

ANDRIJ ROVENCHAK

IVAN FRANKO NATIONAL UNIVERSITY OF LVIV, UKRAINE,
andrij.rovenchak@lnu.edu.ua

DIMITRI PIRBARI

G. TSERETELI INSTITUTE OF ORIENTAL STUDIES OF ILIA STATE UNIVERSITY
dimapir@gmail.com

ERDAL KARACA

INDEPENDENT RESEARCHER, OLDENBURG, GERMANY
erdal.karaca.de@gmail.com

YEZIDI ALPHABET: THE WAY TO UNICODE

Abstract

The Yezidi (Yazidi) script gained widespread recognition in 1911 when Anastase Marie Al-Karmali published two ancient manuscripts: *Maṣḥaf Raš* (Black Scroll) and *Ktēbī Jalweh* (Book of Revelation). The exact time of their creation remains unclear. These manuscripts were written in an original abjad comprising 33 letters. Despite its significance, this early Yezidi script version is inadequately documented within the literature on the history of writing. Furthermore, information regarding its prolonged usage gap – spanning between the aforementioned classical manuscripts and the early 2000s when attempts to revive the Yezidi script commenced – is rather scarce.

It is important to highlight that, for various reasons, the Yezidi clergy do not acknowledge the texts of *Maṣḥaf Raš* and *Ktēbī Jalweh* as authentic sources of faith. Nevertheless, the script itself is recognized and has recently seen utilization among certain Yezidi groups.

A renewed version of the Yezidi script, constituting a full-fledged alphabet, was introduced in 2013 by representatives of the Yezidi community in Georgia. This new alphabet encompasses 42 characters, denoting both consonants and vowels. The classical consonantal alphabet was extended by certain modifications of the existing letters, in particular, by adding diacritical marks and changing some phonetic meanings. Alongside letters, we provide an extensive exploration of Yezidi writing's characteristics, with particular emphasis on aspects such as numbers, punctuation, and diacritical marks within both classical and contemporary contexts. Throughout the article, we incorporate examples from classical manuscripts and modern instances of the Yezidi alphabet's application, offering illustrative support to the text.

The article's content is based on Unicode proposals prepared by the authors between 2018 and 2019. The insights garnered from interactions with the Unicode Technical Committee, as reflected in the article, grant readers an enhanced understanding of the intricacies surrounding the encoding of novel writing systems. Consequently, the Yezidi alphabet was integrated into the Unicode standard, version 13.0, released in March 2020.

Keywords: writing, Yezidis, Yezidi alphabet, diacritical marks, Unicode.

1. Introduction

Work began in April 2018 to include the Yezidi alphabet in Unicode, an international encoding standard that 'provides the basis for processing, storage and interchange of text data in any language in all modern software and information technology protocols' (Unicode FAQ n. d.). Discussions with Unicode representatives, aimed at clarifying certain features of the Yezidi script and addressing technical points, lasted a little more than a year until a positive decision was reached in early May 2019. The alphabet

eventually became part of the Unicode standard, specifically in version 13.0, released in March 2020, and is among the 154 writing systems represented there (Unicode® 13.0.0 n. d.).

The Yezidi writing system is relatively sparsely covered in classical literature on the history of writing. A more detailed analysis can be found in the works of David Diringer (Diringer 1947, 296–298) and Čestmír Loukotka (Loukotka 1946, 95). Hans Jensen briefly mentioned it (Jensen 1969, 325), and Ignaz Gelb also provided a table of signs (Gelb 1963, 142, 144). However, the works of Johannes Friedrich (1966), Thomas Gamkrelidze (1990), and Florian Coulmas (2004), as well as the encyclopedia *Sekai moji jiten* (Kōno, Chino, and Nishida 2001), make no mention of Yezidi writing.

This article will provide a concise overview of the history and current status of the Yezidi alphabet, as well as the interaction process with the Unicode Technical Committee (UTC). The information presented primarily relies on the document by Rovenchak, Pirbari, and Karaca (2019) and supplementary materials we prepared for Unicode. The article is organized as follows: Section 2 offers historical insights into Yezidi's writing; Section 3 delves into an analysis of the writing structure; the intricacies of certain elements – diacritics, numbers, and punctuation – are discussed in more detail in Section 4. The final Section provides brief summaries.

