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

Doc Type:	**Working Group Document**
Title:	**Proposal to add Mayanist Latin letters to the UCS**
Source:	**Michael Everson**
Status:	**Individual Contribution**
Date:	**2006-01-30**

1. Background. In N2931, Lorna Priest and Peter Constable proposed the addition of ɛ LATIN LETTER TRESILLO and 4 LATIN LETTER CUATRILLO to the UCS in support of archaic letters used in 16th-century Guatemala to write Mayan languages such as Cakchiquel, Quiché, and Tzutuhil. Although these two letters were accepted for ballotting in PDAM3 of ISO/IEC 10646, as a set of characters they are inadequate to represent texts in normalized 16th-century orthography which use these letters. Such normalization may be rare—it certainly has been in the past—but it should nevertheless be supported by the UCS.

The letters in question were devised by Brother Francisco de la Parra (†1560 in Guatemala) and were used by a number of early linguist-missionaries to represent sounds occurring in Cakchiquel, Quiché, and Tzutuhil. In his edition of the *Annals of the Cakchiquels*, Brinton 1885 gives a set of four letters (one of which is used as a digraph with *h*) with the following glyphs, alongside descriptions which he attributes to the grammarian Torresano:

ɛ   TRESILLO represented "the only true guttural in the language, being pronounced forcibly from the throat, with a trilling sound (*castañeteando*)". This is now described as [q'], the glottalized uvular stop.

4   CUATRILLO represented "a trilled palatal, between a hard *c* and a *k*". This is now described as [k'], the glottalized velar stop.

4̦   CUATRILLO WITH COMMA represented a sound "somewhat like the *c* with the cedilla, ç, only more quickly and with greater force—*ds* or *dz*". This is now described as [ts'], the glottalized alveolar affricate.

tʒ   TZ "resembles the '4 with comma' but is described as softer, the tongue being brought into contact with the teeth, exactly as *tz* in German". This is now described as [ts], the alveolo-palatal affricate.

4̦h   CUATRILLO WITH COMMA AND H represented "a compound sound produced by combining the cuatrillo with a forcible aspirate". This is now described as [tʃ'], the glottalized alveolo-palatal affricate. Note that the comma is an integral part of the character, not a spacing comma—it is 4̦h, not 4̦,h—nor is it a combining comma below. (Pp 49–50; see Figures 1 and 1.)

Brinton follows this with a discussion of Parra's characters by Otto Stoll; I have given them alongside modern transcriptions:

"The four new signs added to the European alphabet, by some of the old writers on Cakchiquel (Parra, Flores), viz: ɛ, 4, 4̦, 4̦h, are but phonetic modifications of four corresponding signs of the common alphabet. so we get four pairs of sounds, namely:—

c and ꝗ;	[k, k']
k and ε	[q, q']
ch and ꝗh (> ꝗh ?)	[tʃ, tʃ']
tz (i.e. ƺ) and ꝗ	[ts, ts']

> forming two series of consonants, the former of which represents the common letters, and the latter their respective 'cut letters,' which may be described as being pronounced with a shorter and more explosive sound than the corresponding common letter, and separated by a short pause from the preceding or following vowel." (Pp 50–51; see Figure 2.)

Neither Brinton nor Stoll discusses the cameral nature of these "additions to the European alphabet". There is, however, *no reason* to *assume*—as Priest and Constable have done—that the 16th-century devisors considered these letters to be *any* different from any other Latin letters.

2.0. Case. In N2931, Priest and Constable posited that TRESILLO and CUATRILLO were caseless (by naming them without CAPITAL or SMALL, and by apparently assigning them the property "Lo" analogously to U+01C0 LATIN LETTER DENTAL CLICK). No evidence, however, was presented for this assertion. The examples cited in their proposal were only samples of the letters used to show the existence of the letters; such examples have no reference to use. A decision that these letters must be caseless because such charts do not show them with case leads only to a false economy in encoding, which, in turn, leaves the potential user of these characters in UCS encoding without the choice to use them as ordinary Latin letters in normalized texts. It is true that most of the examples using these letters in modern Mayanist literature simply refer to them in discussions of orthography, and do not use them in running text. Indeed, a number of sources note that Brinton 1885a is one of the few editors who made use of them in running text. The fact that these letters are being encoded at all, however, indicates a concern that future scholars be given the tools to use these letters in their work. When the archaic Coptic letters were encoded, they were encoded as casing letters so that Copticists could make use of them in accord with normal scholarly editorial and typographic practice. The Mayanist letters should, *in principle*, also be considered to be casing, so that normal scholarly editorial and typographic practice can be likewise applied to Mayanist studies.

