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 Elbasan script in the SMP of the UCS**
Source: **UC Berkeley Script Encoding Initiative (Universal Scripts Project)**
Authors: **Michael Everson and Robert Elsie**
Status: **Liaison Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2011-02-03**

1. Introduction. From the very first attempts to put the Albanian language to writing, clerics and intellectuals knew that the neighbouring peoples were in possession of distinct writing systems which had helped their respective cultures and literatures to advance. The Greeks had always had a distinct alphabet for their language and the Balkan Slavs had developed their own writing systems: Glagolitic and then Cyrillic, which first flourished at Ohrid, less than one hundred kilometers from Elbasan. The Turkish occupants had also introduced a distinct, new alphabet which they had themselves borrowed from their Arabic and Persian neighbours. It was from the central Albanian Christians that the first original alphabets were created in the period 1750 to 1850. The earliest of these original alphabets, and at the same time the best adapted of them all, was that created for the so-called Elbasan Gospel Manuscript.

The Elbasan Gospel Manuscript, known in Albanian for want of a better term as the *Anonimi i Elbasanit* ('The Anonymous of Elbasan'), is a tiny and quite unique manuscript now preserved at the State Archives in Tiranë which evinces a revolutionary attempt to solve the alphabet dilemma. This 10 x 7 cm manuscript, consisting of 30 unnumbered brown folios, records the earliest-known Albanian-language text in an original alphabet. With the exception of the short fifteenth-century Easter Gospel or Pericope1, it is the oldest work of Albanian Orthodox literature and the oldest Orthodox Bible translation of all. The 59 pages of biblical texts contained in the Elbasan Gospel Manuscript, a total of 6,113 words, were written in an alphabet of forty letters. Thirty-five letters recur normally in the text and five letters can be considered rare or secondary. Though there is a distinctly Greek flavour to some of the characters and a possibly Slavic flavour to others, most of the letters in this alphabet would seem to be new creations, uninfluenced by neighbouring languages and scripts.

2. Processing. Elbasan is a simple alphabetic script written from left to right horizontally. There is no ligation. Three characters have an inherent diacritical dot: ⌣ NDE is used to indicate a pre-nasalized ⌣ /d/; ⋏ LLE is used to indicate geminate ⋏ /l/; ⌣ RRE is used to indicate geminate ⌣ /r/. In many cases the dot on NDE is written like a small N NE. In one instance in the manuscript φ GJE is written with a dot above to indicate prenasalized /ɟ/. Two different letters are used for /n/: N NE is used generally, and ɹ ALPHOID NE is typically used in prenasalized position as in ɹъ /ng/ and ɹφ /nɟ/. Two different letters are used for /ɣ/, ꜰ GHE and ʟ GHAMMA. The manuscript shows some sporadic use of what looks like Greek breathing and/or accent marks. These are not used regularly in the orthography and a complete analysis of their usage has not yet been carried out by scholars. In any case, the generic combining characters should be used for these; they are not, strictly speaking, part of the script itself.

3. Character names. The names 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. The term NA has been introduced on the basis of the character shape ⱥ (because it looks somewhat like a Latin A) to distinguish it from N NE.

4. Numerals. Script-specific numerals are not known. In the manuscript two numbers are given: ιβ̄ 12 appears once, and λ̄ 30 appears twice. It is clear that it is the Greek numeral system in use here, where ι′ = 10, β′ = 2, and λ′ = 30, with the COMBINING OVERLINE used to indicate the numeric use of the numbers.

5. Punctuation. In the manuscript some evidence of spaces and a separating dot can be seen. These would be generic, not script-specific punctuation.

