

ISO/IEC JTC/1 SC/2 WG/2
Universal Multiple-Octet Coded Character Set (UCS)
Secretariat: ANSI

Title: Philippino characters (status report)

Doc. Type: Expert report

Source: Takayuki K. Sato

Project: 02:18.01

Status: FYI interim report

Date: 2000-02-22

Distribution: ISO/IEC JTC/1 SC/2 WG2

Reference: WG2 N1755

Medium:

One of the action items from WG2 to Takayuki K. Sato-Japan is to contact with experts and authority of Philippines for their comments for the N1755. This is a status report of the action item.

Attached paper was presented by Dr. Albacea of University of Philippines Lao Bafios at the MLIT symposium which Sato is acting as an organizer.

Surprisingly the paper is well consistent with the N1755 which was submitted by Mike Everson to the WG2 before (and asking for review).

Unfortunately, Sato also recognized that the source of information for both papers are the same. It might be necessary to verify the idea with other soueces. Therefore, Sato is asking for a review of the n1755 again by the authority of Philippines.

It would require some more time for Sato to have an answer for the action item.

Sato

CODING SCHEMES FOR PHILIPPINE SCRIPTS¹

Eliezer A. ALBACEA²

Institute of Computer Science, University of the Philippines Los Baños

eea@ics.uplb.edu.ph

Introduction

Unknown to many, a writing system is already in place in the Philippines long before the Spaniards came to the islands (long before 1521). The natives in the Philippines wrote on bamboo and specially prepared palm leaves using knives and styli. Antonio Pigafetta, the chronicler who came to the Philippines with Magellan in 1521, found evidence of writing skills among the natives. When Legazpi came to Manila in 1571, he also observed that the natives knew how to read and write. The script used is what is known today as the Tagalog script. This was documented by Pedro Chirino, a Jesuit historian, who wrote in his 1604 *Relacion de las Islas Filipinas*, “*all these islanders are much given to reading and writing, and there is hardly a man, much less a woman, who does not read and write.*” Many other historians found the same. Dr. Antonio Morga, the Senior Judge Advocate of the High Court of Justice and commander of the galleon San Diego, wrote in his 1609 *Sucesos de las Islas Filipinas*, “*almost all the natives, both men and women, write in this language. There are very few who do not write it excellently and correctly*” [9].

The Tagalog script was widely used until the early days of the Spanish regime. By the end of the 17th century its use was almost non-existent and by the late 18th century, it was extinct. The Tagalog script was replaced by the Latin script, a script introduced by the Spaniards. There are several theories proposed to explain the disappearance of the Tagalog script. Two of them are:

1. The Latin alphabet was easily learned by the natives. Proficiency in the new system afforded social and economic benefits.
2. The Tagalog script did not keep step and became inadequate in representing the new sounds of some words that were borrowed from the Spanish language [15].

Although the Tagalog script died centuries ago, there are scripts (variation of the Tagalog script) that survived the test of time and are still being used by some minorities in the Philippines. These scripts are the Hanunoo and Buhids of Mindoro and the Tagbanwas of Palawan. These minorities originally lived in the coastal areas along the ancient trade and migration route between Borneo and Manila on the western flanks of Mindoro and Palawan. These minorities are forced to move inland and by refusing to give up their way

¹ Paper presented at the Fourth Symposium on Standardization of Information Technology – MLIT-4, Yangon, Myanmar, October 27-28, 1999.

² Professor of Computer Science and Director.

of life they were able to preserve their language. Not much were known about them, but late in the 19th century, their scripts and language were discovered and studied [13].

Another script almost unrelated to the Tagalog, Hanunoo, Buhid, and Tagbanwa scripts is the Eskaya script. This script is still being used today by the Eskayas of the Bohol island in the Philippines.

In this paper, we survey the dead and living scripts in the Philippines. A proposal of how to code these scripts will also be given.

The Tagalog Script

The pre-Hispanic Tagalog script is a simple and elegant system. It is better known as *alibata*, a term coined in 1914 by Dean Paul Versoza of the University of Manila. The term came from *alif*, *ba*, and *ta*, the first three letters of the Maguindanao arrangement of the Arabic letters. The term was probably coined due to the Middle East origin of the script.