2.1. Evidence for case in Brinton 1885a. Brinton's usage of Parra's letters in his edition of *The Annals of the Cakchiquels* is notable; indeed he criticizes the Abbé Brasseur de Bourbourg's edition of the *Popol Vuh* for not having used them: the Abbé has, he says, "made use only of the types of the Latin alphabet; and both in this respect and in the fidelity of his translation, he has left much to be desired in the presentation of the work" (p. 52). Having said that, it must be observed that Brinton's typesetters did not favour the Mayanist letters with *any* sort of typographic care: the four letters ε, ꝗ, ꝗ, and ƺ are used indifferently in the text, in both roman and italic contexts, and in both casing and non-casing contexts. Not one of them was designed to harmonize with the 12-point text typeface; indeed they appear to have been cut in 18 points, and these sorts are even used in 9-point footnotes in Brinton 1885a. This *cannot* be considered as evidence that the characters are caseless; it is merely evidence of poor typography.

Nor can the manuscript itself be considered to be entirely definitive as to the question of casing. Brinton states (p. 63): "Capital letters are not often used in the original to distinguish proper names, and as the text has been set up from a close copy of the first text, some irregularities in this respect also must be anticipated." As a modern editor, however, Brinton *does* normalize his texts to distinguish proper names with case for the Latin letters available to him; in both the English and the original Quiché—except where his poor fonts prevent him from setting the Mayanist letters adequately. From Brinton 1885a, pp. 126–129 and 146–148:

66. The chief Citan Qatu ruled, the son of the chief Caynoh, to whom were mystic power and wisdom. Then ruled the chief Qotbalcan. The chief Alinam ruled. Next ruled the chief Xttamer Zaquentol. Then followed in power Qhiyoc Queh Ahgug. In his reign the chief and Galel Xahil Xulu Qatu gathered together the Quiche nation, desiring that war should be declared against those who were attacking the Ginona.… 91. It was on the day 10th Tzy that occurred the destruction of the Quiches at Iximche; but the news of it had not yet reached our ancestors, Oxlahu tzii and Cablahuh Tihax, when the Quiches came to destroy the Zotzil Tukuches.

66. Xahauar ahauh Citan 4atu, ru 4ahol ahauh Caynoh, xa vi 4oh ru puz ru naval ri. Ok xahauar chi4a ahauh 4otbalcan. Xahauar 4a ahauh Alinam xahauar chi4a ahauh, Xttamer Çaquentol. Ok xoc chi4a ahauh 4hiyoc Queh ahƐuƐ. Haok xmolobax el ahauh Ɛalel Xahil Xulu 4atu chire 4echevinak, xax rah ru yac labal ahauh chiree xban vi pa Ɛinona.… 91. Ha 4a chi lahuh 4,ij, rucam ka 4eche vinak chi Yximchee, 4i mani 4a ru tzihol cu4in ka mama Oxlahuh 4,ij, ha Cablahuh Tihax, ok xpeul 4echevinak, camicay richin Ço4,il Tukuchee.

Here we see *Qatu/4atu* beside *son/4ahol* and *Ahgug/ahƐuƐ* (the second should have been *AhƐuƐ*) beside *Galel/Ɛalel*. I am certain that Brinton would have set his text, if the appropriate fonts had been available to him, thus:

66. Xahauar ahauh Citan 4atu, ru 4ahol ahauh Caynoh, xa vi 4oh ru puz ru naval ri. Ok xahauar chi4a ahauh 4otbalcan. Xahauar 4a ahauh Alinam xahauar chi4a ahauh, Xttamer Çaquentol. Ok xoc chi4a ahauh 4hiyoc Queh AhƐuƐ. Haok xmolobax el ahauh Ɛalel Xahil Xulu 4atu chire 4echevinak, xax rah ru yac labal ahauh chiree xban vi pa Ɛinona.… 91. Ha 4a chi lahuh 4,ij, rucam ka 4eche vinak chi Yximchee, 4i mani 4a ru tzihol cu4in ka mama Oxlahuh 4,ij, ha Cablahuh Tihax, ok xpeul 4echevinak, camicay richin Ço4,il Tukuchee.

I say that I am certain that Brinton, had he been able, would have written *Qatu/4atu* beside "*son/4ahol*" and *Ahgug/AhƐuƐ* beside *Galel/Ɛalel*. How can I be certain? The use of 18-pt Ɛ in 12-pt *ahƐuƐ* is an artefact of the fonts available to Brinton. Nothing like that size distinction occurs in the manuscripts. (See also Figures 3 through 5.)