6. Ordering. Ordering is as in the code chart, and follows modern Latin Albanian alphabetic order.

7. Unicode Character Properties

```
10500;ELBASAN LETTER A;Lo;0;L;;;;;N;;;;;
10501;ELBASAN LETTER BE;Lo;0;L;;;;;N;;;;;
10502;ELBASAN LETTER CE;Lo;0;L;;;;;N;;;;;
10503;ELBASAN LETTER CHE;Lo;0;L;;;;;N;;;;;
10504;ELBASAN LETTER DE;Lo;0;L;;;;;N;;;;;
10505;ELBASAN LETTER NDE;Lo;0;L;;;;;N;;;;;
10506;ELBASAN LETTER DHE;Lo;0;L;;;;;N;;;;;
10507;ELBASAN LETTER EI;Lo;0;L;;;;;N;;;;;
10508;ELBASAN LETTER E;Lo;0;L;;;;;N;;;;;
10509;ELBASAN LETTER FE;Lo;0;L;;;;;N;;;;;
1050A;ELBASAN LETTER GE;Lo;0;L;;;;;N;;;;;
1050B;ELBASAN LETTER GJE;Lo;0;L;;;;;N;;;;;
1050C;ELBASAN LETTER HE;Lo;0;L;;;;;N;;;;;
1050D;ELBASAN LETTER I;Lo;0;L;;;;;N;;;;;
1050E;ELBASAN LETTER JE;Lo;0;L;;;;;N;;;;;
1050F;ELBASAN LETTER KE;Lo;0;L;;;;;N;;;;;
10510;ELBASAN LETTER LE;Lo;0;L;;;;;N;;;;;
10511;ELBASAN LETTER LLE;Lo;0;L;;;;;N;;;;;
10512;ELBASAN LETTER ME;Lo;0;L;;;;;N;;;;;
10513;ELBASAN LETTER NE;Lo;0;L;;;;;N;;;;;
10514;ELBASAN LETTER NA;Lo;0;L;;;;;N;;;;;
10515;ELBASAN LETTER NJE;Lo;0;L;;;;;N;;;;;
10516;ELBASAN LETTER O;Lo;0;L;;;;;N;;;;;
10517;ELBASAN LETTER PE;Lo;0;L;;;;;N;;;;;
10518;ELBASAN LETTER QE;Lo;0;L;;;;;N;;;;;
10519;ELBASAN LETTER RE;Lo;0;L;;;;;N;;;;;
1051A;ELBASAN LETTER RRE;Lo;0;L;;;;;N;;;;;
1051B;ELBASAN LETTER SE;Lo;0;L;;;;;N;;;;;
1051C;ELBASAN LETTER SHE;Lo;0;L;;;;;N;;;;;
1051D;ELBASAN LETTER TE;Lo;0;L;;;;;N;;;;;
1051E;ELBASAN LETTER THE;Lo;0;L;;;;;N;;;;;
1051F;ELBASAN LETTER U;Lo;0;L;;;;;N;;;;;
10520;ELBASAN LETTER VE;Lo;0;L;;;;;N;;;;;
10521;ELBASAN LETTER XE;Lo;0;L;;;;;N;;;;;
10522;ELBASAN LETTER Y;Lo;0;L;;;;;N;;;;;
10523;ELBASAN LETTER ZE;Lo;0;L;;;;;N;;;;;
10524;ELBASAN LETTER ZHE;Lo;0;L;;;;;N;;;;;
10525;ELBASAN LETTER GHE;Lo;0;L;;;;;N;;;;;
10526;ELBASAN LETTER GHAMMA;Lo;0;L;;;;;N;;;;;
10527;ELBASAN LETTER KHE;Lo;0;L;;;;;N;;;;;
```

8. Bibliography

Domi, Mahir. 1965. "Rreth autorit dhe kohës së dorëshkrimit elbasanas me shqipërim copash të ungjillit", in *Konferenca e parë e Studimeve Albanologjike*, p. 270-277. Tiranë.

Elsie, Robert. 1995. "The Elbasan Gospel Manuscript (Anonimi i Elbasanit), 1761, and the struggle for an original Albanian alphabet", in *Südost-Forschungen*, Munich, 54, p. 105-159.
http://www.elsie.de/pdf/articles/A1995ElbasanMs_Fig.pdf

Nahtigal, Rajko. 1923. "O elbasanskem pismu in pismenstvu na njem", in *Arhiv za arbanasku starinu, jezik i etnologiju*, Belgrade, 1, p. 160-195.

Pekmezi, Georg. 1901. "Vorläufiger Bericht über das Studium des albanesischen Dialekts von Elbasan", in *Anzeiger der kaiserlichen Akademie der Wissenschaften*. Philosophisch-historische Classe 38, Vienna, 9, p. 39-64

Shuteriqi, Dhimitër. 1949. "Anonimi i Elbasanit. Shkrimi shqip në Elbasan në shekujt XVIII-XIX dhe Dhaskal Todhri", in *Buletin i Institutit të Shkencavet*. Tiranë , 1, p. 33-54.

