

# Armenian Phonetic Characters in Unicode\*

Luc V. Baronian, PhD  
Visiting Scholar, Stanford  
Associate Professor, UQAC

January 19, 2017

## 1 Introduction

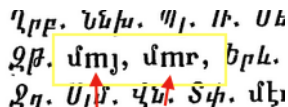
Armenian has a relatively small number of speakers, yet it has a sesquimillennial written tradition and several contemporary academic publications. The numeric disadvantage, combined with the existence of two standard versions of the modern language (Eastern and Western), a unique alphabet and the evolution of Armenia’s universities in the sphere of Russia, rather than that of Western Europe and North America, have all contributed to a poorer knowledge of its language and its academic achievements in the West than it deserves.

For example, there exists an Armenian linguistic tradition, which, like the American, French and other linguistic traditions, has used a phonetic alphabet of its own for about a century, but based on the Armenian alphabet. The works of Hrach’ya Acharyan<sup>1</sup> were crucial in disseminating the use of this phonetic alphabet **still in use today** among Armenian scholars.

## 2 The characters proposed for inclusion in Unicode

While most of the Armenian phonetic characters consist of Armenian letters with or without the use of diacritics such as the umlaut or a combining ring, I have identified two characters that cannot be represented in the current state of Unicode.

The first character consists of ARMENIAN SMALL LETTER AYB (U+0561, **այ**) rotated 180 degrees. The letter could sometimes be confused for a LATIN SMALL LETTER M (U+006D); however, in those Armenian fonts that use square angles for Ayb, the difference is noticeable. Even a rotated curvy Ayb may look different from an *m*, especially in the rendition of the initial *m*-stroke (final Ayb-stroke). The phonetic value attributed to the letter is a mid-low front vowel, similar to IPA [æ]. The example below is taken from the entry for *mayr* ‘mother’ in Acharyan’s 1971 Armenian Root Dictionary. It is used here to indicate the phonetic value of the vowel in the Armenian dialect of Zeytun.



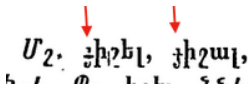
\*I thank Deborah W. Anderson (Script Encoding Initiative at UC Berkeley) for her valued help in the preparation of this document. I also thank Hrach Martirosyan (Austrian Academy of Sciences) and Bert Vaux (Cambridge University) for help with the references, as well as Anaïd Donabédian (INALCO) for useful comments.

<sup>1</sup>Born in Istanbul, Acharyan was a student of phonetician Abbé Rousselot and Indo-Europeanist Antoine Meillet at the Sorbonne and is considered the founder of modern Armenian linguistics. After the Great War, he spent the rest of his career in Soviet Armenia where Acharyan University in Yerevan is now named after him. His etymological dictionary of Armenian (1971) is still considered a standard in the field. The significance of his achievements should not be underestimated: according to Braun (2013), he developed the notion of Voiced Onset Timing (VOT)—without naming it as such—over six decades before Lisker and Abramson (1964). His first publications in French are under Adjarian, which reflects his native Western Armenian pronunciation, rather than the Eastern Armenian pronunciation of his career in Soviet Armenia. In the publications preceding the Armenian orthographic reform (1922-1924), we also find his name transliterated as Hrach’eay Acharean.

Example of a font using a square-angled ARMENIAN SMALL LETTER AYB (default Armenian font on Mac OS X), very distinct from LATIN SMALL LETTER M:



The second character consists of ARMENIAN SMALL LETTER Yi (U+0575, ՚) with a horizontal stroke. The letter looks somewhat like a LATIN SMALL LETTER F (U+0066) rotated 180 degrees. It could be confused with LATIN SMALL LETTER J WITH STROKE (U+0249) or LATIN SMALL LETTER DOTLESS J WITH STROKE (U+025F). The phonetic value attributed to the letter is a voiced version of [h] (IPA [ɦ]). The example below is taken again from Acharyan (1971), from the lexical entry of *yishel* ‘remember’. It is used to indicate the phonetic value of the initial letter/phoneme in the dialect of Mush.



There is no upper case version for either character. Acharyan’s (1971) dictionary is still the standard reference for Armenian etymologies.

### 3 Technical points

#### 3.1 Glyphs, Names and Code points

Here are suggestions for the names and locations, based on similar names already in Unicode and based on the most recent additions to the set of Armenian characters. The glyphs available on the Unicode chart for Armenian (range 0530?058F) were used as a basis for those of this proposal.

- m** ARMENIAN SMALL LETTER TURNED AYB, U+058B
- f** ARMENIAN SMALL LETTER YI WITH STROKE, U+058C

#### 3.2 Properties

The properties should be:

Ll;0;L;;;;N;;;;

#### 3.3 Confusables

Depending on the font style, ARMENIAN SMALL LETTER TURNED AYB is confusable with LATIN SMALL LETTER M (U+006D).

