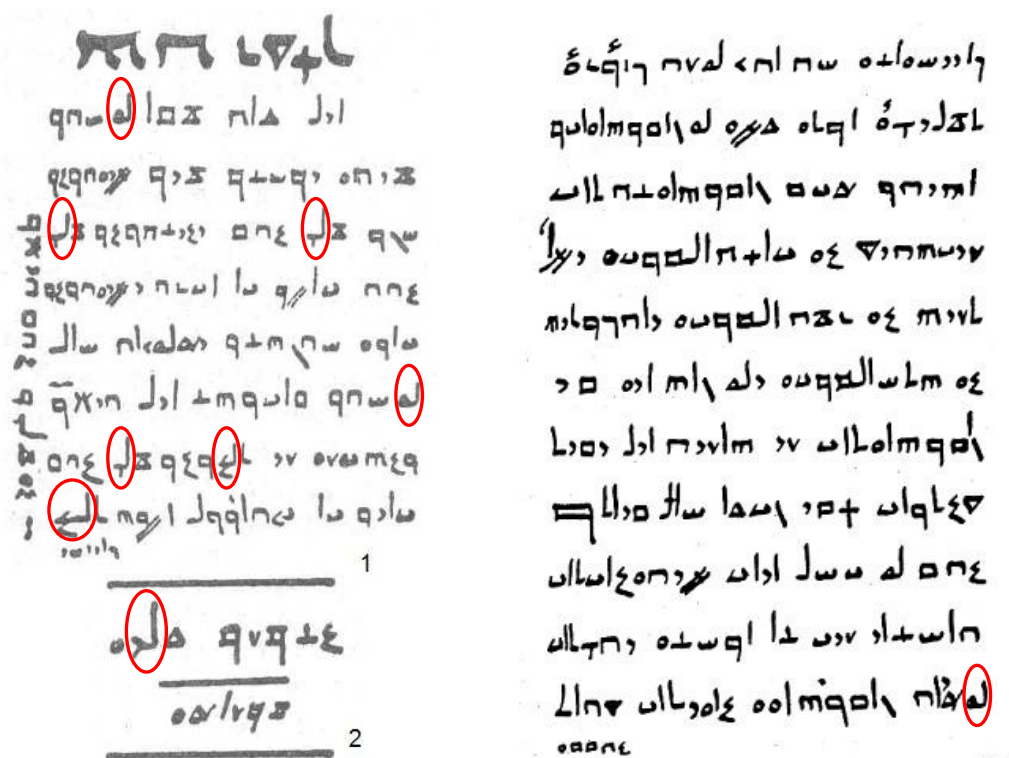


Figure 1: Yezidi manuscript samples (Marie 1911). Ligatures are encircled in red.



Figure 2: Yezidi manuscript sample.

The flower-shaped symbol in the first and last lines is a decorative element for aesthetic purposes only