Brinton's use of case in his Vocabulary and Index of Native Proper Names at the back of the book is also clear; as was common in the 19th century, each entry is title-cased (see Figure 2). The alphabetical order he gives is: A, B, C, Ç, Ch, E, H, I, K, L, M, N, O, P, Qu, R, T, U, V, X, Y, Ɛ, 4, 4h, 4,, Tz. Note that although Brinton describes the use of ꜩ as a separate letter in his introduction, he only uses the digraph *tz* in his text and indices. Doubtless this last was *also* for typographic convenience; *cf.* the *Popol Vuh* manuscript referred to below in §2.2.

Brinton also gives (pp 55–56; see also Figure 4) the following title captions in uppercase (he uses the 18-point outsized letters):

VAE MEMORIA CHIRE 4HAOH
THIS IS THE RECORD FOR THE PROCESS

VAE MEMORIA CHIRE VINAK CHIJ
THIS IS THE STATEMENT OF THE TORTS

VAE MEMORIA ƐANAVINAKIL
THIS IS A RECORD OF THE WITNESSES

Were a facsimile edition of *The Annals of the Cakchiquels* available, we could check the manuscript practice more precisely; but it is not. Recall, however, that Brinton described the letters in question as "four new signs added to the European alphabet". This does not imply the kind of caselessness

that we find for African click letters invented in the nineteenth century. Priest and Constable's apparent belief that these letters should be encoded caselessly because they *were* in fact caseless implies that Spanish missionaries in Guatemala *conceived* of the new letters they devised as specifically caseless. This is not, to my mind, credible. To the missionaries, letters were letters, pure and simple. Whether they *applied* casing consistently to proper names is an orthographic question. Let us look at the *Popol Vuh* manuscript for more on the question of casing.

2.2 Evidence in the *Popol Vuh*. Further evidence for the generalization of case for these letters can be found in Brother Francisco Ximénez' 16th-century bilingual manuscript of the Mayan *Popol Vuh*. UNESCO funded the publication of the first facsimile edition of this work for the International Year of the Book in 1973. On the left-hand pages, the Quiché and Spanish text in the Ximénez' handwriting are given; on the right-hand pages, Agustín Estrada Monroy presents a somewhat normalized transcription of the Spanish text. At the beginning of the work, Estrada has transcribed some of the Quiché text, and there are Quiché names throughout the text. Ximénez' hand is quite spidery, but some of the letters in question can be seen fairly clearly. The shapes of the TRESILLO can be seen, looking often like a two-stroke ligature of *c* and inverted breve, often with the strokes disjointed. Sometimes it looks like two *c*s stacked or like a tall open *e*. Estrada transcribes these as <k> or <c> (without much systematicity) in his transcription of the Quiché at the beginning of the text.

That casing is a feature of orthography in general is clear: a few all-caps titles are given, and many lines and names begin with capital letters. Initial capital Z appears at line 4854 of the Spanish text, and in the corresponding line in the Quiché, in the name *Ztayul*; the form of this letter is similar to the modern EZH. Line 49 of the Spanish text reads as a title: ESTE ES SV SER DICHO QVANDO; the corresponding Quiché text is ARE V ꜩIHOXIC VAE with a very carefully drawn capital *tz* ligature, again, with the EZH shape. The same ꜩ is used at the very beginning of the introduction to the text, lines 1 and 2 of both the Quiché and the Spanish text (see Figures 11–16):

ARE V XE OHER   ESTE ES EL PRINCÍPIO DE LAS
ꜩih varal Quiche vbi   antiguas historias aquí en el quiché.

Here the word *Quiche* is capitalized in the Quiché, and not in the Spanish, in the manuscript. Other examples of inconsistent capitalization of names may be found in this manuscript: *Balam* Quiché, *balam* Spanish; *Mexico* Quiché and Spanish; *rabinal* Quiché, *Rabinal* Spanish. This doesn't signify; it does indicate however that case is an expected feature of the orthography used. We are unlucky that the Popol Vuh manuscript does not seem to have an example of CAPITAL LETTER TRESILLO. But this does not mean that TRESILLO "is" caseless, particularly in view of the evidence of TZ, where its capital is so carefully drawn and its lower-case form is as expected in the script handwriting. Certainly a modern editor using the UCS needs to be able to choose CAPITAL LETTER TRESILLO if he or she wishes to.

