

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

**Doc Type:** Working Group Document

**Title:** Preliminary proposal for encoding the Caucasian Albanian script in the SMP of the UCS

**Source:** UC Berkeley Script Encoding Initiative (Universal Scripts Project)

**Authors:** Michael Everson and Jost Gippert

**Status:** Liaison Contribution

**Action:** For consideration by JTC1/SC2/WG2 and UTC

**Date:** 2011-05-02

**1. Introduction.** The Caucasian Albanian script was used to write a Lezgian language spoken in the Caucasus. (Caucasian Albania is not the same as European Albania.) The script is now well-known due to the discovery of the Sinai palimpsests dating to about the 8th century CE, but previously it was only known through a medieval alphabet list provided in an Armenian manuscript (Erevan Mat. 7117) and a few inscriptions that were unearthed during the construction of the Mingechevir reservoir in North-West Azerbaijan. The phonetic values of the characters helped in the decipherment of the inscriptions and provided the basis of this encoding. It also showed that Caucasian Albanian was a dialect of Old Udi, which is closely related to the Eastern Samur branch of Lezgian (which includes Lezgi proper, Tabasaran, and Aghul). A number of features indicate that the Caucasian Albanian alphabet was based on the Greek, in the same way that the Armenian and Georgian alphabets were. The use of an ou digraph for /u/ is one of these features (compare Greek ΛΟΥΚΑΣ *loukas* ‘Luke’, Armenian *ԼՈՒԿԱՍ* *łowkas*, Old Georgian *ႂႂႂႂႂ* *lowka*, and Caucasian Albanian *ႂႂႂႂႂ* *lowkas*).

**2. Processing.** Caucasian Albanian is a simple alphabetic script written from left to right horizontally. There is no ligation.

**3. Character names.** The names used for the characters here are based on Jost Gippert’s normalized reconstructions of the names.

**4. Numerals.** Script-specific numerals are not known.

**5. Punctuation.** In the manuscript some evidence of spaces and a separating colon can be seen. These would most likely be generic, not script-specific punctuation, though the colon is used as final punctuation. An abbreviation sign is used; further study is required to determine whether a generic or script-specific character should represent this.

**6. Ordering.** Ordering is as in the code chart, and follows the alphabetic order given in the sources.

## 7. Unicode Character Properties

```
10530;CAUCASIAN ALBANIAN LETTER ALT;Lo;0;L;;;;N;;;;;
10531;CAUCASIAN ALBANIAN LETTER BET;Lo;0;L;;;;N;;;;;
10532;CAUCASIAN ALBANIAN LETTER GIM;Lo;0;L;;;;N;;;;;
10533;CAUCASIAN ALBANIAN LETTER DAT;Lo;0;L;;;;N;;;;;
10534;CAUCASIAN ALBANIAN LETTER EB;Lo;0;L;;;;N;;;;;
10535;CAUCASIAN ALBANIAN LETTER ZARL;Lo;0;L;;;;N;;;;;
10536;CAUCASIAN ALBANIAN LETTER EYN;Lo;0;L;;;;N;;;;;
10537;CAUCASIAN ALBANIAN LETTER ZHIL;Lo;0;L;;;;N;;;;;
10538;CAUCASIAN ALBANIAN LETTER TAS;Lo;0;L;;;;N;;;;;
10539;CAUCASIAN ALBANIAN LETTER CHA;Lo;0;L;;;;N;;;;;
```