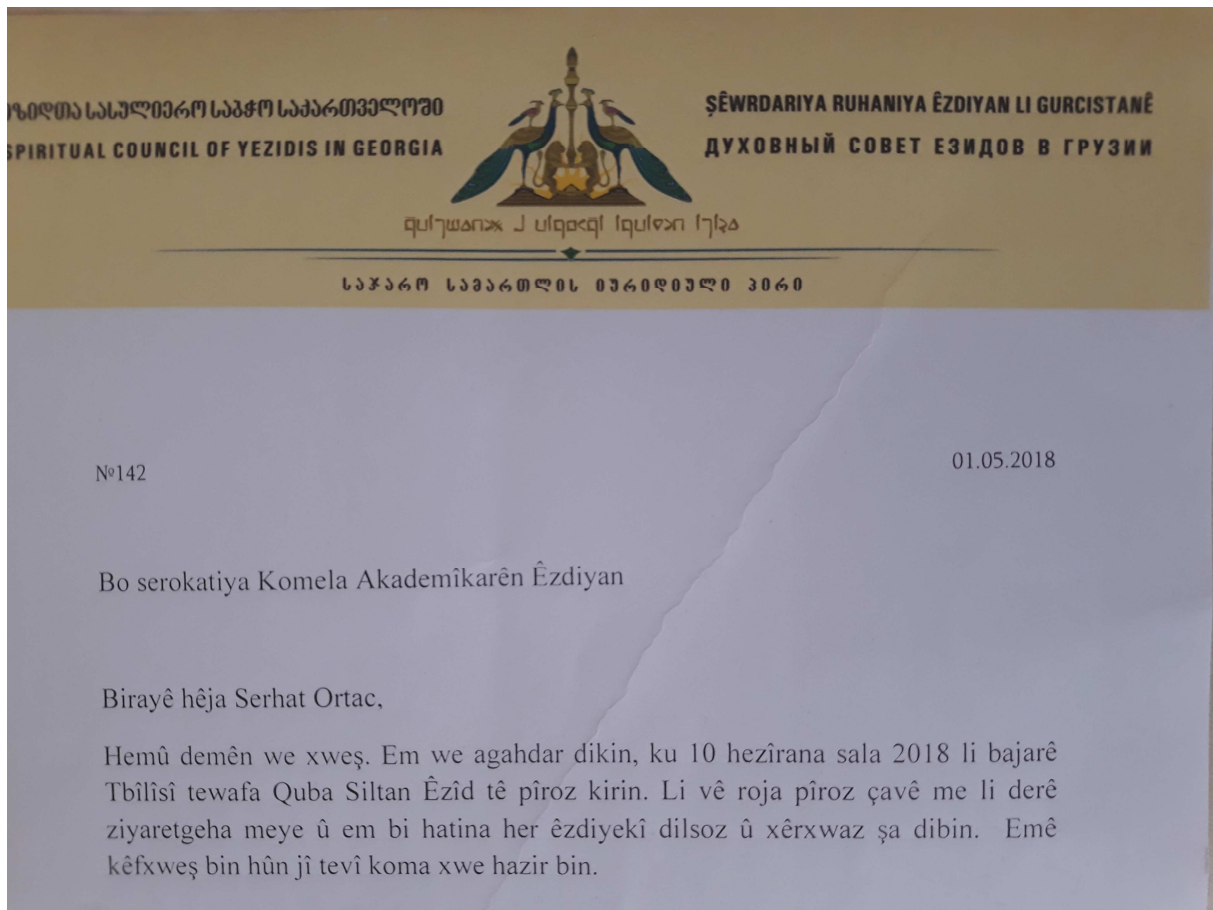


Figure 3: Letterhead of the Spiritual Council of Yezidis in Georgia.



Figure 4: Yezidi inscription on the walls of the Yezidi temple in Tbilisi.



Figure 5: Yezidi inscription in the Yezidi temple in Tbilisi.

