JTC1/SC2/WG2 N5139R2
L2/20-188R2
2020-12-31

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: **2020-12-31**

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 (/ˈtoðri/), 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 Elbasan. 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.

2.1 Diacritical marks. A number of diacritical marks can be observed to be used in the Todhri texts. Some are straight, some curved, some angled to the right or to the left. They are used only with vowels, and remind one of Greek accents, but in Elsie's transcriptions, they are uniformly represented in Latin by the ACUTE ACCENT, and are used to indicate stress. There might be a tendency to use one with narrower characters and the other with wider ones, but in any of the manuscripts there are exceptions. If a concordance of the manuscripts' texts were made, a frequency of choice could be established. In the meantime, however, it is prudent to recommend that U+0301 COMBINING ACUTE ACCENT be used for the vertical mark and that U+0304 COMBINING MACRON be used for the horizontal. Where a macron appears over two vowels (as it does in a number of instances) then U+035E COMBINING DOUBLE MACRON can be used. The remaining mark seems to be found only over the letter OO, which derives from Greek OMEGA; so far a smooth breathing (for Greek U+0313 COMBINING COMMA ABOVE) has been found over OO when used as a vocative particle; this mark is sometimes omitted, and may or may not have a stress mark above it. Fonts should support the following repertoire. In accordance with what is seen in the manuscript. the smooth breathing looks best when represented by U+0311 COMBINING INVERTED BREVE. (Some unusual combinations (which can't be breathings) occur (Figure 4): 𐓣𐓵𐓵𐓍 *prûûm*, 𐓝𐓯𐓝𐓗 *jānë* 'they are'; I list them in parentheses below. Perhaps these are just 𐓣𐓵𐓵𐓍 *prūūm* and 𐓝𐓯𐓝𐓗 *jánë*; in Figure 6 we find 𐓝𐓯𐓝𐓗 *jánë*.)

𐓯 𐓯̄ 𐓯̿ 𐓯́ 𐓯̄́ 𐓯̿́ 𐓯́ 𐓯̄ 𐓯̿ 𐓵 𐓵̄ 𐓵̄ 𐓵̂ 𐓵́ 𐓵̂́ 𐓵̄ 𐓵̄́ 𐓵̄̂ 𐓵̄̂́ 𐓵̄́ 𐓵̄ 𐓵̄̂ 𐓵̄̂́ 𐓯̄́ 𐓵̄́

2.2 Albanian vowels. Albanian has seven vowel phonemes: *a* /a/, *e* /ɛ/, *ë* /ə/, *i* /i/, *o* /ɔ/, *u* /u/, and *y* /y/. The vowels *i* and *o* are represented with a vertical line ɪ and a circle ο; the vowels *e* and *u* have the same base shape with a dot above: i̇ 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 ᴠ *a*, i̇ *e*, ᵮ *ë*, ɪ *i*, ο/ω *o*/*ō*, ȯ *u*, and ᵸ *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 i̇ *e* and ȯ *u* atomically, but make them canonically equivalent to ɪ *i* and ο *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 makes 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, in Group 3 are other additions to the repertoire, and in Group 4 are Greek letters.

1.	A	B	C	Ç	D	DH	E	Ë	F	G	GJ	H	I	J	K	L	LL	M	N	NJ
	ᴠ	ᗰ	ᖴ	ɋ	∧	ᖶ	i	ᵮ	ᘜ	h	ƀ	ϵ	ɪ	ɔ	ᴄ	ɋ	ᖻ	ᑢ	ᴠ	ᵰ
2.	–	MB	–	–	ND	–	–	–	–	NG	NGJ	–	–	–	–	–	–	–	–	–
	–	ℓ	–	–	ⱝ	–	–	–	–	ɦ	ƙ	–	–	–	–	–	–	–	–	–
3.	AS	–	–	–	–	–	–	–	–	–	–	HJ	–	–	–	–	–	–	–	–
	ᵴ	–	–	–	–	–	–	–	–	–	–	ɜ	–	–	–					

1.	O	P	Q	R	RR	S	SH	T	TH	U	V	X	XH	Y	Z	ZH	–	–	–	–
	ο	ᖹ	⅋	ɔ	ſ	ƨ	ʎ	ɋ	ʔ	ȯ	ℓ	ɣ	ɡ	ɦ	ᐯ	ẟ	–	–	–	–
2.	–	–	–	–	–	–	–	–	–	–	–	NX	NXH	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–	–	–	ƶ	ɡ	–	–	–	–	–	–	–
3.	–	–	–	–	–	–	SHT	–	–	–	–	–	–	JY	–	–	–	–	–	–
	–	–	–	–	–	–	ʎ	–	–	–	–	–	–	ƴ	–	–	–	–	–	–
4.	–	–	–	–	–		–	–	–	–	–	–	–	–	GH	ST	SK	KH	PS	OO
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	ᐯ	ƽ	ƨ	X	ɥ	ω

3.1.1 Group 2. The prenasalized characters are by and large modifications of the base letters. Prenasalized *b* looks a bit like *m* (ᑢ M + ᗰ B = ℓ MB); prenasalized *d* fuses both (ᴠ N + ∧ D = ⱝ ND); prenasalized *g* and *gj* both share an extension (ᴠ N + ɦ G = ɦ NG; ᴠ N + ƀ GJ = ƙ NGJ); prenasalized *x* and *xh* share a similar extension (ᴠ N + ɣ X = ƶ NX; ᴠ N + ɡ XH = ɡ NXH).

3.1.2 Group 3. The 4 letters here are a little less clear in their derivation. 𐑎 AS might simply be 𐑊 + 𐑧; 𐑒 HJ is a turned 𐑑 H (von Hahn gives 𐑧 "dark ch" [x] and 𐑒 "light ch" [ç]); 𐑖 SHT is clearly derived from 𐑖 SH + 𐑑 T and is attested many times throughout the texts; and 𐑘 JY (Elsie writes *ju/jy*; 𐑧𐑘̄𐑑 *sjýt* appears on p. 29 of the MS) would appear to be 𐑦 I + 𐑔 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 𐑒 HJ, though this may be an oversight (modifying 𐑑 [h] to 𐑒 [ç] is logical, though /ç/ is not an Albanian phoneme). Elsie's chart also includes 𐑘 JY which von Hahn does not give, but von Hahn gives a letter 𐑯 *të* which is also clearly 𐑑 T + 𐑩 E but is similar to that in shape (see 3.2 below). The text which contains 𐑖 SHT usually writes *të* as 𐑑𐑩 (not 𐑯); see Figures 8 and 9. As to whether SHTA should be encoded on its own, the recommendation is that it should be. Certainly both 𐑖 SHA and 𐑑 TA both exist; I have not yet seen an instance of a sequence "𐑖𐑑 SHT" though the "ligature" is very common (*e.g.*, 𐑣𐑦𐑖𐑩 *hrishtë,* 𐑒𐑦𐑖𐑩 *krishtë* 'Christ'). Compare Glagolitic which has 𐰩 SHA, 𐰢 TE, and 𐰮 SHTA (also clearly a ligature in origin). Users of Todhri are scholars, and they can deal with searching issues revolving around 𐑖𐑑 and 𐑖 , if any, given the relatively small size of the corpus.

