A. Administrative

1. Title: Proposal to add Greek Letter Lowercase Heta and Greek Letter Capital Heta to the UCS

2. Requester's name: Nick Nicholas

3. Requester type: Expert contribution

4. Submission date: 2005–01–01

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? No.

1b. Addition of character(s) to existing block? Name? Yes. Greek or Greek Extended.

2. Number of characters in proposal: Two

3. Proposed category: B.1. Specialized (small collections of characters)

4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): Level 1 noncombining character

Is a rationale provided for the choice? No

5. Is a repertoire including character names provided? Yes

a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? Yes

b. 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? —

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: —

7. References:

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

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

8. Special encoding issues:

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)? Yes

9. Additional Information: See below

C. Technical—Justification

1. Has this proposal for addition of character(s) been submitted before? No

2. 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

If YES, with whom? Subscribers of Unicode Greek and Epigraphical mailing lists.

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? Scholarly community and individuals interested in Greek linguistics and epigraphy

Reference: —

4. The context of use for the proposed characters (type of use; common or rare): Common in Greek epigraphy, occasional in Ancient Greek linguistics

Reference: —

5. Are the proposed characters in current use by the user community? Yes

If YES, where? Reference: Characters are present in various publications on Ancient Greek linguistics, and in publications of epigraphic corpora

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? Yes

If YES, is a rationale provided? Contemporary use, keeping character together with other Greek characters

If YES, reference: —

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? Yes: the two characters differ only in case.

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? No (but see below)

If YES, is a rationale for its inclusion provided? —

If YES, reference: —

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No

If YES, is a rationale for its inclusion provided? —

If YES, reference: —

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? Yes

If YES, is a rationale for its inclusion provided? Yes

If YES, reference: —

11. Does the proposal include use of combining characters and/or use of composite sequences? No

If YES, is a rationale for such use provided? —

If YES, reference: —

Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? —

If YES, reference: —

12. Does the proposal contain characters with any special properties such as control function or similar semantics? No

If YES, describe in detail (include attachment if necessary) —

13. Does the proposal contain any Ideographic compatibility character(s)? No

If YES, is the equivalent corresponding unified ideographic character(s) identified? —

If YES, reference: —

Proposal

This is a proposal for two characters, `GREEK SMALL LETTER HETA`, and `GREEK CAPITAL LETTER HETA`. The reference glyph for the characters is to be Ⱶ ⱶ (the "tack heta") as used in Jeffery (1990).

The tack heta symbol is one of the two transliterations used in Greek Epigraphy to represent the phoneme /h/, corresponding to the use of the diacritic rough breathing (`U+0314 Combining Reversed Comma Above`) in conventional Greek orthography. The other epigraphical transliteration, which is more common especially in present-day use, is Latin capital and small H.

The epigraphical codepoint heta abstracts from the various glyphs used in inscriptions to represent the letter heta. Most regional (*epichoric*) variants of the Greek script had for their heta the same form as the Old Italic codepoint `U+10307 Old Italic Letter He` (𐌇), while a few had the same form as `U+0397 Greek Capital Letter Eta`.

The 'tack' glyph was originally a local variant of heta in the Greek colonies of Heraclea and Tarentum in Italy. The ancient received wisdom was that the H-like heta glyph was broken in two, with the left half representing /h/ (the glyph used in Tarentum), and the right half the absence of /h/. Diacritic versions of the 'left tack' and 'right tack' arose within a few centuries, and eventually evolved into the rough and smooth breathing diacritics of conventional Greek orthography, `U+0314 Combining Reversed Comma Above` and `U+0313 Combining Comma Above`. (The reference glyphs for `U+0485 Combining Cyrillic Dasia Pneumata` and `U+0486 Combining Cyrillic Psili Pneumata` display those tacks used as diacritics.) However, there is no evidence known to me that the 'right tack' was ever actually used as a letter, as distinct from a diacritic; the ancient account connecting the diacritics to the Italian letter may have been a post hoc rationalisation, and has been disputed in the literature.