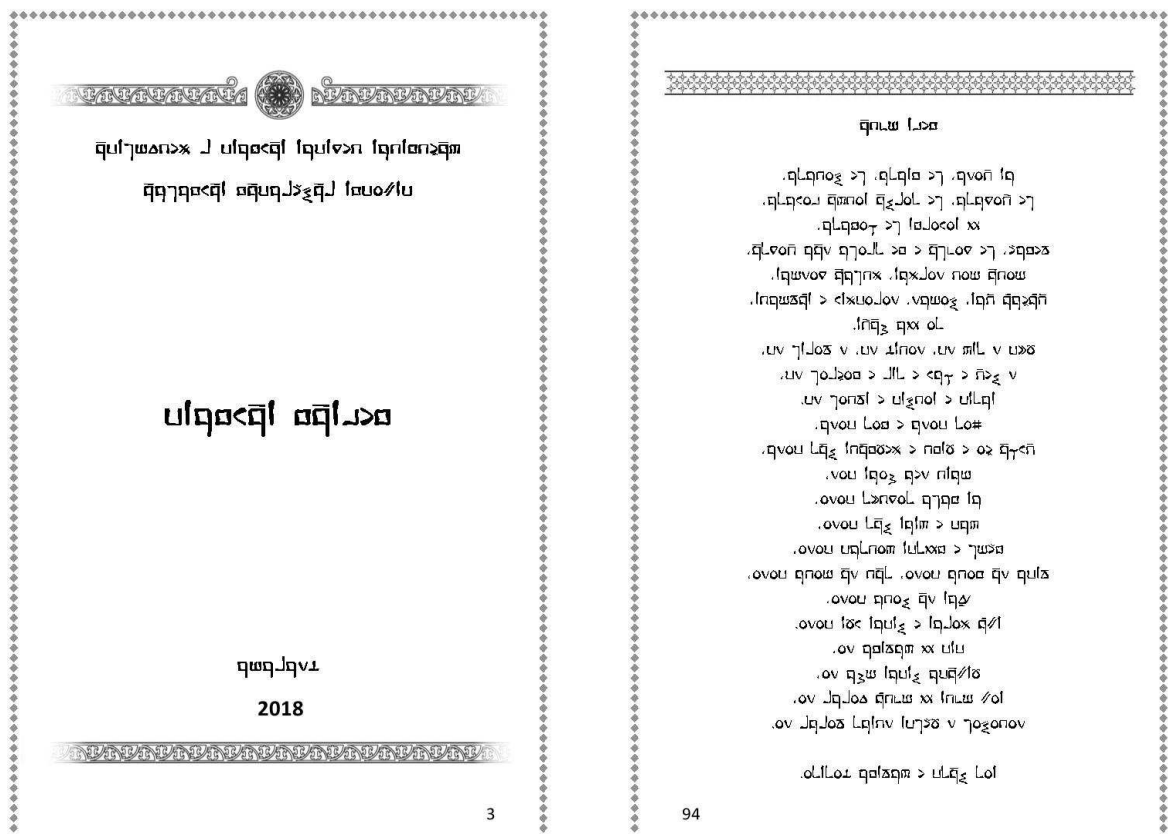


Figure 6: Title page and a sample page from the book of prayers *Dua'yê Êzdiyan* (2018).

בגד 759

[illegible]

Figure 7: Samples from the book of prayers *Dua'yêd Êzdiyan* (2018) in Yezidi.

XVII^e L'alphabet yézidite et sa transcription.

ا	ا	,a	م	ش	s	ن	n
ب	ب	b	+	ص	s	ه	h
ت	ت	t	x	ض	d	و	w,u
ث	ث	g	ط	ت	t	ي	y
ج	ج	j	ظ	ز	z	پ	p
ح	ح	h	ع	ء	e	چ	c
خ	خ	x	غ	ر	r	ث	z
د	د	d	ف	f	f	گ	g
ذ	ذ	z	ق	k	k	ق	v
ر	ر	r	ک	k	k	ق	w
ز	ز	z	ل	ل	l		
س	س	s	م	m	m		

Figure 8: Yezidi alphabet from Marie (1911).