CONSONANTS		VOWELS	
ka	𐑉	a	𐑊
ga	𐑋	i	𐑌
nga	𐑍	u	𐑎
ta	𐑏		
da	𐑑		
na	𐑓		
pa	𐑕		
ba	𐑗		
ma	𐑙		
ya	𐑛		
la	𐑝		
wa	𐑟		
sa	𐑡		
ha	𐑣		
		STOP	//

Figure 1. ALIBATA ALPHABET

The Tagalog script had 17 basic symbols, three of which were the vowels *a*, *i*, and *u* (Figure 1). Each basic consonant symbol had the inherent *a* sound: *ka*, *ga*, *nga*, *ta*, *da*, *na*, *pa*, *ba*, *ma*, *ya*, *la*, *wa*, *sa*, and *ha*.

A diacritical mark called *kudlit* modified the sound of the symbol. The *kudlit* could be a dot, a short line (similar to an apostrophe), or even an arrowhead. When placed above a consonant symbol, it changed the inherent sound of the consonant from *a* to *i*; placed below, the sound became a *u*. See Figure 2 for a complete list of the alphabet and the corresponding modification in sound when a *kudlit* is included.

The Tagalog script was a syllabary, which means that each symbol represents a complete syllable. This is in contrast to the Latin alphabet of the modern Filipino language which is phonetic.

The Tagalog script allowed the representation of two kinds of syllables: V and CV (V for vowel and C for consonant), although the language when spoken had V, CV, VC, and CVC syllables. Hence, only syllables of V and CV types could be written down. For example, syllables like *a*, *bi*, and *mu* could be written down but not the syllables *at*, *kam*, *pit* or *ting*. The language did not have consonant clusters like the CCVC type, e.g., *tram*.

ka	𑀓	ki	𑀓	ku	𑀓
ga	𑀔	gi	𑀔	gu	𑀔
nga	𑀕	ngi	𑀕	ngu	𑀕
ta	𑀖	ti	𑀖	tu	𑀖
da	𑀗	di	𑀗	du	𑀗
na	𑀘	ni	𑀘	nu	𑀘
pa	𑀙	pi	𑀙	pu	𑀙
ba	𑀚	bi	𑀚	bu	𑀚
ma	𑀛	mi	𑀛	mu	𑀛
ya	𑀜	yi	𑀜	yu	𑀜
la	𑀝	li	𑀝	lu	𑀝
wa	𑀞	wi	𑀞	wu	𑀞
sa	𑀟	si	𑀟	su	𑀟
ha	𑀠	hi	𑀠	hu	𑀠

Figure 2. THE POSSIBLE SOUNDS OF THE CONSONANTS

Hanunoo, Buhid, and Tagbanwa Scripts

The scripts of the Hanunoos, Buhids and Tagbanwas bear similarities with the Tagalog script. The symbols and shapes are almost similar. They had the same *kudlits*, had the same orthographic rule about dropping the final consonant in a CVC syllable, and had the same uses for the scripts: writing poetry and personal communications. These similarities reinforce the theory that these minority languages may have originated from the same language where the Tagalog script came from.

Figure 3 shows the two styles of the Buhid alphabet. Both styles can be seen in their writings. The *kudlit* in the Buhid script is in the form of a horizontal line. As in the Tagalog script, the *i* and *u* sound is formed by adding a *kudlit* above or below the symbol for the consonant. Illustrations are given in Figure 3.