The same letter that in Western Greek alphabets was used to write /h/, and called heta, was

used in Eastern Greek alphabets to write /ɛː/, and called eta. (The standard Greek alphabet is of Eastern origin, and its value of (h)eta displaced the Western form in the 4th century B.C.) Nonetheless, epigraphers make a point of keeping eta and heta distinct: lowercase heta is never written in transcription as η, but as Ⱶ or h. (More precisely, when the Ⴗ or Ⴙ letter in an inscription has the phonetic value /h/ rather than /ɛː/, it is transcribed as Ⱶ or h instead of η.) When used in titlecase, Capital Latin H is identical to Capital Greek Eta; the absence of breathing marks is then used to differentiate heta from eta.[1] Thus:

- Inscription (using Old Italic glyphs): ᕼᐯᛰ ᕼᛖᒪᒪᛖᛰOᛋ

- Epigraphical transcription, using 'tack' heta: Ⱶυφ' Ⱶέλλēνος

- Epigraphical transcription, using Latin h: hυφ' Héλλēνος (variants: *hυφ' Héλλēνος*, hυφ' Héλλēνος)

- Conventional orthography: ὑφ' Ἕλληνος.

Epigraphy insists on using a letter rather than a diacritic to transcribe what was a letter in the inscription. This allows for letter-level markup, which would be impractical on a diacritic, e.g. [Ⱶ]έλλēνος (the heta is conjecturally emended by the editor), or Ⱶ̣έλλēνος (the heta is only partly preserved) (Figure 1). Although the linguistic discussion of Ancient dialects tends to use normalised orthography in abstract discussion and lexicography, when inscriptions are cited, they are normally cited with hetas, rather than conventional diacritics.

The tack heta is functionally equivalent to several extant codepoints, but cannot be conflated with any of them.

- Because of the requirement in epigraphy for letter-level markup, it would be impossible to conflate heta with the diacritic, and such a conflation would be impossible in Unicode, even though Héλλēνος and Ἕλληνος are the same word (and in fact are typically collated as the same word).

- The Old Italic codepoint U+10307 Old Italic Letter He originates in the usual form of heta in Greek antiquity, but it is not the glyph conventionally used in modern epigraphy (see below), and would not be recognised as a permissible alternative transcription by epigraphers. At any rate, Old Italic script should be kept distinct from Greek unless there is no alternative, and the Old Italic codepoint does not allow for a casing contrast.

- The Latin H and tack heta are glyph variants of the same underlying character in the context of Greek epigraphy, as is very clear in Figures 4 and 5. Nonetheless, it would be impractical to implement tack heta as a presentation variant of a Latin character, especially when that variant occurs not in a Latin script context at all, but a Greek

---

script context. It may be desirable in fonts to have H h be glyph variants of Ⱶ Ⱶ, however, since they are fully equivalent in the context of Greek epigraphy, and there would be obvious advantages in epigraphers using a distinct heta codepoint, whatever the glyph displayed.

- Although in most epichoric alphabets eta and heta were mutually exclusive, this was not universally the case; e.g. θοῖναι δὲ ταίδ[ε νόμιμ]οι· Ἀπέλλαι καὶ Β[ουκά]τια, Ηηραῖα, Δαιδαφ[όρια], Ποιτρόπια, Βυσίου [μην]ὸς τὰν ηεβδέμαν καὶ [τ]ὰν ηενάταν … (Buck 1955:242), from Delphi (Ꮒ = /h/, H = /ɛ:/), where the word *hɛ:raîa* (Ηηραῖα = Ἡραῖα) 'festival of Hera' combines heta and eta. (Note that Buck transcribes heta as Latin H.) This also occurred in Heraclea and Tarentum, with Ⱶ = /h/, H = /ɛ:/. In epichoric alphabets featuring both eta and heta, there is additional motivation for the glyph H for heta to be avoided.