3.1.3 Group 4. The 6 letters here answer to Greek characters: 𐑜 OO = ω OMEGA; 𐑙 PS = ψ PSI; 𐑔 ST = ς STIGMA; 𐑝 GH = γ GAMMA; 𐑟 KH = χ CHI. The letter 𐑨 SK may or may not be related to ξ XI but certainly was /sk/ and not /ks/—compare the Greek loanwords 𐑧𐑒𐑯𐑤𐑯𐑛𐑰𐑧 *skandhalís* 'to stumble' (σκανδάλης) and 𐑤𐑣𐑯𐑧𐑦 *dhidháski* 'teaching' (διδάσκαλος). Note that despite the fact that the printed source charts (Figures 1, 2, and 3) show the OO ω with a horizontal bar above (ω̄), 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 𐑜 𐑝𐑷𐑑 at the beginning of line 15 of Figures 4 and 5; it is given as 𐑜 𐑝𐑷𐑑 *ō zót* in Figure 6 and as 𐑪 𐑝𐑷𐑑 *o zót* without an omega in Figure 8.

3.2 Ligature *të.* A ligature 𐑯 *të* 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 𐑑𐑩 is common. Fonts may choose to include an optional ligature for 𐑯 *të*.

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 EI, and *ë* written E. Here are the correspondences:

Todhri	*UCS*	*von Hahn*	*Todhri*	*UCS*	*von Hahn*	*Todhri*	*UCS*	*von Hahn*	*Todhri*	*UCS*	*von Hahn*
𐑊	A	*a*	𐑞	NGA	*ngha*	𐑪	O	*o*	𐑲	THA	*θa*
𐑎	AS	*as*	𐑚	GJA	*gja*	𐑜	OO	*ω*	𐑫	U	*u*
𐑐	BA	*ba*	𐑡	NGJA	*ngja*	𐑤	PA	*pa*	𐑝	VA	*wa*
𐑚	MBA	*mba*	𐑒	HA	*ha*	𐑙	PSA	*psa*	𐑗	XA	*dsa*
𐑑	CA	*tsa*	𐑓	HJA	*chja*	𐑕	QA	*kja*	𐑨	NXA	*ndsa*
𐑕	CHA (ÇA)	*tscha*	𐑦	I	*i*	𐑮	RA	*ra*	𐑡	XHA	*dscha*
𐑛	DA	*da*	𐑘	JA	*ja (deutsch. j)*	𐑠	RRA	*rra*	𐑩	NXHA	*ndscha*
𐑯	NDA	*nda*	𐑒	KA	*ka*	𐑧	SA	*sa*	𐑔	Y	*ü*
𐑣	DHA	*δa*	𐑤	LA	*lja*	𐑨	SKAN	*xan*	𐑘	JY	**jü*
𐑦	EI	*e (offen e)*	𐑣	LLA	*la*	𐑔	STA	*sta*	𐑟	ZA	*ζa*
𐑩	E (Ë)	*e (stumm e)*	𐑒	MA	*ma*	𐑖	SHA	*scha*	𐑘	ZHA	*ja (franz. j)*
𐑓	FA	*fa*	𐑮	NA	*na*	𐑖	SHTA	*schta*	𐑝	GHA	*γa*
𐑣	GA	*gha*	𐑯	NJAN	*njan*	𐑑	TA	*ta*	𐑟	KHA	*cha*

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. For consistency with the already-encoded Elbasan script, the letters GHA and KHA are given at the end of the alphabet, along with the other Greek letter borrowings in their expected alphabetical order. All Todhri letters are sorted distinctly at the first level:

𐗀 < 𐗁 < 𐗂 < 𐗃 < 𐗄 < 𐗅 < 𐗆 < 𐗇 < 𐗈 < 𐗉 < 𐗊 < 𐗋 < 𐗌 < 𐗍 < 𐗎 <
𐗏 < 𐗐 < 𐗑 < 𐗒 < 𐗓 < 𐗔 < 𐗕 < 𐗖 < 𐗗 < 𐗘 < 𐗙 < 𐗚 < 𐗛 < 𐗜 < 𐗝 <
𐗞 < 𐗟 < 𐗠 < 𐗡 < 𐗢 < 𐗣 < 𐗤 < 𐗥 < 𐗦 < 𐗧 < 𐗨 < 𐗩 < 𐗪 < 𐗫 < 𐗬 <
𐗭 < < 𐗯 < 𐗰

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 EI;Ll;0;L;;;;;N;;;;;
105CA;TODHRI LETTER E;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 HA;Ll;0;L;;;;;N;;;;;
105D1;TODHRI LETTER HJA;Ll;0;L;;;;;N;;;;;
105D2;TODHRI LETTER I;Ll;0;L;;;;;N;;;;;
105D3;TODHRI LETTER JA;Ll;0;L;;;;;N;;;;;
105D4;TODHRI LETTER KA;Ll;0;L;;;;;N;;;;;
105D5;TODHRI LETTER LA;Ll;0;L;;;;;N;;;;;
105D6;TODHRI LETTER LLA;Ll;0;L;;;;;N;;;;;
105D7;TODHRI LETTER MA;Ll;0;L;;;;;N;;;;;
105D8;TODHRI LETTER NA;Ll;0;L;;;;;N;;;;;
105D9;TODHRI LETTER NJAN;Ll;0;L;;;;;N;;;;;
105DA;TODHRI LETTER O;Ll;0;L;;;;;N;;;;;
105DB;TODHRI LETTER PA;Ll;0;L;;;;;N;;;;;
105DC;TODHRI LETTER QA;Ll;0;L;;;;;N;;;;;
105DD;TODHRI LETTER RA;Ll;0;L;;;;;N;;;;;
105DE;TODHRI LETTER RRA;Ll;0;L;;;;;N;;;;;
105DF;TODHRI LETTER SA;Ll;0;L;;;;;N;;;;;
105E0;TODHRI LETTER SHA;Ll;0;L;;;;;N;;;;;
105E1;TODHRI LETTER SHTA;Ll;0;L;;;;;N;;;;;
105E2;TODHRI LETTER TA;Ll;0;L;;;;;N;;;;;
105E3;TODHRI LETTER THA;Ll;0;L;;;;;N;;;;;
105E4;TODHRI LETTER U;Ll;0;L;;;;;N;;;;;
105E5;TODHRI LETTER VA;Ll;0;L;;;;;N;;;;;
105E6;TODHRI LETTER XA;Ll;0;L;;;;;N;;;;;
105E7;TODHRI LETTER NXA;Ll;0;L;;;;;N;;;;;
105E8;TODHRI LETTER XHA;Ll;0;L;;;;;N;;;;;
105E9;TODHRI LETTER NXHA;Ll;0;L;;;;;N;;;;;
105EA;TODHRI LETTER Y;Ll;0;L;;;;;N;;;;;
105EB;TODHRI LETTER JY;Ll;0;L;;;;;N;;;;;
105EC;TODHRI LETTER ZA;Ll;0;L;;;;;N;;;;;
105ED;TODHRI LETTER ZHA;Ll;0;L;;;;;N;;;;;
105EE;TODHRI LETTER GHA;Ll;0;L;;;;;N;;;;;
105EF;TODHRI LETTER STA;Ll;0;L;;;;;N;;;;;
105F0;TODHRI LETTER SKAN;Ll;0;L;;;;;N;;;;;
105F1;TODHRI LETTER KHA;Ll;0;L;;;;;N;;;;;
105F2;TODHRI LETTER PSA;Ll;0;L;;;;;N;;;;;
105F3;TODHRI LETTER OO;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.

