

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 Toghri 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: 2020-10-01**

1. Introduction. A number of 18th- and 19th-century orthographies exist for Albanian; one of the more significant ones given its corpus is called Toghri (/ˈtoðri/), from the nickname of the man who devised it, Dhaskal Toghri ‘teacher Toghri’, 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 Toghri 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 Toghri (“teacher Toghri”), whose full name was Theodor Haxhifilipi (ca. 1730-1805), and who stemmed from Elbasan.

The biblical texts in this Toghri 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 Toghri script, Hahn explains:

“As far as I have been able to ascertain, this alphabet seems to be native only to the town of Elbasan. It is unknown in Durrës, 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 [Voskopojë], 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;

2.2 Albanian vowels. Albanian has seven vowel phonemes: *a* /a/, *e* /ɛ/, *ë* /ɐ/, *i* /i/, *o* /ɔ/, *u* /u/, and *y* /y/. The vowels *i* and *o* are represented with a vertical line *ı* and a circle *o*; the vowels *e* and *u* have the same base shape with a dot above: *ı̇* and *ȯ*. It has been suggested that options for representing these be presented to SC2 and to the UTC.

2.2.1 Atomic encoding. Encode *v a, ı̇ e, ɹ̇ ë, ı̇ i, o/ω o/ō, ȯ u,* and *ɸ y* individually. This is the simplest model, and it is one which has been used successfully with a great many scripts. For a historical script like Todhri, the simplest model is the best, particularly as diacritics may be used to indicate stress.

2.2.2 Atomic encoding with canonical decomposition. Encode *ı̇ e* and *ȯ u* atomically, but make them canonically equivalent to *ı i* and *o o* followed by COMBINING DOT ABOVE. There does not seem to be much rationale for this. Just because a variety of diacritics are used for stress does not mean that diacritics should also be permitted spellings for *e* and *u*. In fact for textual analysis the stress diacritics might be considered “noise” and it could be preferred to strip them out. In such a case the distinction between *e* and *i* on the one hand and *o* and *u* on the other could be lost. This encoding would work, but it would make font development and inputting needlessly complicated.

2.2.3 Decomposed encoding. This proposal would have only five vowels encoded, and require the use of COMBINING DOT ABOVE to indicate two of the seven Albanian vowels. There is little to recommend this. In light of the use of a range of diacritics with all of the vowels, adding combining dot above for *e* and *u* would just complicate text representation and analysis, and invite multiple or erroneous spellings. This is by no means the right way to encode Todhri.

3.1 Repertoire. The repertoire of letters represents not only the basic phonemes of Albanian, but also a number of sequenced phonemes, such as letters for prenasalized consonants. In the table below, Group 1 gives the basic Albanian phonemes as expressed in the modern orthography; in Group 2 are given the prenasalized equivalents, and in Group 3 are Greek letters, and in Group 4 are 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
	Ɑ	Ɱ	Ɐ	ç	ⱪ	Ⱬ	ı̇	ɹ̇	Ɔ	h	ɸ	ε	ı	ɹ	κ	λ	ll	μ	ν	ɲ
2.	—	MB	—	—	ND	—	—	—	—	NG	NGJ	—	—	—	—	—	—	—	—	—
	—	ℓ	—	—	ɲ	—	—	—	—	h	ɸ	—	—	—	—	—	—	—	—	—
3.	—	—	—	—	—	—	—	—	—	GH	—	—	—	—	KH	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	ʋ	—	—	—	—	χ	—	—	—	—	—
4.	AS	—	—	—	—	—	—	—	—	—	—	HJ	—	—	—	—	—	—	—	—
	ʋ	—	—	—	—	—	—	—	—	—	—	ə	—	—	—	—	—	—	—	—
1.	O	P	Q	R	RR	S	SH	T	TH	U	V	X	XH	Y	Z	ZH				
	o	Ɑ	ç	ɹ	ʃ	z	λ	q	ɹ̇	ȯ	ℓ	ɸ	ɸ	ɸ	ɸ	ʋ	ɸ			
2.	—	—	—	—	—	—	—	—	—	—	—	NX	NXH	—	—	—				
	—	—	—	—	—	—	—	—	—	—	—	z	ɸ	—	—	—				
3.	OO	PS	—	—	—	(SK)	—	ST	—	—	—	—	—	—	—	—				
	ω	ɸ	—	—	—	(s)	—	5	—	—	—	—	—	—	—	—				
4.	—	—	—	—	—	—	SHT	—	—	—	—	—	—	JY	—	—				
	—	—	—	—	—	—	λ	—	—	—	—	—	—	ɸ	—	—				

3.1.1 Group 2. The prenasalized characters are by and large modifications of the base letters. Prenasalized *b* looks a bit like *m* (*ç* M + *Ɑ* B = *ℓ* MB); prenasalized *d* fuses both (*ı̇* N + *ⱪ* D = *ɲ* ND);

prenasalized *g* and *gj* both share an extension ($v N + h G = h NG$; $v N + h GJ = h NGJ$); prenasalized *x* and *xh* share a similar extension ($v N + x X = x NX$; $v N + g XH = g NXH$).