Zamputi, Injac. 1949. "Disa shënime rreth alfabetit të dorëshkrimit të Anonimit elbasanas", in *Buletin i Institutit të Shkencavet*. Tiranë, 1, p. 55-57.

Zamputi, Injac. 1951. "Dorëshkrimi i Anonimit t'Elbasanit, Transliterim, transkriptim dhe koment", in *Buletin i Institutit të Shkencavet* Tiranë. 3-4, p. 64-130.

	1050	1051	1052
0	ⴹ 10500	∧ 10510	Ⱶ 10520
1	ⵝ 10501	ⴷ 10511	ჩ 10521
2	ⴖ 10502	Ⱨ 10512	ⴜ 10522
3	⌢ 10503	Ⴈ 10513	Ⴝ 10523
4	⊔ 10504	ⴰ 10514	�,10524
5	⊔ 10505	⌐ 10515	Ⴔ 10525
6	▽ 10506	Ο 10516	Ⴑ 10526
7	ⵒ 10507	Π 10517	✕ 10527
8	Ⴠ 10508	Κ 10518	
9	Ⴆ 10509	Ⴄ 10519	
A	Ⴧⴀ 1050A	Ⴄ 1051A	
B	ⴔ 1050B	Ⅽ 1051B	
C	Ⴚ 1050C	Ⴉ 1051C	
D	Ⰻ 1050D	∨ 1051D	
E	Ⅰ 1050E	θ 1051E	
F	Ⴚ 1050F	Ⴟ 1051F	

Letters

10500   ⴹ   ELBASAN LETTER A
10501   ⵝ   ELBASAN LETTER BE
10502   ⴖ   ELBASAN LETTER CE
10503   ⌢   ELBASAN LETTER CHE
10504   ⊔   ELBASAN LETTER DE
10505   ⊔   ELBASAN LETTER NDE
10506   ▽   ELBASAN LETTER DHE
10507   ⵒ   ELBASAN LETTER EI
10508   Ⴠ   ELBASAN LETTER E
10509   Ⴆ   ELBASAN LETTER FE
1050A   Ⴀ   ELBASAN LETTER GE
1050B   ⴔ   ELBASAN LETTER GJE
1050C   Ⴚ   ELBASAN LETTER HE
1050D   Ⰻ   ELBASAN LETTER I
1050E   Ⅰ   ELBASAN LETTER JE
1050F   Ⴚ   ELBASAN LETTER KE
10510   ∧   ELBASAN LETTER LE
10511   ⴷ   ELBASAN LETTER LLE
10512   Ⱨ   ELBASAN LETTER ME
10513   Ⴈ   ELBASAN LETTER NE
10514   ⴰ   ELBASAN LETTER NA
10515   ⌐   ELBASAN LETTER NJE
10516   Ο   ELBASAN LETTER O
10517   Π   ELBASAN LETTER PE
10518   Κ   ELBASAN LETTER QE
10519   Ⴄ   ELBASAN LETTER RE
1051A   Ⴄ   ELBASAN LETTER RRE
1051B   Ⅽ   ELBASAN LETTER SE
1051C   Ⴉ   ELBASAN LETTER SHE
1051D   ∨   ELBASAN LETTER TE
1051E   θ   ELBASAN LETTER THE
1051F   Ⴟ   ELBASAN LETTER U
10520   Ⱶ   ELBASAN LETTER VE
10521   ჩ   ELBASAN LETTER XE
10522   ⴜ   ELBASAN LETTER Y
10523   Ⴝ   ELBASAN LETTER ZE
10524   Ⴤ   ELBASAN LETTER ZHE
10525   Ⴔ   ELBASAN LETTER GHE
10526   Ⴑ   ELBASAN LETTER GHAMMA
10527   ✕   ELBASAN LETTER KHE

10. Figures.

Elbasan Gospel Manuscript					
Albanian Alphabet, 1761					
a	Ɛϲ			n	N	n(d)	٩
b	X			nj	˥		
c	P			o	o		
ç (+xh)	˜			p	Π		
d	⊔	nd	iɟ	q	K		
dh	▽			r	Ч		
e	Z			rr	Ӵ		
ë	ϛ			s	c		
f	Ь			sh	Я		
g	ᕊ			t	V		
gj	ϕ			th	θ		
h	ϛ			u	૪		
i	И			v	μ		
j	ı			x	h		
k	ᒥ			y	૪		
l	ᴧ			z	ϛ		
ll	∴ᴧ			zh	y		
m	ᑋ			Greek kh	X		
				Greek ğ	F		
				one missing	٦ another Greek g		

