

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

Doc Type: Working Group Document

Title: Revised proposal for encoding the Philippine scripts in the UCS

Source: Michael Everson, EGT (IE)

Status: Expert Contribution

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

Date: 1998-11-23

This document is based on the proposal written by Rick McGowan and published in UTR#3, and the proposal written by me in N1755. It is a revision of N1755 based on expert input, and contains the proposal summary.

A. Administrative

1. Title

Revised proposal for encoding the Philippine scripts in the UCS.

2. Requester's name

Michael Everson, EGT (WG2 member for Ireland).

3. Requester type

Expert contribution.

4. Submission date

1998-11-23.

5. Requester's reference

6a. Completion

This is a complete proposal.

6b. More information to be provided?

No.

B. Technical -- General

1a. New script? Name?

Yes. Four related scripts (Tagalog, Hanunóo, Buhid, and Tagbanwa) are proposed in this document.

1b. Addition of characters to existing block? Name?

No.

2. Number of characters

81 (20, 23, 20, 18).

3. Proposed category

Category A.

4. Proposed level of implementation and rationale

The Philippine scripts require Level 2 implementation as other Brahmic scripts do.

5a. Character names included in proposal?

Yes.

5b. Character names in accordance with guidelines?

Yes.

5c. Character shapes reviewable?

Yes (see below).

6a. Who will provide computerized font?

Hector Santos via Michael Everson.

6b. Font currently available?

Yes.

6c. Font format?

TrueType.

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes.

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

N1755 (the WG2 hardcopy, not the online version).

8. Does the proposal address other aspects of character data processing?

Yes (see below).

C. Technical -- Justification

1. Contact with the user community?

Yes: Hector Santos. See <http://www.bibingka.com/dahon>.

2. Information on the user community?

See below.

3a. The context of use for the proposed characters?

Tagalog was formerly (until the mid-1700s) used to write Tagalog and Ilocano languages; Hanunóo, Buhid, and Tagbanwa are used today to write the Hanunóo, Buhid, and Tagbanwa languages.

3b. Reference

See bibliography below.

4a. Proposed characters in current use?

Yes

4b. Where?

In the Philippines. The scripts also enjoy some use in North America by Philippine communities for various purposes.

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Contemporary use and accordance with the Roadmap.

6. Should characters be kept in a continuous range?

Yes, they should be encoded in a single block as presented here.

7a. Can the characters be considered a presentation form of an existing character or character sequence?

No.

7b. Where?

7c. Reference

8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?

No. There is some similarity between the four scripts but the differences are striking enough to warrant against their unification.

8b. Where?

8c. Reference

9a. Combining characters or use of composite sequences included?

Yes.

9b. List of composite sequences and their corresponding glyph images provided?

No. Some information is in the hardcopy samples circulated to WG2 in N1755. Vowel signs combine with all letter characters; in some cases unique logotypes result from the combination.

10. Characters with any special properties such as control function, etc. included?

No.

E. Proposal

User community

Tagalog is a script of the Philippines. It was formerly used to write the Tagalog, Bisaya, Ilocano, and other languages. The Tagalog language now utilizes the Latin script. The Tagalog script is distantly related to the southern Indian scripts, but the exact route by which they were brought to the Philippines is not certain. It seems that they may have been transported by way of the palaeographic scripts of Western Java between the 10th and 14th centuries. Written accounts of the Tagalog script by Spanish missionaries, and documents in Tagalog, are known from the period of initial Spanish incursion (mid-1500s). The Tagalog script had fallen out of normal use by the mid-1700s. It has three living descendants: the Hanunóo, Buhid (also called Mangyan), and Tagbanwa (also called Bisaya) scripts, also part of this proposal.

Structure

Vowel signs are used in a manner similar to that employed by other Brahmic scripts. The vowel I is written with a mark above, and the vowel U with an identical mark below, the associated consonant. The mark is known as *kudlit* or *tulbok* in Buhid and *ulitan* in Tagbanwa. The script has only the two vowel signs I and U, which are also used respectively to stand for the vowels E and O. Though all languages normally written with the Tagalog script have syllables possessing final consonants, they cannot normally be expressed in this script. Reforms to express final consonants with a virama character were proposed for the Tagalog script, but were rejected by native users, who considered the script adequate without it. A similar reform for the Hanunóo script seems to have been better received. These signs were not proposed for all of the scripts; because they are found in existing character sets (by Hector Santos), they are encoded here for Tagalog and Hanunóo. The HANUNOO PAMUDPOD has been used by the Hanunóo themselves; the inclusion of the Tagalog permits, at least, the representation of the texts proposing its addition (documents of historical interest) in UCS encoding. Other reforms, such as the addition of E and O vowel signs or the letter FA, have not been included here as they are wanting attestation (even in existing character sets). There is room in the tables for their later addition should it prove necessary.