3.1.2 Group 3. The 6 letters here answer to Greek characters: $\Psi GH = \gamma$ GAMMA; $\chi KH = \chi$ CHI; $\omega OO = \omega$ OMEGA; $\psi PS = \psi$ PSI; $\zeta ST = \zeta$ STIGMA. The letter σSK may or may not be related to ξXI but certainly was /sk/ and not /ks/—compare the Greek loanwords $\sigma\kappa\alpha\upsilon\delta\alpha\lambda\eta\varsigma$ *skandhalis* ‘to stumble’ (σκανδάλης) and $\delta\iota\delta\acute{\alpha}\sigma\kappa\alpha\lambda\omicron\varsigma$ *dhidhaski* ‘teaching’ (διδάσκαλος). Note that despite the fact that the printed source charts (Figures 1, 2, and 3) show the $oo \omega$ with a horizontal bar above ($\bar{\omega}$), a variety of shapes are found in the actual manuscript, including as mentioned above. The form without any diacritic does occur more than once in the manuscripts, so it is recommended that this be the form used in the code charts. The phrase *o zot* ‘O God’ can be seen as $\bar{\omega} \nu\acute{o}\lambda$ at the beginning of line 15 of Figures 4 and 5; it is given as $\omega \nu\acute{o}\lambda \bar{o} z\acute{o}t$ in Figure 6 and as $o \nu\acute{o}\lambda o z\acute{o}t$ without an omega in Figure 8.

3.1.3 Group 4. The 4 letters here are a little less clear in their derivation. νAS might simply be $v + z$; $\eth HJ$ is a turned ϵH (von Hahn gives χ “dark ch” [x] and \eth “light ch” [ç]); λSHT is clearly derived from $\lambda SH + \tau T$ and is attested many times throughout the texts; and $\mathfrak{J} JY$ (Elsie writes *ju/jy*; $\mathfrak{J} JY$ appears on p. 29 of the MS) would appear to be $i I + j Y$, and is common enough in the texts. It is impossible to talk about “typographic ligatures” for Todhri as there was no traditional typography.

Note that Elsie’s chart omits von Hahn’s $\eth HJ$, though this may be an oversight (modifying $\epsilon [h]$ to $\eth [\ç]$ is logical). Elsie’s chart also includes $\mathfrak{J} JY$ which von Hahn does not give, but von Hahn gives a letter $\Psi t\ddot{e}$ which is also clearly $\tau T + \eta EH$ but is similar to that in shape (see 3.2 below). The text which contains λSHT usually writes *të* as \mathfrak{J} (not Ψ); see Figures 8 and 9. As to whether $SHTA$ should be encoded on its own, the recommendation is that it should be. Certainly both λSHA and τTA both exist; I have not yet seen an instance of a sequence “ $\lambda\tau SHT$ ” though the “ligature” is very common (e.g., $\epsilon\mathfrak{J}\lambda\tau$ *hrishtë*, $\epsilon\mathfrak{J}\lambda\tau$ *krishtë* ‘Christ’). Compare Glagolitic which has Ш SHA , Т TE , and ШТ SHTA (also clearly a ligature in origin). Users of Todhri are scholars, and they can deal with searching issues revolving around $\lambda\tau$ and λ , if any, given the relatively small size of the corpus.

3.2 Ligature të. A ligature $\Psi t\ddot{e}$ is given in von Hahn’s chart. Perhaps von Hahn saw an example of a logograph for *të* ‘of’. It has not been seen in any of the manuscripts I have access to, where the unligated form is common. Fonts may choose to include an optional ligature for $\Psi t\ddot{e}$.

4. Character names. Letter-names end in -A following von Hahn 1854; two of von Hahn’s letters have -AN and this 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 \ddot{e} written EH. Here are the correspondences:

Todhri	UCS	von Hahn	Todhri	UCS	von Hahn	Todhri	UCS	von Hahn	Todhri	UCS	von Hahn
v	A	a	h	NGA	ngha	v	NA	na	l	SHTA	shta
ν	AS	as	h̄	GJA	gja	ν	NJAN	njan	l̄	TA	ta
ϑ	BA	ba	h̄̄	NGJA	ngja	o	O	o	ʔ	THA	tha
ℓ	MBA	mba	v	GHA	ga	ω	OO	ω	ó	U	u
l̄	CA	tsa	ε	HA	ha	ι	PA	pa	ℓ	VA	va
q	CHA (ÇA)	tscha	ε̄	HJA	chja	ψ	PSA	psa	ʔ	XA	dsa
l̄	DA	da	i	I	i	z	QA	kja	z	NXA	nda
x	NDA	nda	ʔ	JA	ja (deutsch. j)	ʔ	RA	ra	g	XHA	dscha
l̄	DHA	da	ç	KA	ka	ʃ	RRA	rra	ḡ	NXHA	ndscha
i	E	e (offen e)	χ	KHA	cha	z	SA	sa	h̄	Y	ü
ʔ	EH (Ë)	e (stumm e)	q	LA	lja	8	SKAN	xan	h̄	JY	*jü
ϐ	FA	fa	h̄	LLA	la	5	STA	sta	ν	ZA	za
h	GA	gha	ç̄	MA	ma	λ	SHA	scha	δ	ZHA	ja (franz. j)

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

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*.
- 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.

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

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

9. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Todhri encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

Figures.

§. 1. Das Alphabet.

Nr.	Zeichen.	Benennung.	Werth.	Nr.	Zeichen.	Benennung.	Werth.	Nr.	Zeichen.	Benennung.	Werth.
1.	v	A	a	19.	ʋ	Ra	r	37.	ʒ	θa	griech θ
2.	i	E	offen e	20.	f	Rra	rr	38.	ʋ	Ba	b
3.	l	I	i	21.	ʋ	Fa	f	39.	ʋ	Mba	mb
4.	o	O	o	22.	ʋʋ	Δa	griech δ	40.	ʋ	Pa	p
5.	o	U	u	23.	ʋ	Ma	m	41.	v	Na	n
6.	ʋ	Ü	ü	24.	ʋ	Ja	deutsch j	42.	q	Tscha	tsch
7.	ʋʋ	E	stumm e	25.	h	Gha	gh	43.	g	Dscha	dsch
8.	z	Sa	s	26.	k	Ngha	ngh	44.	g	Ndscha	ndsch
9.	ʋ	Ca	griech c	27.	ʋ	Gja	gj	45.	5	Sta	st
10.	ʋ	Tsa	ts	28.	ʋ	Ngja	ngj	46.	ʋ	Scha	sch
11.	ʋ	Dsa	ds	29.	v	Γa	griech γ	47.	ʋ	Ja	franz. j
12.	ʋ	Ndsa	nds	30.	ʋ	Psa	ps	48.	ʋ	Schta	scht
13.	ʋ	Wa	w	31.	ε	Ha	h	49.	ʋ	Te	te
14.	H	La	l	32.	X	Cha	dunkel ch	50.	ʋ	Njan	nj
15.	ʋ	Lja	lj	33.	ʋ	Chja	hell ch	51.	ʋ	As	as
16.	ʋ	Kja	kj	34.	ʋʋ	Ta	t	52.	ʋ	Ω	griech ω
17.	c	Ka	k	35.	ʋ	Da	d				
18.	8	Xan	x	36.	ʋʋ	Nda	nd				

Figure 1. Todhri characters as given in von Hahn 1854. See §4 above.

THE ALBANIAN ALPHABETS.

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

THE LIGATURES.

Elbasan.				Büthakukye.	
<i>as</i>	ⲕ = α + σ	<i>dsh</i>	ⲙ = θ + σσ	<i>ky</i>	ⲗ ⲉ = κ + ι
<i>ps</i>	ⲙ = π + σ	<i>ndsh</i>	ⲙ = ν + θσσ	<i>ks</i>	ⲗ ⲑ = κ + σ
<i>mb</i>	ⲗ = μ + β	<i>nd</i>	ⲙ = ν + δ	<i>dz</i>	ⲗ ⲟ = δ + σ
<i>st</i>	ⲙ = σ + τ	<i>ds</i>	ⲙ = δ + σ	<i>ds</i>	ⲗ ⲑ = θ + σ
<i>sh</i>	ⲗ = σ + σ	<i>nds</i>	ⲙ = ν + δσ	<i>tzy</i>	ⲗ ⲟ = τ + σ + ι
<i>shy</i>	ⲗ = σσ + ε	<i>ngh</i>	ⲗ = ν + κ	<i>ng</i>	ⲗ ⲑ = η + ν
<i>sht</i>	ⲗ = σσ + τ	<i>gy</i>	ⲗ = κ + ε		
<i>te</i>	ⲙ = τ + ε	<i>ngy</i>	ⲗ = ν + κε		
<i>ts</i>	ⲙ = τ + σ	<i>ly</i>	ⲗ = λ + ε		
		<i>ky</i>	ⲗ = κ + ι		