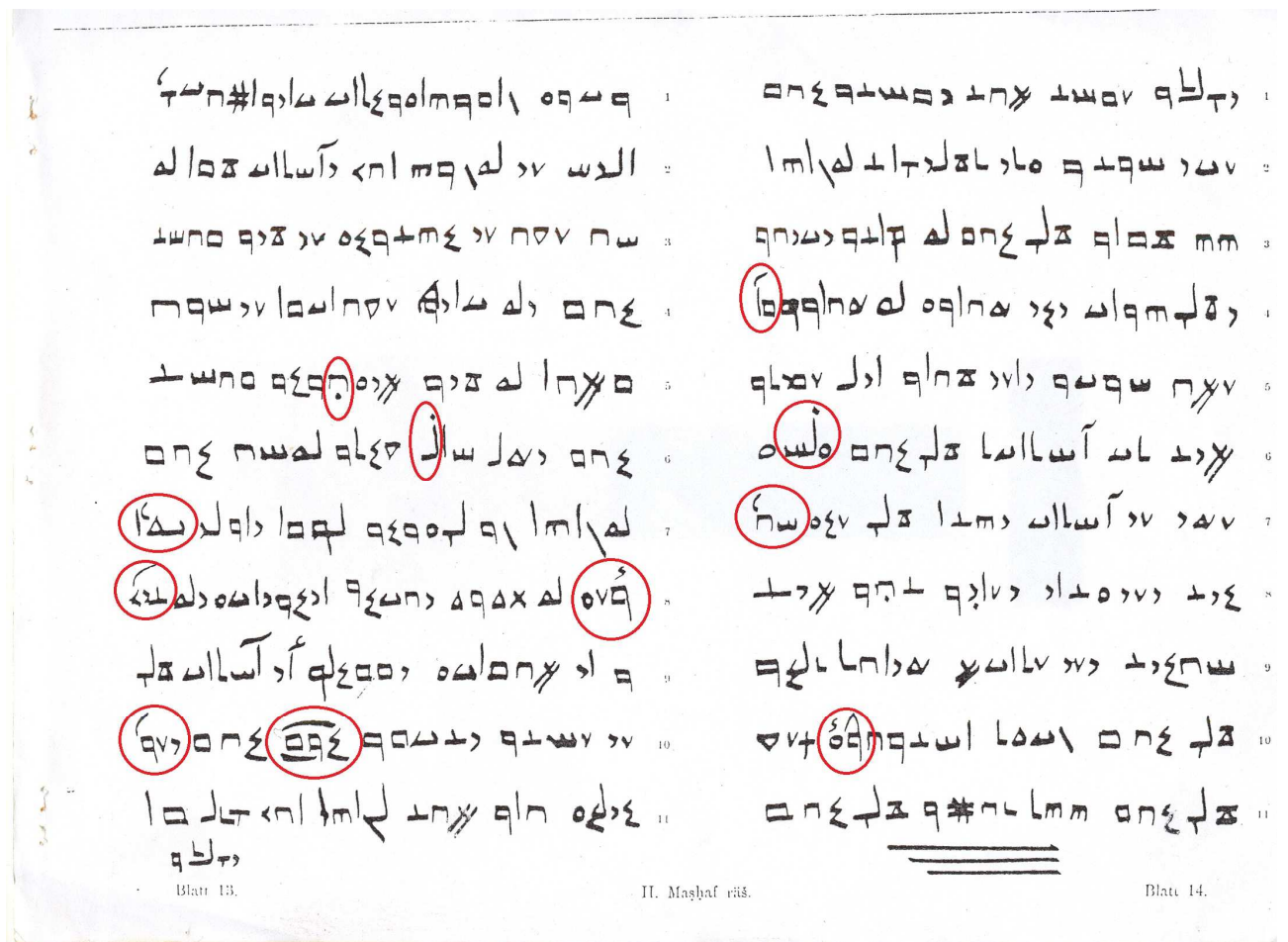


Figure 9: Diacritical marks in *Masḥaf Raš*. Encircled are word fragments with the following marks:
 right page, top to bottom, ELIF MADDA, LAM WITH DOT ABOVE, RA with hyphenation mark, YOT WITH CIRCUMFLEX ABOVE followed by ET with HAMZA above;
 left page, top to bottom and right to left within a line, RA with dot below, LAM WITH DOT ABOVE, CIM followed by the hyphenation mark, YOT with HAMZA above, ZA with hyphenation mark, decorative horizontal parentheses in *k(id)*, YOT with hyphenation mark.