Directionality

The Philippine scripts are read from left to right in horizontal lines running from top to bottom. They may be *written* either in that manner, or in vertical lines running from bottom to top, moving from left to right. In the latter case, the letters are *written* sideways so they may be *read* horizontally. This method of writing is probably due to the medium and writing implements used. Text is often scratched with a sharp instrument onto beaten strips of bamboo which are held pointing away from the body and worked from the proximal to distal ends, from left to right.

Ordering

UTR#3 states: “The alphabetical order of Tagalog is known from Tagbanwa speakers and is described in folktales. This order is used in the accompanying charts. The two vowel signs are added at the end of the

alphabet.” The names list in UTR#3, however, is (except for the vowel signs) given in Latin alphabetical order (*a, i, u, ba, da, ga, ha, ka, la, ma, na, nga, pa, sa, ta, wa, ya, -i, -u*). Daniels & Bright give another ordering, “based on the 16th-century Tagalog sequence” (*a, i, u, ha, pa, ka, sa, la, ra, ta, na, ba, ma, ga, da, ya, nga, wa*). This proposal gives the characters in the traditional Brahmic order (*a, i, u, ka, ga, nga, ta, da, na, pa, ba, ma, ya, ra, la, wa, sa, ha*), which is followed in many sources, including Santos 1994 and 1995 (source of the fonts used in this proposal). The accompanying chart is divided into four segments, from left to right: Tagalog, Hanunóo, Buhid, Tagbanwa. Each of these 2-column segments should be given a separate collection ID in annex A of ISO/IEC 10646.

Processing

The Philippine scripts are written from left to right and follow the usual Brahmic pattern. Consonants have an inherent /a/ vowel sound, and can be written with either a vowel sign or (in the case of Tagalog and Hanunóo) a null “vanishing vowel” sign. In some cases, the vowel signs simply rest over or under the consonants. In Hanunóo and Buhid, however, special conjoined glyphs are formed (Santos 1995).

Punctuation

Punctuation has been unified for the Philippine scripts. In the Hanunóo block, PHILIPPINE SINGLE PUNCTUATION and PHILIPPINE DOUBLE PUNCTUATION are encoded. Tagalog makes use only of the latter; Hanunóo, Buhid, and Tagbanwa make use of both of them.

Code positions proposed in the table are based on my Roadmap to the BMP, version 2.10.

Unicode Character Properties

Spacing letters, category “Lo”, bidi category “L” (strong left to right)

1380-138C, 138E-1391, 2960-2971, 29C0-29D1, 29E0-29EC, 29EE-29F0

Non-spacing marks, category “Mn”, bidi category “ON” (other neutral); combining priorities in parentheses:

1392, 2972, 29D2, 29F2 (232)

1393, 2973, 29D3, 29F3 (222)

1394 (9)

Spacing marks, category “Mc”, bidi category “ON” (other neutral); combining priorities in parentheses:

2974 (9)

Symbols, category “Po”, bidi category “L” (strong left to right)

2975-2976

Bibliography

- Kuipers, Joel C., and Ray McDermott. 1996. “Insular Southeast Asian Scripts”, in Peter T. Daniels and William Bright, eds. *The world’s writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0
- Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8
- Francisco, Juan R. 1973. *Philippine Palaeography*. Philippine Journal of Linguistics, Special Monograph Issue Number 3. Quezon City: Linguistic Society of the Philippines.
- Haarmann, Harald. 1990. *Die Universalgeschichte der Schrift*. Frankfurt: Campus. ISBN 3-593-34346-0
- Nakanishi, Akira. 1990. *Writing systems of the world: alphabets, syllabaries, pictograms*. Rutland, VT: Charles E. Tuttle. ISBN 0-8048-1654-9
- Santos, Hector. 1994. *The Tagalog script*. (Ancient Philippine Scripts Series; 1). Los Angeles: Sushi Dog Graphics.
- Santos, Hector. 1995. *The living scripts*. (Ancient Philippine Scripts Series; 2). Los Angeles: Sushi Dog

Graphics.