Figure 2. Todhri (“Elbasan”) and Vithkuqi (“Büthakukye”) characters as given in Taylor 1883. Taylor’s “ligatures” seem to be his (not entirely successful) attempts to show the derivation of the letters:

v as (ν α + ε σ), *ψ ps* (ψ π + ε σ), *ⲗ mb* (ⲗ μ + ρ β), *ⲙ st* (ⲙ ε σ + ϑ τ—correctly Greek ζ as derived from ζ), *ⲗ sh* (ⲗ ε σ + ε σ), *ⲗ zh* (ⲗ λ sh + ϑ ε̇), *ⲗ sht* (ⲗ λ sh + ϑ t), *ⲗ c* (ⲗ ϑ t + ε s), *ⲗ gh* (ⲗ ϑ d + λ sh), *ⲗ nxh* (ν n + ϑ xh), *ⲗ nd* (ν n + λ d), *ⲗ x* (λ d + ε s), *ⲗ nx* (ν n + ϑ x), *ⲗ ng* (ν n + κ—correctly ν n + h g), *ⲗ gj* (κ k + ϑ ε̇—correctly perhaps h g + ϑ ε̇), *ⲗ ngj* (ν n + ⲗ gj), *ⲗ l* (ⲗ ll + ϑ ε̇—l is /l/ and ll is /l/), *ⲗ q* (κ k + ι i).

The Todhri Alphabet

Modern Alb.	Todhri	Modern Alb.	Todhri
a	∇	t	Ɔ
b	Ɔ	th	Ɔ
c	∟	u	◦
ç	Ɔ	v	Ɔ
d	∧	x	Ɔ
dh	Ɔ	xh	Ɔ
e	i	y	Ɔ
ë	Ɔ	z	Ɔ
f	⊙		
g	h	Ligatures	
gj	h	ju/jy	Ɔ
h	Ɔ	mb	Ɔ
i	i	nd	xx
j	j	ng	h
k	c	ngj	h
l	d	nx	Ɔ
ll	H	ps	Ɔ
m	c	sht	Ɔ
n	v	sk	8
nj	h		
o	o	Rare letters	
ō	⊙	nxh	Ɔ (?)
p	∟	st	5
q	Ɔ	as	Ɔ
r	Ɔ	γ (Greek)	Ɔ
rr	Ɔ	χ (Greek)	Ɔ
s	Ɔ	zh	Ɔ
sh	Ɔ		

Figure 3. Todhri alphabet chart from Elsie 2016:155.

