To: Unicode Technical Committee
From: Deborah Anderson, SEI
Date: 31 January 2005
RE: Feedback on Proposals for Tack Hetas and Greek Epigraphical Letters
[Nicholas, L2/05-002 and L2/05-003]

Below are four expert responses regarding the proposals for the Tack Hetas and the Greek Epigraphical letters.

In sum, Tack Heta enjoys clear support amongst those queried, all scholars involved in the electronic publication of Greek materials. The other epigraphic symbols seemed to also have support for separate encoding, though one respondent (#4) felt a name-change for both the Corinthian EI and TSAN would be advisable.

Respondents:
Message 1: Elaine Matthews, Fellow, St Hilda's College, Oxford
Editor/Director, *Lexicon of Greek Personal Names* (print and online edition), Oxford University.

Message 2: Dr R. J. E. Thompson
Tutor and Director of Studies in Classics
Director of Computer-Assisted TextReading Project (which produces electronic versions of ancient texts suitable for students of the classical languages)
Selwyn College, Cambridge

Message 3: John Mansfield, Co-Director, Cornell Greek Epigraphy Project
(The Cornell Greek Epigraphy Project produces and distributes CD-ROMs of Greek epigraphical texts.)

Message 4: Dr Peter Haarer
Lecturer, Ancient History, Oxford University
(researcher with the Centre for the Study of Ancient Documents, Oxford, 2004)

1. -----Original Message-----
From: Elaine Matthews [mailto:elaine.matthews@classics.oxford.ac.uk]
Sent: Sunday, January 30, 2005 1:12 PM
To: Deborah W. Anderson
Subject: Re: Question about heta / epigraphical letters

Dear Deborah,

Background
The LGPN [Lexicon of Greek Personal Names] has been in operation since 1973. A successful
migration of our data across several generations of systems has taken place, and today
all our data (published) reside in identical database systems.
Unpublished data reside in identical text files. i..e from the point of
view of data preservation, we are fine.

Our route for the future (expressed in funding applications, no way we
will depart) is XML and unicode; if we use another database system, it
will be (probably) Postgres. We belong to the open-source community.

In the meantime, we still employ the transliteration system which we
devised in 1974, very much like Betacode as far as letters go, but not
as far as accentuation goes. But of course this transliteration can be
'transliterated' on to other formats. Currently our web site offers:
LGPN-code, Symbol, SpIonic, and (on our most advanced searching
facility) Betacaode, Perseus transcription, Monotonic encoding
Windows-1253 (aimed at modern Greeks, who are regular visitors to our
site), GreekKeys and Unicode. All start from our own transliteration.

To focus now on your questions:

Heta is important – sometimes it is the only indicator that a name had a
rough or smooth breathing. We use heta but not in the main name field,
only in the 'final bracket'information given at the end of the entry
i.e. we record it as a *variant*. At present we give only main name
forms on our web site, so have not faced up to heta; but it will be a
BIG ISSUE when we deliver all our data online, which we are committed to

do in an agreement with La Trobe University (for LGPN II online).

The questions
1. Would the encoding of these characters be useful to you and others?

YES. (But we don't distinguish lower case and capital. I think what we use in all cases you would call lower case.)

2. How are you currently encoding the hetas?
By the letter 7, in transliteration system devised in the early 70s. However, we are poised to move from this system to unicode in the near future, both for storage and manipulation of our data and for output to the web. We need unicode heta.

3. Are you in favor of encoding them, or would encoding them cause problems?
Yes, I am in favour.

Footnote in case it is helpful: LGPN is a unique project i.e. no-one else will ever collect together all Greek names from all sources and in all places. Updating and maintaining our datasets is vital. We (will soon) have published one quarter of a million individuals, and have about 200,000 more awaiting editing and publication. In any online systems, LGPN will be a main research tool. We use hetas.

I have now met a number of American colleagues working in these fields.

Best wishes,

Elaine

2. -----Original Message-----
From: rjet1@hermes.cam.ac.uk On Behalf Of Dr R J E Thompson
Sent: Sunday, January 30, 2005 11:53 AM
To: Deborah W. Anderson
Subject: Re: FW: Question about heta / epigraphical letters

At 18:02 28/01/2005, you wrote:

>1. Would the encoding of these characters be useful to you and others?

Yes, encoding of heta and the epichoric characters would be useful to anyone whose preference for encoding ancient Greek is Unicode, and who works with "non-standard" (i.e. non-classical or non-Attic) Greek.

>2. How are you currently encoding the epigraphical/hetas letters?

For heta, I use Roman h and H, and thus retain a Unicode font, but in a broken way. Similarly for the epichoric latter san/Pamphylian digamma, I use the homomorphic Cyrillic character, remaining with Unicode but brokenly. The situation is worse for archaic sampi, where I switch into a home-built, non-standard, non-Unicode font.

>3. Are you in favor of encoding them, or would encoding them cause problems?

I am in favour of encoding these glyphs. I cannot see how doing so could cause any problems.