2. Historical Background

The earliest instances of Yezidi script application are evident in two manuscripts: *Maṣḥaf Raš* (The Black Scroll) and *Ktēbī Jalweh* (The Book of Revelation), which were initially unveiled by Anastas Marie Al-Karmali in 1911 (Marie 1911). It is worth noting that preceding these, there existed manuscripts in Arabic script bearing the same titles (Joseph 1909).

These manuscripts were penned onto sheets of delicate parchment derived from treated gazelle skin. The pages of *Ktēbī Jalweh* encompass 16 lines of text, while *Maṣḥaf Raš* comprises 11 lines. The former contains a monologue by *Tawisī Melek*, the principal angel within the Yezidi religion. *Maṣḥaf Raš*, on the other hand, narrates a cosmogony that transcends conventional Yezidi beliefs (Amoev 1999; Pirbari and Amoev 2013). The language employed in *Maṣḥaf Raš* is comparably simpler than that of *Ktēbī Jalweh*.

Regrettably, both the time of manuscript creation and their authorship remain shrouded in uncertainty, prompting diverse viewpoints on the matter. Plausible origins span from the 12th to 14th centuries (Marie 1911; Omarkhali 2017, 20), with a stronger likelihood of leaning towards a later period, possibly the 19th century (Mingana 1916; Pirbari and Amoev 2013). However, this does not negate the existence of Yezidi's writing. Yezidis hold the historical belief in the existence of sacred manuscripts titled *Maṣḥaf Raš* and *Ktēbī Jalweh*, albeit their original versions are presumed lost. Subsequently, copies transcribed in a distinct Yezidi alphabet emerged, though regrettably, their content suffered distortions. As a result, a segment of the Yezidi clergy acknowledges the script itself yet does not endorse the manuscripts' content as authentic sources of Yezidi faith. Thus, *Maṣḥaf Raš* and *Ktēbī Jalweh* can be considered Yezidi apocryphal literature. Various reasons underlie their non-recognition of the manuscripts' authenticity (Rodziewicz 2022). The origin and substance of these manuscripts, however, form an independent subject of studies, exceeding the confines of our present article's scope.

These classical manuscripts are composed in the Sorani language (Central Kurdish, ISO 693-3 code: ckb), related to the modern Yezidi language. The latter is referred to by a part of Yezidis as *Êzdikî*. From a linguistic standpoint, this classification places it within the Indo-European > Indo-Iranian > Iranian > West Iranian > North-Western Iranian > Kurdish languages > Kurmanji (ISO 693-3 code: kmr) (Eberhard, Simons, and Fennig 2021). Presently, the language employs the Latin, Cyrillic, and Arabic scripts for writing.

Notably, the sacred Yezidi manuscripts *mišūr*, dating back to the 13th-14th centuries, are inscribed in Arabic (Pirbari, Mossaki, and Yezdin 2020). In the late 2000s, efforts to rejuvenate writing practices were undertaken in Iraq. This endeavor included the publication of a Yezidi-Arabic dictionary. Regrettably, due to the challenging political and social circumstances endured by Iraq's Yezidi community, this initiative did not progress further. It is also important to underline that the Yezidi alphabet was not substantially employed for an extended duration, spanning from the era of the aforementioned two manuscripts until the early 2000s.

In 2013, the Yezidi Spiritual Council of Georgia decided to revive the Yezidi alphabet and use it for writing prayers, for sacred books, on the letterhead of the organization (see Fig. 1), in Yezidi heraldry, etc. For this purpose, Dimitri Pirbari and Kerim Amoev modernized the classical Yezidi alphabet and adapted it to the phonetic features of the modern Yezidi language. This choice is justified by the special

role of classical manuscripts in Yezidi's identity and takes into account the proximity of Kurmanji and Sorani. The book *Yezidi Script* was published (Pirbari and Amoev 2013), which describes each letter and its phonetic meaning in detail. Today, the Yezidi script is used by the clergy in the Yezidi temple in Tbilisi. The names of saints are also written in this alphabet on the walls of the temple (see Fig. 2). The book of prayers, *Dua 'yed Êzdiyan*, in the Yezidi script was recently published (see Fig. 3).