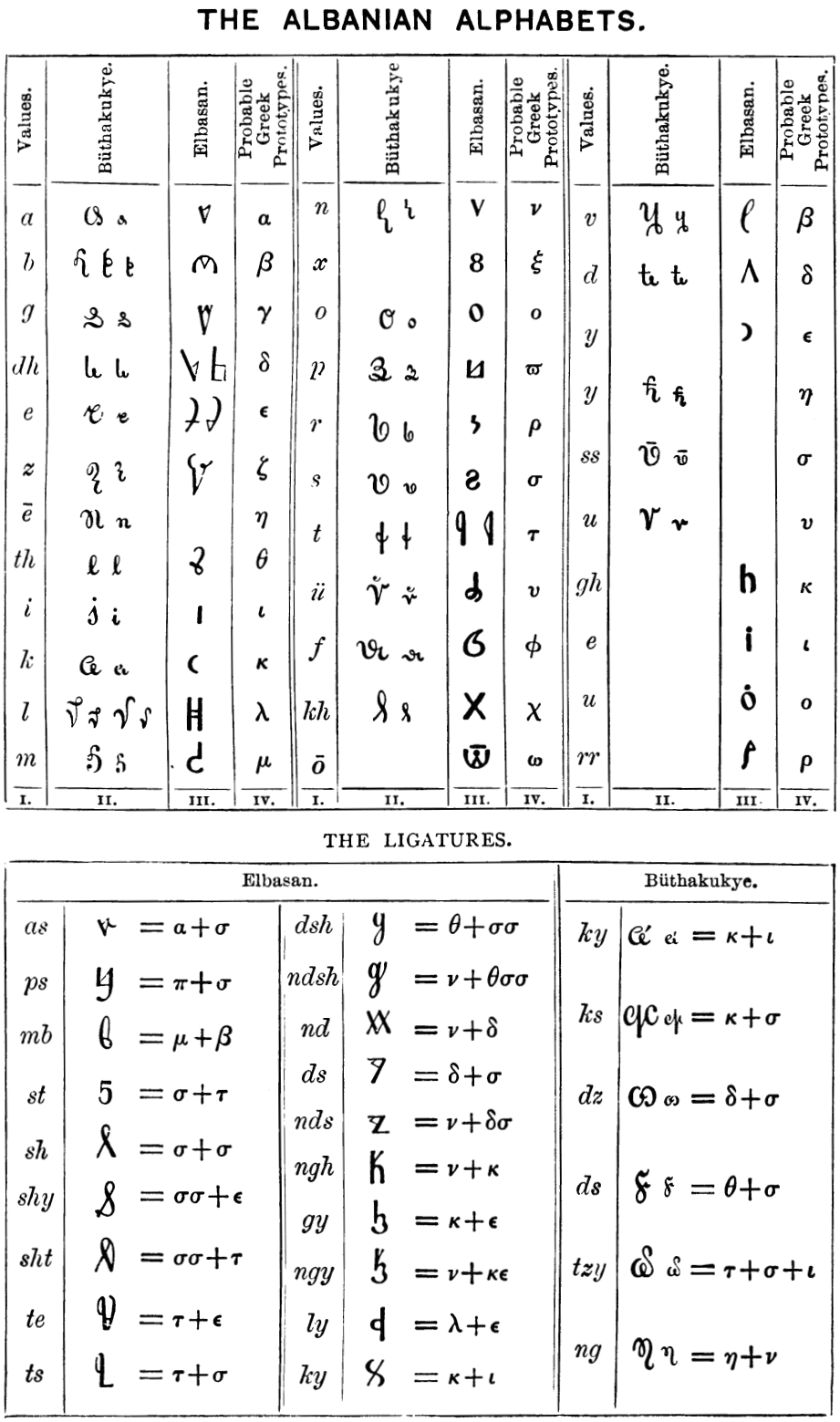
9. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Todhri encoding. Any views, findings, conclusions or recommendations 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.		*A*	a	19.		*Ra*	r	37.		*θa*	griech. θ
2.		*E*	offen e	20.		*Rra*	rr	38.		*Ba*	b
3.		*I*	i	21.		*Fa*	f	39.		*Mba*	mb
4.		*O*	o	22.		*Δa*	griech. δ	40.		*Pa*	p
5.		*U*	u	23.		*Ma*	m	41.		*Na*	n
6.		*Ü*	ü	24.		*Ja*	deutsch. j	42.		*Tscha*	tsch
7.		*E*	stumm e	25.		*Gha*	gh	43.		*Dscha*	dsch
8.		*Sa*	s	26.		*Ngha*	ngh	44.		*Ndscha*	ndsch
9.		*ζa*	griech. ζ	27.		*Gja*	gj	45.		*Sta*	st
10.		*Tsa*	ts	28.		*Ngja*	ngj	46.		*Scha*	sch
11.		*Dsa*	ds	29.		*Γa*	griech. γ	47.		*Ja*	franz. j
12.		*Ndsa*	nds	30.		*Psa*	ps	48.		*Schta*	scht
13.		*Wa*	w	31.		*Ha*	h	49.		*Te*	te
14.		*La*	l	32.		*Cha*	dunkel ch	50.		*Njan*	nj
15.		*Lja*	lj	33.		*Chja*	hell ch	51.		*As*	as
16.		*Kja*	kj	34.		*Ta*	t	52.		*Ω*	griech. ω
17.		*Ka*	k	35.		*Da*	d				
18.		*Xan*	x	36.		*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			α	*n*			ν	*v*			β
b			β	*x*			ξ	*d*			δ
g			γ	*o*			ο	*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*			ο
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	= α+σ	*dsh*	= θ+σσ	*ky*	= κ+ι
ps	= π+σ	*ndsh*	= ν+θσσ	*ks*	= κ+σ
mb	= μ+β	*nd*	= ν+δ	*dz*	= δ+σ
st	= σ+τ	*ds*	= δ+σ	*ds*	= θ+σ
sh	= σ+σ	*nds*	= ν+δσ	*tzy*	= τ+σ+ι
shy	= σσ+ε	*ngh*	= ν+κ	*ng*	= η+ν
sht	= σσ+τ	*gy*	= κ+ε		
te	= τ+ε	*ngy*	= ν+κε		
ts	= τ+σ	*ly*	= λ+ε		
		ky	= 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: *as* (*a*+ *s*), *ps* (*p* + *s*), *mb* (*m* + *b*), *st* (*s* + *t*—correctly Greek ς as derived from ς), *sh* (*s* + *s*), *zh* (*sh* + *ë*), *sht* (*sh* + *t*), *të* (*t* + *ë*), *c* (*t* + *s*), *xh* (*d* + *sh*), *nxh* (*n* + *xh*), *nd* (*n* + *d*), *x* (*d* + *s*), *nx* (*n* + *x*), *ng* (*n* + *k*—correctly *n* + *g*), *gj* (*k* + *ë*—correctly perhaps *g* + *ë*), *ngj* (*n* + *gj*), *l* (*ll* + *ë*—*l* is /l/ and *ll* is /ł/), *q* (*k* + *i*).

