# On the decision to rename and change the glyphs of Greek phonetic letters
## Eduardo Marin Silva
## 26/Dic/2024

This is a response on decision 4.7 of the script proposal recommendations (L2/24-228). As context, IPA made a proposal to add to add Greek letters with palatal hook as well as some modifier version of basic Greek letters. It was decided it was best for them to have Greek names and a Greek script property but then to assign them to the Latin block just so they can be along with other related letters. I only disagreed with the codepoint assignment so I suggested them to be reassigned to spaces in the Greek blocks, but then the Script Encoding Working Group decided to rename the letters to Latin (as well as the script property). I highly disagree with that decision and I shall elaborate on why.

But before I continue I would like to implore with the group to actually respond to the arguments I'm raising. This document took a lot of effort to make and it would be ashamed if it good the same response as L2/24-088 which only received this response in L2/24-068

> "This short doc argues against the SAH response (L2/24-013r) to the author's comments in L2/23-233 regarding changes to the arrow names used in Egyptology. (The arrows were proposed in L2/23-185.) We briefly reviewed L2/24-088. Our previous comments stand. No action by the UTC is required."

Which in turn was a response from this:

> "On the arrows for Egyptology: Changing the current descriptive names of the arrow characters does not add any value. No change is needed."

Neither response even attempted to address my detailed rationale in favor. I'm not saying that the working group has to agree with me, but if all that is required for a response is an opaque rejection, it really discourages anybody from contributing.

So once again, I IMPLORE THE WORKING GROUP TO READ AND RESPOND TO THE ARGUMENTS THIS TIME.

**Preamble) The IPA alphabet has very different properties than the Greek or Latin alphabets:**

a) IPA does not have casing relations; indeed, selecting IPA text and making it all uppercase would hinder its legibility for its users (COMPARE THIS TO ENGLISH TEXT THAT CAN BE READ FINE IF IT'S ALL CAPS).

b) Superscript characters in regular orthographies tend to have many uses: to indicate a reference, exponentiation, an isotope, *or even to indicate whispering*. While in IPA, superscripts characters always indicate a modifier version of a sound.

c) Regular orthographies are not concerned with universality. The same letter may have different pronunciations in different contexts, the same sound may be represented with different letters or there may be letters that don't change the pronunciation, but any of these differences may sometimes carry a semantic distinction. IPA was created to sidestep those issues, by creating a worldwide agreed upon way to represent sounds independent of language or semantics. indeed INDEPENDENT OF ANY NOTION OF SCRIPT.

So even though this points to IPA notation being a distinct entity beyond either Greek or Latin, because the creators and users did not care about the concept of script boundaries or of creating their own script, they would use Greek letter-forms out of convenience, and they would not perceive them as out of place. To this day, there is so much overlap between the letter-forms of Cyrillic, Greek and Latin (not related to IPA) it's often easier for me to think of them jointly as a 3-way continuum.

Indeed, many natural orthographies would do the same glyph borrowing, which is why we now encounter in Unicode many disunifications that would have been considered glyph variants had the three scripts been treated jointly, but required disunification under a combination of semantic and technical grounds.

These disunifications have obfuscated the issue in question, unnecessarily.

**1) The main argument in favor of changing the names is faulty:**

In document L2/24-202 the following argument in favor of the name change is made:

> *"It may be consistent with Greek-derived modifier letters ᵝ, ᵡ, ᵦ, ᵪ (1D5D, 1D61, 1D66, 1D6A) [Unicode 4.0, L2/02-141] and ᶿ (1DBF) [Unicode 4.1, L2/04-044), which decompose to Greek characters β, θ, χ (03B2, 03B8, 03C7) [Unicode 1.0], which themselves unify Latin and Greek forms used in phonetics transcriptions like the IPA, and added before Latin forms β (A7B2) [Unicode 8.0, L2/12-270] and Latin χ (AB53) [Unicode 7.0, N4296] were disunified for casing. However, this is not consistent with Greek-derived letters with attached diacritics, closed forms or turned forms that do not decompose, which are typically encoded in Latin script as they are not used in Greek script.*
> *For example [Unicode 1.1 characters]:*
> *- ƛ 019B LATIN SMALL LETTER LAMBDA WITH STROKE*
> *- ⍵ 0277 LATIN SMALL LETTER CLOSED OMEGA*
> *- ƍ 018D LATIN SMALL LETTER TURNED DELTA"*

I would like to point out that while previous encoding decisions can serve as precedent for subsequent decisions, they should never be used (by themselves) as arguments in favor of new encoding decisions, as there are plenty of examples of less than ideal decisions that cannot be reversed due to stability policies.

Making decisions just to be consistent with previous ones is not a good argument by itself. We should strive to take the best encoding decisions for the given case and only use previous decisions as references, not rules.