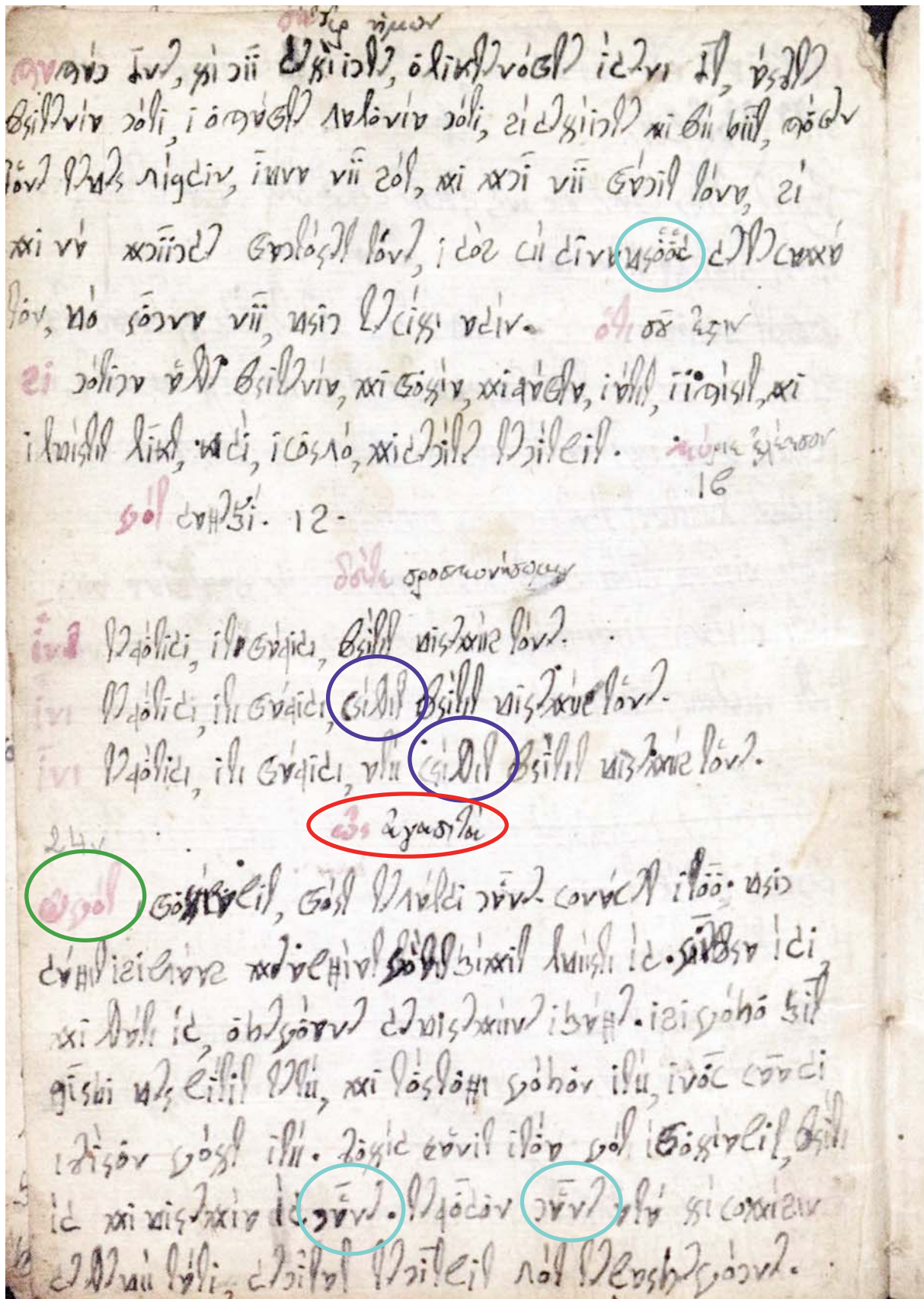


Figure 4. The Trisagion Prayers; The Lord’s Prayer; Psalms 84: 1-4 from Elsie 2016:98. Circled is $\bar{\omega}$ $\bar{v}\bar{o}\bar{l}$ \bar{o} $\bar{z}\bar{o}t$ ‘O God’ (in green). In the Greek heading (in red) above Psalm 84 reads $\omega\varsigma$ $\acute{\alpha}\gamma\alpha\pi\iota\tau\acute{\alpha}$ (evidently for $\omega\varsigma$ $\acute{\alpha}\gamma\alpha\pi\eta\tau\acute{\alpha}$ ‘as dear’); the scribe writes both rough and smooth breathings identically. In purple are two instances of *krishtit* showing the letter SHTA; in cyan are two instances of *jānë* showing written as $\bar{\omega}\bar{v}\bar{v}$ $\bar{j}\bar{a}\bar{n}\bar{e}$ and one instance of *prūūm* written as $\bar{\omega}\bar{v}\bar{o}\bar{c}$ $\bar{p}\bar{r}\bar{u}\bar{u}\bar{m}$.