Unicode Consortium. 1992. *Unicode Technical Report #3: exploratory proposals*.

Wolf, Edwin, II, ed. 1947. *Doctrina christiana: the first book printed in the Philippines, Manila 1593. A facsimile of the copy in the Lessing J. Rosenwald Collection*. Washington, DC: Library of Congress.

Sample from Wolf 1947:85

lalo salabat. Ang ycalua ybig
in mo naman ang capova mo tavó
parangang catauan mo. Amen.

Jesus.

Handwritten text in a cursive script.



Handwritten text in a cursive script.

Handwritten text in a cursive script.

Handwritten text in a cursive script.

Handwritten text in a cursive script.

Handwritten text in a cursive script.

Handwritten text in a cursive script.

Handwritten text in a cursive script.

Handwritten text in a cursive script.

Handwritten text in a cursive script.

Sample from Santos 1994, illustrating a page from Francisco Lopez' *Libro a naisuratan amin ti bagas ti Doctrina Cristiana*, 1620. Note his implementation of the cruciform VIRAMA.

සර් සබ්වුඤ්ඤා

භවං සර් භවං භවං සර් භවං භවං

සර් භවං භවං භවං භවං

භවං භවං

භවං භවං

භවං භවං සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

|| සර් සර් ||

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

සර් සර් සර් සර් සර් සර් සර් සර් සර් සර්

Proposal to encode Philippine scripts in the UCS

Sample from Santos 1995. On the left, Hanunóo; on the right, Buhid, showing the consonants in combination with vowel signs, including special logotypes in the with the vowel signs I and U.

	A	I	U		A	I	U
B	✓	✓̇	✓̈	B	✓	✓̇	✓̈
D	7	7̇	7̈	D	7	7̇	7̈
G	∩	∩̇	∩̈	G	∩	∩̇	∩̈
H	∨	∨̇	∨̈	H	∨	∨̇	∨̈
K	∅	∅̇	∅̈	K	∅	∅̇	∅̈
L	∩ ₃	∩̇ ₃	∩̈ ₃	L	∩	∩̇	∩̈
M	∩	∩̇	∩̈	M	∩	∩̇	∩̈
N	∩	∩̇	∩̈	N	∩	∩̇	∩̈
P	∩	∩̇	∩̈	P	∩	∩̇	∩̈
S	∩ ₄	∩̇ ₄	∩̈ ₄	S	∩	∩̇	∩̈
T *	∩	∩̇	∩̈	T *	∩	∩̇	∩̈
W	∩	∩̇	∩̈	W	∩	∩̇	∩̈
Y	∩	∩̇	∩̈	Y	∩	∩̇	∩̈
NG	∩	∩̇	∩̈	NG	∩	∩̇	∩̈
R	—	f	t	R	—	f	t

Proposal for Tagalog, Hanunóo, Buhid and Tagbanwa Scripts

	138	139	296	297	29C	29D	29E	29F
0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								
B								
C								
D								
E								
F								