The Todhri Alphabet

Modern Alb.	Todhri
a	
b	
c	
ç	
d	
dh	
e	
ë	
f	
g	
gj	
h	
i	
j	
k	
l	
ll	
m	
n	
nj	
o	
ō	
p	
q	
r	
rr	
s	
sh	

Modern Alb.	Todhri
t	
th	
u	
v	
x	
xh	
y	
z	

Ligatures

ju/jy	
mb	
nd	
ng	
ngj	
nx	
ps	
sht	
sk	

Rare letters

nxh	
st	
as	
γ (Greek)	
χ (Greek)	
zh	

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

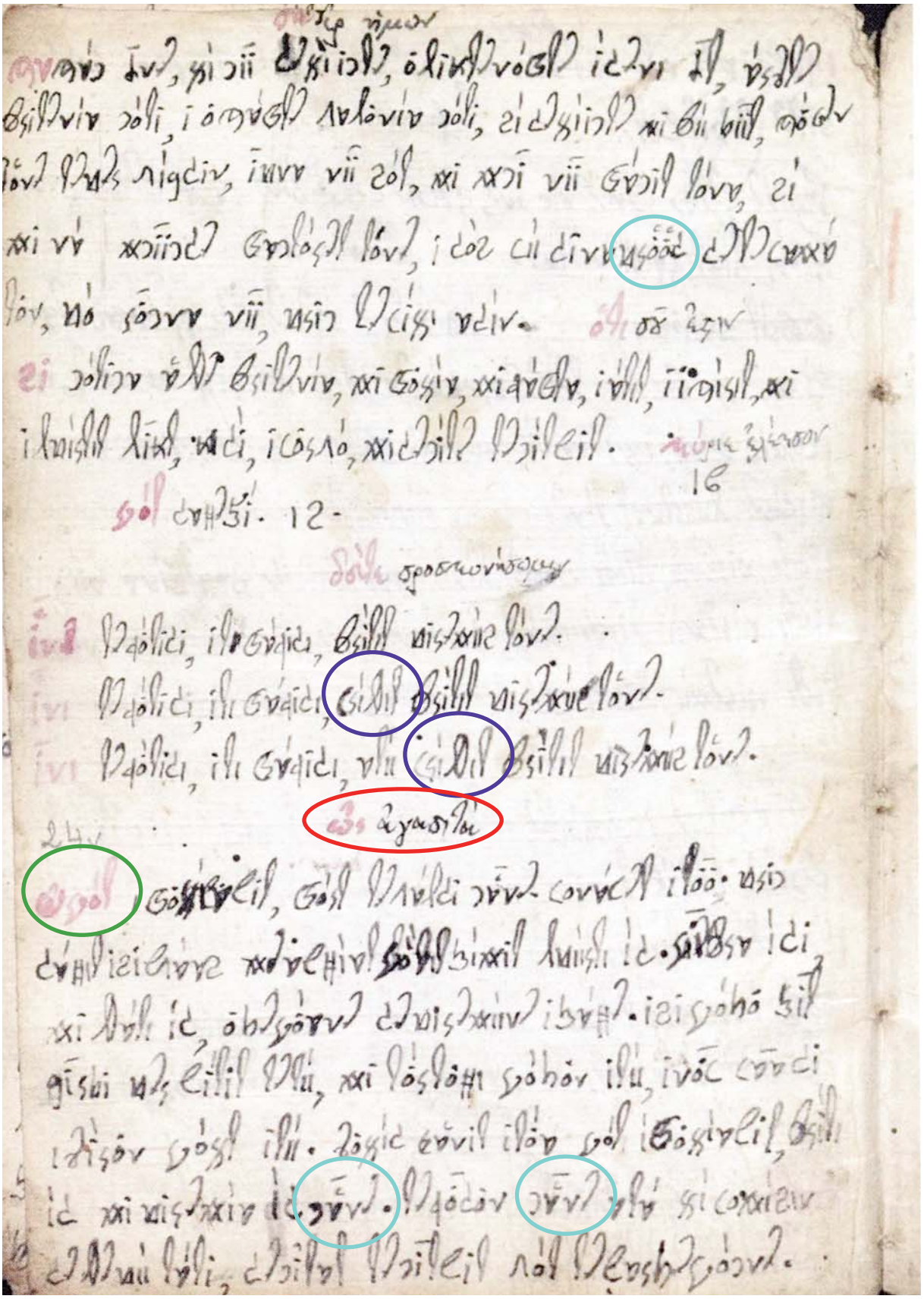


Figure 4. The Trisagion Prayers; The Lord's Prayer; Psalms 84: 1-4 from Elsie 2016:98. Circled is ѡ̂ ѵ́ѳ́ *ō zót* 'O God" (in green). In the Greek heading (in red) above Psalm 84 reads ὡς ἀγαπιτὰ (evidently for ὡς ἀγαπητὰ 'as dear'); the scribe writes both rough and smooth breathings identically. In purple are two instances of *kríshtit* showing the letter SHTA; in cyan are two instances of *jánë* showing written as ꙗ̄ѵ̄ѵ *jā̄në* and one instance of *prūūm* written as ꙗꙋ̀ꙋ̀ꙋ̀ꙋ *prûûm*.