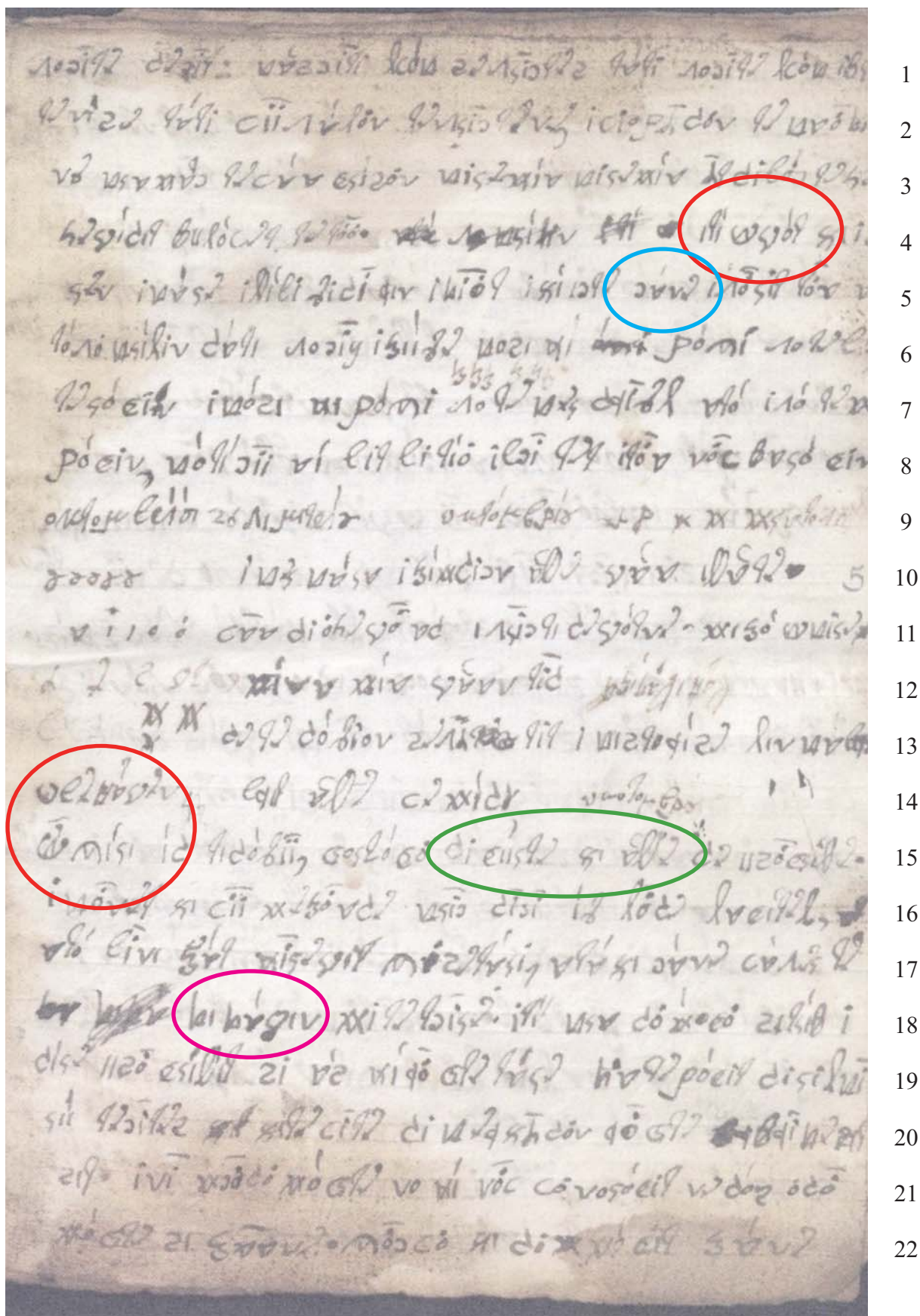


Figure 6. Hebrews 1:8–12; 2 Timothy 2:1–6 from Elsie 2016:30. Circled in red are ω ὦλ ὀ zót ‘O God’, ω ἁἡῦν ὀ vellázèn ‘O brothers’, and ω ἁἡῦν ὀ biri ‘O my son’; in green ci éis ʔ ʒi ṽlʔ me hiir tē qi ashtē ‘in the grace that is’ with TA followed by EH (ʔ) and SHTA followed by EH (ʔ). Circled in cyan is ἁἡῦν jánē ‘they are’ and circled in magenta is liláʔiv dhidházin ‘teaching’ (recte liláʔiv dhidhaskin?).

01. do jëtë më jët pás jëte shkóp së dréjtës táte do jëtë shkóp e mbre-
02. tënisë táte, kée dashún të dréjtënë e ke orrymun të paúdhë-
03. në, prandáj të káa hrésun perëndía perëndía ýt me vujtë gë-
04. gëzimit mbii shókët të túu. [...] i o zót i e hé-
05. rën e páre e shtíve themélin e dheut e qiejtë jánë e dóret tua, a-
06. tó du prishen má ti do jéç e gjiíthë posi një rrobe do të vote vje-
07. tëróhen e pó si një rrobe do të përmléthsh ató e dó të ndë-
08. rróhen, pó tí jée aí vetvetiu e vjëtët e túa nuk mbaróhen.
09. [...]
10. [...] e për pára e gjíndmeja ástë zaa i shátë
11. [...] káa me u gëzúam i drejti më zótnë, ndigjó o perë-
12. ndíaa ndia zaaan tém.
13. [...] më Tëmótheon së dytet epistolísë shin Pávlit
14. o vëllázën [r]vlit ástë këndími
15. o bíri ím tímhóthée, forcóh me hiir të qi ástë në Iisú Hristë
16. e púnët qi cée ndëgjúamë prej ínjé íth shúmë shahítësh,
17. ató vémi ngját njérëzet besëtáre, atá qi jánë kádër të
18. dhidházin nde të tjéré, e tí pra mundóhu si trím e
19. míre Iisu Hristit se ás një luftëtáre ngatërróhet me reshpe-
20. ríi të jëtës qi të ketë me pëlqýmun luftë mblédhë-
21. sit, e né ndju mundóftë no një núk kunoróhet në mós u mú-
22. ndóftë si gjáané bujku qi mundóhet gjánë

Figure 7. Transcription of figure 6 above from Elsie 2016:31. In line 18 Elsie transcribed ללל'ז' as *dhidházin* but this may be a scribal error for *dhidháskin* (not ללל'ז').