საპატრიარქოს ეპიშოპოს სასულიერო საბჭო
YEZIDI SPIRITUAL COUNCIL OF GEORGIA



CIVATA RUHANIYA ÊZDIYAN YA GURCISTANÊ
ДУХОВНЫЙ СОВЕТ ЕЗИДОВ ГРУЗИИ

გაქათანი ✱ იკ ასეცქი იკუსანი რქა

0163, თბილისი, ვიქტორ კუპრაძის №60 Tel.: (+995 32) 219 08 33 0163, Georgia, Tbilisi, Viktor Kupradze str. №60
yezidi.ge@gmail.com
www.yezidi.ge

Fig. 1: Letterhead of the Spiritual Council of Yezidis in Georgia.



Fig. 2: Yezidi inscriptions on the walls of the Yezidi temple in Tbilisi.

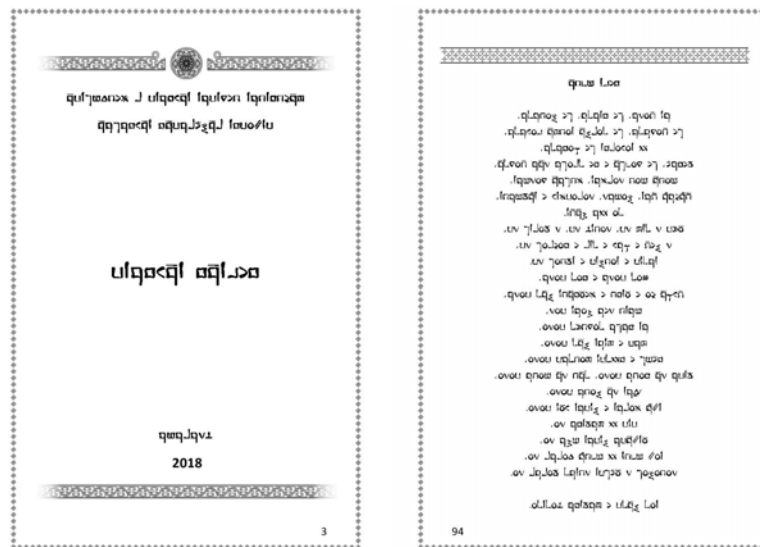


Fig. 3: Title and sample page from *Dua'yêd Êzdiyan* prayer book (2018).

3. Structure of Writing

Yezidi's writing is alphabetic, the direction of writing is horizontal, from right to left, and lines are placed sequentially from top to bottom. In its classical form, the Yezidi script encompasses 33 letters, and its ancestral inspiration appears to stem from the Perso-Arabic lineage (Diringer 1947 with reference to Furlani 1930; 1932). This ancestral connection is evident in the letter sequence (as depicted in Fig. 4). According to the classification of Peter Daniels (1990), it is an *abjad* (from the Arabic أبجد), indicating a script lacking characters to represent all vowels.¹

