Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal for encoding the Todhri script in the SMP of the UCS**
Source: **UC Berkeley Script Encoding Initiative (Universal Scripts Project)**
Authors: **Michael Everson**
Status: **Liaison Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2020-07-10**

1. Introduction. A number of 18th-and 19th-century orthographies exist for Albanian; one of the more significant ones given its corpus is called Todhri (['toðri]), from the nickname of the man who devised it, Dhaskal Todhri 'teacher Todhri', whose name was Theodor Haxhifilipi (§1730–1805). Unfortunately, I did not get a chance to work with the late Robert Elsie on this proposal, which never benefitted from his expertise due to his untimely death. I give below here the introduction to the script from Elsie 2017:

> The Todhri alphabet is an original writing system that was used for Albanian in the central Albanian region of Elbasan in the eighteenth and nineteenth centuries. It consists of 36 letters (plus nine ligatures and six rare graphemes). This script is attributed to one Dhaskal Todhri ("teacher Todhri"), whose full name was Theodor Haxhifilipi (ca. 1730-1805), and who stemmed from Elbasan.
>
> The biblical texts in this Todhri alphabet were discovered by the Austrian consul in Janina, Johann Georg von Hahn (1811-1869), who announced his finding in an article entitled *Bericht über die Auffindung eines uralten albanesischen Alphabetes* (Report on the Discovery of an Ancient Albanian Alphabet), submitted to the Austrian Academy of Sciences in November 1849. Hahn published a small portion of this material, in what he regarded as "the original" Albanian alphabet, in his seminal three-volume *Albanesische Studien* (Albanian Studies), Jena 1854, and regarded it, rather naively, as stemming from ancient Phoenician script. On the author and origin of the material in Todhri script, Hahn explains:

>> *"As far as I have been able to ascertain, this alphabet seems to be native only to the town of Elbas an. It is unknown in Durres, Shkodra and in all of Toskeria (Epirus), but it seems to have been understood and used a bit in the neighbouring town of Berat to the south.*
>>
>> *In Elbasan, usage of this alphabet is not limited to ecclesiastical texts. It is used there by several merchants for correspondence with their compatriots abroad. Some of them, indeed, keep their accounts in this script, but I doubt whether there are more than fifty people who know this script nowadays.*
>>
>> *This alphabet is traditionally attributed to a certain Theodor, a teacher at the Greek school there, who was not only a town preacher but also a well-educated man. It cannot be ascertained, however, whether he invented the alphabet or was simply the first person to bring it to Elbasan. He is said to have translated into Albanian not only the New Testament, but also the Old Testament (from the Septuagint [Greek translation} because he did not know Hebrew). Unfortunately all of his writings were burnt by his relatives during a major outbreak of the plague, out of fear of infection. Judging from*

the age of his living relatives, he probably died around the end of the last century. He had studied in Moschopolis [Voskopoja], ten hours to the east of Berat. [...]

I searched in vain in Elbasan for manuscripts in this Albanian alphabet. All I could find were two copybooks, one of seven pages and the other of twelve pages in quarto; fragments of a Gheg translation of the Greek Horologion; and a Gheg translation of a fragment of the Gospel of St John written on a folio sheet. The man who wrote the older of these works in his youth was still alive. As such, none of these manuscripts was over fifty years old. "

Todhri stemmed from a family of Orthodox silversmiths. For periods in the second half of the eighteenth century, he apparently kept the register of the Orthodox Church of Saint Mary's in the fortress of Elbasan, though he was not a clergyman himself. In these records, he called himself Theodor, son of Haxhi, or Theodor Dhaskalos (teacher). He seems to have died around 1805.

After Hahn, the study of Todhri script was taken up by Leopold Geitler (1847-1885), professor at the University of Zagreb, in his work *Die albanesischen und slavischen Schriften* (Albanian and Slavic Scripts), Vienna 1883, who regarded the Todhri alphabet as derived from Greek and Roman cursive. He had access to the copybooks that Hahn had found in Elbasan, and in 1883 he deposited them in the Imperial Library, now the Austrian National Library, in Vienna. The thirty sheets (46 pages) are still preserved in the manuscript division of the Library. This manuscript, kindly made available by the manuscript division of the said library, forms the basis of this publication. Other texts in Todhri script, of which there may have been many, seem, tragically, to have been lost.