3. Glyph design. Capital and small ꜩ and ꜩ are both attested in the Popol Vuh manuscript, and the t-ezh ligature shape seems appropriate enough. The CUATRILLOs are also easy enough to design: Take capital J and small j and attach the flag of a 4 to it, extending the horizontal bar far enough to nestle a small comma inside of it for the CUATRILLO WITH COMMA, thus: 4 J ꝯ ꝯ j ꝯ ꝯ. This approach is supported well by all of the examples, even the clumsier ones. The TRESILLO is the most problematic. Brinton's ꜭ is strange in the first place because it goes below the line, but then it is clearly not designed in harmony with the text font he is using. In the manuscripts, the lower-case tresillo sits on the same baseline as does the letter *c*, and the examples show either a sort of two-stroke tall open-*e*/double-*c* shape, or a *c* with a kind of inverted breve attached to it—sometimes indeed not attached to it. I favour the latter and propose <ꜯ> for the normalized shape for the lower-case TRESILLO. Both loops there are made out of the upper stroke of the *c*. I have taken the same approach

in designing the capital TRESILLO, taking the top loop from the upper stroke of a reversed 3, thus: Ɛ Ꞓ. This may be a bit less felicitous, and I would welcome other ideas; repeating the top stroke of capital C does not seem satisfactory: Ꞓ. Note that these are quite distinct from the Africanist hooked Cs: compare Ꞓ ꞓ with Ƈ ƈ.

4. Proposal. I propose the addition of the following letters to the UCS:

2C78 Ꞓ   LATIN CAPITAL LETTER TRESILLO
2C79 ꞓ   LATIN SMALL LETTER TRESILLO
2C7A Ꝺ   LATIN CAPITAL LETTER CUATRILLO
2C7B ꝺ   LATIN SMALL LETTER CUATRILLO
2C7C Ꝺ̦   LATIN CAPITAL LETTER CUATRILLO WITH COMMA
2C7D ꝺ̦   LATIN SMALL LETTER CUATRILLO WITH COMMA
2C7E Ꜩ   LATIN CAPITAL LETTER TZ
2C7F ꜩ   LATIN SMALL LETTER TZ

Note that this entails deleting 2C6F LATIN LETTER TRESILLO and 2C70 LATIN LETTER CUATRILLO from PDAM 3.

Unicode Character Properties
```
2C78;LATIN CAPITAL LETTER TRESILLO;Lu;0;L;;;;;N;;;;2C79;
2C79;LATIN SMALL LETTER TRESILLO;Ll;0;L;;;;;N;;;2C78;;2C78
2C7A;LATIN CAPITAL LETTER CUATRILLO;Lu;0;L;;;;;N;;;;2C7B;
2C7B;LATIN SMALL LETTER CUATRILLO;Ll;0;L;;;;;N;;;2C7A;;2C7A
2C7C;LATIN CAPITAL LETTER CUATRILLO WITH COMMA;Lu;0;L;;;;;N;;;;2C7D;
2C7D;LATIN SMALL LETTER CUATRILLO WITH COMMA;Ll;0;L;;;;;N;;;2C7C;;2C7C
2C7E;LATIN CAPITAL LETTER TZ;Lu;0;L;;;;;N;;;;2C7F;
2C7F;LATIN SMALL LETTER TZ;Ll;0;L;;;;;N;;;2C7E;;2C7E
```

Bibliography.
Brinton, Daniel G. 1885a. *The annals of the Cakchiquels: the original text, with a translation, notes and introduction*. (Library of Aboriginal American Literature; 6) Philadelphia: Brinton's Library of Aboriginal American Literature.
Brinton, Daniel G. 1885b. "Supplementary remarks to the grammar of the Cakchiquel language", in *Proceedings of the American Philosophical Society*. Cited in Brinton 1885a above.
Campbell, Lyle. 1977. *Quichean linguistic prehistory*. (University of California publications in linguistics, 81.) Berkeley: University of California Press.
Estrada Monroy, Agustín, ed. 1973. *Empiezan las historias del origen de los indios de esta provincia de Guatemala Popol Vuh*. Traducido de la lengua quiché a la castellana por el R. P. fray Francisco Ximénez. Edición facsimilar. Palaeografía parcialmente modernizada y notas por Agustin Estrada Monroy. Guatemala: José de Pineda Ibarra.
Robertson, John S. 1984. "Colonial evidence for a pre-Quiche, ergative 3sg *ru-.", in *International Journal of American Linguistics* 50.452-5.
Robertson, John S. 1986. "A reconstruction and evolutionary statement of the Mayan numerals from twenty to four hundred", in *International Journal of American Linguistics* 52.227-41.
Robertson, John S. 1999. "The history of first-person singular in the Mayan languages", in *International Journal of American Linguistics* 65.449-65.
Tedlock, Dennis. 1996. *Popol Vuh: The definitive edition of the Mayan book of the dawn of life and the glories of gods and kings*. New York: Touchstone. ISBN 0-684-81845-0
Torresano, Estevan. [s.d.] *Arte de la langua Cakchiquel*. MS in the possession of Daniel Brinton in 1885. Cited in Brinton 1885a above.

Figures.