--
Dr R. J. E. Thompson
Tutor and Director of Studies in Classics
Selwyn College, Cambridge, CB3 9DQ   College : +44 1223 335649
Faculty of Classics, Cambridge, CB3 9DA   Faculty : +44 1223 331957
                                    Mobile : +44 7818 074484
                                    http://www.doccyt.org
3. -----Original Message-----
From: John Mansfield [mailto:jmm12@cornell.edu]
Sent: Sunday, January 30, 2005 11:26 PM
To: Deborah W. Anderson
Cc: John Mansfield
Subject: Re: Nick Unicode

Dear Deborah,

Some positive, but not very organized, comments on Nick's Unicode proposals:

(FYI. In our current system at Cornell, the underlying Betacode is displayed in a Unicode font, either as an available character or a stand-in Betacode escape.)

Prop. 1. Small/Capital Heta. PHI's implementation of Betacode on the epigraphical CDs includes a code for Heta (*J) [Don't ask me why the code is tagged as a capital letter.], displayed as a sans-serif Roman h. We have oodles of aitches in our texts. But we have never been able to distinguish between lowercase and capital, as desired for proper names. (Betacode lacks a capital qoppa, too.) The 'tack heta' glyph, although not standard in printed publications, is a reasonable solution to this problem. The natural font pair h/H is confusing. We need, at least, a lowercase heta-daseia. Some editors write (e.g.) hAbro-, making the following vowel bear the burden of capitalization.

Prop. 2(a). Sampi. Generally we would use the Betacode text symbol #5 for both numbers and (on the rare occasion) letters. [When used as a letter, we add a regularized gloss with -ss- in the database for searching.] But for Pamphylian inscriptions (the texts from Brixhe's corpus are in our database at CU) we use a made-up code *Y2, which literally means "big weird psi," in the same way that S2 and S3 signify variant forms of sigma.

Prop. 2(b). Corinthian/Boeotian E(I). I was surprised to see these two flavors mixed. I cannot speak with any authority about the Boeotian uses. But when I was at Berkeley, I worked as a graduate assistant for Prof. Amyx when he was finishing "Corinthian Vase-Painting of the Archaic Period." The Corinthian E(I) is an essential codepoint. Look at the four corpora of Corinthian vase-inscriptions (Payne, Arena, Lorber, Amyx): each author negotiated a special E(I) character with the printer for use in transcriptions; the rest of the letters are standard Greek minuscules, even though the actual written forms (B = e, M = s, S = i) look very different. By the way, Betacode also lacks this code, so, in the Inscriptiones Graecae IV, for this character we have been forced to resort to <4EI>4 = "EI in ligature" (not true).

Prop.2(c). Tsan/Digamma. Again, I was surprised by the conflation of

Arcadian and Pamphylian -- but I'm not saying I know better. For the uncommon Arcadian form, on the PHI CD we used the made-up code S5, which means "s/t-ish sort of thingy." When we come to Pamphylia we transcribe N-reverse in Betacode as &V$ (literally a Roman font capital V) and substitute this for any editor's transcription with Greek digamma. The coin-inscription cited by Nick at the bottom of p. 9 (also including the Pamphylian sampi [above]) appears in our database as: &V$ANA/*Y2AS {26VANA/SSAS}26 *PREIIAS.

My small point. When the material forces us Betacode folk to make up non-standard, unimplemented codes or to stand on our heads codewise, that's a pretty good argument for a separate Unicode codepoint.

Best wishes,

John

4. -----Original Message-----
From: Peter Haarer [mailto:peter.haarer@classics.oxford.ac.uk]
Sent: Monday, January 31, 2005 12:00 PM
To: Deborah W. Anderson
Subject: Proposal on tack heta

Dear Debbie,

This looks good and would be useful. I have, in particular, had difficulties before when trying to represent upper-case Heta. I'm not sure where this should sit in the alphabetic row - I'm almost tempted to say it should come before Eta, as Eta could perhaps be described as a derivative. One other small point - I was a little surprised you have chosen an illustrative example in Old Italic where the letters are retrograde (i.e. backwards within the prevailing direction of writing) - but this is hardly fatal.

===
Sampi, Ei and Tsan.

I offer the following brief comments, though would stress the limits of my expertise in this area. As ever, Charles Crowther and / or Alan Johnston would be likely to give you a more informed view.

1. "SAMPI"

I agree with the ideal that there should be two signs, one each for Ionian and Pamphylian sampi, but also with the observation that the corpus of texts containing such signs is likely to remain very small. A sign which differentiates Archaic sampi from the later numeral sampi is certainly a good idea.

2. "EI"

I think there is a case for having the two new glyphs (though again, the corpus of texts seems likely to remain small). However, the name of the glyph(s) does need to be something other than "ei", which would certainly risk the confusion noted with the earliest attested name for epsilon. Further, The signs proposed for the two glyphs are perhaps not ideal for the reasons noted in the proposal.

3. "TSAN"

I think the proposal spells out the weight of reservations clearly. "Tsan" is perhaps an unwise name for the letter, given the problematic nature of its relationship with San.

4. "Pamphylian Digamma"

I feel out of my depth here, and that I should therefore not comment.

Dr. Peter Haarer
Corpus Christi College, Oxford OX1 4JF
tel. +44 1865 276759