From: Peter Kirk

Date: 24 January 2004

Re: Request for Change to Greek Collation Order for KOPPA

Current Situation

There appear to be four varieties and/or uses of the Greek letter *koppa*:

1. An archaic glyph used as a letter in early (pre-classical) Greek inscriptions;

2. The same archaic glyph used as a numeral, 90, in classical and Hellenistic Greek, and sometimes in modern Greek;

3. A transitional ("uncial" or "missing link") glyph used as a numeral in post-classical pre-modern Greek.

4. A non-archaic glyph used as a numeral in modern Greek.

According to Unicode as currently defined, code points 03D8 and 03D9 are intended for variety 1 and probably variety 2; and code points 03DE and 03DF are intended for variety 4, with the intention for variety 3 undefined. In an earlier version of Unicode there was only one *koppa*, 03DE.

The current default collation (DUCET, allkeys.txt) has the following order and weights for *koppa* and the surrounding characters:

```
1D28  ;  [.10FD.0020.0002.1D28]  # GREEK LETTER SMALL CAPITAL PI
03DF  ;  [.10FE.0020.0002.03DF]  # GREEK SMALL LETTER KOPPA
03DE  ;  [.10FE.0020.0008.03DE]  # GREEK LETTER KOPPA
03D9  ;  [.10FF.0020.0002.03D9]  # GREEK SMALL LETTER ARCHAIC KOPPA
03D8  ;  [.10FF.0020.0008.03D8]  # GREEK LETTER ARCHAIC KOPPA
03C1  ;  [.1100.0020.0002.03C1]  # GREEK SMALL LETTER RHO
```

Proposed Change

This is a proposal to change the DUCET weights for U+03D8 and U+03D9 to reflect the fact that archaic *koppa* is not a distinct character but a variant form of (non-archaic) *koppa*, and so should not be collated separately at the first level. Compare how for example U+03F2 lunate *sigma* (an archaic variant form of *sigma*) is collated with regular *sigma* at the first level and distinguished only at the third level. This proposal would treat the distinction between archaic *koppa* and regular *koppa* also as a third level distinction, and so with the following suggested order and weights:

```
1D28  ;  [.10FD.0020.0002.1D28]  # GREEK LETTER SMALL CAPITAL PI
03DF  ;  [.10FE.0020.0002.03DF]  # GREEK SMALL LETTER KOPPA
03D9  ;  [.10FE.0020.0004.03D9]  # GREEK SMALL LETTER ARCHAIC KOPPA
03DE  ;  [.10FE.0020.0008.03DE]  # GREEK LETTER KOPPA
03D8  ;  [.10FE.0020.000A.03D8]  # GREEK LETTER ARCHAIC KOPPA
03C1  ;  [.1100.0020.0002.03C1]  # GREEK SMALL LETTER RHO
```

Background

See http://www.tlg.uci.edu/~opoudjis/unicode/numerals.html#koppa. See also Michael Everson's 1998 proposal for separate encoding of archaic *koppa*, http://anubis.dkuug.dk/jtc1/sc2/wg2/docs/n1938.pdf (note that Everson confuses *san* and *sampi*). I note that Everson proposed disunification of alphabetic *koppa* (variety 1) from numeric *koppa* (varieties 2-4); but the decision taken by Unicode and ISO/IEC JTC1/SC2/WG2 was to disunify archaic *koppa* (varieties 1-2) from modern *koppa* (variety 4).

Justification 1: Lexical Usage

In the standard Liddell-Scott-Jones lexicon of classical Greek (Henry George Liddell. Robert Scott. A Greek-English Lexicon. revised and augmented throughout by. Sir Henry Stuart Jones. with the assistance of. Roderick McKenzie. Oxford. Clarendon Press. 1940. ISBN: 0198642261) there is a single entry for *koppa*, which is the only one between the last word starting with *pi* and the entry for *rho*. In this entry use as a letter and a numeral is described, and both the archaic and the transitional glyphs are shown. The following is a scan of the entry from the printed lexicon:

This indicates that in the standard lexicon the numeric and alphabetic *koppa* are considered to be different uses of the same character, and that this character has different forms. There is no trace of these characters being separately alphabetised. The same is true of the following entry from the modern Greek *Ilios Encyclopædic Lexicon*, copied from Everson's proposal N1938, in which the archaic and modern glyphs are clearly considered variants of the same letter:

Justification 2: Confusion in Encoding of Texts

Correctly encoded texts are unlikely to contain both archaic *koppa* and regular *koppa*, because according to current Unicode specifications they are used in two different contexts. However, in previous versions of Unicode, before archaic *koppa* was separately defined, the regular *koppa* code point was used in both contexts. 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. 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*. In order that simple searches work correctly on such materials, it is advisable that the regular and archaic *koppa* are collated together at the top level.

The confusion of those who have tried to represent koppa in Unicode is illustrated by the form in the online edition in the well-respected Perseus corpus of the entry in the Liddell-Scott-Jones lexicon reproduced above. The following is an image captured from the display of this entry on the page http://perseus.uchicago.edu/cgi-bin/ptext?doc=Perseus%3Atext%3A1999.04.0057&query=head%3D%2329 with the Mozilla 1.6 browser:

ϙ

ϙ ϙ, **koppa** (q.v.), nineteenth letter in the Etruscan abecedaria (*IG* 14.2420), occurring in *IG* 9(1).334.1, al. (Locr., v B.C.), etc.; as numeral = 90,

The following is the same text as copied into my word processor and reformatted with the font Code2000, chosen because it is known to use the correct glyphs for the variants of *koppa*:

ꟼ

ꟼ ꟼ, koppa (q.v.), nineteenth letter in the Etruscan abecedaria (IG 14.2420), occurring in IG 9(1).334.1, al. (Locr., v B.C.), etc.; as numeral = 90,

The following are the Unicode code points for the first part of the main line: 03DE 0020 03DF 002C 0020 006B 006F 0070 0070 0061. 03DE and 03DF are the codes for the normal *koppas*; the archaic *koppa* codes 03D8 and 03D9 are not used. But this text is presented by the website with archaic *koppa* glyphs, which is presumably the intention at least for the first *koppa*, following the printed edition and because this is intended to describe classical rather than modern usage. Archaic *koppa* glyphs have appeared only because the font selected has archaic glyphs at the modern *koppa* code points.

In fact the second glyph should be not an archaic but an intermediate *koppa*. It is not clearly defined which code point should be used for this variety.

This is one example of a possibly large number of texts which have been encoded with the wrong variety of *koppa*. Some such texts may be corrected in due course; others will remain uncorrected indefinitely. The uncertainty of how to encode the intermediate variety will remain. The implication is that complete consistency of encoding can never be expected.

If texts remain inconsistent, there is a problem for those collating texts and searching for *koppa*, if archaic *koppa* and modern *koppa* are collated separately at the first level. The problem largely disappears if they are distinguished only at the third collation level, as then a basic (first level) search for one form of *koppa* will match the other form.

Conclusion

Archaic and modern *koppa* should be collated as variants of the same character, and so distinguished only at the third level, both because this is the established lexical practice and because this facilitates collation and searching of possibly inconsistent texts.