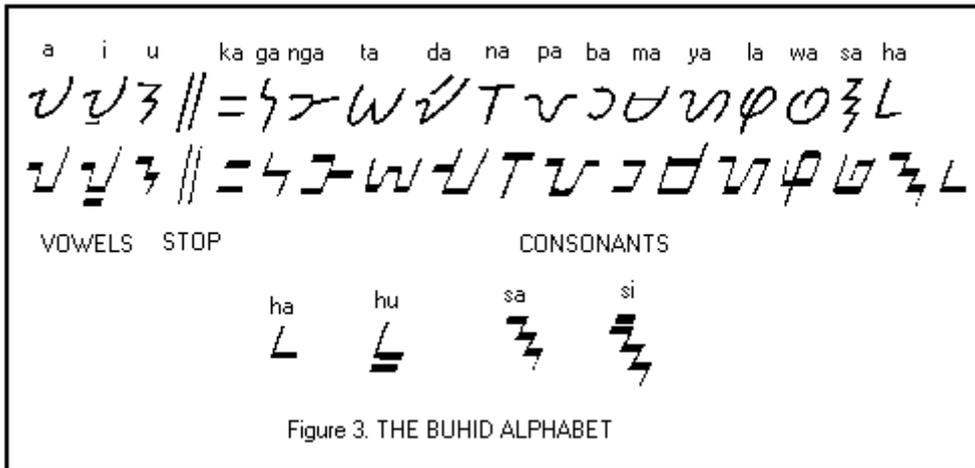


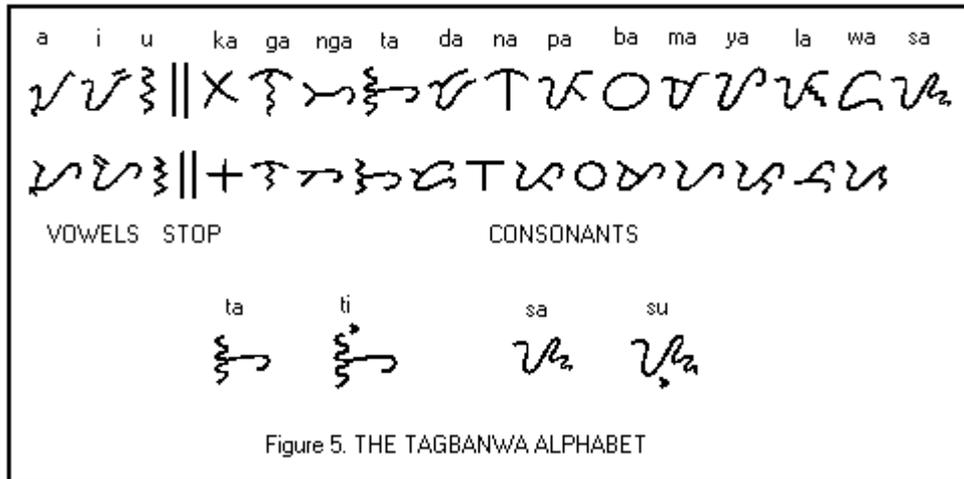
Figure 3. THE BUHID ALPHABET

Figure 4 shows the two styles of the Hanunoo alphabet. Similarly, the two styles can be observed from the writings of Hanunoo of Palawan. The *kudlit* in this case is in the form of a short diagonal line. The same rule as in the Tagalog script applies for the kudlit.



Figure 4. THE HANUNOO ALPHABET

Finally, Figure 5 shows the two styles of the Tagbanwa alphabet. Unlike the Tagalog, Buhid and Hanunoo where each has 3 vowels and 14 consonants in the alphabet, the Tagbanwa alphabet has 3 vowels and 13 consonants. Closer examination of some documents written using the Tagbanwa script reveal that the *ha* sound is not present in the Tagbanwa language. The *kudlit* is also different. It comes in the form of a small arrowhead.

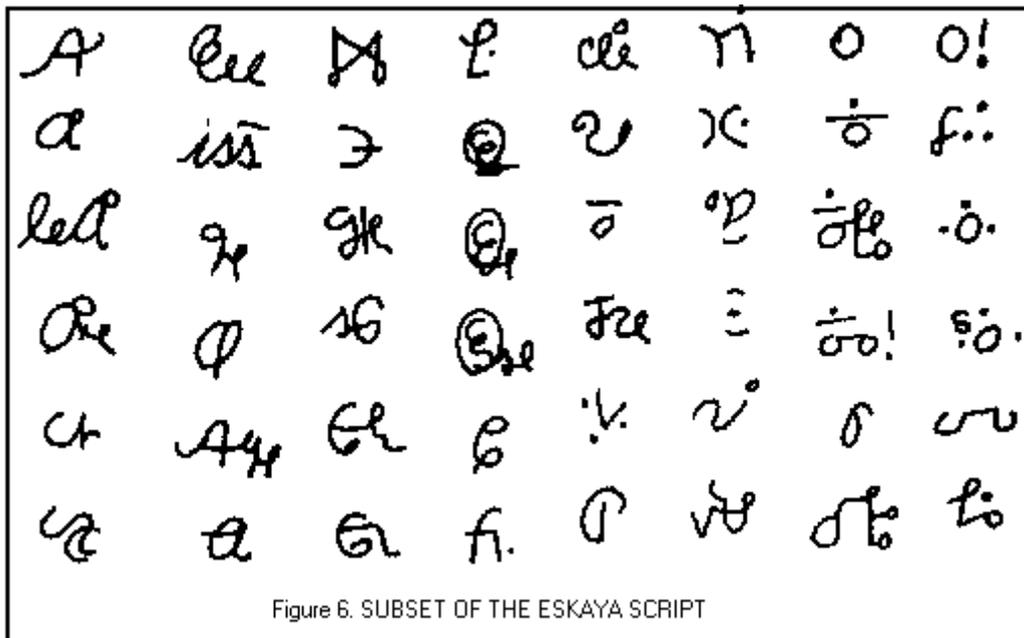


The Eskaya Script

The Eskaya of Bohol is a group of minorities who claim to be direct descendants of the people of the kingdoms of Butuan, Sumatra, and the Middle East. Although they look like the other natives of Bohol (and the Philippines), they have their own writing system.