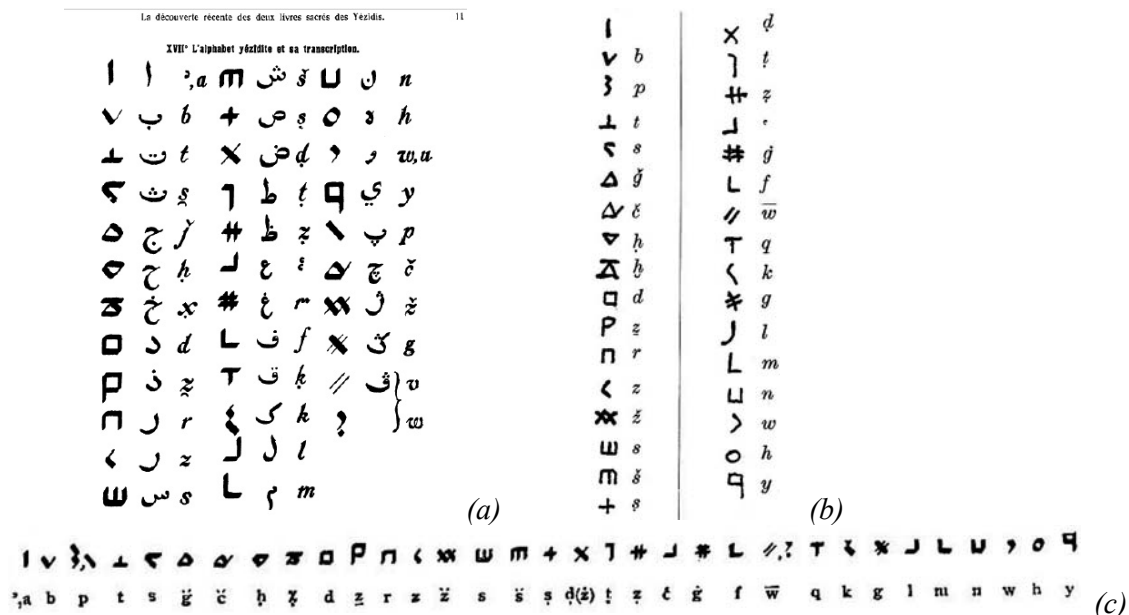


Fig. 4: Classical Yezidi *abjad* according to different authors
 (a) — Marie (1911), (b) — Jensen (1969, 325), (c) — Diringer (1947, 299).

¹ This type of writing used to be referred to as a consonantal alphabet and might also be called a syllabic script. However, it is different from a syllabary proper, as exemplified by scripts like Hiragana, Cherokee, or Vai, where the shapes of symbols corresponding to the same initial consonant but different vowels are unrelated. It also differs from an *abugida* or alphasyllabary, as exemplified by Devanagari and Ge'ez, where different vowels for the same initial consonant are indicated by more or less regular shape changes.

A certain similarity can be observed in the forms of Yezidi and Arabic letters ‘ / a, h, u / w (Diringer 1947, 298), as well as s, n, and l, and to some extent, b. Additionally, a resemblance to the Arabic Kufic script can be noted, which tips the historical scales in favor of the Yezidi script having an older origin. Generally, however, the depiction of Yezidi letters displays distinct and simple geometric features, setting them apart from Arabic script. Interestingly, the utilization of a simple vertical bar for elif (‘ / a) may find further justification considering the frequent appearance of this letter in texts. A similar observation can be made in the context of Nko, an alphabet created in 1949 for the Maninka language in Guinea (Rovenchak 2011; 2019).

Unlike Arabic writing, Yezidi letters are rendered individually, and their shapes remain constant regardless of their position within a word. Nevertheless, it is evident that in handwritten versions, letters can be combined, akin to other alphabetic scripts such as Latin, Cyrillic, Hebrew, Armenian, and Georgian. An illustrative example is provided in Fig. 5.

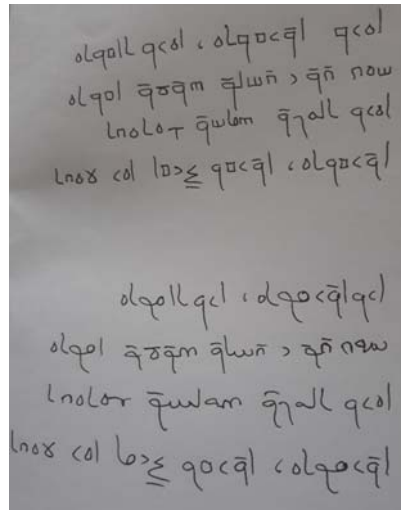


Fig. 5: A sample of a modern Yezidi handwritten text. The “printed” letters are shown in the upper part; the cursive version is shown in the lower part.

Certain ligatures emerge in classical manuscripts, for instance: 𐌗𐌕 (l-t), 𐌗𐌘 (l-c), 𐌗𐌙 (l-q), 𐌗𐌚 (l-k), 𐌗𐌛 (l-y), 𐌗𐌜 (l-v),² as depicted in Fig. 6. However, in the contemporary version of the alphabet, ligatures are no longer employed.