In 1901, the Albanian linguist Gjergj Pekmezi (1872-1938), in his *Vorläufiger Bericht über das Studium des albanischen Dialektes van Elbasan* (Preliminary Report on the Study of the Albanian Dialect of Elbasan) sent to the Austrian Academy of Science, reported that he had come across texts in Todhri script on his journey to Elbasan in 1900:

"It was quite by coincidence that during my stay in Elbasan, an over eighty-year-old monk called Parthenius came to see me for health reasons from the Monastery of St. John Vladimir which is situated an hour away from Elbasan. He is no doubt the oldest man alive who knows how to write in this alphabet. He told me that he had learned it from a student of the Theodor who was said to have invented it. This student, called Camblak, succeeded Theodor as a teacher in Elbasan and later became Hegumen [abbot] of the said monastery. According to him, Theodor invented the script and brought it with him from Moschopolis where he had studied. The script is now dying out. It was pushed aside by the Kristoforidhi alphabet, whereas now it is the Fashëri alphabet that is being used. Very few people (some merchants and a woman who gave me a sample of her writing) still use the old script for their letters. Church hymns can also be found written in it. Not too long ago, it was being used by Demetrios Andreas Pina, a native of Elbasan and one-time bishop of central Albania. [.. .}

Parthenius told me that there was a chest in the church choir of the monastery with some unused church utensils, including books and manuscripts, in it. He thought there might still be something of Theodor's there, even though, according to Hahn, almost everything was burnt. I took up his suggestion and found the chest, but it was in a terrible state of neglect. There and in the side altar to the left, there were not only sheets of Greek parchment and hymnbooks, but also a few pages and copybooks written in this Elbasan script. They seemed to stem from various ages because, among them, were some of Pina 's sermons.

There were three copybooks, containing 6-20 sheets, of the size of those discovered by Hahn, which were probably the oldest - but just how old is difficult to say. They were written in various hands. One of them with some church hymns probably stemmed from

the eighteenth century. A second one contained some ancient Greek fables with interlinear modern Greek and Albanian translations. There were also pages from the New Testament in other handwriting. The third copybook and various separate sheets were much the same. The most interesting of all was a rather dirty page of thick paper, torn in the middle, that seemed to contain the oldest Albanian fragment. It was most certainly from the eighteenth century, and Parthenius seemed to believe it was written by Theodor. In the corner of the front page was the number 55 and on the back page was 56, an indication that it derived from a larger body of writing. The contents also made this clear - two sections from the First Epistle of Paul to the Romans. [.. .]"

Together with this information, Pekmezi published two facsimiles of the aforementioned manuscripts in Todhri script.

In 1918, the first six pages of the manuscript of the Divine Liturgy of Saint John Chrysostom were then published by the Elbasan scholar Lef Nosi (1877-1946). Nosi's substantial archives and collections were confiscated by the communists after his execution and some of the material found its way into the state archives. It was the Elbasan scholar of the communist period, Dhimitër Shuteriqi (1915- 2003), who made an inventory of Todhri writings and published several articles in the late 1940s and 1950s. He was the first to make this fascinating component of Albania's cultural heritage widely known in Albania.

Todhri script was also the subject of an article by the Slovenian scholar Rajko Nahtigal (1877-1958) published in 1923 and entitled *O elbasanskem pismu in pismenstvu na njem* (On the Elbasan Script and Writing in it), and was touched upon briefly by the German scholar Armin Hetzer (b. 1941) in 1987 in another article entitled *Armenier und Albaner* (Armenians and Albanians).

Todhri script seems to have been used sporadically in Elbasan well into the twentieth century. The family chronicle of Popaj family of Elbasan, kept from 1785 to 1933, was written in Todhri. There are also indications that the last person to write in Todhri script was a woman in Elbasan called Dona Xhuvani who died in the 1930s. However, nothing can be said for certain until we have a full inventory of all texts written in Todhri, and this is unlikely to occur anytime soon. Most of the material has vanished. For the full text, see: Robert Elsie, *Early Albanian Bible Translations in Todhri Script.* Albanian Studies, Vol. 30. (London: Centre for Albanian Studies, 2016) 162 pp.