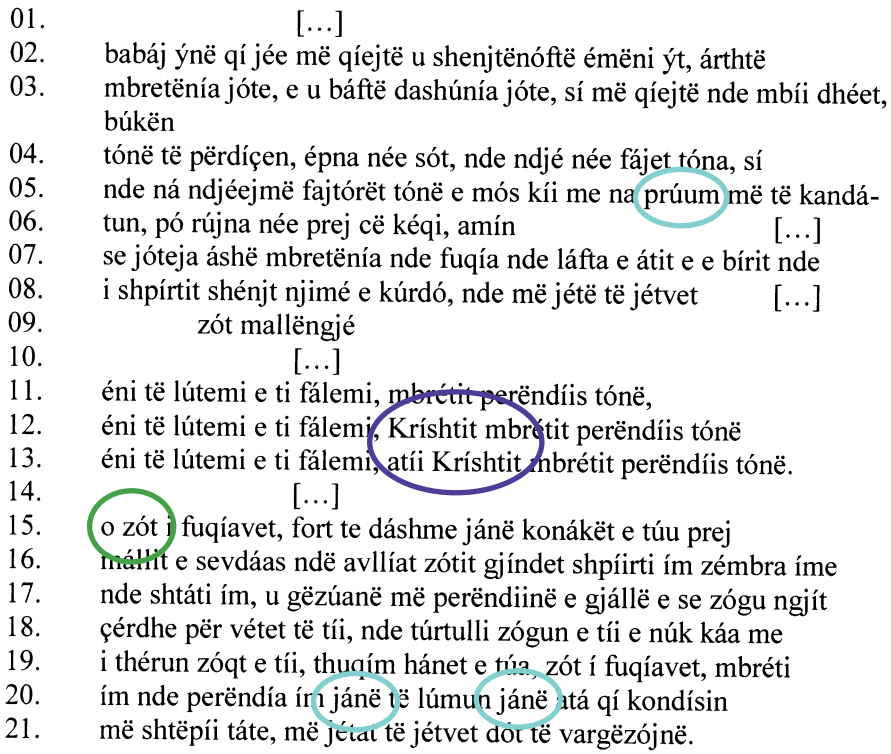

01. […]
02. babáj ýnë qí jée më qíejtë u shenjtënóftë émëni ýt, árthtë
03. mbretënía jóte, e u báftë dashúnía jóte, sí më qíejtë nde mbíi dhéet, búkën
04. tónë të përdíçen, épna née sót, nde ndjé née fájet tóna, sí
05. nde ná ndjéejmë fajtórët tónë e mós kíi me na prúum më të kandá-
06. tun, pó rújna née prej së 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, Krishtit mbrétit perëndíis tónë
13. éni të lútemi e ti fálemi, atíi Krishtit 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 dot të vargëzójnë.

Figure 5a. Transcription of Figure 4 above from Elsie 2016:99.