Besides the above, there are five sounds occurring in the Cakchiquel, Quiche and Tzutuhil, for which five special characters were invented, or rather adopted, by the early missionary Francisco de la Parra, who died in Guatemala, in 1560. They are the following :—

ȝ ȝʰ ȝ Ɛ ꞩ

The origin and phonetic value of these, as given by the grammarian Torresano, are as follows :[1]—

Ɛ This is called the *tresillo*, from its shape, it being an old form of the figure three, reversed, thus, Ɛ. It is the only true guttural in the language, being pronounced forcibly from the throat, with a trilling sound (*castañeteando*).

ȝ From its shape this is called the *cuatrillo*, Parra having

[1] Fr. Esteyan Torresano, *Arte de la Lengua Cakchiquel*, MS., in my possession.

Figure 1. Sample from Brinton 1885a, showing the "five special characters" (that is, four characters and one digraph with -*h*) and describing them. Note how in lead type he has simply inverted a 3 in his description of the origin of TRESILLO.

ȝ The name applied to this is, the *cuatrillo con coma*, or the 4 with a comma. It is pronounced somewhat like the *c* with the cedilla, ç, only more quickly and with greater force— *ds* or *dz*.

ꞩ This resembles the "4 with a comma," but is described as softer, the tongue being brought into contact with the teeth, exactly as *tz* in German.

ȝh A compound sound produced by combining the cuatrillo with a forcible aspirate, is represented by this sign.

Naturally, no description in words can convey a correct notion of these sounds. To learn them, one must hear them spoken by those to the manner-born.

Dr. Otto Stoll, who recently made a careful study of the Cakchiquel when in Guatemala, says of Parra's characters :—

" The four new signs added to the European alphabet, by some of the old writers on Cakchiquel (Parra, Flores), viz : Ɛ, ȝ, ȝ, ȝh, are but phonetic modifications of four corresponding signs of the common alphabet. So we get four pairs of sounds, namely :—

c and ȝ ;
k and Ɛ
ch and ȝh
tz and ȝ

forming two series of consonants, the former of which represents the common letters, and the latter their respective "cut letters," which may be described as being pronounced with a shorter and more explosive sound than the corresponding

Figure 2. Sample from Brinton 1885a continuing the discussion. He shows, as I have above on the top of page 2, the pairings of the plain and the glottal sounds; he does not use his ꞩ here, though he ought to, given his discussion of this immediately above.

Figure 3. Sample from the index of proper names in Brinton 1885a.
There is no reason to think that *Xe Caka Abah* is capitalized differently than *4hiyoc Queh AhĆuĆ*,
or indeed that *XĆeka4uch* does not properly contrast with *Ćeka4uch*.
If *Ć* were truly caseless, we might expect *Ć*Eka4uch* in titlecasing, mightn't we?

1. VAE XTINU4IBAH HALAL QUITZIH
HE NABEY

Ka tata ka mama, heri xeboço vinak oher mahaniok
ti laÇabex vae huyu taÇah; 4a ruyon ok umul 4iquin
4oh, que cha, ha ok ki xquilaÇabeh huyu taÇah he 4a
ka tata ka mama, yx nu4ahol, pa Tulan.

2. Xtinu4ibah 4a quitzih ri ki he nabey ka tata ka
mama ÇaÇavitz rubi, Çactecauh ru bi hunchic, he 4oh
quitzih que cha 4a 4haka palouh xoh pevi, pa Tulan ru
bi huyu, xoh alax xoh 4aholax vi pe ruma ka tee, ka
tata, yxka4ahol, quecha ri oher tata mama, ÇaÇavitz,
Çactecauh qui bi, ri ki xepe pa Tulan he cay chi achij
heri xoh boço, oh Xahila.

Figure 4. Sample from Brinton 1885a showing CUATRILLO WITH COMMA used in an all-caps and a plain context. This is surely *XTINU4IBAH* (and not *XTINU4IBAH*) contrasting with *Xtinu4ibah*.

4ambal richin Ykoma𝛏 vae.

The Conquest of· the Ikomagi.

34. Ok xet chi4a chinaht ri Cakixahay 4ubulahay ru bi, rikam Ykoma𝛏i, cani x4amar rokotaxic cuma, runah 4a xilitah Chi𝛏alibal rubi huyu; xuya vi ri hoye vi quivach, ok xilitah, xaxu 𝛏aba chic rij. Quere4a xubinaah vi huyu, Chi𝛏alibal ri. Xcha 4a ok xu ya ri: Xa yn acha𝛏 animal, xa mixi4hacatah, xaquin ikan a tem, a 4hacat, yn huvi chi vinak 4o vikan. Xcha ri Ykoma𝛏i, he 4a rikan ri Cakixahay, 4ubulahay; que-re4a ru4amic Ykoma𝛏 ri, xere 4a xcolotah. Chic ri xebo𝛓o chic Ço4il vinak, qui tata qui mama ri Ahpo𝛓o-4il Qulavi çochoh, 4ula vi 4anti quibi; xaqui vinakil xeel chic mani chic quikan.