2. Processing. Todhri is a simple, unicameral alphabetic script written from left to right horizontally. There is no real ligation, though Elsie describes some of its characters as "ligatures" but this seems more a likely description of their derivations than of a typographic quality. A number of diacritical marks are used in Todhri; though their usage and distribution has not been analysed thoroughly, they appear to be used to indicate stress, and do not appear to be script-specific. (They remind one of Greek accents.) Some of the glyphs which have been identified are: v̇ v́ v̄ ī í ī í ō ó ō̄ ō̄ ō̄ ō̄ ō̄ ō̄ ō̄ ō̄.

3. Character names. Letter-names end in -A following von Hahn 1854; where has -AN for a few of them the form has been retained. The spellings used for the characters here are based on those of the modern Albanian alphabet, with *ç* written CH, *e* written E, and *ë* written EH.

4. Numerals and punctuation. One supposes that European digits are used with Todhri; COMMA and FULL STOP have been seen in the manuscripts.

5. Repertoire. The repertoire of letters represents not only the basic phonemes of Albanian, but also a number of sequences, such as letters for prenasalized consonants. In the table below, the first line gives the basic Albanian phonemes as expressed in the modern orthography; in the second are given the prenasalized equivalents, and in the third are Greek letters and other additions to the repertoire.

	1.	A	B	C	Ç	D	DH	E	Ë	F	G	GJ	H	I	J	K	L	LL	M	N	NJ
	2.	–	MB	–	–	ND	–	–	–	–	NG	NGJ	–	–	–	–	–	–	–	–	–
	4.	AS	–	–	–	–	–	–	–	–	GH	–	HJ	–	–	KH	–	–	–	–	–

	1.	O	P	Q	R	RR	S	SH	T	TH	U	V	X	XH	Y	Z	ZH
	2.	–	–	–	–	–	–	–	–	–	–	–	NX	NXH	–	–	–
	3.	OO	PS	–	–	–	SK	SHT	ST	–	–	–	–	–	JY	–	–

A. The prenasalized characters are by and large modifications of the base letters. Prenasalized *b* looks a bit like *m* (Ϲ M + ℳ B = ℓ MB); prenasalized *d* fuses both (ν N + Λ D = ϰ ND); prenasalized *g* and *gj* both share an extension (ν N + ℎ G = ℏ NG; ν N + ₅ GJ = ₅ NGJ); prenasalized *x* and *xh* share a similar extension (ν N + ℽ X = ℥ NX; ν N + ϭ XH = ℘ NXH).

B. Of the 10 letters here 6 of them answer to Greek characters: Ѵ GH = γ GAMMA; Χ KH = χ CHI; ω OO = ω OMEGA; ч PS = ψ PSI; ȣ SK ?= ξ XI; ϛ ST = ς STIGMA.

C. The remaining 4 letters are a little less clear. ᴠ AS might simply be *v* + ᴈ; ᴈ HJ is a turned ϵ H (von Hahn gives Χ "dark ch" [x] and ᴈ "light ch" [ç]); λ SHT is clearly λ SH + ℐ T and is attested several times in at least one of the texts; and ⅄ JY (Elsie writes *ju/jy*) would appear to be ι I + ♃ Y.

Note that Elsie's chart omits von Hahn's ᴈ HJ, though this may be an oversight (modifying Χ [c] to ᴈ [ç] is logical). Elsie's chart also includes ⅄ JY which von Hahn does not give, but von Hahn gives a letter Ѱ *të* which is also clearly ℐ T + ? EH but is similar to that in shape. Perhaps this letter should be encoded, though it does not appear in the present chart. The text which contains λ SHT writes *të* as ℐ?.

6. Ordering. As for the Vithkuqi script, it is likely that modern Albanians will wish the alphabetical order for Todhri to be similar to that of the Latin Albanian alphabet. (The order given by von Hahn is not particularly practical for modern users.) This order has been followed in the code chart, with prenasalized consonants following their bases and other characters inserted where appropriate.