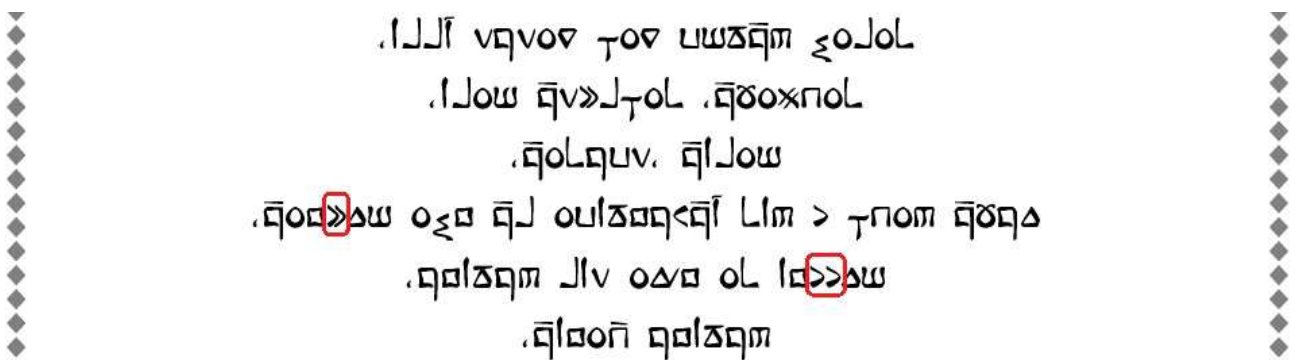


Figure 10: Sample text with both kerned and unkerne d ligature to represent letter UUM used in the same word: scûdeê "worship" / scûda "worship" in Ezāfe (izafet, Arabic *Idāfah*). This page fragment is taken from the book of prayers *Dua'yêd Êzdiyan* (2018)

	10E8	10E9	10EA	10EB
0	𐌰 10E80	𐌱 10E90	𐌲 10EA0	𐌳 10EB0
1	𐌴 10E81	𐌵 10E91	𐌶 10EA1	𐌷 10EB1
2	𐌸 10E82	𐌹 10E92	𐌺 10EA2	
3	𐌻 10E83	𐌼 10E93	𐌽 10EA3	
4	𐌾 10E84	𐌿 10E94	𐍀 10EA4	
5	𐍁 10E85	𐍂 10E95	𐍃 10EA5	
6	𐍅 10E86	𐍆 10E96	𐍇 10EA6	
7	𐍈 10E87	𐍉 10E97	𐍊 10EA7	
8	𐍋 10E88	𐍌 10E98	𐍍 10EA8	
9	𐍎 10E89	𐍏 10E99	𐍐 10EA9	
A	𐍑 10E8A	𐍒 10E9A		
B	𐍔 10E8B	𐍕 10E9B	𐍖 10EAB	
C	𐍘 10E8C	𐍙 10E9C	𐍚 10EAC	
D	𐍛 10E8D	𐍜 10E9D	𐍝 10EAD	
E	𐍞 10E8E	𐍟 10E9E		
F	𐍡 10E8F	𐍢 10E9F		