34. Then they saw at a distance those called the Caki-xahay and the Qubulahay, subjects of the Ikomagi. They were captured after they had been routed by a surprise, when they were not far from a place called Chigalibal. They were pardoned when they arrived, and our warriors extended their hands to them. Hence that place was called Chigalibal. They said, in yielding: "I am your brother, your elder. You are the conquer-ors. We are the subjects of your throne and your power. I swear it before these who are my subjects." Thus spoke the Ikomagi, and thus their subjects, the Cakixa-hay and the Qubulahay. Thus did Ikomag submit and save his life. With them the Zotzils brought forth those fathers and elders, the Ahpozotzils named Qulavi Zochoh and Qulavi Qanti. But only their families, not their vassals, proceeded therefrom.

Figure 5. Sample from Brinton 1885a showing the capitalization of proper names. In English he gives *Cakixahay*, *Qubulahay*, *Ahpozotzil*, *Qulavi Zochoh*, and *Qulavi Qanti*. His Quiché for these reads *Cakixahay*, *4ubulahay*, *Ahpoço4il*, *4ulavi çochoh*, and *4ula vi 4anti* [sic, but compare *4ulavi cochoh* and *4ulavi 4anti* in the index shown in Figure 3 above]. In normalized and corrected form these must be *Cakixahay*, *4ubulahay*, *Ahpoço4il*, *4ulavi Çochoh*, and *4ulavi 4anti*.

For Cakchiquel we have the following progression. The Solana Cakchiquel diction-ary (said to have been widely used in the 1500s) had no palatalized velars. Some examples are:

4ak	"flea"	(k'aq)
cak	"red"	(kaq)
queh	"deer"	(ke·x)
icaĉ	"sling (honda)"	(ikaq')
rochoch queh	"stable" (literally "its-house deer")	(roĉoĉ ke·x)

The Vare(1)a dictionary was compiled ca. 1600, but was based on earlier work by Parra (discussed above) and Pedro de Betanzos (who wrote his Arte de la lengua de Guatemala, ca. 1545), both lost. Therefore, it is not possible to be certain whether Varela presents Cakchiquel as spoken in 1600 or as spoken in the mid 1500s in the works he used. In either case, Varela lacks palatalization:

icah	"axe"	(ikax)
icaĉ	"sling (honda)"	(ikaq')
queh	"deer"	(ke·x)
x4ak	"fingernails (uñas)"	(išk'aq)

Figure 6. Sample from Campbell 1977, showing rather ghastly typographic forms for both CUATRILLO and TRESILLO. The former stands high on the baseline but is otherwise unobjectionable. The latter is a fusion of c and ^ circumflex, which is not unlike what actually occurs in the manuscript, though here the letter seems to have been achieved by kerning (as its representation repeated in differs in the two words *icaĉ*: *icac^* and *icaĉ*).
(Professor Campbell has, it must be noted, given his support to this proposal.)

various grammars. I have preserved the original orthography. The earliest Quiche (and Cakchiquel) grammarians did superlative linguistic work in making up symbols for pho-nemes which went beyond Spanish phonology. For example, Anleo's *3* (actually, a reversed ɜ: ɛ) corresponded to *q'*, *4* to *k'*, *k* to *q*. The problem is that later grammarians were less accurate in a consistent recording of the uniquely Mayan phonological distinctions. Below, Anleo's work is phonemically accurate, as is the Kekchi. The others are not.

Figure 7. Sample from Robertson 1984, where the author equates DIGIT 3 with REVERSED OPEN E, and suggests that TRESILLO is OPEN E, which it is not. He substitutes DIGIT 4 for CUATRILLO.

> Rather than normalizing the spellings, I maintain the original orthographies. In the reconstructions, however, I use typical American linguistic conventions. In Colonial highland Guatemalan languages, the following orthography was adopted: /s/ = s, z, or c; /š/ = x; /x/ = h; /k/ = c or qu; /k'/ = Ꝗ; /q/ = k; /q'/ = ε; /¢'/ = Ꝗ ; /č'/ = Ꝗ h. It must be pointed out, however, that these symbols were inconsistently used, particularly Moran (1720).

Figure 8. Sample from Robertson 1986, where the author has a better CUATRILLO, though still high on the baseline. He also continues to equate TRESILLO and OPEN E.