It seems that what is being argued here, is that since we have previously encoded graphically modified Greek letters as Latin characters, then there is precedent for doing the same for these letters. This is true, but is it the best encoding decision for these characters in particular?

Like we discussed already, the IPA community clearly had a preference for Latin letter-forms given they were familiar with them, but would use Greek ones without any qualms, because in their mind, their notation was first and foremost meant to be universal, not Latin based. Indeed, it's not out of the question they would have gone further by using Cyrillic, Armenian or Georgian letter-forms.

The author does argue that

> *"The 3 Greek-derived symbols are only used within a Latin transcription system, having them in a different script creates shaping issues with implementations that handle scripts separately. For example, ρ͡β or t͡θ may be split per script and font features adjusting the ligature tie would not be applied, or non-IPA forms β, θ, χ may be used."*

While this is a good technical consideration, I find it dubious this is actual a problem for text processors. If keeping everything under the same script was so important for IPA rendering, then the usage of the Greek codepoints would have been deprecated long ago in favor of Latin replacements.

Those other characters cited, like the closed omega or turned delta were in fact encoded as Latin characters and the reason was probably because they were only expected to be used with other Latin characters. While that may be enough reason for some, my point is that they could have easily been made Greek characters and having them mix with Latin characters would not be any more problematic than the current situation.

Let's for a moment limit the discussion to the Closed Omega. If I was around when this character was being proposed, I would have advocated for it to be Greek, as the current name makes the implication that there is a distinct but related Latin Omega that in turn is the parent of the Closed Omega when it's clear the true parent is the Greek codepoint (03C9 ω) and the current Latin Omega (A7B7 ꭷ) corresponds to a different branch of the family tree (see the figure on section 2 to see what I mean).

**2) The working group's argument regarding Chi is even weaker:**

In the rationale for the decision for the name change I quote:

> *"We have other Greek-derived non-modifier phonetic transcription letters like U+AB54 ꭔ LATIN SMALL LETTER CHI WITH LOW RIGHT RING and U+AB55 ꭕ LATIN SMALL LETTER CHI WITH LOW LEFT SERIF, which are named as Latin letters."*

But they forget to mention that the character U+AB53 ꭓ LATIN SMALL LETTER CHI, is the true parent of those characters and this character in turn is derived from U+03C7 χ GREEK SMALL LETTER CHI.

Latin Chi was disunified under good rationale from Greek Chi and this is independent on its use for phonetics. While in sans serif they may look confusable, they clearly have different glyphs in serif fonts with the Latin one looking like a stretched version of the small x (Greek Chi, Latin Chi and Latin x side-by side: χꭓx). We can also see that the Latin Chi uppercase is very different, by having both ascender and descender parts while the Greek uppercase looks identical to Latin X (Greek Chi, Latin Chi and Latin X pairs: Xχ, Ꭓꭓ, Xx).

The fact that the two Greek derived letters they cite are named "Latin" is just evidence they are one step extra removed from the original Greek letter, not that there is precedent for naming letters as Latin just because of their use in a primarily Latin phonetic notation (see my diagram in section 3)

It would be better evidence for their point if AB53 was never disunified or if at least the serif glyphs of AB54 and AB55 matched the usual glyph of 03C7 better.

**3) The actual usage of the letters shows Greek forms:**

Perhaps the working group will now point to the current glyphs of Beta and Chi of the proposed letters as evidence that they should be considered Latin and not Greek (theta has distinct considerations better discussed separately), but the glyphs were modified in a decision based on document L2/24-146 and the reason for it was:

> "There is some discussion among phoneticians over whether the IPA letters beta, theta and chi should be Latin or Greek in form, but IPA letters with a palatal hook were retired with the Kiel convention in 1989, and it's unlikely that such a distinction will need to be made for them."

So if I understood the argument correctly, because it was unlikely there would be both Greek and Latin letter-forms with palatal hook that would need to be distinguished, the use of either form was not problematic. This in hindsight was an erroneous decision, and I regret not raising an issue with it before. If one looks at the figures in the SAME DOCUMENT, it's clear they are using Greek forms; AND WHY WOULDN'T THEY?

Latin Chi was created for Teuthonista which is a completely different notation than IPA, while IPA has always used the Greek form, without using the uppercase. Latin Beta was created for natural orthographies of African languages (primarily in Gabon), and has similar disunification rationale than Latin Chi; once again IPA has only used the Greek form and only the lowercase.

The fact that the Latin codepoints have uppercase variants at all, is evidence that they CANNOT be the parent for the characters with palatal hook, as IPA would have not had any use for the uppercase.