7. Unicode Character Properties

```
105C0;TODHRI LETTER A;Ll;0;L;;;;;N;;;;;
105C1;TODHRI LETTER AS;Ll;0;L;;;;;N;;;;;
105C2;TODHRI LETTER BA;Ll;0;L;;;;;N;;;;;
105C3;TODHRI LETTER MBA;Ll;0;L;;;;;N;;;;;
105C4;TODHRI LETTER CA;Ll;0;L;;;;;N;;;;;
105C5;TODHRI LETTER CHA;Ll;0;L;;;;;N;;;;;
105C6;TODHRI LETTER DA;Ll;0;L;;;;;N;;;;;
105C7;TODHRI LETTER NDA;Ll;0;L;;;;;N;;;;;
105C8;TODHRI LETTER DHA;Ll;0;L;;;;;N;;;;;
105C9;TODHRI LETTER E;Ll;0;L;;;;;N;;;;;
105CA;TODHRI LETTER EH;Ll;0;L;;;;;N;;;;;
105CB;TODHRI LETTER FA;Ll;0;L;;;;;N;;;;;
105CC;TODHRI LETTER GA;Ll;0;L;;;;;N;;;;;
105CD;TODHRI LETTER NGA;Ll;0;L;;;;;N;;;;;
105CE;TODHRI LETTER GJA;Ll;0;L;;;;;N;;;;;
105CF;TODHRI LETTER NGJA;Ll;0;L;;;;;N;;;;;
105D0;TODHRI LETTER GHA;Ll;0;L;;;;;N;;;;;
105D1;TODHRI LETTER HA;Ll;0;L;;;;;N;;;;;
105D2;TODHRI LETTER HJA;Ll;0;L;;;;;N;;;;;
105D3;TODHRI LETTER I;Ll;0;L;;;;;N;;;;;
105D4;TODHRI LETTER JA;Ll;0;L;;;;;N;;;;;
```

```
105D5;TODHRI LETTER KA;Ll;0;L;;;;;N;;;;;
105D6;TODHRI LETTER KHA;Ll;0;L;;;;;N;;;;;
105D7;TODHRI LETTER LA;Ll;0;L;;;;;N;;;;;
105D8;TODHRI LETTER LLA;Ll;0;L;;;;;N;;;;;
105D9;TODHRI LETTER MA;Ll;0;L;;;;;N;;;;;
105DA;TODHRI LETTER NA;Ll;0;L;;;;;N;;;;;
105DB;TODHRI LETTER NJAN;Ll;0;L;;;;;N;;;;;
105DC;TODHRI LETTER O;Ll;0;L;;;;;N;;;;;
105DD;TODHRI LETTER OO;Ll;0;L;;;;;N;;;;;
105DE;TODHRI LETTER PA;Ll;0;L;;;;;N;;;;;
105DF;TODHRI LETTER PSA;Ll;0;L;;;;;N;;;;;
105E0;TODHRI LETTER QA;Ll;0;L;;;;;N;;;;;
105E1;TODHRI LETTER RA;Ll;0;L;;;;;N;;;;;
105E2;TODHRI LETTER RRA;Ll;0;L;;;;;N;;;;;
105E3;TODHRI LETTER SA;Ll;0;L;;;;;N;;;;;
105E4;TODHRI LETTER SKAN;Ll;0;L;;;;;N;;;;;
105E5;TODHRI LETTER STA;Ll;0;L;;;;;N;;;;;
105E6;TODHRI LETTER SHA;Ll;0;L;;;;;N;;;;;
105E7;TODHRI LETTER SHTA;Ll;0;L;;;;;N;;;;;
105E8;TODHRI LETTER TA;Ll;0;L;;;;;N;;;;;
105E9;TODHRI LETTER THA;Ll;0;L;;;;;N;;;;;
105EA;TODHRI LETTER U;Ll;0;L;;;;;N;;;;;
105EB;TODHRI LETTER VA;Ll;0;L;;;;;N;;;;;
105EC;TODHRI LETTER XA;Ll;0;L;;;;;N;;;;;
105ED;TODHRI LETTER NXA;Ll;0;L;;;;;N;;;;;
105EE;TODHRI LETTER XHA;Ll;0;L;;;;;N;;;;;
105EF;TODHRI LETTER NXHA;Ll;0;L;;;;;N;;;;;
105F0;TODHRI LETTER Y;Ll;0;L;;;;;N;;;;;
105F1;TODHRI LETTER JY;Ll;0;L;;;;;N;;;;;
105F2;TODHRI LETTER ZA;Ll;0;L;;;;;N;;;;;
105F3;TODHRI LETTER ZHA;Ll;0;L;;;;;N;;;;;
```

8. Bibliography

Diringer, David. 1949. The Alphabet.