1053A;CAUCASIAN ALBANIAN LETTER YOWD;Lo;0;L;;;;N;;;;;  
 1053B;CAUCASIAN ALBANIAN LETTER ZHA;Lo;0;L;;;;N;;;;;  
 1053C;CAUCASIAN ALBANIAN LETTER IRB;Lo;0;L;;;;N;;;;;  
 1053D;CAUCASIAN ALBANIAN LETTER SHA;Lo;0;L;;;;N;;;;;  
 1053E;CAUCASIAN ALBANIAN LETTER LAN;Lo;0;L;;;;N;;;;;  
 1053F;CAUCASIAN ALBANIAN LETTER INYA;Lo;0;L;;;;N;;;;;  
 10540;CAUCASIAN ALBANIAN LETTER XEYN;Lo;0;L;;;;N;;;;;  
 10541;CAUCASIAN ALBANIAN LETTER DYAN;Lo;0;L;;;;N;;;;;  
 10542;CAUCASIAN ALBANIAN LETTER CAR;Lo;0;L;;;;N;;;;;  
 10543;CAUCASIAN ALBANIAN LETTER JHOX;Lo;0;L;;;;N;;;;;  
 10544;CAUCASIAN ALBANIAN LETTER KAR;Lo;0;L;;;;N;;;;;  
 10545;CAUCASIAN ALBANIAN LETTER LYIT;Lo;0;L;;;;N;;;;;  
 10546;CAUCASIAN ALBANIAN LETTER HEYT;Lo;0;L;;;;N;;;;;  
 10547;CAUCASIAN ALBANIAN LETTER QAY;Lo;0;L;;;;N;;;;;  
 10548;CAUCASIAN ALBANIAN LETTER AOR;Lo;0;L;;;;N;;;;;  
 10549;CAUCASIAN ALBANIAN LETTER CHOY;Lo;0;L;;;;N;;;;;  
 1054A;CAUCASIAN ALBANIAN LETTER CHI;Lo;0;L;;;;N;;;;;  
 1054B;CAUCASIAN ALBANIAN LETTER CYAY;Lo;0;L;;;;N;;;;;  
 1054C;CAUCASIAN ALBANIAN LETTER MAQ;Lo;0;L;;;;N;;;;;  
 1054D;CAUCASIAN ALBANIAN LETTER QAR;Lo;0;L;;;;N;;;;;  
 1054E;CAUCASIAN ALBANIAN LETTER NOWC;Lo;0;L;;;;N;;;;;  
 1054F;CAUCASIAN ALBANIAN LETTER DZYAY;Lo;0;L;;;;N;;;;;  
 10550;CAUCASIAN ALBANIAN LETTER SHAK;Lo;0;L;;;;N;;;;;  
 10551;CAUCASIAN ALBANIAN LETTER JAYN;Lo;0;L;;;;N;;;;;  
 10552;CAUCASIAN ALBANIAN LETTER ON;Lo;0;L;;;;N;;;;;  
 10553;CAUCASIAN ALBANIAN LETTER TYAY;Lo;0;L;;;;N;;;;;  
 10554;CAUCASIAN ALBANIAN LETTER FAM;Lo;0;L;;;;N;;;;;  
 10555;CAUCASIAN ALBANIAN LETTER DZAY;Lo;0;L;;;;N;;;;;  
 10556;CAUCASIAN ALBANIAN LETTER CHAT;Lo;0;L;;;;N;;;;;  
 10557;CAUCASIAN ALBANIAN LETTER PEN;Lo;0;L;;;;N;;;;;  
 10558;CAUCASIAN ALBANIAN LETTER GHEYS;Lo;0;L;;;;N;;;;;  
 10559;CAUCASIAN ALBANIAN LETTER RAT;Lo;0;L;;;;N;;;;;  
 1055A;CAUCASIAN ALBANIAN LETTER SEYK;Lo;0;L;;;;N;;;;;  
 1055B;CAUCASIAN ALBANIAN LETTER VEYZ;Lo;0;L;;;;N;;;;;  
 1055C;CAUCASIAN ALBANIAN LETTER TUWR;Lo;0;L;;;;N;;;;;  
 1055D;CAUCASIAN ALBANIAN LETTER SHOY;Lo;0;L;;;;N;;;;;  
 1055E;CAUCASIAN ALBANIAN LETTER IWR;Lo;0;L;;;;N;;;;;  
 1055F;CAUCASIAN ALBANIAN LETTER CYAW;Lo;0;L;;;;N;;;;;  
 10560;CAUCASIAN ALBANIAN LETTER CAYN;Lo;0;L;;;;N;;;;;  
 10561;CAUCASIAN ALBANIAN LETTER YAYD;Lo;0;L;;;;N;;;;;  
 10562;CAUCASIAN ALBANIAN LETTER PIWR;Lo;0;L;;;;N;;;;;  
 10563;CAUCASIAN ALBANIAN LETTER KIW;Lo;0;L;;;;N;;;;;

## 8. Bibliography

To be provided

**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 Caucasian Albanian encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

	1053	1054	1055	1056
0	Ლ 10530	Ბ 10540	Მ 10550	Ყ 10560
1	Ზ 10531	Ბ 10541	Ბ 10551	Ყ 10561
2	Ქ 10532	Ლ 10542	Ა 10552	Ს 10562
3	Ს 10533	Ბ 10543	Მ 10553	Მ 10563
4	Მ 10534	Მ 10544	Მ 10554	
5	Მ 10535	Ყ 10545	Მ 10555	
6	Ქ 10536	Მ 10546	Მ 10556	
7	Ს 10537	Მ 10547	Მ 10557	
8	Მ 10538	Მ 10548	Მ 10558	
9	Მ 10539	Მ 10549	Მ 10559	
A	Მ 1053A	Მ 1054A	Მ 1055A	
B	Მ 1053B	Მ 1054B	Მ 1055B	
C	Მ 1053C	Მ 1054C	Მ 1055C	
D	Მ 1053D	Მ 1054D	Მ 1055D	
E	Მ 1053E	Მ 1054E	Მ 1055E	
F	Მ 1053F	Მ 1054F	Მ 1055F	