The Eskaya writing system is reported to be composed of close to a thousand characters. A subset of these characters is given in Figure 6. Some of the characters represent sounds that do not exist in Philippine languages nor in most Austronesian languages. Some symbols are consonant clusters that are not natural sounds in any language spoken in the region.

The basic structure of the script is that of a syllabary. Like Indic scripts, the basic value of certain symbols are modified by ligatures. There are characters for V, CV, CVC, CCV, CCVC, VC, VCC, CVCC, CCVCC, and diphthongs. This large variety of composite characters is the reason why there are close to a thousand characters in the Eskaya script. For example, the script has different character representations for *ba*, *bi*, and *be*. Also, the language is reported to have the characteristics of logograms, with some symbols doubling as representations for words and ideas while at the same time representing sounds. This is similar to the Chinese, Egyptian and Mayan writing systems.



The Latin Script

With more than three hundred years of presence in the Philippines, the Spaniards greatly influenced the writing system in the Philippines. In fact, the Spaniards were successful in replacing the native Tagalog script with the Latin script. Admittedly, the Latin script can represent a greater number of sounds compared to the Tagalog script. This greatly contributed to the acceptance of the Latin script.

The character set (Figure 7) of the present Latin script is exactly similar to that of the English language plus the two characters *Ñ (ñ)* and *Ng (ng)*. The character *Ñ (ñ)* was borrowed from the Spanish character set and *Ng (ng)* was included as one character due to the numerous Filipino native words using these two letters as one. The *Ng* character almost surely was adopted from the pre-Hispanic Tagalog script.

Filipino Alphabet	A B C D E F G H I J K L M N Ñ Ng O P Q R S T U V W X Y Z
English Alphabet	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Figure 7. THE LATIN FILIPINO ALPHABET

The present character set (Figure 7) evolved from earlier character sets that were extended to accommodate new words that are of foreign origin. For example, only about thirty or so years ago, the character set of the Latin script did not include the letters *C, F, J, Q, V, X,* and *Z*. This earlier character set of the Latin script is given in Figure 8.

Filipino Alphabet	A B K D E G H I L M N N g O P R S T U W Y
English Alphabet	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Figure 8. THE LATIN FILIPINO ALPHABET (older version)

With this character set, a C sound is handled by either K or S (e.g., cement – semento, castle – kastilyo), an F by P (e.g., family – pamilya), a J by H (e.g., Japanese – Hapones), a Q by a K (e.g., quality – kalidad), a V by B (e.g., vacant – bakante), X by KS (e.g., boxing – boksing), and Z by S (e.g., zone – sona).

Coding Schemes for the Scripts

Since most of the characters in the Latin script are already defined in the ISO standards, the simplest approach is to simply add two new codes, one for Ng and another for Ñ. This will allow all the characters in the Latin script to be defined as one keystroke.

The Tagalog, Hanunoo, Buhid and Tagbanwa have two types of symbols:

1. basic character symbol; and
2. basic character symbol with a *kudlit* either on top or at the bottom.

With this characteristic of scripts, coding can be handled using any of the three methods mentioned in Kobayahi [1]. We discuss below how each method can be used to code most of the Philippine scripts.

The simplest coding scheme is to use one code for each basic symbol, one code for each basic symbol with the *kudlit* at the top, and one code for each basic symbol with the *kudlit* at the bottom. Since the character sets of the Tagalog, Hanunoos, Buhids, Tagbanwas, and the Latin scripts are quite small in number, this method is a practical choice. The Tagalog, Hanunoo and Buhid scripts have 45 symbols each, and the Tagbanwa has 42 symbols.

However, the number of codes can still be reduced by adopting the ISO 6937 method. In this method a symbol with a *kudlit* may be produced by typing the *kudlit* (nonspacing) plus

the basic symbol. Or, it can be handled using the ISO/IEC 10646 method where a symbol with a *kudlit* may be produced by typing the symbol plus a backspace plus the *kudlit*.