Десницкая, Агния Васильевна. 1968. *Албанский язык и его диалекты*. Leningrad: Nauka.

Десницкая, Агния Васильевна. 1987. *Албанская литература и албанский язык*. Leningrad: Nauka.

Elsie, Robert. 2016. *Early Albanian Bible translations in Todhri Script*. (Albanian Studies; 30) London: Centre for Albanian Studies. ISBN 978-1-53970333-4.

Elsie, Robert. 2017. "The alphabet of Dhaskal Todhri", in *Albanian Alphabets: Borrowed and invented*. (Albanian Studies; 35) London: Centre for Albanian Studies. ISBN 978-1-54429409-4.

Faulmann, Karl. 1880. *Das Buch der Schrift*. Wien.

Skendi, Stavro. 1960. "The history of the Albanian alphabet: a case of complex cultural and political development", in *Südost-Forschungen: Internationale Zeitschrift für Geschichte, Kultur und Landeskunde Südosteuropas* 19:263-284.

Trix, Frances. 1997. "Alphabet conflict in the Balkans: Albanian and the congress of Monastir", in *International Journal of the Sociology of Language* 128:1-23.

Veqilharxhi, Naum. 1845. *Evëtori Shqip Fort i Shkurtër*. Todhri.

§. 1. Das Alphabet.

Nr.	Zeichen.	Benennung.	Werth.	Nr.	Zeichen.	Benennung.	Werth.	Nr.	Zeichen.	Benennung.	Werth.
1.	Ⴟ	A	a	19.	ꜙ	Ra	r	37.	ꝫ	θa	griech. ϑ
2.	i	E	offen e	20.	ꝑ	Rra	rr	38.	ᴁ	Ba	b
3.	l	I	i	21.	Ꮯ	Fa	f	39.	Ɵ	Mba	mb
4.	O	O	o	22.	Ʌb	Δa	griech. δ	40.	Ⴚ	Pa	p
5.	O	U	u	23.	ᴄ	Ma	m	41.	V	Na	n
6.	♂	Ü	ü	24.	)	Ja	deutsch j	42.	q	Tscha	tsch
7.	ꝑꝑ	E	stumm e	25.	h	Gha	gh	43.	g	Dscha	dsch
8.	ꝫ	Sa	s	26.	Ᵽ	Ngha	ngh	44.	ꝙ	Ndscha	ndsch
9.	Ⴘ	Ça	griech. ζ	27.	b	Gja	gj	45.	5	Sta	st
10.	Ⴑ	Tsa	ts	28.	b	Ngja	ngj	46.	Ⴗ	Scha	sch
11.	Ꙇ	Dsa	ds	29.	Ⴟ	Γa	griech. γ	47.	Ⴝ	Ja	trans. j
12.	Ⴍ	Ndsa	nds	30.	ꙮ	Psa	ps	48.	Ⴞ	Schta	scht
13.	⸒	Wa	w	31.	ϵ	Ha	h	49.	Ⴈ	Te	te
14.	Ⱨ	La	l	32.	X	Cha	dunkel ch	50.	ꝕ	Njan	nj
15.	ꝑ	Lja	lj	33.	ꝫ	Chja	hell ch	51.	Ⴘ	As	as
16.	Ⴥ	Kja	kj	34.	ꝗꝗ	Ta	t	52.	Ⴡ	Ω	griech. ω
17.	ᴄ	Ka	k	35.	Λ	Da	d				
18.	8	Xan	x	36.	XX	Nda	nd				

Figure 1. Todhri characters as given in von Hahn 1854.

THE ALBANIAN ALPHABETS.

Values.	Büthakukye.	Elbasan.	Probable Greek Prototypes.	Values.	Büthakukye	Elbasan.	Probable Greek Prototypes.	Values.	Büthakukye.	Elbasan.	Probable Greek Prototypes.
a			α	n		ν	ν	v			β
b			β	x		8	ξ	d			δ
g			γ	o		0	o	y			ϵ
dh			δ	p			ϖ	y			η
e			ϵ	r			ρ	ss			σ
z			ζ	s			σ	u			υ
ē			η	t			τ	gh			κ
th			θ	ü			υ	e			ι
i			ι	f			ϕ	u			o
k			κ	kh			χ	rr			ρ
l			λ	ō			ω				
m			μ								
I.	II.	III.	IV.	I.	II.	III.	IV.	I.	II.	III.	IV.