G = 00
P = 00

Proposal for Tagalog, Hanunóo, Buhid and Tagbanwa Scripts

dec	hex	Name
	80	TAGALOG LETTER A
	81	TAGALOG LETTER I
	82	TAGALOG LETTER U
	83	TAGALOG LETTER KA
	84	TAGALOG LETTER GA
	85	TAGALOG LETTER NGA
	86	TAGALOG LETTER TA
	87	TAGALOG LETTER DA
	88	TAGALOG LETTER NA
	89	TAGALOG LETTER PA
	8A	TAGALOG LETTER BA
	8B	TAGALOG LETTER MA
	8C	TAGALOG LETTER YA
	8D	(This position shall not be used)
	8E	TAGALOG LETTER LA
	8F	TAGALOG LETTER WA
	90	TAGALOG LETTER SA
	91	TAGALOG LETTER HA
	92	TAGALOG VOWEL SIGN I
	93	TAGALOG VOWEL SIGN U
	94	TAGALOG SIGN VIRAMA
	95	(This position shall not be used)
	96	(This position shall not be used)
	97	(This position shall not be used)
	98	(This position shall not be used)
	99	(This position shall not be used)
	9A	(This position shall not be used)
	9B	(This position shall not be used)
	9C	(This position shall not be used)
	9D	(This position shall not be used)
	9E	(This position shall not be used)
	9F	(This position shall not be used)
	60	HANUNOO LETTER A
	61	HANUNOO LETTER I
	62	HANUNOO LETTER U
	63	HANUNOO LETTER KA
	64	HANUNOO LETTER GA
	65	HANUNOO LETTER NGA
	66	HANUNOO LETTER TA
	67	HANUNOO LETTER DA
	68	HANUNOO LETTER NA
	69	HANUNOO LETTER PA
	6A	HANUNOO LETTER BA
	6B	HANUNOO LETTER MA
	6C	HANUNOO LETTER YA
	6D	HANUNOO LETTER RA
	6E	HANUNOO LETTER LA
	6F	HANUNOO LETTER WA
	70	HANUNOO LETTER SA
	71	HANUNOO LETTER HA
	72	HANUNOO VOWEL SIGN I
	73	HANUNOO VOWEL SIGN U
	74	HANUNOO SIGN PAMUDPOD
	75	PHILIPPINE SINGLE PUNCTUATION
	76	PHILIPPINE DOUBLE PUNCTUATION
	77	(This position shall not be used)
	78	(This position shall not be used)
	79	(This position shall not be used)
	7A	(This position shall not be used)
	7B	(This position shall not be used)
	7C	(This position shall not be used)
	7D	(This position shall not be used)
	7E	(This position shall not be used)
	7F	(This position shall not be used)
	C0	BUHID LETTER A
	C1	BUHID LETTER I
	C2	BUHID LETTER U
	C3	BUHID LETTER KA
	C4	BUHID LETTER GA
	C5	BUHID LETTER NGA
	C6	BUHID LETTER TA
	C7	BUHID LETTER DA
	C8	BUHID LETTER NA
	C9	BUHID LETTER PA
	CA	BUHID LETTER BA
	CB	BUHID LETTER MA
	CC	BUHID LETTER YA
	CD	BUHID LETTER RA
	CE	BUHID LETTER LA
	CF	BUHID LETTER WA
	D0	BUHID LETTER SA
	D1	BUHID LETTER HA
	D2	BUHID VOWEL SIGN I
	D3	BUHID VOWEL SIGN U
	D4	(This position shall not be used)
	D5	(This position shall not be used)
	D6	(This position shall not be used)

Group 00

dec	hex	Name
	D7	(This position shall not be used)
	D8	(This position shall not be used)
	D9	(This position shall not be used)
	DA	(This position shall not be used)
	DB	(This position shall not be used)
	DC	(This position shall not be used)
	DD	(This position shall not be used)
	DE	(This position shall not be used)
	DF	(This position shall not be used)
	E0	TAGBANWA LETTER A
	E1	TAGBANWA LETTER I
	E2	TAGBANWA LETTER U
	E3	TAGBANWA LETTER KA
	E4	TAGBANWA LETTER GA
	E5	TAGBANWA LETTER NGA
	E6	TAGBANWA LETTER TA
	E7	TAGBANWA LETTER DA
	E8	TAGBANWA LETTER NA
	E9	TAGBANWA LETTER PA
	EA	TAGBANWA LETTER BA
	EB	TAGBANWA LETTER MA
	EC	TAGBANWA LETTER YA
	ED	(This position shall not be used)
	EE	TAGBANWA LETTER LA
	EF	TAGBANWA LETTER WA
	F0	TAGBANWA LETTER SA
	F1	(This position shall not be used)
	F2	TAGBANWA VOWEL SIGN I
	F3	TAGBANWA VOWEL SIGN U
	F4	(This position shall not be used)
	F5	(This position shall not be used)
	F6	(This position shall not be used)
	F7	(This position shall not be used)
	F8	(This position shall not be used)
	F9	(This position shall not be used)
	FA	(This position shall not be used)
	FB	(This position shall not be used)
	FC	(This position shall not be used)
	FD	(This position shall not be used)
	FE	(This position shall not be used)
	FF	(This position shall not be used)

Plane 00

Row XX