- The non-combining U+1FFE Greek Dasia is functionally similar to heta, though it belongs to a different typographical tradition, and could be reused for the heta, despite the dissimilarity in glyphs. This would not solve the requirement of an uppercase form, however; and the dasia is currently encoded with the general category of Sk (Symbol, Modifier).

- The tack glyphs are already present in Unicode as U+22A2 Right Tack and U+22A6 Assertion, and any current use of tack heta in computer typesetting presumably uses those glyphs. The typographical requirements of mathematics makes them unsuited for use as Greek letters, as do their character properties; and these codepoints also would not allow a case distinction.

There is occasional use of case contrast in heta, as in Jeffery (1990) (Figure 2).

Unlike the conventional capital eta, from which tack heta is arguably derived, there is no tradition of serifs on tack heta. This is because epigraphical Greek is traditionally printed in

a 'sans-serif' font (cf. the font New Athena Unicode: Ⱶυφ᾽ Ⱶέλλēνος); at any rate, Greek typography does not employ serifs in lowercase letters. On occasion, in order to illustrate the precise form of letter used, the epichoric form of Greek letters is used (e.g. ⯃ᴇLLᴇᴍᴏϚ). Although this is arguably more akin to illustration than transcription and thus does not count as plaintext, it would be acceptable in this context to treat ⯃ as a glyph variant of heta, just as it would be to treat Ⳑ as a glyph variant of lambda.

The tack heta should be encoded as a letter, and it should sort either immediately after or immediately before eta; since the characters were almost always mutually exclusive (with the exceptions of Delphi, Heraclea/Tarentum, and Cnidus), there is no established ordering between them. Most indexes containing heta (e.g. Buck 1955) give it secondary weighting, as if the heta was a rough breathing in conventional orthography; such indexes also ignore digamma, so that the words are sorted as if they are in standard Greek orthography (digamma was dropped in Attic, the classical standard). Such weighting should not be enforced in the Default Collation, however, and the sorting of heta as a distinct letter is found in histories of the Greek script like Jeffery's.

Note that although the heta is a distinct letter, and its representation as a 'tack' cannot be

handled by current Unicode, its encoding as Latin H is prevalent. The Latin H is so prevalent that Prof Klaus Hallof, director of the *Inscriptiones Graecae* corpus, has insisted to me that the tack heta is to be dispreferred and is avoided in the *IG*—which indicates there may well be resistance to a proposal for a distinct heta codepoint. He has also pointed out that the tack heta sign is identical to the numeric sign `U+10142 Attic Acrophonic Symbol One Drachma` (provisionally accepted in Unicode), although contextual disambiguation of the two would usually be easy.

There are occasional attempts to differentiate the glyph for <h> = heta from <h> = Latin H (Figure 4, Figure 6); but for the most part the heta <h> is identical to the Latin <h>, being at most italicised (Figure 5). The tack glyph is used in Jeffery's (1990) influential history of the Greek script, and sporadically in grammars and epigraphy; it is more frequent in publications on Heraclea and Tarentum, where the tack was native and there is a need to differentiate eta from heta (Figure 3, 4, 5). But standard epigraphical corpora have used the Latin h for well over a century (including inscriptions from those two colonies).

Further background on the history of heta and eta is provided in Nicholas (2003).

Examples

μιαρ[ο]ν [ε]ιμε[ν?] *b*: α[ι] ανθρōπον �haⲅ[- - - | - - -]αντα
χρημα μηθ|εν **:** μιαρον ειμεν [- - - | - - - χρημ]ατōν
μηθεν ⲅ[ι]λασ] *c*: μον ειμεν **:** αι [- - - | - - - ανθ]ρōπōι
μιαρōι **:** ·κα|θαρσιν δε ειμεν ⲅ[- - - | - - - α]ποθανοι
καθαρα|μενον **:** κατα νομ[ον - - - | - - -] ⲅιαρο δαμο |
τε[- -].

Figure 1. Jeffery (1990:405). Instances of letter-level markup on tack heta: dot under heta on 3rd and 5th line.