THE LIGATURES.

Elbasan.						Büthakukye.	
as	$= \alpha + \sigma$	dsh	$= \theta + \sigma\sigma$			ky	$= \kappa + \iota$
ps	$= \pi + \sigma$	ndsh	$= \nu + \theta\sigma\sigma$			ks	$= \kappa + \sigma$
mb	$= \mu + \beta$	nd	$= \nu + \delta$			dz	$= \delta + \sigma$
st	$= \sigma + \tau$	ds	$= \delta + \sigma$				
sh	$= \sigma + \sigma$	nds	$= \nu + \delta\sigma$			ds	$= \theta + \sigma$
shy	$= \sigma\sigma + \epsilon$	ngh	$= \nu + \kappa$			tzy	$= \tau + \sigma + \iota$
sht	$= \sigma\sigma + \tau$	gy	$= \kappa + \epsilon$				
te	$= \tau + \epsilon$	ngy	$= \nu + \kappa\epsilon$			ng	$= \eta + \nu$
ts	$= \tau + \sigma$	ly	$= \lambda + \epsilon$				
		ky	$= \kappa + \iota$				

Figure 2. Todhri characters as given in Taylor 1883. Taylor's "ligatures" seem to be his attempts to show the derivation of the letters.

The Todhri Alphabet

Modern Alb.	Todhri	Modern Alb.	Todhri
a		t	
b		th	
c		u	
ç		v	
d		x	
dh		xh	
e		y	
ë		z	
f			
g		**Ligatures**	
gj		ju/jy	
h		mb	
i		nd	
j		ng	
k		ngj	
l		nx	
ll		ps	
m		sht	
n		sk	
nj			
o		**Rare letters**	
ō		nxh	
p		st	
q		as	
r		γ (Greek)	
rr		χ (Greek)	
s		zh	
sh			

Figure 3. Todhri alphabet chart from Elsie 2016.

Figure 4. The The Trisagion Prayers; The Lord's Prayer; Psalms 84: 1-4 from Elsie 2016:98.

01.   [...]
02. babáj ýnë qí jée më qíejtë u shenjtënóftë émëni ýt, árthtë
03. mbretënía jóte, e u báftë dashúnia jóte, sí më qíejtë nde mbíi dhéet, búkën
04. tónë të përdíçen, épna née sót, nde ndjé née fájet tóna, sí
05. nde ná ndjéejmë fajtórët tónë e mós kíi me na prúum më të kandá-
06. tun, pó rújna née prej cë kéqi, amín   [...]
07. se jóteja áshë mbretënía nde fuqía nde láfta e átit e e bírit nde
08. i shpírtit shénjt njimé e kúrdó, nde më jétë të jétvet   [...]
09.   zót mallëngjé
10.   [...]
11. éni të lútemi e ti fálemi, mbrétit perëndíis tónë,
12. éni të lútemi e ti fálemi, Kríshtit mbrétit perëndíis tónë
13. éni të lútemi e ti fálemi, atíi Kríshtit mbrétit perëndíis tónë.
14.   [...]
15. o zót i fuqíavet, fort te dáshme jánë konákët e túu prej
16. mállit e sevdáas ndë avllíat zótit gjíndet shpíirti ím zémbra íme
17. nde shtáti ím, u gëzúanë më perëndiinë e gjállë e se zógu ngjít
18. çérdhe për vétet të tíi, nde túrtulli zógun e tíi e núk káa me
19. i thérun zóqt e tíi, thuqím hánet e túa, zót í fuqíavet, mbréti
20. ím nde perëndía ím jánë të lúmun jánë atá qí kondísin
21. më shtëpíi táte, më jétat të jétvet dót të vargëzójnë.

Figure 5. Transcription of figure 4 above from Elsie 2016.

1.   Πάτερ ἡμῶν

2.

3.

4.

5.

6.   [...]

7.

8.   [...]

9.

10.   [Greek]

11.

12.

13.

14.   [Greek]

15.

16.

17.

18.

19.

20.

21.

