

TO: Unicode Technical Committee

FROM: Deborah Anderson

DATE: 30 January 2004

RE: Comments on Request for Change to Greek Collation Order for Koppa L2/04-030

In document L2/04-030 Peter Kirk proposes a change in the Greek collation order for koppa. I wish to add a few additional comments.

1. Description of the use of the two koppas

The archaic koppa (hereafter "q-koppa," after its shape), a fairly rare letter, is found primarily on epigraphical materials in pre-Classical Greece. As noted in Peter Kirk's document, it continued on in use after pre-Classical times as a numeral with the value of '90'. The q-shaped koppa is used amongst Classicists and the letter is frequently included in Classical Greek fonts.

Modern Greek uses the non-archaic koppa (hereafter "lightning-bolt koppa"), strictly with the numerical value of '90'. It has officially been used in Greek legislative texts since 1821 (Michael Everson's document "On GREEK LETTER KOPPA" N1938, p. 5). As observed by Nick Nicholas (email dated 18 January 2004), "in my experience, modern Greek writers don't even know what the archaic koppa is," though presumably the modern Greek Classicists would be able to identify the two.

Generally, one would expect that the Classicists would be interested in searches for the q-koppa, and modern Greek users the lightning-bolt koppa. Since the two koppas have been formally distinguished since Unicode 4.0, the current system in which the different koppas are collated separately at the first level seems to address these needs. If the change were made to the collation algorithm as proposed, users may be surprised to get the lightning bolt when they are just looking for q-shaped koppa.

2. Comment on Justification 2: Confusion of Encoding the Texts (p. 2):

"Correctly encoded texts are unlikely to contain both archaic koppa and regular koppa..."

It is conceivable that the two could co-occur in more modern Greek materials where ancient Greek is cited. If indeed the two did co-occur in the same text, one could customize the collation table.

Richard Peevers of the Thesaurus Linguae Graecae found no example of the two ever co-occurring in the same text in the TLG text corpus, which has over 89 million words of Greek. However, since the lightning-bolt koppa is attested ca. 1821 and the TLG includes texts dating only until AD 1453, this may be expected. One should note that the intermediate form (similar to the shape Ϡ), perhaps the "missing link" between the change from q-koppa to the lightning bolt, appears in texts dating from the ninth century CE

onwards, again with the numerical value of '90' (Everson N1938, p.7; this is "Type #3" on Peter Kirk's proposal, p. 1). How these "intermediary koppas" are handled (i.e., are they encoded as q-koppas or lightning-bolt koppas), may impact searches.

3. Comment on Justification 2: Confusion of Encoding the Texts (p. 2):

"However, in previous versions of Unicode, before archaic koppa was separately defined, the regular koppa code point was used in both contexts."

For clarification, the history of the koppa is summarized in the following table (based on information provided by Ken Whistler). Ken added that the Greek national body was not ready to acknowledge the lowercase archaic forms as part of the 10646-1:1993 merger, but did finally agree to the q-koppa forms as distinct from the lightning-bolt koppas, and this disunification is reflected in Unicode 4.0.

Fig. 1: Koppa and its History in Unicode

Unicode	03DE	03DF	03D8	03D9
1.0	GREEK CAPITAL LETTER KOPPA (Ϟ and Ϙ) (Since both glyphs were provided, a font could pick either.)	GREEK SMALL LETTER KOPPA (ϙ and ϙ)		
1.1	GREEK LETTER KOPPA (no glyphs, but 10646-1: 1993 had the [upper case] Ϟ glyph)	(deleted)		
2.0	GREEK LETTER KOPPA (Ϟ [upper case])			
3.0	GREEK LETTER KOPPA KOPPA (Ϟ [upper case])			
4.0	GREEK LETTER KOPPA (Ϟ [upper case])	GREEK SMALL LETTER KOPPA (ϙ)	GREEK LETTER ARCHAIC KOPPA (Ϙ)	GREEK SMALL LETTER ARCHAIC KOPPA (ϙ)

4 . Comment on Justification 2: Confusion of Encoding the Texts (p. 2):

"A number of fonts were made with archaic koppa glyphs at what are now the regular koppa code points, and some of these fonts, e.g. Arial Unicode MS, are still being distributed unmodified."