The Eskaya script, on the other hand, contains composite characters that are formed from basic symbols. Hence, it will have to be handled in a manner similar to other Asian scripts like the Thai, Myanmar, Nepali, and Devanagari scripts. Further studies of the Eskaya script must be done to really identify all the symbols in the script. At present, what is reported in the literature is just a subset of the script.

Computer Implementations of Scripts

Except for \tilde{N} and *Ng*, all the characters in the Latin script can be produced by pressing a key in the English keyboard. The character *Ng* (*ng*) can basically be produced using the same English keyboard using two keys (i.e., by pressing N followed by g) and \tilde{N} (\tilde{n}) can be produced using a combination of the control key + shift key + ~ and followed by *N* or *n* in the English keyboard. This is how \tilde{N} (\tilde{n}) is produced in software like Microsoft Word and other Windows-based software.

For the other scripts, there are commercially available software for producing them. There are packages for the Tagalog, Hanunoo, Buhid, and Tagbanwa scripts. One is the Sushi Dog Graphics font package for IBM and Macintosh platforms which are sold for a reasonable amount. This Sushi Dog Graphics package may run on all Windows-based software and on all Macintosh applications [12,14].

References

1. Kobayashi, T.L. Input Method API Makes Applications Free!!, Proceedings of MLIT-3, Hanoi, Vietnam, October 6-7, 1998, 175-200.
2. Tashiro, S. Report from output-processing system group, Proceedings of MLIT-3, Hanoi, Vietnam, October 6-7, 1998, 177-188.
3. Tuladhar, A.B. Nepali Font Standards, Proceedings of MLIT-3, Hanoi, Vietnam, October 6-7, 1998, 79-106.
4. Albacea, E.A. Multilingual Needs: The Case of the Philippines, Proceedings of MLIT-3, Hanoi, Vietnam, October 6-7, 1998, 107-110.
5. Mikami, Y. Towards Multilingual Information Processing, Proceedings of MLIT-3, Hanoi, Vietnam, October 6-7, 1998, 145-158.
6. Koanantakool, T., Tanprasert, C., Vivaran, C., and Meknavin, S. Current Status of Thai Language Processing and Multilingual Processing, Proceedings MLIT-2, Tokyo, Japan, November 7-8, 1997.
7. Albacea, E.A. Current Status of Language Processing and Multilingual Processing in the Philippines. Proceedings 2nd International Symposium on Standardization of Multilingual Information Technology, Tokyo, Japan, November 7-8, 1997, 153-163.
8. Albacea, E.A. and Topacio, C.A. Philippine Language and Multilingual Processing: A Country Report. Proceedings 1st International Symposium on Standardization of Multilingual Information Technology - MLIT 97, Singapore, May 26-28, 1997, 139-142.

9. Santos, H. "Literacy in Pre-Hispanic Philippines" in A Philippine Leaf at <http://www.bibingka.com/dahon/literacy/literacy.htm>, USA, October 26, 1996.
10. Santos, H. "Mystery Scripts in the Philippines" in A Philippine Leaf at <http://www.bibingka.com/dahon/mystery/mystery.htm>, USA, October 26, 1996.
11. Santos, H. "The Tagalog Script" in A Philippine Leaf at <http://www.bibingka.com/dahon/tagalog/tagalog.htm>, USA, October 26, 1996.
12. Santos, H. "Computer Fonts, Tagalog Script" in A Philippine Leaf at <http://www.bibingka.com/dahon/misc/tagfonts.htm>, USA, October 26, 1996.
13. Santos, H. "Our Living Scripts" in A Philippine Leaf at <http://www.bibingka.com/dahon/living/living.htm>, USA, Jan 31, 1997.
14. Santos, H. "Computer Fonts, Living Scripts" in A Philippine Leaf at <http://www.bibingka.com/dahon/misc/livfonts.htm>, USA, October 26, 1996.
15. Santos, H. "Extinction of a Philippine Script" in A Philippine Leaf at <http://www.bibingka.com/dahon/extinct/extinct.htm>, USA, October 26, 1996.
16. Santos, H. "The Eskaya Script" in A Philippine Leaf at <http://www.bibingka.com/dahon/mystery/eskaya.htm>, USA, January 25, 1997.
17. Tirol, J.B. ESKAYA OF BOHOL: Its Writing System, The Bohol Chronicle XL, Number 9, July 4, 1993.