Figure 6. The text of John 16:1–23 transcribed by von Hahn, from Elsie 2016:104.

1. …këtó fóla júve
2. qi të mos skandalisi, do tju bájnë syrgjýn júve, po do ví kóho, qí kushdó qi tju vrási júve, do ti dúket sikur i bán kur-
3. bán perëndísë, e këtó do jua bájnë júve se nuk më ngjófën ás múu, as babánë, pó këtó fóla júve qí kur të vínjë
4. kóhoja ti kujtóni ató qi ju tháshë únë, këtó núk jua tháshë për pára se íshje me júu, ma nashtí po shkój ke aí qi më dërgój, e
5. njeríi nuk më pyet, kú shkon, pó pse ju tháshë këtó fjálë júve, u helmútë fórtë, pó únë të dréjtën thémë júve, qi ju báhet
6. mírë júve qi të shkój únë, se në mós shkófsha únë, híri nuk vjéne më júu, má në vofsh do te çój hírin më júu, pra si të
7. víi aí híir, do shájë dynjánë për gjynáftë, ne për drejtënítë, ne për gjýqët, e për gjynáftë, psé núk mbesójnë më múu,
8. për drejtëníte, se ke babái do shkój, nde núk më shífshni ma, e për gjýqët, se arhóndi kësáj dynjáje u sháa, kám ne shúm
9. për tju thánë júve, po núk mundi mbáni méndë nashtí, pó kúr të vínjë aí shpírti drejtënësë, do ju t hék më çdó vërtetníi e
10. núk ka mé fólë prej vetíu, pó sá të ndëgjojë do flásë nde do rrëféjë të për t árdhmet júve, aí múu do laftójë, se prej
11. méje do dálë, nde do kallëzój júve, gjíth qísh ká babái të míat janë, prandáj tháshë qí prej méje do dálë e tju kall-
12. ëzójë júve, ne pák hérë, prá núk më shífni, përsí ne pák hérë, prá do më shífni, se do shkój ke babái, prá thánë çirá-
13. këtë shóq me shóq ç asht këjó qi na thótë, ne nji çík prá nuk më shífni, prá ne nji çík do më shífni, nde se dó shkój ke babái,
14. pra thóshin ça[shtë] këjó çikë që na thótë, nuk e dimë se qish na thótë kështú. u kujtúa Kríshti se do te pýsin e ju tha vetë, po pýtni
15. pse tháshë qí ne nji çík prá núk më shífni, përsíi ne nji çík prá do më shifni; vërté, vërté thémi júve, se do qáni e dó váji jú,
16. má dynjája dó gëzójë, e jú dó helmóchi, pó hélmi jújë dó kthéhet më gëzím, grúja kúr dó líndi, idhënóhet si i érthë
17. saháti sajë, ma si të líndi fëmínë, núk e kujtón ma të fështirët, prej gëzímit se léu njerí më dynjátë, ashtu ne ju, nashtí
18. idhëníi kéni, po përsíi dukem ndër júu, e dó ju gëzóhet zémbra jújë, qí atë gëzím súnd ju a márri njeríi, pra m at hére mú mós

Figure 7. Transcription of figure 4 above from Elsie 2016.

	105C	105D	105E	105F
0	Ʋ 105C0	Ʋ 105D0	⸖ 105E0	♂ 105F0
1	Ʋ 105C1	ε 105D1	⸞ 105E1	Ꮍ 105F1
2	⋒ 105C2	Ǝ 105D2	ſ 105E2	Ʋ 105F2
3	ℓ 105C3	ı 105D3	ȿ 105E3	ꝇ 105F3
4	Ꝗ 105C4	) 105D4	8 105E4	
5	Ꝗ 105C5	⸜ 105D5	5 105E5	
6	Ʌ 105C6	Ҳ 105D6	ℓ 105E6	
7	ҳҳ 105C7	ꝗ 105D7	Ꝉ 105E7	
8	Ꮂ 105C8	Ҥ 105D8	ꝙ 105E8	
9	i 105C9	ꝺ 105D9	₴ 105E9	
A	Ꟊ 105CA	Ʌ 105DA	ȯ 105EA	
B	Ϭ 105CB	Ʋ 105DB	ℓ 105EB	
C	ℎ 105CC	o 105DC	Ɉ 105EC	
D	Ꮒ 105CD	Ɯ 105DD	Ƶ 105ED	
E	ꝭ 105CE	ꜳ 105DE	ꝗ 105EE	
F	ꝭ 105CF	Ꮍ 105DF	ꝗ 105EF	