## Letters

10530	Ლ	CAUCASIAN ALBANIAN LETTER ALT
10531	Ზ	CAUCASIAN ALBANIAN LETTER BET
10532	Ქ	CAUCASIAN ALBANIAN LETTER GIM
10533	Ს	CAUCASIAN ALBANIAN LETTER DAT
10534	Მ	CAUCASIAN ALBANIAN LETTER EB
10535	Მ	CAUCASIAN ALBANIAN LETTER ZARL
10536	Ქ	CAUCASIAN ALBANIAN LETTER EYN
10537	Ს	CAUCASIAN ALBANIAN LETTER ZHIL
10538	Მ	CAUCASIAN ALBANIAN LETTER TAS
10539	Მ	CAUCASIAN ALBANIAN LETTER CHA
1053A	Მ	CAUCASIAN ALBANIAN LETTER YOWD
1053B	Მ	CAUCASIAN ALBANIAN LETTER ZHA
1053C	Მ	CAUCASIAN ALBANIAN LETTER IRB
1053D	Მ	CAUCASIAN ALBANIAN LETTER SHA
1053E	Მ	CAUCASIAN ALBANIAN LETTER LAN
1053F	Მ	CAUCASIAN ALBANIAN LETTER INYA
10540	Მ	CAUCASIAN ALBANIAN LETTER XEYN
10541	Მ	CAUCASIAN ALBANIAN LETTER DYAN
10542	Ლ	CAUCASIAN ALBANIAN LETTER CAR
10543	Ბ	CAUCASIAN ALBANIAN LETTER JHOX
10544	Მ	CAUCASIAN ALBANIAN LETTER KAR
10545	Ყ	CAUCASIAN ALBANIAN LETTER LYIT
10546	Მ	CAUCASIAN ALBANIAN LETTER HEYT
10547	Მ	CAUCASIAN ALBANIAN LETTER QAY
10548	Მ	CAUCASIAN ALBANIAN LETTER AOR
10549	Მ	CAUCASIAN ALBANIAN LETTER CHOY
1054A	Მ	CAUCASIAN ALBANIAN LETTER CHI
1054B	Მ	CAUCASIAN ALBANIAN LETTER CYAY
1054C	Მ	CAUCASIAN ALBANIAN LETTER MAQ
1054D	Მ	CAUCASIAN ALBANIAN LETTER QAR
1054E	Მ	CAUCASIAN ALBANIAN LETTER NOWC
1054F	Მ	CAUCASIAN ALBANIAN LETTER DZYAY
10550	Მ	CAUCASIAN ALBANIAN LETTER SHAK
10551	Ბ	CAUCASIAN ALBANIAN LETTER JAYN
10552	Ა	CAUCASIAN ALBANIAN LETTER ON
10553	Მ	CAUCASIAN ALBANIAN LETTER TYAY
10554	Მ	CAUCASIAN ALBANIAN LETTER FAM
10555	Მ	CAUCASIAN ALBANIAN LETTER DZAY
10556	Მ	CAUCASIAN ALBANIAN LETTER CHAT
10557	Მ	CAUCASIAN ALBANIAN LETTER PEN
10558	Მ	CAUCASIAN ALBANIAN LETTER GHEYS
10559	Მ	CAUCASIAN ALBANIAN LETTER RAT
1055A	Მ	CAUCASIAN ALBANIAN LETTER SEYK
1055B	Მ	CAUCASIAN ALBANIAN LETTER VEYZ
1055C	Მ	CAUCASIAN ALBANIAN LETTER TUWR
1055D	Მ	CAUCASIAN ALBANIAN LETTER SHOY
1055E	Მ	CAUCASIAN ALBANIAN LETTER IWR
1055F	Მ	CAUCASIAN ALBANIAN LETTER CYAW
10560	Ყ	CAUCASIAN ALBANIAN LETTER CAYN
10561	Ყ	CAUCASIAN ALBANIAN LETTER YAYD
10562	Ს	CAUCASIAN ALBANIAN LETTER PIWR
10563	Მ	CAUCASIAN ALBANIAN LETTER KIW



## A. Administrative

1. Title

**Preliminary proposal for encoding the Caucasian Albanian script in the SMP of the UCS**

2. Requester's name

**UC Berkeley Script Encoding Initiative (Universal Scripts Project)**

**(Authors: Michael Everson, Jost Gippert)**

3. Requester type (Member body/Liaison/Individual contribution)

**Liaison contribution.**

4. Submission date

**2011-05-02**

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

**Caucasian Albanian.**

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

**52.**

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?

**Jost Gippert.**

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 Old Udi 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?