43*a**. Μⲅηγαρις μ' ανηθε|κεν ⲅερηι. *b.* Επιγνοτη μ'
ανηθεκην τηι ⲅερηι.

Figure 2. Jeffery (1990:412). Case contrast between hetas: *Mhε:garis m' anε:t^heken Herε:i* "Mhegaris raised me to Hera"; *Epignotε: m' anε:t^heken tε:i Herε:i* "Epignote raised me to Hera"

ΘΕ ο . | . Τ Υ Χ Α Ι Α Γ Α ⊙ Α Ι
ⲅΑ Γ ο | Λ Ι Σ + Α Τ ⌒ Ν Τ Α Ρ Α Ν
Τ ο Ν Δ | Ι Α Τ ο Ν Ν Α Ι ο Ν Κ Α Ι Τ
Γ Ε Ρ Ι | Γ Α Ν Τ Υ Χ Ι Α Σ Κ Α Ι Γ
Τ Α Χ . | . Ρ ⌒ Ι Κ Α Ι Γ Ε Ρ Ι Τ ⌒ Ν

θεὸ[ς] | τύχα ἀγαθᾶ [ἐπερωτῆ]
ⲅα πό|λις [ⲅ]α τῶν Ταραν[τίνων]
τὸν Δ|ία τὸν Νάιον καὶ τ[ὰν Διώναν]
περὶ | παντυχίας καὶ π . . .
τάχει | δρῶ καὶ περὶ τῶν . . . .

ⲅΕⲅΡΑΚⳑΕϚ
ⲅΗρακλῆς

Figure 3. Roberts (1887:275; 208). Tack heta used for inscription from Tarentum (left); conventional rough breathing used for inscription from elsewhere in Italy (right).

1519. Tarentum. Small terracotta objects. F.Ghinatti, *Sileno* 23 (1997) 120-126. discusses the function of the many small terracotta disks found in Tarentum; they contain proper names, single letters, or letters in ligature, and small relief-figures. He focuses on the most frequent inscriptions (Ϝημιω(δέλιον), τετάρ(τιον), τριταῖα, τρίτω and δυοτρι(ταῖα); see *SEG* XXXVI 892-906 and 1538. He prefers to interpret the figures as indicators of taxes or customs dues paid on goods in Tarentum's harbor; the names are those of manufacturers or merchants. He refers to similar objects from Lokroi, with the letters ΠΡΑ, interpreted by G. as πρά(κτο ρες). In Tarentum the πράκτορες may have been customs-officials, who levied taxes on goods in transit. The goods were divided into at least four categories of payment, based on the value of the goods: those paying a tetartemorion/tetartion (1/48 stater), half an obol (Ϝημιωδέλιον; 1/2-stater), a tritemorion (τριταῖα; 1/16 stater), and a triemiobolos (δυοτριταῖα; 1/8 stater); the corresponding tax-rates were 2, 4, 6 and 12%.

A   hἶσα μένε τιννύνα(ι)
B   αβγδεϜhϚ (dextrorsum) ǀ ϚhϜδγβ (sinistrorsum)

B. 2. read as ϚδϜhρβ by G.Buchner in an addendum on 231.

Figure 4. *Supplementum Epigraphicum Graecum* 47 (1997): 420, 401. Tack heta used for inscriptions in Tarentum (top); Latin h (shrunk to match lowercase Greek) used for inscription from Cumae, also in Italy (bottom).

18. Der Hauchlaut wird durch ein besonderes, aus H differenziertes Zeichen Ϝ wiedergegeben, das der Vorfahr des Spiritus asper (Ϝ Ϝ ‘) ist. Heraklea teilt diese spezifische Eigentümlichkeit mit Tarent (z. B. Ϝηράκλητος auf einer Münze); die Aussprache des Hauchlautes scheint kräftiger als sonst gewesen zu sein. Abweichend vom Attischen steht das Hauchzeichen in hἄκρο(-σκιρίαις), hἶσος (auch in der Κοινή!), s. unter 10. (vielleicht unter dem Einfluß von hόμο-), hόκτώ, hέννέα (und Ableitungen dieser beiden durch Analogie nach ἑπτά), hάρνησις, hοίσοντι

Figure 5. Thumb (1932:97). Tack heta used to illustrate the epichoric alphabet of Tarentum and Heraclea (line 4); but Latin h used to discuss the distribution of /h/ as a phoneme, in normalised Greek orthography (lines 7-10). The tack is thus used as a glyphic variant of underlying h.