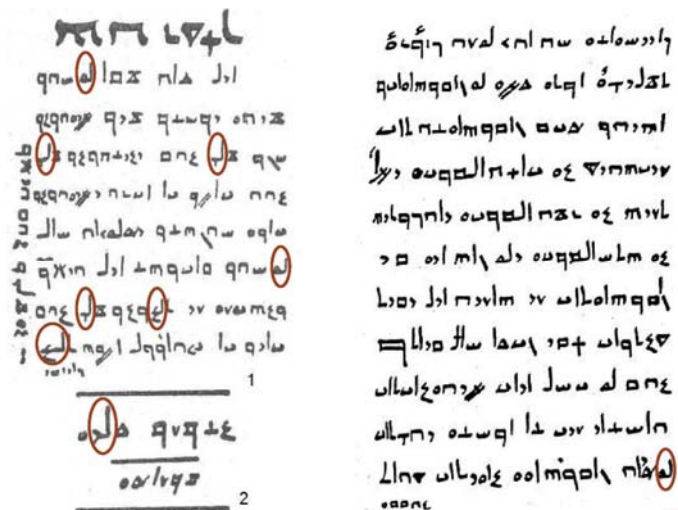


Fig. 6: Fragment of the Maṣḥaf Raš manuscript (Marie 1911); ligatures are circled in red.

² The JG Yezidi font by Jason Glavy is used here for the ligatures. The rest of the letters of the Yezidi alphabet in the text are typed in the PirbaryiHistory font, the author of which is Dimitri Pirbari.

Table 1: New Yezidi Alphabet and its Representation in Other Scripts

Yezidi	Arabic	Latin	Cyrillic	IPA (approximate)	Letter name
ا	ا	A a	А а	a	<i>elif</i>
ب	ب	B b	Б б	b	<i>be</i>
پ	پ	P p	П п	p	<i>pe</i>
په	په	P' p'	П' п'	p ^h	<i>p'e</i>
ت	ت	T' t'	Т' т'	t ^h	<i>t'e</i>
س	س	S		θ ~ s	<i>se</i>
سج	سج	C c	Ц ц	ɕ	<i>cim</i>
سچ	سچ	Ç ç	Ч ч	tʃ	<i>çim</i>
سچه	سچه	Ç' ç'	Ч' ч'	tʃ ^h	<i>ç'im</i>
ح	ح	H' h'	Һ' һ'	h	<i>h'a</i>
خ	خ	X x	Х х	x	<i>xa</i>
د	د	D d	Д д	d	<i>dal</i>
ذ	ذ	z		ð ~ z ^f	<i>zal</i>
ر	ر	R r	Р р	r	<i>ra</i>
ره	ره	R' r'	Р' р'	r	<i>r'a</i>
ز	ز	Z z	З з	z	<i>za</i>
ژ	ژ	J j	Ж ж	ʒ	<i>ja</i>
س	س	S s	С с	s	<i>sin</i>
ش	ش	Ş ş	Ш ш	ʃ	<i>şin</i>
ص	ص	ş		s ^f	<i>şad</i>
ض	ض	ḍ		d ^f	<i>ḍad</i>
ط	ط	T t	Т т	t	<i>ta</i>
ظ	ظ	z'		z ^y	<i>z'e</i>
ع	ع	' / E' e'	Ә' ә'	ʔ	<i>'eyn</i>
غ	غ	X' x'	Г' г'	ɣ	<i>x'eyn</i>
ف	ف	F f	Ф ф	f	<i>fa</i>
ف	ف	V v	В в	v	<i>va</i>
ف	ف	V v	В в	v	<i>va (variant)</i>
ق	ق	Q q	Q q	q	<i>qaf</i>
ك	ك	K k	К к	k	<i>kaf</i>
كه	كه	K' k'	К' к'	k ^h	<i>k'af</i>
گ	گ	G g	Г г	g	<i>gaf</i>
ل	ل	L l	Л л	l	<i>lam</i>
م	م	M m	М м	m	<i>mim</i>
ن	ن	N n	Н н	n	<i>nun</i>
و	و	U u	Ӯ ö	u	<i>um</i>
وو	وو	Û û	У у	u:	<i>uum (see Note 1)</i>

2. On the left page, from top to bottom and right to left within the line: *ra* with a dot below, *lam* with a dot above, *cim* with a hyphen, *yot-hamza*, *za* with a hyphen, decorative (?) horizontal brackets in *k(id)*, *yot* with a hyphen.