The situation with regard to fonts is improving. Many newer fonts include the glyphs in their correct positions. Nick Nicholas has fully documented the level of compliance of other Greek fonts on a webpage

accessible from the TLG homepage, <http://www.tlg.uci.edu>, and provides links to the fonts, many of which are free. Approximately 14 include the q-koppa in the correct position.

Figure 2: Recent Greek Unicode Fonts

	03DE	03DF	03D8	03D9
Arial Unicode MS	Ϟ			
Payne Unbode	Ϟ	ϟ		
TITUS Cyberbit	Ϡ	ϡ	Ϟ	ϟ
Lucida Sans Unicode	Ϡ			
New Athena Unicode	Ϡ	ϡ	Ϟ	ϟ
Palatino Linotype	Ϡ			

Regarding Arial Unicode MS, I notified Paul Nelson (18 March 2003) of the error and he said it would be corrected in the next version of the font. I also informed Jost Gippert of the TITUS project of the error in the TITUS Cyberbit Basic font (17 January 2003), which is now corrected. I inquired about the error in the Linguist’s Software font, “LaserGreek in Unicode,” and they replied that the placement of the q-koppa was “deliberate” because it was felt this is was the more common form; Linguist’s Software has promised to put the q-koppa glyph in its proper location in their next version. I have since sent messages to the designers of the two remaining fonts with the q-koppa glyph in the lightning-bolt koppa’s place, Aisa and Vusilius Old Face, hoping these too can be corrected.

5 . Comment on Justification 2: Confusion of Encoding the Texts (p. 2):

“There is thus likely to be a significant body of texts, which will remain in archives indefinitely, in which the regular koppa code points are used for what is now defined as the separate archaic koppa”

There are two big digital libraries of Greek texts, TLG and Perseus. I checked with Richard Peevers (26 January 2004): the TLG has all the Unicode values for koppa correctly encoded, for they store their data in Beta Code and the different koppas are distinguished. I wrote to Anne Mahoney at Perseus, who replied (26 January 2004) that she would check on the encoding for koppa and make the correction if needed.

It is my experience that many Classicists tend to want to stick with their old (often non-standard) fonts. Note that both Perseus and TLG still provide users the option to display the Greek texts with non-standard font encodings (i.e., based on WinGreek, GreekKeys, Sgreek, SP Ionic, and a few others),

alongside a Unicode option. Fortunately, the American Philological Association has been promoting two free Unicode compliant fonts, Athena Unicode (PC) and New Athena Unicode (Mac), which include the q-koppas in the correct positions. The increasing availability of good Unicode compliant fonts that offer Classicists the glyph repertoire they require (with an acceptable appearance and proper diacritic stacking) is helping to gradually change the situation.

While it is possible that users may have copied Greek texts--with koppa incorrectly encoded--from the Perseus digital library, for example, Perseus has willingly agreed to make corrections, if needed. Anne Mahoney reported, however, that koppa was rare in the Perseus corpus. There no doubt exist wrongly encoded q-koppas in texts (at 03DE), because it was only in Unicode 4.0 that the forms were finally distinguished. I would suggest, however, that the far greater problem lies with text data on scholars' computers that were created with a wide array of non-standard fonts, which were able to provide Classicists at the time with the full repertoire they needed.

6. There was a discussion of the pros and cons of the proposal to change the Greek collation order for koppa on a separate email list hosted by the Stoa Consortium. The participants in the email interchange (carried on by four members, out of about 40+ on the email list) did not come to a consensus in support of this proposal. Maria Pantelia of the TLG deferred to Nick Nicholas, who was inclined to leave the current collation algorithm alone. Some of the comments made in the discussion by Nick Nicholas and Jeremy March have been incorporated in this document.

I would recommend the ramifications of the proposed change should be discussed and then returned back to the Classicists with an interest in Greek computing and collation to discuss it further.