

Revised: Proposal to encode three Armenian superscript characters

Hossep Dolatian Յովսէփ Տէօվլէթեան
hossep.dolatian@alumni.stonybrook.edu

Independent Linguist

2024-03-15

1 Introduction

Armenian is an Indo-European language written in the Armenian script. Conventional characters in the script are available in Unicode. However, there are some Armenian symbols that are used in Armenian dialectology, but which currently do not have a Unicode implementation.

We focus on three characters which are superscript versions of the Armenian letters <է>, <ի>, <յ>. Unicode correctly provides support for the normal version of these characters, meaning when the character is not a superscript.

Normal version			Superscript version	
0567	է	Armenian Small Letter Eh	է̇	Not supported
056B	ի	Armenian Small Letter Ini	ի̇	Not supported
0575	յ	Armenian Small Letter Yi	յ̇	Not supported

The superscript version of these letters are used to mark different phonetic articulations. They are used in dictionaries and grammars of non-standard dialects. We provide examples of their usage in the **Justification** section.

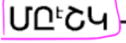
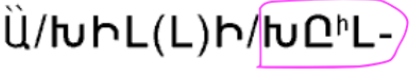

2 Request

This proposal requests the addition of two new characters in the Armenian block with the following names:

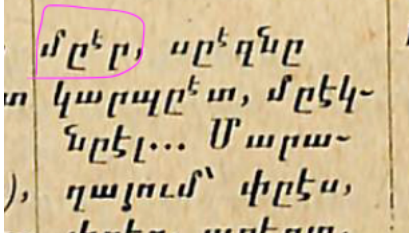
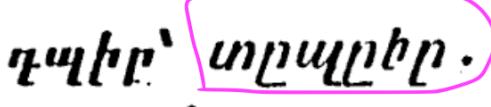
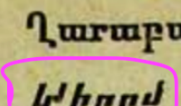
է̇	U+0558 MODIFIER LETTER ARMENIAN SMALL EH
ի̇	U+058B MODIFIER LETTER ARMENIAN SMALL INI
յ̇	U+058C MODIFIER LETTER ARMENIAN SMALL YI

3 Justification

These superscript symbols are quite prominently used in dialectological work of Karabakh Armenian or Artsakh Armenian. This is the dialect of Armenian that was spoken in the disputed territory of Karabakh (in present-day Azerbaijan), up until the mass exodus and ethnic cleansing of Armenians in September 2023. We show below example entries from a Karabakh dictionary (Մարգարյան 2013), along with an IPA re-transcription from us. In our screenshots from the dictionary, we use pink boxes to demarcate a word that has the superscript. In our textual rendering, we highlight in yellow.

Symbol	Phonetic value	Example from Մարգարյան 2013	Page #, example, and IPA re-transcription
x ^t	Fronted vowel /x ^e /	մըննան ըղնում ա (ԲՅԲ, 189):  - տե՛ն ՄԱՇԿ: ՄԸՇՆԱԿ - տե՛ն ՄԱՇՆԱԿ:	503 <ՄԸՇԿ> /məʃk/
x ^h	Raised vowel /x ⁱ /	 ած. Զգալի, շատ, ժո յօսու ճննողում ա.	284 <ԽԸԻԼ> /χəi/
x ^l	Patalized consonant	 - տե ԿԱԼԱՎ/ԿԱԼ	374 <ԿԱԼԱՎ> /kʲalav/

We likewise provide examples of these symbols from foundational readings in Armenian dialectology.

Symbol	Example	Source #, example, and IPA re-transcription
x ^t		Դավթյան 1944: 16 <մըլեր> /məɽ/
x ^h		Աճառեան 1911: 282 <տըպըր> /təpəɽ/
x ^l		Դարիբյան 1941: 182 <կիրըմ> /kʲirim/

4 Alternative Encoding Possibilities

In our textual renderings above, we wrote the superscripts using Microsoft Word's superscript functionality. Visually, people can see the superscript forms. But, if one tries to select and copy-paste the text <Է¹> into any other format such as a text file, then the resulting output undoes the superscript form <Է¹>. This means that the superscript forms are not recoverable from text. An automatic Unicode representation is needed so that dictionaries and other linguistic material that use these symbols can be adequately digitized.

5 Properties

- Line_Break: AL
- Diacritic: Yes
- Other_Lowercase: Yes
- Script: Armenian
- Sentence_Break: Lower
- Word_Break: Lower
- UnicodeData.txt:
 - 0558;MODIFIER LETTER ARMENIAN SMALL EH;Lm;0;L;<super> 0567;;;N;;;;;
 - 055B;MODIFIER LETTER ARMENIAN SMALL INI;Lm;0;L;<super> 056B;;;N;;;;;
 - 055C;MODIFIER LETTER ARMENIAN SMALL YI;Lm;0;L;<super> 0575;;;N;;;;;

6 References

- Աճառեան, Հրաչեայ. 1911. Հայ Բարբառագիտութիւն [Armenian dialectology]. Մոսկուա-Նոր-Նախիջեան: Լազարեան Ճեմարան Արեւելեան Լեզուաց.
- Դավթյան, Կարո Ստեփանի. 1966. Լեռնային Ղարաբաղի Բարբառային Քարտեզը [Dialectal map of Mountainous Karabakh]. Երևան: Հայկական ՄՍՀ Գիտությունների Ակադեմիայի Հրատարակչություն.
- Ղարիբյան, Արարատ. 1941. Համառոտություն հայ բարբառագիտության [Summary of Armenian dialectology]. Երևան: Պետհամալսարանի տպ.
- Սարգսյան, Արմեն Յուրիկի. 2013. Ղարաբաղի բարբառի բառարան [Dictionary of the Karabakh dialect]. Երևան: Էդիթ Պրինտ.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title: *Proposal to encode three Armenian superscript characters of <ł>, <h>, <g>*
2. Requester's name: *Hossep Dolatian*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *02/14/2024*
5. Requester's reference (if applicable): *Ben Yang*
6. Choose one of the following:
- This is a complete proposal: *Yes*
- (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
- a. This proposal is for a new script (set of characters):
- Proposed name of script: _____
- b. The proposal is for addition of character(s) to an existing block: *yes*
- Name of the existing block: *Armenian*
2. Number of characters in proposal: *3*
3. Proposed category (select one from below - see section 2.2 of P&P document):
- | | | | | |
|---------------------------------------|------------------------------------|-------------------------------------|---|--|
| A-Contemporary | B.1-Specialized (small collection) | <input checked="" type="checkbox"/> | 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 | |
4. Is a repertoire including character names provided?
- a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *Yes*
- b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
- a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? _____
- b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): _____
6. References:
- a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *yes*
- b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *yes*
7. Special encoding issues:
- 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. Additional Information:

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/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>no</i>
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	<i>no</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>yes</i> <i>We provide references to dictionaries that use these symbols</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>rare</i> <i>We provide references to grammars that use these symbols</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>yes</i> <i>We provide references to books that use these symbols</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>No</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<i>No</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>No</i>