At least one diacritical mark carries a well-established meaning: the dot over *lam* < ڤ > altered the phonetic value from soft [l] to hard [t].

The meanings of some diacritical marks in classical manuscripts are not always evident. The significance of the horizontal brackets in Fig. 9 remains uncertain, and for instance, the dot between the letters **هه م ا ق ا ل** in the lower right part of Fig. 6 might be a random artifact. The circumflex over *yot* < ڤ > could indicate its incorrect usage (akin to “crossing out”) (Marie 1911), or it could modify the sound from [i] to [ê]. The dot under the letter *ra* might correspond to a “weakened” pronunciation, “le *r* très doux” (Marie 1911), probably / *r* / as opposed to /*r*/.

The presence of this diacritical mark also sparked a debate about whether the letter *va* < ڤ > in the new alphabet can be interpreted as *um* with a diacritic. However, this interpretation lacks sufficient grounds, so *va* was retained in the final code table.

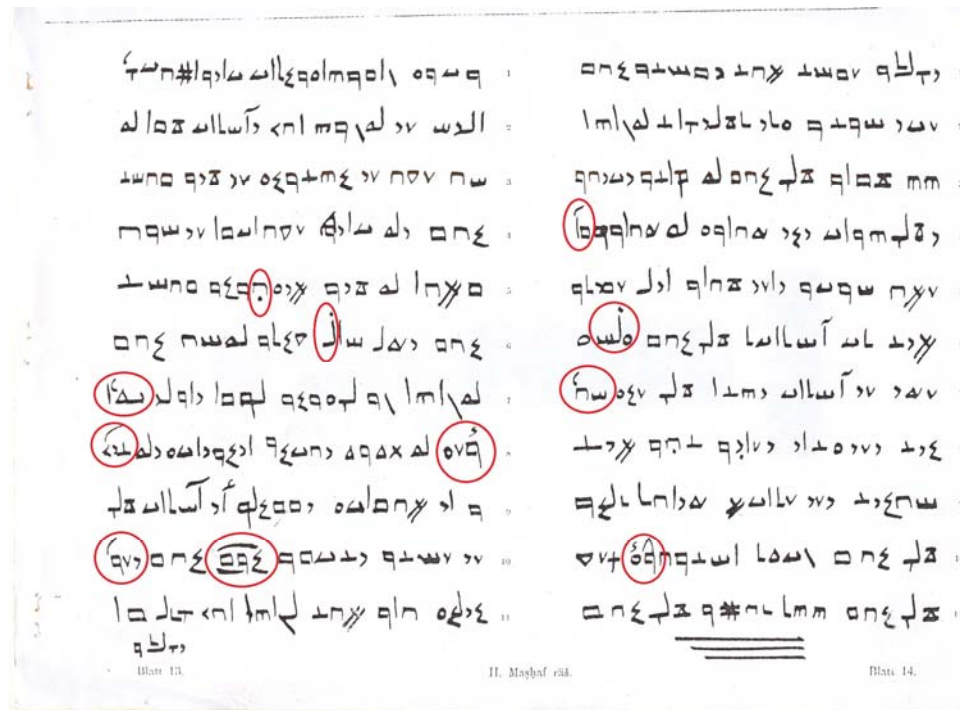


Fig. 9: Diacritical marks in *Maṣḥaf Raš*.
For a detailed discussion, see the corresponding paragraphs in the text.

It should be noted that during discussions with the UTC, a decision was made to exclude, at least in the current stage, the incorporation of diacritical marks from classical manuscripts into the standard. The inclusion was limited to only two letters with diacritics, < ڤ > and < ڤ >.

The hyphen (presented in the form of an inverted apostrophe < ‘ >) was placed twice in the classical manuscripts: at the end of a line where a word was split and at the beginning of the next line before the second part of the word (refer to Fig. 10). In modern orthography, hyphens are used solely at the end of a line.

Furthermore, it is worth noting that the same sign in Fig. 9 can be observed in a non-final position, **هه م ا ق ا ل**. This use might correspond to the duplication of the preceding letter (*cim*). However, such a function is not accounted for in modern writing.