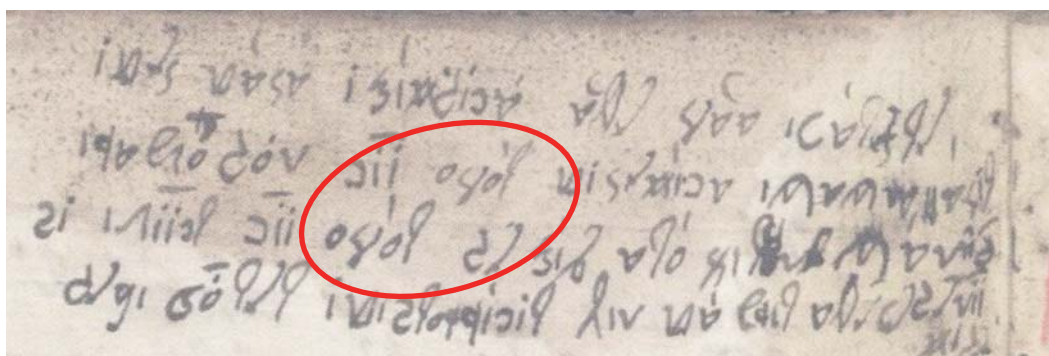


Figure 8. Hebrews 13:7–16 from Elsie 2016:20.

Circled are two instances of o ז'ó' o zót 'O God' with an omicron rather than an omega.

01. e për páara e gjíndmeja ástë zaa i katërtë.
02. e lavdómun jée o zot perëndíja i baballáret
03. se i d[r]éejt jée o zót në gjith ató qi bán më nee.
04. më çifútët epistolíjet shin Pávlit ástë këndími:

Figure 9. Transcription of figure 8 above from Elsie 2016:21.

1. ...këtó fóla júve
2. qi të mós skandalisi, do tju bájñe syrgjýn júve, po do ví kóho, qí
3. kushdó qi tju vrási júve, do ti dúket sikur i bán kur-
4. bán perëndisë, e këtó do jua bájñe júve se nuk më ngjófën ás múu, as
5. babánë, pó këtó fóla júve qí kur të vínjë
6. kóhoja ti kujtóni ató qi ju tháshë únë, këtó núk jua tháshë për pára se
7. íshje me júu, ma nashtí po shkój ke aí qi më dërgój, e
8. njerii nuk më pyet, kú shkon, pó pse ju tháshë këtó fjálë júve, u
9. helmútë fórtë, pó únë të dréjtën thémë júve, qí ju báhet
10. mírë júve qí të shkój únë, se në mós shkófsha únë, híri nuk vjéne më
11. júu, má në vofsh do te çój hírin më júu, pra si të
12. víi aí híir, do shájë dynjánë për gjynáftë, ne për drejténítë, ne për
13. gjýqët, e për gjynáftë, psé núk mbesójñe më múu,
14. për drejténite, se ke babái do shkój, nde núk më shífshni ma, e për
15. gjýqët, se arhóndi kësáj dynjájë u sháa, kám ne shúm
16. për tju thánë júve, po núk mundi mbáni méndë nashtí, pó kúr të vínjë
17. aí shpírti drejténësë, do ju t hék më çdó vërtetní e
18. núk ka mé fólë prej vetiu, pó sá të ndëgjojë do flásë nde do rrëfjë të
19. për t árdhmet júve, aí múu do laftójë, se prej
20. méje do dálë, nde do kallëzój júve, gjíth qísh ká babái të míat janë,
21. prandáj tháshë qí prej méje do dálë e tju kall-
22. ëzójë júve, ne pák hérë, prá núk më shífni, përsí ne pák hérë, prá do
23. më shífni, se do shkój ke babái, prá thánë çirá-
24. këtë shóq me shóq ç asht këjó qí na thótë, ne nji çík prá nuk më shífni,
25. prá ne nji çík do më shífni, nde se dó shkój ke babái,
26. prá thóshin ça[shtë] këjó çikë që na thótë, nuk e dimë se qísh na thótë
27. kështú. u kujtúa Kríshtí se do te pýsin e ju tha vetë, po pýtni
28. pse tháshë qí ne nji çík prá núk më shífni, përsíi ne nji çík prá do më
29. shífni; vërtë, vërtë thémí júve, se do qáni e dó váji jú,
30. má dynjája dó gëzójë, e jú dó helmóchi, pó hélmi jújé dó kthéhet më
31. gëzím, grúja kúr dó líndi, idhënohet si i értë
32. sahátí sajë, ma si të líndi fëmínë, núk e kujtón ma të fështirët, prej
33. gëzímít se léu njerí më dynjátë, ashtu ne ju, nashtí
34. idhënií kéni, po përsíi dukem ndër júu, e dó ju gëzóhet zémбра jújé, qí
35. atë gëzím súnd ju a márrí njerii, pra m at hére mú mós

Figure 11. Transcription of figure 10 above from Elsie 2016:105.