Polyandria

7

[ὁ ξένε, εὔhυδρ]όν ποχ' ἐναίομες ἄστυ Ϟορίνθο·
[νῦν δ' hαμὲ Αἴα]ντος [νᾶσος ἔχ]ει Σ[αλαμίς].

Figure 6. Peek (1955:3). Latin script <h> used for heta, in distinction to the normal plain Latin script of the German commentary.

Properties

UCD Entry

```
aaaa;GREEK SMALL LETTER HETA;Ll;0;L;;;;;N;;;bbbb;;
bbbb;GREEK CAPITAL LETTER HETA;Lu;0;L;;;;;N;;;;aaaa;
```

DUCET Entry

```
03B7  ; [.10F1.0020.0002.03B7] # GREEK SMALL LETTER ETA
aaaa  ; [.10F1.0020.0002.aaa] # GREEK SMALL LETTER HETA
1D6C8 ; [.10F1.0020.0005.1D6C8] # MATHEMATICAL BOLD SMALL ETA; QQK
1D702 ; [.10F1.0020.0005.1D702] # MATHEMATICAL ITALIC SMALL ETA; QQK
1D73C ; [.10F1.0020.0005.1D73C] # MATHEMATICAL BOLD ITALIC SMALL
ETA; QQK
1D776 ; [.10F1.0020.0005.1D776] # MATHEMATICAL SANS-SERIF BOLD SMALL
ETA; QQK
1D7B0 ; [.10F1.0020.0005.1D7B0] # MATHEMATICAL SANS-SERIF BOLD
ITALIC SMALL ETA; QQK
0397  ; [.10F1.0020.0008.0397] # GREEK CAPITAL LETTER ETA
bbbb  ; [.10F1.0020.0008.bbbb] # GREEK CAPITAL LETTER HETA
1D6AE ; [.10F1.0020.000B.1D6AE] # MATHEMATICAL BOLD CAPITAL ETA; QQK
1D6E8 ; [.10F1.0020.000B.1D6E8] # MATHEMATICAL ITALIC CAPITAL ETA;
QQK
1D722 ; [.10F1.0020.000B.1D722] # MATHEMATICAL BOLD ITALIC CAPITAL
ETA; QQK
1D75C ; [.10F1.0020.000B.1D75C] # MATHEMATICAL SANS-SERIF BOLD
CAPITAL ETA; QQK
1D796 ; [.10F1.0020.000B.1D796] # MATHEMATICAL SANS-SERIF BOLD
ITALIC CAPITAL ETA; QQK
```

References

Buck, C.D. 1955. *The Greek Dialects.* 2nd ed. Chicago: University of Chicago Press.

Jeffery, L.H. 1990. *The Local Scripts of Archaic Greece.* 2nd ed. Oxford: Clarendon Press.

Méndez Dosuna, J. 2003. Deconstructing 'height dissimilation' in Modern Greek. *Journal of Greek Linguistics* 3:83–114. Available as PDF: http://www.benjamins.com/jbp/series/JGL/3/art/0004a.pdf

Nicholas, N. 2003. http://www.tlg.uci.edu/~opoudjis/unicode/unicode_aitch.html

Peek, W. 1955. *Griechische Vers-Inschriften. Band I: Grab-Epigramme.* Berlin: Akademie-Verlag.

Roberts, E.S. 1887. *An Introduction to Greek Epigraphy, Part I.* Cambridge: Cambridge University Press.

Thumb, A. 1932. *Handbuch der Griechischen Dialekte, Erster Teil.* Ed. Kieckers, E. Heidelberg: Carl Winter.