1 Πάτερ ἡμῶν
2. ⱺⱲⱺⲟ ⰹⲛⲋ ⳾ⲓ ⲟⲓ ⲇⲋ ⳾ⲓⲟⲩⲋ ⲟ ⲗⲓⱳⲩⲋⲛⲟⳭⲩ ⲓⲇⲋⲛⲓ ⰹⲩ, ⲛⲟⲋⲩⲋ
3. ⳿ⲟⲓⲩⲋⲛⲓⲛ ⲟⲟⲩi, i ⲟ ⱺⲛⳬⲩ ⲁⲛⲗⲟⲛⲓⲛ ⲟⲟⲩi, ⲟⲓ ⲇⲋ ⳾ⲓⲟⲩⲋ ⲭi ⳿ⲓⲓ ⲃⲓⲩ, ⱺⲟⲇⲋⲛ
4. ⲩⲟⲛⲋ ⲩⲋ ⲃⲋⲟⲗⲓⲫiⲛ, ⲓⲃⲛⲛ ⲛⲓⲓ ⲟⲟⲩ, ⲭi ⲭⲟⲓ ⲛⲓⲓ Ⳍⲛⲟⲓⲩ ⲩⲟⲛⲛ, ⲟⲓ
5. ⲭi ⲛⲛ ⲭⲟⲓⲓⲟⲇⲋ Ⳍⲛⲟⲩⲟⲟⲋⲩ ⲩⲟⲛⲋ i ⲇⲟⲟ ⲇⲓ ⲇi ⲛⲛ ⲃⲟⲟⲇ ⲇⲋ ⲩⲋ ⲥⲛⲭⲛ-
6. ⲩⲟⲛ, ⲃⲟ ⲟⲟⲛⲛ ⲛⲓⲓ ⲃⲟⲓⲟ ⲩⲋ ⲇⲟⳭⲓ, ⲛⲇⲓⲛ [Greek]
7. ⲟⲓ ⲟⲟⲩⲓⲟⲛ ⲛⲙⲋ ⳿ⲟⲓⲩⲋⲛⲓⲛ ⲭi Ⳍⲟⲟ⳾ⲓⲛ ⲭi ⲫⲛⳬⲩⲛ i ⲛⲩⲓⲩ i i ⲟⲓⲟⲓⲩ ⲭi
8. ⲓ ⲗⲃⲓⲟⲓⲩ ⲗⲓⲛⲩ ⲛⲓⲇⲓ i ⲥⲟⲟⲗⲟ, ⲭi ⲇⲋ ⲟⲓⲩⲋ ⲩⲋ ⲟⲓⲟⲓⲩ [κύριε ἐλέισον (*for* ἐλέησον)]
9. ⲙⲟⲩ ⲇⲛⲂⲋⲃⲓ
10. [Greek]
11. ⲓⲛⲓ ⲩⲋ ⲫⲟⲩⲓⲇⲓ i ⲩⲓ Ⳍⲛⲫⲓⲇⲓ, ⳿ⲟⲓⲩⲓ ⲃⲓⲟⲋⲭⲓⲓⲃ ⲩⲟⲛⲋ,
12. ⲓⲛⲓ ⲩⲋ ⲫⲟⲩⲓⲇⲓ i ⲩⲓ Ⳍⲛⲫⲓⲇⲓ, ⲥⲓⲗⲓⲩ ⳿ⲟⲓⲩⲓ ⲃⲓⲟⲋⲭⲓⲓⲃ ⲩⲟⲛⲋ
13. ⲓⲛⲓ ⲩⲋ ⲫⲟⲩⲓⲇⲓ i ⲩⲓ Ⳍⲛⲫⲓⲇⲓ, ⲛⲩⲓ ⲥⲓⲗⲓⲩ ⳿ⲟⲓⲩⲓ ⲃⲓⲟⲋⲭⲓⲓⲃ ⲩⲟⲛⲋ.
14. [ὡς ἀγαπητὰ]
15. ⲱ ⲙⲟⲩ ⲓ Ⳍⲟⲟ⳾ⲓⲙ⳿ⲓⲩ, Ⳍⲟⲃⲩ ⲩⲓ ⲗⲙⲗⲇⲓ ⲟⲛⲛⲙⲇⲋⲩ i ⲩⲟⲟ ⲃⲟⲓ
16. ⲇⲙⲂⲓⲩ i ⲃⲓ⳿ⲗⲙⲙⲃ ⲭⲋ ⲛ⳿Ⲃⲓⲛⲩ ⲙⲟⲩⲓⲩ Ⳮⲟⲩⲓⲩ ⲗⲃⲓⲟⲃⲩⲓ ⲓⲇ ⲙⲓⲃⲃ ⲓⲇⲓ
17. ⲭi ⲗⲙⲩⲓ ⲓⲇ, ⲟ ⲏⲋⲙⲟⲙⲛⲋ ⲇⲋ ⲃⲓⲟⲋⲭⲓⲓⲛⲋ i ⳌⲙⲂⲋ i ⲃⲓ ⲙⲟⲏⲟ Ⳍⲓⲩ
18. ⲩⲓⲗⲓ ⲃⲋⲟ ⳿ⲓⲩⲓⲩ ⲩⲋ ⲩⲓⲓ, ⲭi ⲩⲟⲟⲟⲩⲟⲫⲓ ⲙⲟⲏⲟⲛ i ⲩⲓⲓ i ⲛⲟⲇ ⲥⲛⲛⲛ ⲇⲓ
19. ⲓ ⲋⲓⲟⲟⲛ ⲙⲟ⳾ⲩ i ⲩⲓⲓ, ⲋⲟ⳾ⲓⲇ ⲉⲙⲛⲓⲩ i ⲩⲟⲛ, ⲙⲟⲩ ⲓ Ⳍⲟⲟ⳾ⲓⲙ⳿ⲓⲩ, ⳿ⲟⲓⲩ
20. ⲓⲇ ⲭi ⲛⲓⲟⲋⲭⲓⲙ ⲓⲇ ⲟⲛⲛⲋ ⲩⲋ ⲫⲟⲇⲟⲛⲛ ⲟⲛⲛⲋ ⲛⲩⲙ ⳾ⲓ ⲥⲟⲭⲓⲟⲓⲛ
21. ⲇⲋ ⲗⲋⲛⲓⲓ ⲩⲙⲩⲓ, ⲇⲋ ⲟⲓⲩⲛⲩ ⲩⲋ ⲟⲓⲩⲓⲩ ⲗⲟⲩ ⲩⲋ ⲉⲟⲟⲏⲋⲙⲙⲟⲛⲋ.

Figure 5b. Acute-accented vowels are marked in magenta; vowels with macron are marked in blue.

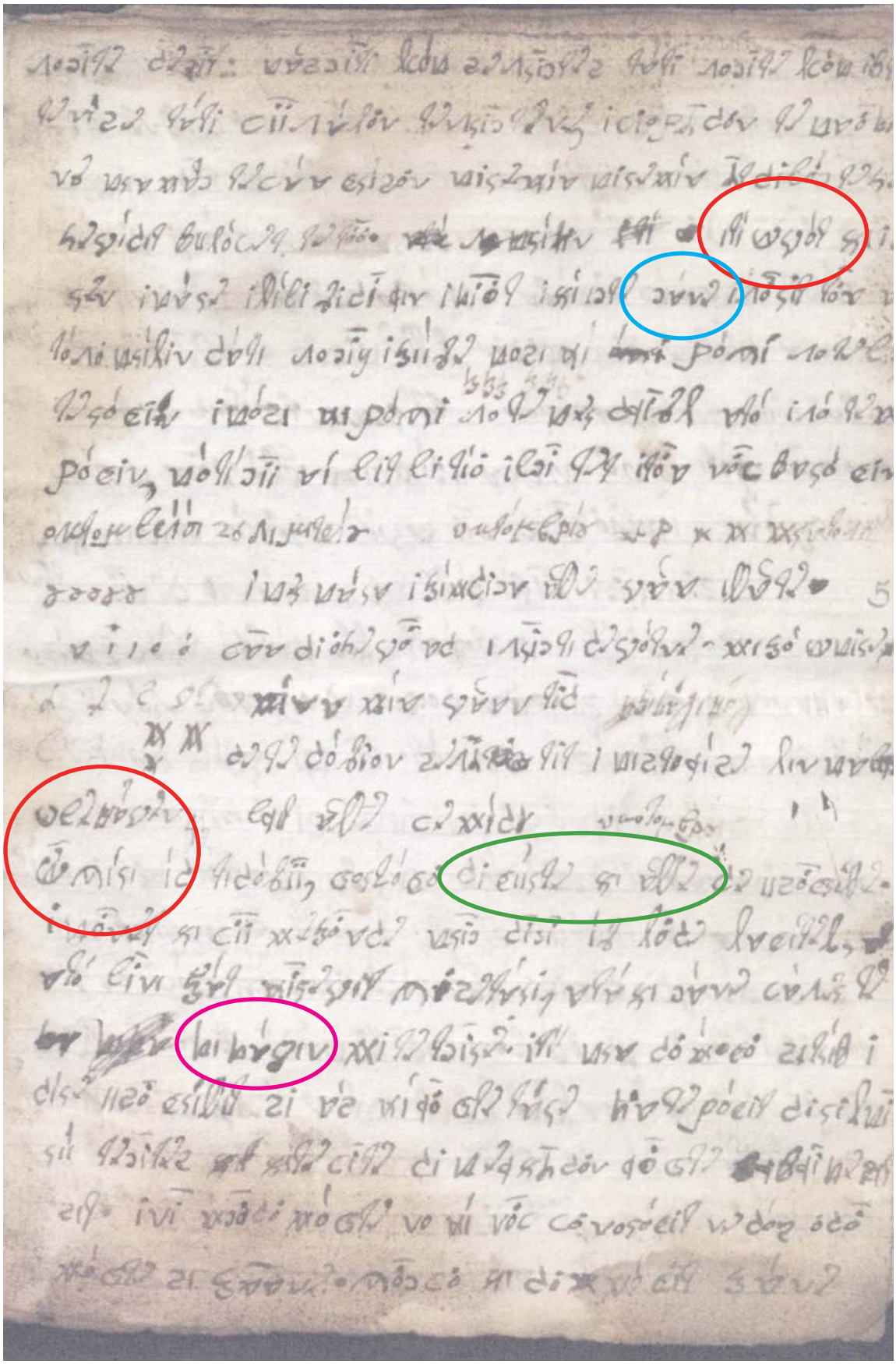
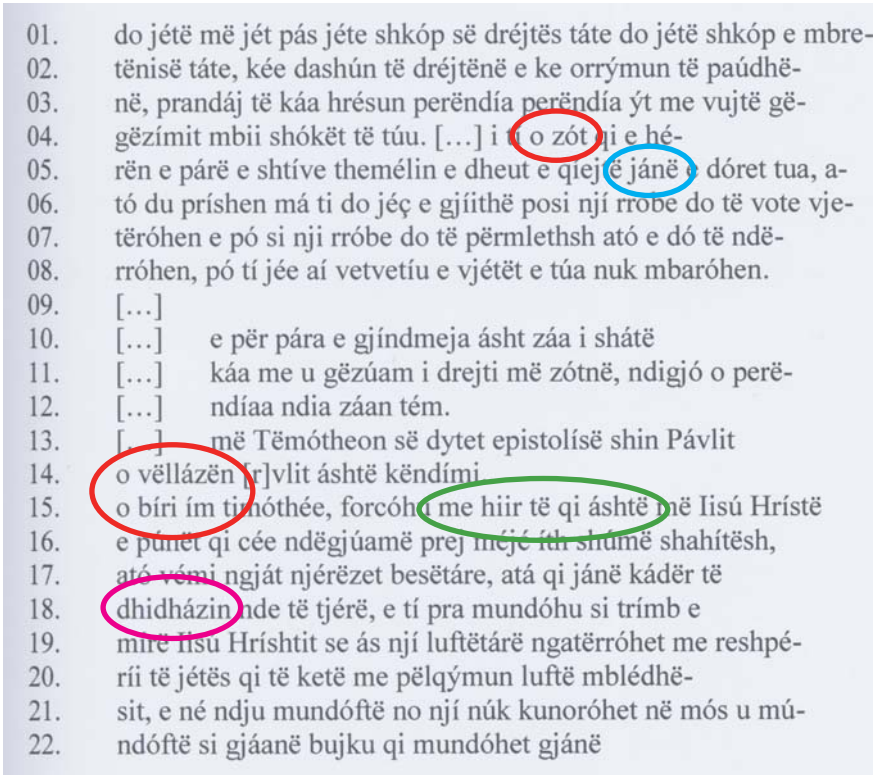