The decision to change the glyphs has obfuscated the origin of these letters. I have illustrated this in the following diagram (incorrect version on the right and more nuanced version on the left):
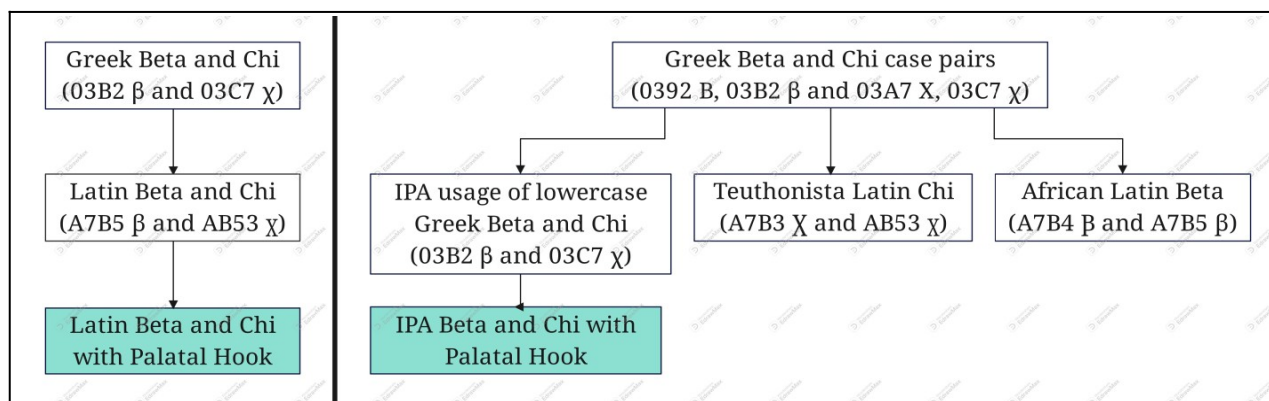


The glyphs that are interpreted as Latin Beta with Palatal Hook and Latin Chi with Palatal Hook are basically Unicode inventions with no attestation in any actual phonetic notation.

Furthermore, if using either form for the glyphs was not meant to be problematic, then why was there a preference for the Latin forms at all? Wouldn't it make more sense to use whichever form matched the actual usage regardless of the name or script?

One gets the impression, that whoever was behind that decision REALLY wanted these letters to be next to the other related IPA letters (that are definitely Latin) and was wiling to change the glyphs so they wouldn't look out of place in a Latin context.

**4) None of this explains why the modifier Psi and Omega are still in the Latin block:**

I hope that I have convinced some people that naming these letters as Latin is basically inventing letters that have not been attested; forcing the existing characters into another branch of a family tree to fit a conclusion. Perhaps this is done to assign as many phonetic characters as possible the Latin script property out of convenience; but right now I'm speculating, so I would prefer if the people behind that decision speak up.

Let's assume that we grant the argument to make these characters Latin. In L2/24-228 the Working group argues:

> *"However, the existing Greek-derived phonetic transcription letters in the Latin Extended blocks are modifier letters, like U+1D5D MODIFIER LETTER SMALL BETA. Modifier letters have an additional constraint: they must under NFkD produce the original letter, and with that in mind it makes more sense for them to be termed as Greek letters"*

While I agree with what is being said, why are these GREEK modifier letter still in a Latin block then?

**5) Latin Theta discussion:**

The existence of a distinct Latin Theta is known and the best argument for disunification is in L2/12-138 but for some reason or another it has not been disunified. I would like to point out that I consider this letter to belong to a separate branch than the Theta with palatal hook and they both have the Greek codepoints as a parent. Meaning I still consider Theta with palatal hook to be Greek, just like Beta and Chi, regardless of the disunification status of the Latin Theta.

The lowercase glyphs of the Greek and Latin Theta are identical in some contexts, but all that means for this discussion, is that the glyph for the IPA character was not unduly modified like Beta and Chi were.

However if the Theta with Palatal hook were encoded as Latin that would imply the existence of an already encoded Latin Theta (jut like the existence of the Latin Chi with a ring implies the existence of an unmodified distinct Latin Chi).

So if the persons that refuse to disunify the Latin Theta are the same ones arguing that the Greek letters with palatal hook should be considered Latin, then they need to recognize this contradiction.

**6) The best resolution:**

After thinking about the best resolution I arrived at the following compromise. The letters with Palatal Hooks would change their glyphs, names and script property to reflect the Greek identity I worked so hard to show. But the can remain in the Latin Extended G block. There could even be a header at the start explaining that some Greek letters are included here to keep them with the rest of the related Latin letters. The Greek codepoints can have cross references and vice-versa.

Not moving the modifier Psi and Omega still seems odd, but the script property and name are being kept so if it means so much, they can remain there as well.

If all else fails, just changing the glyphs to reflect actual usage should not be problematic at all.