- COMPL: 1ST on 3D: *xin/nuloɛoh* [š-in/nu-loq'ox] COMPL-ERG1SG-hit 'I hit it'
- INCOMPL: 1ST on 3D: *canuloɛoh* [ka-in/nu-loq'ox] INCOMPL-ERG1SG-love 'I love it'
- INCOMPL: 1ST on 2D: *catnuloɛoh* [k-at-in/nu-loq'o-x] INCOMPL-ABS2-ERG1SG-love-AFF.TR 'I love you'
- POSSESSIVE: *nu-Ꝗahol* [k'axol] 'my [male] son'

Figure 9. Sample from Robertson 1999. Here OPEN E is still used for TRESILLO, but an improving CUATRILLO is found, hanging below the baseline as it should.

In the Popol Vuh and other early alphabetic documents in Quiché and other Quichean languages, vowels followed by glottal stops were written *aa, ee, ii, oo, uu;* in the new spellings these become *a', e', i', o', u'.* Among the glottalized consonants, *b* becomes *b',* *tt* becomes *t',* *4h* becomes *ch',* and *4,* becomes *tz'.* Among k sounds, the plain front variety formerly written *c* or *qu* is now *k,* while the plain back variety formerly written *k* is now *q.* The glottalized forms once written *4* and *3* are now *k'* and *q',* respectively. And finally, *h* becomes *j,* while *z* and *ç* become *s.* Where untranslated proper names appear in roman type, they follow the original spellings of the documents. Otherwise, original spellings (in roman type enclosed in brackets) are given only where the new spellings of the same words (in italics) include corrections of phonetic or scribal errors.

The names of the Guatemalan Mayan peoples and languages mentioned in this work, as written in the new official alphabets, are *K'iche', Kaqchikel, Tz'utujil, Poqomchi, Q'eqchi', Ixil, Mam,* and *Jakalteko.* The spellings used in

Figure 10. Sample from the popular translation of the Popol Vuh in Tedlock 1996. The author uses DIGIT 4 and DIGIT 3 for CUATRILLO and TRESILLO. The names given in modern orthography in the last paragraph can be given in normalized orthography according to normal modern editorial practice if casing pairs for the Mayanist letters are available in the UCS: *K'iche'/Ꝗichee, Kaqchikel/Cakchiquel, Tz'utujil/Ꝗutuhil, Poqomchi/Pokomchi, Q'eqchi'/Čekchii, Ixil/Ixil, Mam/Mam, Jakalteko/Hacalteco.*

Figure 11. Sample from the Popol Vuh manuscript (p 24). The CAPITAL LETTER TZ is shown in the second line alongside CAPITAL LETTER Q in Quiche, though in Spanish SMALL LETTER Q is used. Further down the SMALL LETTER TZ is used; the word is *Ꜩih* or *tȝih* 'word, speech' in both cases: *quiche tȝih = historias quicheas*.

Figure 12. Sample from the Popol Vuh manuscript showing general inconsistency in capitalization practice. On this page of the manuscript the name *Balam* is written with a capital three times in Quiché, but written *balam* once in Quiché and four times in Spanish—I give only part of the page showing two capitals in Quiché and two smalls in Spanish, to save space. It is not, certainly, evidence that Spanish has no CAPITAL LETTER B. It is simply a feature of the scribe's practice.

Figure 13. Sample from the Popol Vuh manuscript, showing a number of instances of TRESILLO in use in the word *ƇaƇ* 'fire' (see *fuego* in the Spanish). The *Balam/balam* inconsistency is also found here, and the phrase *oher ʒih* 'ancient traditions' is found, and *Vuestro* has a capital *V*.

Figure 14. Sample from the Popol Vuh manuscript. The word phrase *Ƈalel ahtʒic vinac* occurs thrice; it means 'the prominent speaker' ('man of words'), rendered once in Estrada Monroy's edition as *Calel y Ahtzih Vinac* and twice as *principales*.

Figure 15. Sample from the Popol Vuh manuscript. The CAPITAL LETTER TZ is written carefully in the first line. Paragraph initials are capitalized in Quiché and in Spanish. The TRESILLO is found in a couple of words in the Quiché.

Figure 16. Sample from the Popol Vuh manuscript. *Mexico* is written in both Quiché and Spanish with a capital letter, as is *Dan*. In Quiché, *rabinal* is written where in Spanish *Rabinal* is written, and in both lower case is used in the name *Caƈchequeleb* in Quiché and *Caƈchiqueles* in Spanish (transliterated as *cacchiqueles* in Estrada Monroy's typeset edition). We are simply unlucky that the capital doesn't appear, however, as is clear from the general use of casing throughout the document. In all caps, the word would have to be *ƇAƇCHEQUELEB*, not *ƇAƇCHEQUELEB*.