10E80	𐌰	YEZIDI LETTER ELIF
10E81	𐌴	YEZIDI LETTER BE
10E82	𐌸	YEZIDI LETTER PE
10E83	𐌻	YEZIDI LETTER PHE
10E84	𐌾	YEZIDI LETTER THE
10E85	𐍁	YEZIDI LETTER SE
10E86	𐍅	YEZIDI LETTER CIM
10E87	𐍈	YEZIDI LETTER CHIM
10E88	𐍋	YEZIDI LETTER CHHIM
10E89	𐍎	YEZIDI LETTER HHA
10E8A	𐍑	YEZIDI LETTER XA
10E8B	𐍔	YEZIDI LETTER DAL
10E8C	𐍘	YEZIDI LETTER ZAL
10E8D	𐍛	YEZIDI LETTER RA
10E8E	𐍞	YEZIDI LETTER RHA
10E8F	𐍡	YEZIDI LETTER ZA
10E90	𐌱	YEZIDI LETTER JA
10E91	𐌵	YEZIDI LETTER SIN
10E92	𐌹	YEZIDI LETTER SHIN
10E93	𐍂	YEZIDI LETTER SAD
10E94	𐍀	YEZIDI LETTER DAD
10E95	𐍃	YEZIDI LETTER TA
10E96	𐍆	YEZIDI LETTER ZE
10E97	𐍉	YEZIDI LETTER EYN
10E98	𐍊	YEZIDI LETTER XHEYN
10E99	𐍇	YEZIDI LETTER FA
10E9A	𐍒	YEZIDI LETTER VA
10E9B	𐍕	YEZIDI LETTER VA ALTERNATE FORM
10E9C	𐍙	YEZIDI LETTER QAF
10E9D	𐍜	YEZIDI LETTER KAF
10E9E	𐍟	YEZIDI LETTER KHAF
10E9F	𐍢	YEZIDI LETTER GAF
10EA0	𐌲	YEZIDI LETTER LAM
10EA1	𐌶	YEZIDI LETTER MIM
10EA2	𐌺	YEZIDI LETTER NUN
10EA3	𐌽	YEZIDI LETTER UM
10EA4	𐍀	YEZIDI LETTER WAW
10EA5	𐍃	YEZIDI LETTER OW
10EA6	𐍇	YEZIDI LETTER EW
10EA7	𐍊	YEZIDI LETTER HAY
10EA8	𐍍	YEZIDI LETTER YOT
10EA9	𐍐	YEZIDI LETTER ET
10EAB	𐍖	YEZIDI COMBINING HAMZA MARK
10EAC	𐍚	YEZIDI COMBINING MADDA MARK
10EAD	𐍝	YEZIDI HYPHENATION MARK
10EB0	𐌳	YEZIDI LETTER LAM WITH DOT ABOVE
10EB1	𐌷	YEZIDI LETTER YOT WITH CIRCUMFLEX ABOVE

**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:	<i>Proposal for encoding the Yezidi script in the SMP of the UCS</i>		
2. Requester's name:	<i>Andrij Rovenchak, Dimitri Pirbari, and Erdal Karaca</i>		
3. Requester type (Member body/Liaison/Individual contribution):	<i>Liaison (SEI)</i>		
4. Submission date:	<i>2019-03-11</i>		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:			<i>YES</i>
(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):			<i>YES</i>
Proposed name of script:	<i>Yezidi</i>		
b. The proposal is for addition of character(s) to an existing block:			
Name of the existing block:			
2. Number of characters in proposal:			<i>47</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary ☒ B.1-Specialized (small collection) ☐ B.2-Specialized (large collection) ☐			
C-Major extinct ☐ D-Attested extinct ☐ E-Minor extinct ☐			
F-Archaic Hieroglyphic or Ideographic ☐ G-Obscure or questionable usage symbols ☐			
4. Is a repertoire including character names provided?			<i>YES</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?			<i>YES</i>
b. Are the character shapes attached in a legible form suitable for review?			<i>YES</i>
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Dimitri Pirbar via Andrij Rovenchak</i>		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Dimitri Pirbar via, Andrij Rovenchak</i>		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			<i>YES</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?			<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)?			<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 <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.

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?	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.)?	YES
If YES, with whom? <i>representatives of the community, including the authors: D.Pirbani & E.Karaca</i>	
If YES, available relevant documents: YES	
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>see page 1</i>	
4. The context of use for the proposed characters (type of use; common or rare)	rare
Reference:	
5. Are the proposed characters in current use by the user community?	YES
If YES, where? Reference: <i>by members of the Yezidi community in Georgia, Armenia, Iraq, Germany.</i>	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	NO
If YES, is a rationale provided?	
If YES, reference:	
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?	YES
If YES, is a rationale for its inclusion provided?	
If YES, reference: <i>see Section 7 of the Proposal, page 2</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?	NO
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?	YES
If YES, is a rationale for such use provided?	
If YES, reference: <i>see Section 7 of the Proposal, page 2</i>	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference: NO	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	NO
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	NO
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	