

ISO
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

ISO/IEC JTC1/SC2/WG2
Universal Multiple-Octet Coded Character Set
(UCS)

ISO/IEC JTC1/SC2/WG2 N
Date: Feb 5th 1999

Title: Encoding of the Armenian script in ISO/IEC 10646, answer to SC2 N3222
Source: US
Status: National Body contribution
Action: For the consideration of WG2
References: ISO/IEC JTC1/SC2/WG2 N 1395, N1446, N1560, N1616 and SC2 N3222
Distribution: ISO/IEC JTC1/SC2/WG2 members

Summary

This document addresses concerns expressed in document SC2 N 3222 by the Armenian National Body and should demonstrate that ISO/IEC 10646 covers exhaustively the Armenian writing system. Each of the concern has been answered individually.

1. *Armenian characters in ISO 10585 and ISO/IEC 10646-1 do not conform to the Armenian national standard AST 34.002-97 and to the Armenian alphabet and grammatical system.*

Without having a copy of the Armenian standard it is difficult to assess 'conformance'. The 2 ISO standards specify both a repertoire and an encoding, so they can't be strictly conformant to another encoding. It is however the goal of ISO/IEC 10646 to cover the repertoire of existing national standards. It should be noted however that the Armenian standard was created after the current version of 10646, so minor updates could be required. More on this on the answer to issue 2.

It is also important to note that all characters required to encode the Armenian writing systems do not have to be included in the Armenian block of ISO/IEC 10646 (0530-058F). As with most other writing systems, it is expected that characters from other blocks (for example the Basic Latin block) be used to that effect. ISO/IEC 10646 doesn't address grammatical issue, so this particular point about being non-conformant to the grammatical system seems out of scope.

2. *About 15 Armenian characters are not included in the ISO standards.*

We will only consider ISO/IEC 10646 inclusion. A quick check of these characters shows in fact that MOST OF these characters, although not encoded in the Armenian block of ISO/IEC 10646 (0530-058F) are in fact encoded in various part of the standard. 10646 unifies symbols used among various writing systems, i.e. it doesn't encode several time a symbol required within various writing systems. This principle is explained in clause 18 of the standard. In addition, grammatical distinctions in punctuation are never encoded in SC2 standards. Some AST 34.002-97 names prefixed with 'TEXTS' suggesting that this could have been the intention. The following table shows the proposed mappings:

Character	AST 34.002-97 name	ISO/IEC 10646 name	AST 34.002-97 (8 bit)	ISO/IEC 10646 code position
	ETERNITY SIGN	(1)	10/01	
§	PARAGRAPH	SECTION SIGN	10/02	00A7
)	TEXTS RIGHT PARENTHESIS	RIGHT PARENTHESIS	10/04	0029

Character	AST 34.002-97 name	ISO/IEC 10646 name	AST 34.002-97 (8 bit)	ISO/IEC 10646 code position
(TEXTS LEFT PARENTHESIS	LEFT PARENTHESIS	10/05	0028
»	RIGHT QUOTATION MARK	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK	10/06	00BB
«	LEFT QUOTATION MARK	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK	10/07	00AB
–	DASH	EN DASH	10/08	2013
•	MIDDLE DOT	MIDDLE DOT	10/09	00B7
,	TEXTS COMMA	COMMA	10/11	002D
–	JOINED LINE	ARMENIAN HYPHEN (2)	10/12	
–	HYPHEN SIGN	HYPHEN	07/09	2010
...	ELLIPSIS POINTS	HORIZONTAL ELLIPSIS	10/14	2026
՛	CAPITAL APOSTROPHE	(3)	15/14	
՛	SMALL APOSTROPHE	(3)	15/15	

Notes:

1. The ETERNITY symbol could be a potential for inclusion, however earlier discussions with Armenian experts could not reach a consensus on this.
2. The character 058A ARMENIAN HYPHEN was proposed by document WG1 N1616, following discussions among experts, including several Armenian people. It was later accepted by WG2 and included in a FPDAM. There are many other symbols that could mean 'JOINED LINE' in ISO 10646, among those: 002D HYPHEN MINUS, 00AD SOFT HYPHEN, etc.
3. It is not clear that symbols need a capital and small form. The exact shape and positioning of symbols is not specified by ISO/IEC 10646, and it is the responsibility of the display process to render these symbols according to the context.

3. In ISO 10585 and ISO 10646-1 Armenian character sets are different which is unacceptable.

Two different groups with different usage created these 2 ISO standards, so it can be expected to see minor variations. The bulk of the Armenian repertoire is identical, including all the commonly used Armenian letters. Furthermore ISO 10585 is a 7-bit repertoire and ISO/IEC 10646 is a 16/32 bit encoded repertoire (16 bit in the BMP). All efforts have been made to check that the repertoire covered by ISO 10585 is included in ISO/IEC 10646 and most experts agree that this is the case.

4. The codes defined in the ISO 10585 and ISO/IEC 10646-1 were never used in Armenia. We are using the codes which are defined in the Armenian national standard AST 34.002-97 and are in practical use.

We cannot comment much on the assertion of ISO 10585 usage as it was designed outside this ISO Sub Committee (SC2). It should be noted however that usage of Armenian characters is not restricted to Armenia. They are widely used around the world in computer and bibliographic communities. ISO/IEC 10646 and the Unicode Standard are already in widespread use by operating systems and applications and they allow much better international interoperability than the limited 7 or 8 bit standards. We can only encourage the Armenian national body to promote the usage of ISO/IEC 10646. If there are minor additions to be made that is still possible, WG2 welcomes requests for character additions that are well documented.

5. ISO 10585 and ISO/IEC 10646-1 do not conform to the logic of ISO 2022 standard.

We cannot comment on ISO 10685, but concerning ISO/IEC 10646-1 the assertion above is simply not true. Please refer to clause 17 of the standard (page 9 and 10) for the technical details.

Again, we are open to discussion to understand better the Armenian national body requirements and would welcome additional technical rationale for the characters 'not included'.