Figure 6. Hebrews 1:8–12; 2 Timothy 2:1–6 from Elsie 2016:30. Circled in red are ꙍ ꙋ́ꙋ́ꙗ *ō zót* ‘O God’, ꙍ ꙋꙗꙗ́ꙋ́ꙗꙋ *ō vëllázën* ‘O brothers’, and ꙍ̄ ꙍ́ꙋꙋ *ṓ bíri* ‘O my son’; in green ꙇ́ꙋ ꙋꙋ ꙋꙋ ꙋ́ꙋꙗ *me hiir të qi ashtë* ‘in the grace that is’ with TA followed by E (ꙗꙗ) and SHTA followed by E (ꙋꙗ). Circled in cyan is ꙋ́ꙋꙋꙗ *jánë* ‘they are’ and circled in magenta is ꙋꙋꙋ́ꙋꙋꙋ *dhidházin* ‘teaching’ (recte ꙋꙋꙋ́ꙋꙋꙋ *dhidhaskin*?).



01. do jétë më jét pás jéte shkóp së dréjtës táte do jétë shkóp e mbre-
02. tënisë táte, kée dashún të dréjtënë e ke orrýmun të paúdhë-
03. në, prandáj të káa hrésun perëndía perëndía ýt me vujtë gë-
04. gëzímit mbii shókët të túu. […] i ti o zót qi e hé-
05. rën e párë e shtíve themélin e dheut e qiejtë jánë e dóret tua, a-
06. tó du príshen má ti do jéç e gjíithë posi njí rróbe do të vote vje-
07. tëróhen e pó si nji rróbe do të përmlethsh ató e dó të ndë-
08. rróhen, pó tí jée aí vetvetíu e vjétët e túa nuk mbaróhen.
09. […]
10. […] e për pára e gjíndmeja ásht záa i shátë
11. […] káa me u gëzúam i drejti më zótnë, ndigjó o perë-
12. […] ndíaa ndia záan tém.
13. […] më Tëmótheon së dytet epistolísë shin Pávlit
14. o vëllázën [r]vlit áshtë këndími
15. o bíri ím timóthée, forcóhu me hiir të qi áshtë më Iisú Hrístë
16. e púnët qi cée ndëgjúamë prej méjé íth shúmë shahítësh,
17. ató vémi ngját njérëzet besëtáre, atá qi jánë kádër të
18. dhidházin nde të tjérë, e tí pra mundóhu si trímb e
19. mirë Iisú Hríshtit se ás njí luftëtárë ngatërróhet me reshpé-
20. ríi të jétës qi të ketë me pëlqýmun luftë mblédhë-
21. sit, e né ndju mundóftë no njí núk kunoróhet në mós u mú-
22. ndóftë si gjáanë bujku qi mundóhet gjánë

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

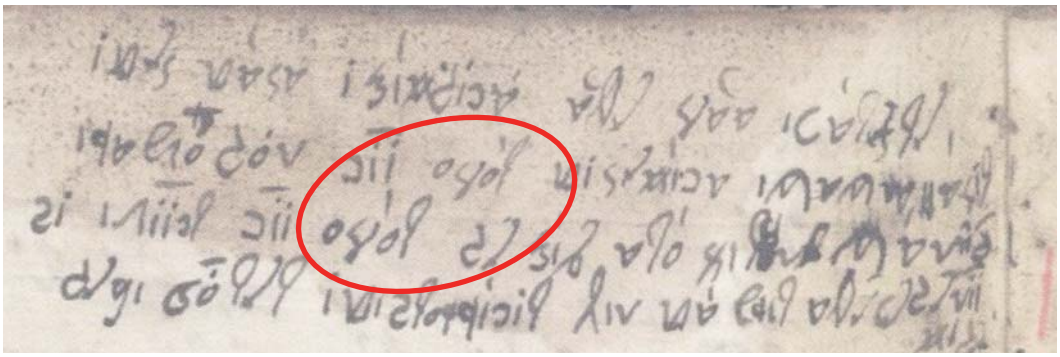
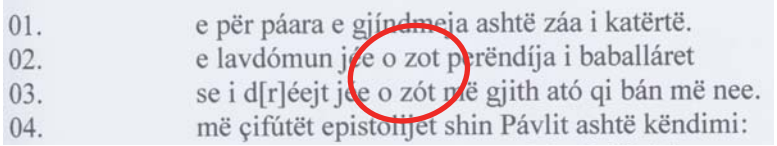


Figure 8. Hebrews 13:7–16 from Elsie 2016:20.
Circled are two instances of ο ϒόϥ *o zót* 'O God' with an omicron rather than an omega.