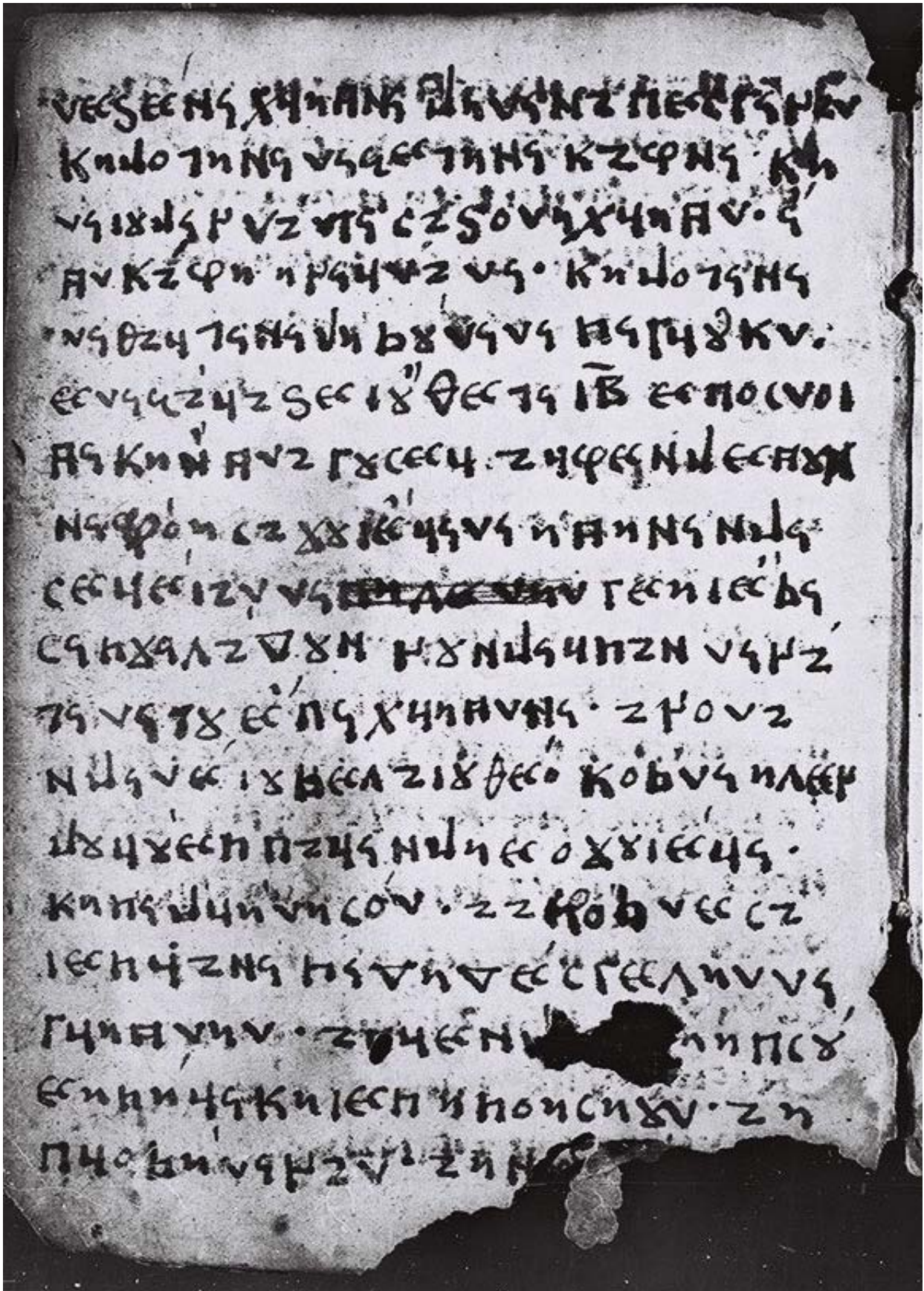**Figure 1.** Chart of the Elbasan alphabet by Robert Elsie.