105C0	Ʋ	TODHRI LETTER A
105C1	Ʋ	TODHRI LETTER AS
105C2	⋒	TODHRI LETTER BA
105C3	ℓ	TODHRI LETTER MBA
105C4	Ꝗ	TODHRI LETTER CA
105C5	Ꝗ	TODHRI LETTER CHA
105C6	Ʌ	TODHRI LETTER DA
105C7	ҳҳ	TODHRI LETTER NDA
105C8	Ꮂ	TODHRI LETTER DHA
105C9	i	TODHRI LETTER E
105CA	Ꟊ	TODHRI LETTER EH
105CB	Ϭ	TODHRI LETTER FA
105CC	ℎ	TODHRI LETTER GA
105CD	Ꮒ	TODHRI LETTER NGA
105CE	ꝭ	TODHRI LETTER GJA
105CF	ꝭ	TODHRI LETTER NGJA
105D0	Ʋ	TODHRI LETTER GHA
105D1	ε	TODHRI LETTER HA
105D2	Ǝ	TODHRI LETTER HJA
105D3	ı	TODHRI LETTER I
105D4	)	TODHRI LETTER JA
105D5	⸜	TODHRI LETTER KA
105D6	Ҳ	TODHRI LETTER KHA
105D7	ꝗ	TODHRI LETTER LA
105D8	Ҥ	TODHRI LETTER LLA
105D9	ꝺ	TODHRI LETTER MA
105DA	Ʌ	TODHRI LETTER NA
105DB	Ʋ	TODHRI LETTER NJAN
105DC	o	TODHRI LETTER O
105DD	Ɯ	TODHRI LETTER OO
105DE	ꜳ	TODHRI LETTER PA
105DF	Ꮍ	TODHRI LETTER PSA
105E0	⸖	TODHRI LETTER QA
105E1	⸞	TODHRI LETTER RA
105E2	ſ	TODHRI LETTER RRA
105E3	ȿ	TODHRI LETTER SA
105E4	8	TODHRI LETTER SKAN
105E5	5	TODHRI LETTER STA
105E6	ℓ	TODHRI LETTER SHA
105E7	Ꝉ	TODHRI LETTER SHTA
105E8	ꝙ	TODHRI LETTER TA
105E9	₴	TODHRI LETTER THA
105EA	ȯ	TODHRI LETTER U
105EB	ℓ	TODHRI LETTER VA
105EC	Ɉ	TODHRI LETTER XA
105ED	Ƶ	TODHRI LETTER NXA
105EE	ꝗ	TODHRI LETTER XHA
105EF	ꝗ	TODHRI LETTER NXHA
105F0	♂	TODHRI LETTER Y
105F1	Ꮍ	TODHRI LETTER JY
105F2	Ʋ	TODHRI LETTER ZA
105F3	ꝇ	TODHRI LETTER ZHA

A. Administrative

1. Title
Preliminary proposal for encoding the Todhri script in the SMP of the UCS
2. Requester's name
UC Berkeley Script Encoding Initiative (Universal Scripts Project)
3. Requester type (Member body/Liaison/Individual contribution)
Liaison contribution.
4. Submission date
2020-07-10
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
No.
6b. More information will be provided later
Yes.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
Yes.
1b. Proposed name of script
Todhri.
1c. The proposal is for addition of character(s) to an existing block
No.
1d. Name of the existing block
2. Number of characters in proposal
78.
3. Proposed category (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)
Category E.
4a. Is a repertoire including character names provided?
Yes.
4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
Yes.
4c. Are the character shapes attached in a legible form suitable for review?
Yes.
5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Jaon Glavy via Michael Everson.
5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson, FontLab.
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.
7. 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. 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/Public/UNIDATA/ UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
No.
2a. 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, in L2/09-328.
2b. If YES, with whom?
Robert Elsie, Stiven Llupa, Edon Muhaxheri.
2c. 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?
See above.

4a. The context of use for the proposed characters (type of use; common or rare)

To write the Albanian language.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In scholarly publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?