	105C	105D	105E	105F
0	ᵛ 105C0	ᵛ 105D0	Ɔ 105E0	ᵔ 105F0
1	ᵛ 105C1	€ 105D1	Ɔ 105E1	ᵔ 105F1
2	ᵛ 105C2	€ 105D2	Ɔ 105E2	ᵔ 105F2
3	ᵛ 105C3	€ 105D3	Ɔ 105E3	ᵔ 105F3
4	ᵛ 105C4	€ 105D4	Ɔ 105E4	
5	ᵛ 105C5	€ 105D5	Ɔ 105E5	
6	ᵛ 105C6	€ 105D6	Ɔ 105E6	
7	ᵛ 105C7	€ 105D7	Ɔ 105E7	
8	ᵛ 105C8	€ 105D8	Ɔ 105E8	
9	ᵛ 105C9	€ 105D9	Ɔ 105E9	
A	ᵛ 105CA	€ 105DA	Ɔ 105EA	
B	ᵛ 105CB	€ 105DB	Ɔ 105EB	
C	ᵛ 105CC	€ 105DC	Ɔ 105EC	
D	ᵛ 105CD	€ 105DD	Ɔ 105ED	
E	ᵛ 105CE	€ 105DE	Ɔ 105EE	
F	ᵛ 105CF	€ 105DF	Ɔ 105EF	

- 105C0 ᵛ TODHRI LETTER A
- 105C1 ᵛ TODHRI LETTER AS
- 105C2 Ɔ TODHRI LETTER BA
- 105C3 Ɔ TODHRI LETTER MBA
- 105C4 Ɔ TODHRI LETTER CA
- 105C5 Ɔ TODHRI LETTER CHA
- 105C6 Ɔ TODHRI LETTER DA
- 105C7 Ɔ TODHRI LETTER NDA
- 105C8 Ɔ TODHRI LETTER DHA
- 105C9 Ɔ TODHRI LETTER E
- 105CA Ɔ TODHRI LETTER EH
- 105CB Ɔ TODHRI LETTER FA
- 105CC Ɔ TODHRI LETTER GA
- 105CD Ɔ TODHRI LETTER NGA
- 105CE Ɔ TODHRI LETTER GJA
- 105CF Ɔ TODHRI LETTER NGJA
- 105D0 ᵛ TODHRI LETTER GHA
→ 03B3 γ greek small letter gamma
- 105D1 € TODHRI LETTER HA
- 105D2 € TODHRI LETTER HJA
- 105D3 Ɔ TODHRI LETTER I
- 105D4 Ɔ TODHRI LETTER JA
- 105D5 Ɔ TODHRI LETTER KA
- 105D6 Ɔ TODHRI LETTER KHA
→ 03C7 χ greek small letter chi
- 105D7 Ɔ TODHRI LETTER LA
- 105D8 Ɔ TODHRI LETTER LLA
- 105D9 Ɔ TODHRI LETTER MA
- 105DA ᵛ TODHRI LETTER NA
- 105DB ᵛ TODHRI LETTER NJAN
- 105DC Ɔ TODHRI LETTER O
- 105DD Ɔ TODHRI LETTER OO
→ 03C9 ω greek small letter omega
- 105DE Ɔ TODHRI LETTER PA
- 105DF Ɔ TODHRI LETTER PSA
→ 03C8 ψ greek small letter psi
- 105E0 Ɔ TODHRI LETTER QA
- 105E1 Ɔ TODHRI LETTER RA
- 105E2 Ɔ TODHRI LETTER RRA
- 105E3 Ɔ TODHRI LETTER SA
- 105E4 8 TODHRI LETTER SKAN
→ 03BE ξ greek small letter ksi
- 105E5 5 TODHRI LETTER STA
→ 03DB ζ greek small letter stigma
- 105E6 Ɔ TODHRI LETTER SHA
- 105E7 Ɔ TODHRI LETTER SHTA
• in origin a ligature of sha and ta
- 105E8 Ɔ TODHRI LETTER TA
- 105E9 Ɔ TODHRI LETTER THA
- 105EA Ɔ TODHRI LETTER U
- 105EB Ɔ TODHRI LETTER VA
- 105EC Ɔ TODHRI LETTER XA
- 105ED Ɔ TODHRI LETTER NXA
- 105EE Ɔ TODHRI LETTER XHA
- 105EF Ɔ TODHRI LETTER NXHA
- 105F0 Ɔ TODHRI LETTER Y
- 105F1 Ɔ TODHRI LETTER JY
- 105F2 ᵛ TODHRI LETTER ZA
- 105F3 Ɔ TODHRI LETTER ZHA

A. Administrative

1. Title

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

2020-10-01

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

53.

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?