Figure 2. Page 3 of the Elbasan Gospel Manuscript. The number 12 can be seen in the 6th line.
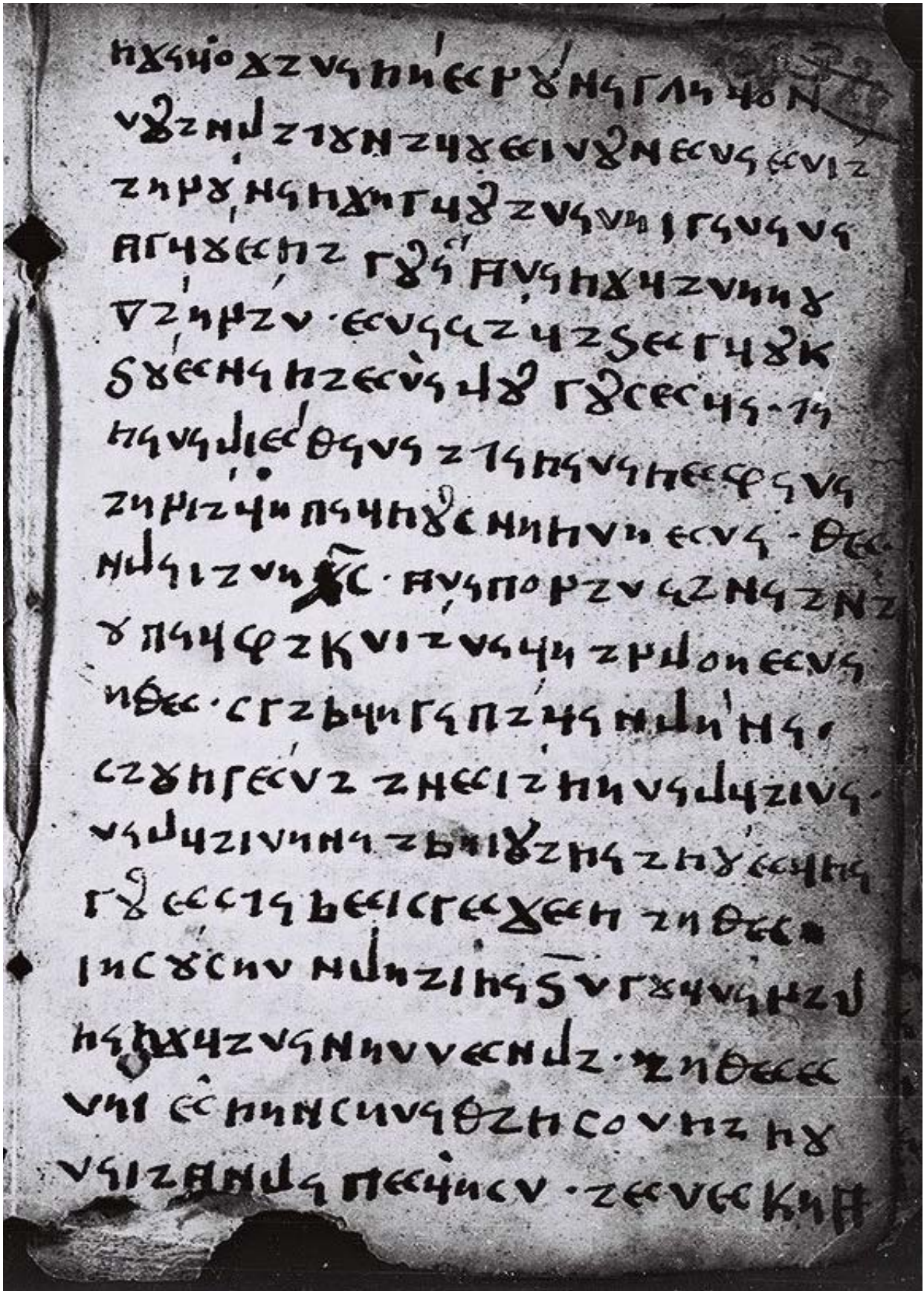
Figure 3. Page 48 of the Elbasan Gospel Manuscript. Some Greek breathings can be seen in line 4, and the combining overline can be seen in line 9 in the abbreviation for χριστος.

A. Administrative

1. Title
Preliminary proposal for encoding the Elbasan 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
2011-02-03
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
Elbasan.
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
40.
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?
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.
2b. If YES, with whom?
Robert Elsie.
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?