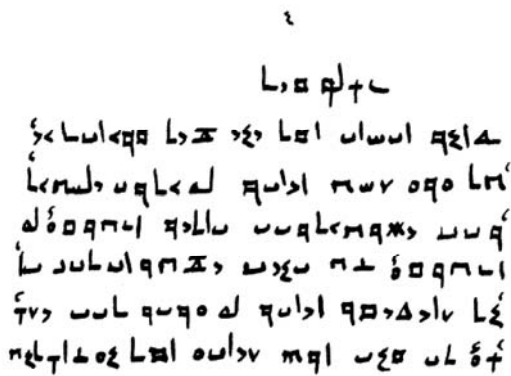


Fig. 10: A fragment of the text of the *Ktēbī Jalweh* manuscript, where most of the lines contain hyphenated words. You can also see its number 4 (Arabic ٤) at the top of the page.

To the best of our knowledge, separate symbols for numbers were never used in the Yezidi script. In ancient manuscripts, (Eastern) Arabic numerals (٠, ١, ٢, ٣, ٤, ٥, ٦, ٧, ٨, ٩) were employed, cf. Marie’s article (Marie 1911), while in contemporary texts, the traditional 0...9 are used.

Punctuation marks align with Arabic conventions: comma (,), semicolon (;), and question mark (؟). Periods, colons, and exclamation marks retain their traditional forms (. , : , !) because they are direction-independent.

5. Conclusions

We have described the characteristics of Yezidi writing, with particular focus on the new version of the alphabet proposed in 2013 by members of the Yezidi community in Georgia. This modernized alphabet comprises 42 letters denoting both consonants and vowels. It emerged from the abjad, which consisted of 33 letters, and underwent specific modifications, notably the inclusion of diacritical marks.

Alongside letters, we have meticulously examined other components of writing: numerals, punctuation, and diacritical marks in classical and contemporary writing. Furthermore, we have highlighted the inquiries raised by Unicode representatives subsequent to the submission of a proposal to incorporate the Yezidi alphabet into this international standard. Information concerning the interaction with the Unicode Technical Committee serves to enhance comprehension of the procedures entailed in the development of novel writing systems there.

We also expect that technical work will be completed in the near future, which will make full use of the Yezidi alphabet available on computers and other devices running under various operating systems (Windows, Android, Unix, MacOS).

As a follow-up to the Unicode coverage, Google created a font for the Yezidi script as part of the Noto font collection (Noto Serif Yezidi n. d.). However, this event was overshadowed by the font’s description, which initially labeled the language as “Kurdish”, a very sensitive issue for the Yezidis, who consider such a description to be very vulnerable. Following a request, the description was corrected to “Yezidi, Kurmanji”, but this change did not last long; it was subsequently modified to “Kurdish, Yezidi”. Unfortunately, despite further requests from the Yezidi community to remove “Kurdish” from the description, their efforts did not succeed.

We hope that the material presented in this article will captivate individuals interested in Yezidi culture and history, writing-related matters in a broader context, and technical aspects that facilitate information exchange in the contemporary digital world.

Acknowledgment

This study 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.

Our special thanks go to Deborah Anderson for her constant attention to the work on the standardization of the Yezidi alphabet and her organizational support.

References:

- Amoev, Kerim. 1999. *Ezidebis ts'minda ts'ignebi (Yezidi Holy Books)*. Tbilisi.
- Coulmas, Florian. 2004. *The Blackwell Encyclopedia of Writing Systems*. Blackwell, Malden, MA: Blackwell.
- Daniels, Peter. 1990. "Fundamentals of Grammatology." *Journal of the American Oriental Society* 110 (4): 727-731. <https://doi.org/10.2307/602899>.
- Diringer, David. 1947. *The Alphabet: A Key to the History of Mankind*. London: Hutchinson.
- Dua'yêd Êzdiyan*. 2018. Tbilisi: Universal.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2021. *Ethnologue: Languages of the World*, 24th edition. Dallas, TX: SIL International. Online version: <http://www.ethnologue.com>.
- Friedrich, Johannes. 1966. *Geschichte der Schrift: unter besonderer Berücksichtigung ihrer geistigen Entwicklung*. Heidelberg: C. Winter.
- Furlani, Gino. 1930. "Religione dei yezidi: Testi religiosi dei yezidi." In *Testi e documenti per la storia delle religioni*, vol. III. Bologna: Zanichelli.
- Furlani, Gino. 1932. "Sui yezidi". *Rivista Degli Studi Orientali* 13 (2): 97-132. <http://www.jstor.org/stable/41863678>.
- Gamkrelidze, Thomas. 1990. *Ts'eris anbanuri sistema da dzveli qartuli damts'erloba: anbanuri ts'eris tipologia da ts'armomavloba (Alphabetic Writing and the Old Georgian Script: A Typology and Provenience of Alphabetic Writing Systems)*. Tbilisi: Tbilisi University Press.
- Gelb, Ignace J. 1963. *A Study of Writing*, revised edition. Chicago: University of Chicago Press.
- Jensen, Hans. 1969. *Die Schrift in Vergangenheit und Gegenwart*, 3rd edition. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Joseph, Isya. 1909. "Yezidi Texts". *The American Journal of Semitic Languages and Literatures* 25 (2): 111-156. <https://www.jstor.org/stable/527914>.
- Kōno, Rokurō, Eiichi Chino, and Tatsuo Nishida [河野, 六郎; 千野, 栄一; 西田, 龍雄], eds. 2001. *世界文字辞典 [Sekai moji jiten] (Scripts and Writing Systems of the World)*. [東京: 三省堂] Tokyo: Sanseido.
- Loukotka, Čestmír. 1946. *Vývoj písma*. Praha: Orbis.
- Marie, Anastase. 1911. "La découverte récente des deux livres sacrés des Yézidis". *Anthropos* 6 (1): 1-39. <https://www.jstor.org/stable/40444071>.
- Mingana, Alphonse. 1916. "Devil-Worshippers: Their Beliefs and Their Sacred Books." *The Journal of the Royal Asiatic Society of Great Britain and Ireland* 48 (July): 505-526. <https://doi.org/10.1017/S0035869X00049364>.
- Noto Serif Yezidi*. n. d. Available at: <https://fonts.google.com/noto/specimen/Noto+Serif+Yezidi>. Accessed August 01, 2023.
- Omarkhali, Khanna. 2017. *The Yezidi Religious Textual Tradition: From Oral to Written*. Wiesbaden: Harrassowitz.
- Pirbari, Dimitri, and Kerim Amoev. 2013. *Ezidskaya pis'mennost'*. Tbilisi: "Nekeri".
- Pirbari, Dimitri, Nodar Mossaki, and Mirza Sileman Yezdin. 2020. "A Yezidi Manuscript: – Mišūr of P'īr Sīnī Bahrī/P'īr Sīnī Dārānī, its Study and Critical Analysis." *Iranian Studies* 53 (1-2): 223-257. <https://doi.org/10.1080/00210862.2019.1669118>.
- Rodziewicz, Artur. 2022. "Between Orthopraxy and Orthodoxy: International Yezidi Theological Academy in Tbilisi." *Kulturní studia*, no. 1 (18): 81-116. <https://doi.org/10.7160/ks.2022.180104>.
- Rovenchak, Andrij. 2011. "Phoneme Distribution, Syllabic Structure, and Tonal Patterns in Nko Texts." *Mandenkan: Bulletin semestriel d'études linguistiques mandé* 47: 77-96.
- Rovenchak, Andrij. 2019. "Kil'kisnyī analiz system pys'ma: alfavit nko." *Ukraina moderna* 27: 139-152. <https://doi.org/10.30970/uam.2019.27.1066>.

Rovenchak, Andrij, Dimitri Pirbari, and Erdal Karaca. 2019. *L2/19-051R: Proposal for Encoding the Yezidi Script in the SMP of the UCS*. Available at: <https://www.unicode.org/L2/L2019/19051r-yezidi.pdf>. Accessed August 01, 2023.

Unicode® 13.0.0. n. d. Available at <https://www.unicode.org/versions/Unicode13.0.0/>. Accessed August 01, 2023.

Unicode FAQ. n. d. Available at <https://home.unicode.org/basic-info/faq/>. Accessed August 01, 2023.