Figure 17. Sample from the Popol Vuh manuscript. Circled are words using the TRESILLO. The manuscript gives *Ćaquićot*; Estrada Monroy transliterates this as *kaquicot*. The editor is not consistent, wavering between *k* and *c* because he has no TRESILLO: *ćut/cut*, *calunić/calunic*, *hoć/hoc*, *petić/petik*, *ćolić/kolic*, *ćavah/kavah*, *puać/puak*, *locoxić/locoxic*, *quihiloxić/quihiloxik*. The disjointed *c* + inverted breve form of the TRESILLO is seen throughout.

TABLE xx - Row 2C: LATIN EXTENDED-C

	2C6	2C7
0		
1		
2		
3		
4		
5		
6		
7		
8		Ɔ̃
9		ɔ̃
A		4
B		4
C		4,
D		4,
E		Ꜩ
F		ꜩ

G = 00
P = 00

TABLE XXX - Row 2C: LATIN EXTENDED-C

hex	Name	hex	Name
60	(This position shall not be used)		
61	(This position shall not be used)		
62	(This position shall not be used)		
63	(This position shall not be used)		
64	(This position shall not be used)		
65	(This position shall not be used)		
66	(This position shall not be used)		
67	(This position shall not be used)		
68	(This position shall not be used)		
69	(This position shall not be used)		
6A	(This position shall not be used)		
6B	(This position shall not be used)		
6C	(This position shall not be used)		
6D	(This position shall not be used)		
6E	(This position shall not be used)		
6F	(This position shall not be used)		
70	(This position shall not be used)		
71	(This position shall not be used)		
72	(This position shall not be used)		
73	(This position shall not be used)		
74	(This position shall not be used)		
75	(This position shall not be used)		
76	(This position shall not be used)		
77	(This position shall not be used)		
78	LATIN CAPITAL LETTER TRESILLO		
79	LATIN SMALL LETTER TRESILLO		
7A	LATIN CAPITAL LETTER CUATRILLO		
7B	LATIN SMALL LETTER CUATRILLO		
7C	LATIN CAPITAL LETTER CUATRILLO WITH COMMA		
7D	LATIN SMALL LETTER CUATRILLO WITH COMMA		
7E	LATIN CAPITAL LETTER TZ		
7F	LATIN SMALL LETTER TZ		

A. Administrative

1. Title
Proposal to add Mayanist Latin letters to the UCS.
2. Requester's name
Michael Everson
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution.
4. Submission date
2006-01-30
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
Yes.
6b. More information will be provided later
No.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
No.
Proposed name of script
1b. The proposal is for addition of character(s) to an existing block
Yes.
1b. Name of the existing block
Latin Extended-C.
2. Number of characters in proposal
4
3. Proposed category (see section II, Character Categories)
Category A.
4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)
Level 1.
4b. Is a rationale provided for the choice?
Yes.
4c. If YES, reference
Spacing letters.
5a. Is a repertoire including character names provided?
Yes.
5b. If YES, are the names in accordance with the naming guidelines in Annex L of ISO/IEC 10646-1: 2000?
Yes.
5c. Are the character shapes attached in a legible form suitable for review?
Yes.
6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Michael Everson. TrueType.
6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson. Fontographer.
7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.
7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.
8. 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)?
Casing is addressed.
9. 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.
Functions and properties are like all Latin capital and small letters.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
Yes. The Claudian letters are part of a preliminary proposal in N2957, which does not contain the proposal summary form. Here, however, they are proposed on their own.
2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
Yes.

2b. If YES, with whom?
Lyle Campbell, Professor of Linguistics, Director of the Center for American Indian Languages, Department of Linguistics, University of Utah.
2c. If YES, available relevant documents
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
No.
4a. The context of use for the proposed characters (type of use; common or rare)
These are rarely-used characters used in Latin epigraphy.
4b. Reference
5a. Are the proposed characters in current use by the user community?
There are font implementations used by specialists.
5b. If YES, where?
See the figures above.
6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?
Yes.
6b. If YES, is a rationale provided?
Yes.
6c. If YES, reference
Keep with other Latin letters.
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
If possible.
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No.
8b. If YES, is a rationale for its inclusion provided?
8c. If YES, reference
9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
No.
9b. If YES, is a rationale for its inclusion provided?
9c. If YES, reference
10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.
10b. If YES, is a rationale for its inclusion provided?
10c. If YES, reference
11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?
No.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
12b. If YES, reference
13a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
13b. If YES, describe in detail (include attachment if necessary)
14a. Does the proposal contain any Ideographic compatibility character(s)?
No.
14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?
14c. If YES, reference