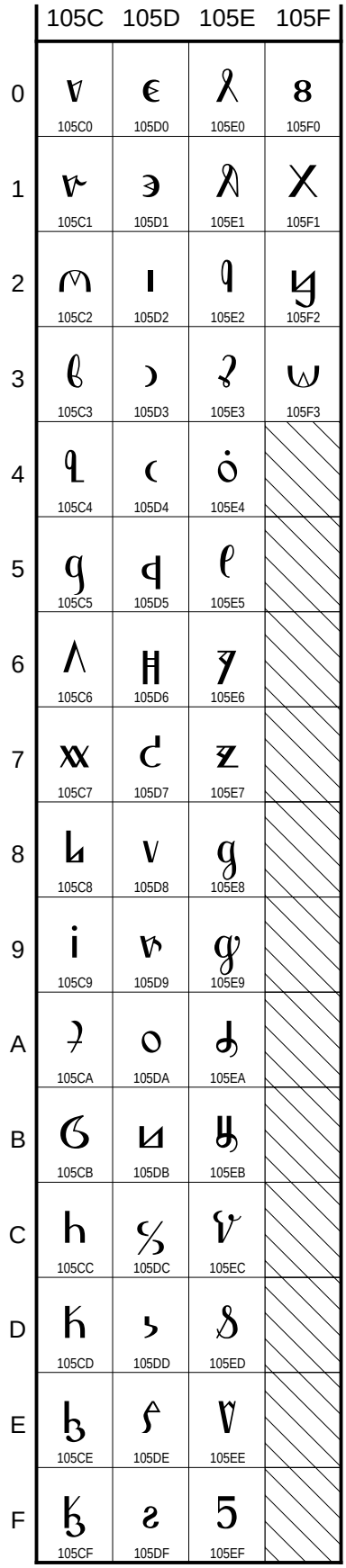
01. e për páara e gjíndmeja ashtë záa i katërtë.
02. e lavdómun jée o zot perëndíja i baballáret
03. se i d[r]éejt jée o zót më gjith ató qi bán më nee.
04. më çifútët epistolíjet shin Pávlit ashtë këndimi:

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

Figure 10. The text of John 16:1–23 transcribed by von Hahn, from Elsie 2016:104.
Circled is the Greek loanword ⲥⲕⲁⲛⲇⲁⲗⲓⲍⲓ *skandalízi* ‘to stumble’.
Also circled is ⲭⲁⲛⲏ *jânë* which might have a sort of breathing over it.

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 Krísht 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 júje 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 júje, qí atë gëzím súnd ju a márri njerií, pra m at hére mú mós

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



Albanian letters

105C0 𐞀 TODHRI LETTER A
105C1 𐞁 TODHRI LETTER AS
105C2 𐞂 TODHRI LETTER BA
105C3 𐞃 TODHRI LETTER MBA
105C4 𐞄 TODHRI LETTER CA
105C5 𐞅 TODHRI LETTER CHA
105C6 𐞆 TODHRI LETTER DA
105C7 𐞇 TODHRI LETTER NDA
105C8 𐞈 TODHRI LETTER DHA
105C9 𐞉 TODHRI LETTER EI
105CA 𐞊 TODHRI LETTER E
105CB 𐞋 TODHRI LETTER FA
105CC 𐞌 TODHRI LETTER GA
105CD 𐞍 TODHRI LETTER NGA
105CE 𐞎 TODHRI LETTER GJA
105CF 𐞏 TODHRI LETTER NGJA
105D0 𐞐 TODHRI LETTER HA
105D1 𐞑 TODHRI LETTER HJA
105D2 𐞒 TODHRI LETTER I
105D3 𐞓 TODHRI LETTER JA
105D4 𐞔 TODHRI LETTER KA
105D5 𐞕 TODHRI LETTER LA
105D6 𐞖 TODHRI LETTER LLA
105D7 𐞗 TODHRI LETTER MA
105D8 𐞘 TODHRI LETTER NA
105D9 𐞙 TODHRI LETTER NJAN
105DA 𐞚 TODHRI LETTER O
105DB 𐞛 TODHRI LETTER PA
105DC 𐞜 TODHRI LETTER QA
105DD 𐞝 TODHRI LETTER RA
105DE 𐞞 TODHRI LETTER RRA
105DF 𐞟 TODHRI LETTER SA
105E0 𐞠 TODHRI LETTER SHA
105E1 𐞡 TODHRI LETTER SHTA
• in origin a ligature of sha and ta
105E2 𐞢 TODHRI LETTER TA
105E3 𐞣 TODHRI LETTER THA
105E4 𐞤 TODHRI LETTER U
105E5 𐞥 TODHRI LETTER VA
105E6 𐞦 TODHRI LETTER XA
105E7 𐞧 TODHRI LETTER NXA
105E8 𐞨 TODHRI LETTER XHA
105E9 𐞩 TODHRI LETTER NXHA
105EA 𐞪 TODHRI LETTER Y
105EB 𐞫 TODHRI LETTER JY
105EC 𐞬 TODHRI LETTER ZA
105ED 𐞭 TODHRI LETTER ZHA

Greek letters

105EE 𐞮 TODHRI LETTER GHA
→ 03B3 γ greek small letter gamma
105EF 𐞯 TODHRI LETTER STA
→ 03DB ϛ greek small letter stigma
105F0 𐞰 TODHRI LETTER SKAN
→ 03BE ξ greek small letter ksi
105F1 𐞱 TODHRI LETTER KHA
→ 03C7 χ greek small letter chi
105F2 𐞲 TODHRI LETTER PSA
→ 03C8 ψ greek small letter psi
105F3 𐞳 TODHRI LETTER OO
→ 03C9 ω greek small letter omega

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

2020-12-31

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?