ARMENIAN SMALL LETTER YI WITH STROKE is confusable with LATIN SMALL LETTER DOTLESS J WITH STROKE (U+025F) and, though it never has a dot, with LATIN SMALL LETTER J WITH STROKE (U+0249).

#### 3.4 Examples

The following are from Acharean’s (1913) legend explaining the phonetic value of each of the two characters and the dialects in which the corresponding phones are found. These earlier attestations used a less uniform style than the standardized 1971 attestations.

m = ä (ա և Ի ձայներուն միջին  
 աստիճանն է. կը գտնուի  
 Ղրբ. Աղլ. Կր. Վն. Խրբ. Զթ.  
 - Հճ. ՄրՂ. Արբ. Ասլ. Մրղ.):

ժ – հ-էն աւելի տկար շունչ մ'է.  
 ՄՂ. ԱլՂ.

## 4 Significance of the characters

Adopting these two characters in Unicode will improve the digitization process of the century-old corpus of standard and specialized works in Armenian linguistics. The website *Nayiri.com* already makes several Armenian dictionaries available, including Acharyan's 1913 and 1971<sup>2</sup> dictionaries that use these two letters. The National Library of Armenia continuously puts online Armenian publications from the 20th century that are very difficult to find in the West and the absence of these two characters is an obstacle to this digitization effort.

Furthermore, while Armenian linguists publishing in Europe and North America favor the use of the IPA, several linguists based in Armenia continue to use Armenian-based characters in field transcriptions and in their publications. Their use for indicating pronunciation is also common in bilingual Armenian dictionaries<sup>3</sup> for the millions of Armenian speakers wanting to learn a second language and who do not necessarily master the Roman and Greek alphabets on which the IPA is primarily based. When not using the IPA, scholars do also use diacritic alternatives to these two letters,<sup>4</sup> perhaps encouraged by their absence in Unicode. Regardless, the existing dictionaries and published dialect studies using these characters should be justification enough to include them in Unicode.

## 5 References and websites

Acharean, Hr. 1913. *Hayerēn Gawaṙakan Baṙaran*. Tiflis: Lazarean Chemaran Arewelean Lezowats'. [Armenian Dialectal Dictionary: pages 11-12 include a key to the phonetic symbols.]

Acharyan, Hr. 1971. *Hayeren Armatakan Baṙaran*. Erewan: Erewani Hamalsarani Hratarakch'owt'yown. [Armenian Root Dictionary: contains pronunciation notes for various dialects.]

Aghabekyan, M., A. Haneyan and R̄. Odowdyan. 2010. *Hayereni Hnch'atarberakneri taṙadardzman nor hamakargē*. In: *Hayereni barbaṙagitakan atlas*, prak 3, pp. 5-22. [Armenian phonetic transliteration: new correspondants. In: the Armenian dialectal atlas, pp. 5-22.]

Braun, Angelika. 2013. An early case of VOT. *Interspeech 2013*, pp. 119-122.

Gowshagchean, Martiros and Rev. Tigran Khndrowni. 1970. *Anglerēn-Hayerēn Ardi Baṙaran*. *Hretarakch'atown K. Tonikean ew Ordik'*: Beirut.

<sup>2</sup>The website uses the year 1926 in error for the latter dictionary. The edition they put online is the 1971 one, as made clear at the bottom of the second digital page.

<sup>3</sup>Examples include Gowshagchean and Khndrowni (1970) and Gowyowmchean (2002)

<sup>4</sup>Aghabekyan et al. (2010) is a recent proposal for a renewed system.

Gowyowmchean, Mesrop K. 2002. Grpani Nor Bařaran Anglerēn-Hayerēn. Beirut: Shirak. [New English-Armenian Pocket Dictionary.]

Lisker, L. and Abramson, A. 1964. A cross-language study of voicing in initial stops: acoustical measurements, *Word* 20, pp. 384-422.

Nayiri : currently contains 93 digitized Armenian dictionaries.  
<http://www.nayiri.com>

National Library of Armenia  
<http://www.nla.am>

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

**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:	<i>Proposal to Include Two Armenian Phonetic Characters</i>
2. Requester's name:	<i>Luc V. Baronian (Stanford, UQAC)</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2017-01-19</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(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):	<i>No</i>	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>	
Name of the existing block:	<i>Armenian (Range: 0530-058F)</i>	
2. Number of characters in proposal:	<i>2</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<i>yes</i>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Michael Everson</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Evertime</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>yes</i>	
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)?	<i>no</i>	

**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.

<sup>1</sup> 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	No
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:	Yes <i>Anaid Donabedian (INALCO) Bert Vaux (Cambridge) Hr. Martirosyan (AAS)</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	No <i>(Armenian linguists)</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes <i>Acharyan 1971 dictionary still standard in the field</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:	Yes No
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
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:	No
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:	No
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:	Yes Yes <i>Small differences and different alphabet, see document.</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:	No
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)	No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No