

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Additional Greek characters for the UCS
Source: Michael Everson
Status: Working Document
Date: 1998-05-25

This document proposes several Greek characters for inclusion in the UCS, and discusses the rationale for their inclusion.

CAPITAL LETTERS STIGMA, DIGAMMA, KOPPA, and SAMPI (Ϝ ϝ Ϟ ϟ) are already present in ISO/IEC 10646-1. SMALL LETTER equivalents of these (ϝ Ϟ ϟ Ϡ) are found in ISO 5428:1984 but were not included in the UCS due to comments from the Greek national body, ΕΛΟΤ. If for nothing else, for compatibility of text interchange these four characters should be added to the UCS.

ΕΛΟΤ's view that the SMALL LETTERS are anachronistic is well understood, but it ignores the facts of publishing in the Greek language since the advent of printing five centuries ago. Eminent scholars worldwide have made use of these archaic characters and have applied case to them in normalized texts without regard to the theoretical authenticity of casing for the characters; they cased them as it was natural to do so. In short, while lower-case variants of these letters did not develop *at the same time* as did lower-case variants of other Greek letters, they developed nonetheless *at a later date*. Character set technology is not the place to "prescribe" the caselessness of character which have been given case, even anachronistically. The absence of these characters from the UCS only serves to prevent modern works on the Greek language from being represented correctly in UCS encoding.

In B. F. C. Atkinson's book *The Greek language* (London: Faber & Faber 1932), a text in the Elean dialect is given:

Ἄ φράτρα τοῖς φαλείοις. πατρίαν θαρρῆν καὶ γενεὰν καὶ ταῦτῳ. αἱ ζέ τις
κατιαραύσειε φάρρηνορ φαλείο, αἱ ζέ μὲ ἴπιθειαν τὰ ζίκαια ὄρ μέγιστον τέλος ἔχοι
καὶ τοὶ βασιλᾶες, ζέκα μναῖς κα ἀποτίνοι φέκαστος τῶν μὲ ἴπιποεόντων κα(θ)θυταῖς
τοῖ Ζι Ὀλυνπίοι.

CAPITAL LETTER DIGAMMA and SMALL LETTER DIGAMMA were distinguished by the editor of this text in the normal way, to differentiate ordinary words from proper nouns. The font used in Atkinson's book is Neo Hellenic, designed by Victor Scholderer in 1927.

Lower-case letters are used in Herbert Weir Smyth's *Greek grammar* (Cambridge: Harvard University Press, 1920). In the chart of Greek numerals on pp. 103–104, all letters, including these four, are given in lower case.

SMALL LETTER STIGMA was used widely in Greek typography throughout the history of printing Greek. In Ἱερατεινὴ Κομμαριανοῦ, Λουκία Δρούλια, and Evro Layton's beautiful book *Τὸ ἑλληνικὸ βιβλίον* (Ἀθήνα: Ἐθνικὴ Τραπεζα τῆς Ἑλλάδος, 1986), plates 28 (Βενετία 1573), 38 (Φλωρεντία 1496), 42 (Φλωρεντία 1496), 48 (Βενετία 1497), 63 (Βενετία 1560), 116 (Βενετία 1805), 125 (Κωνσταντινούπολη 1799), 132, 133 (Βιέννα 1797), 159 (Βενετία 1755), 181 (Φλωρεντία 1496), 184 (Λονδίνο 1590), 185 (Βενετία 1540), 200 (1626), 201 (Βενετία 1739), 202 (1765), and 232 (Κωνσταντινούπολη 1627) all make use of the SMALL LETTER STIGMA as a ligature of SIGMA and TAU. After each plate number here I have given the year and the city in which the work was published where this information was available the cities are Venice, Florence, Constantinople, Vienna, and London. Note that Κωνσταντινούπολη is written Κονσαντινούπολη in these sources.

To reiterate, of the letters Ϝ ϝ Ϟ ϟ Ϡ ϡ, at least F and ϝ are definitely differentiated as ordinary letters of the alphabet in texts I have to hand. While the other letters Ϟ ϟ Ϡ ϡ are chiefly used as

numerals, all appear with lower-case forms (as in Atkinson 1932), and the lower-case ς has particularly wide use as a ligature in Greek typography. These four missing SMALL LETTERS should be added to the UCS.









Two other characters are also proposed here for inclusion. While there are many ligatures used in early Greek typography, not all of them need be encoded in the UCS. But *two* of them should be considered because of their uniqueness, their ubiquitousness, and their *contemporary* and *public* use.

Several other special typographical variants of Greek letters are already encoded in the UCS. Among these are β̄ GREEK BETA SYMBOL, θ̄ GREEK THETA SYMBOL, and π̄ GREEK PI SYMBOL. These may be used as technical symbols, or as typographical variants in Greek texts, particularly in texts of a couple of centuries ago. I propose the addition of ϣ̄ GREEK KAI SYMBOL, Ő GREEK CAPITAL LETTER OU, and 8̄ GREEK SMALL LETTER OU.

The GREEK KAI SYMBOL was used in early Greek typography as an abbreviation (not a ligature) for καὶ ‘and’. Just as the Latin word *et* ‘and’ can be represented by the abbreviation &, so καὶ was represented by ϣ (or ϣ̄). Apart from its frequent appearance in early Greek typography, I recently noticed that this sign is in *common* use in contemporary Athens. Many, many shops use the character & to denote καὶ on their shop signs, but a considerable number of them eschew this Latin sign and prefer to use ϣ. As a foreigner, although admittedly one with some typographical knowledge, I found the use of & to look rather odd in Greek text, and was pleased to discover that ϣ was preferred by many. It was only *after* I had seen it on many public shop signs that I also discovered that ϣ was a standard typographical sign in Greek printing, as evidenced in *Τὸ ἑλληνικὸ βιβλίον*.

I believe that if ϣ were found in coded character sets, Greeks would prefer it to &, and I suspect that the prevalence of the Latin & on Greek signage today is *solely* an artifact of limited 8-bit technology. I hope that EΛOT and WG2 will choose to support its inclusion in the UCS.

The GREEK CAPITAL LETTER OU, and GREEK SMALL LETTER OU are also in use in contemporary Greece, though to my knowledge there has never been character set support for them. While somewhat rarer than the GREEK KAI SYMBOL, I saw at least one shop sign (an auto and machine parts shop, as I recall) which used Ő, and in the hallways of the School of Architecture at the Πολυτεχνεῖο on 28 Οκτώβριου πατησίων in Athens there appears quite a lot of rather artistic grafitti which also contains this character. In early Greek typography, it is possible to say that 8 is by far and away the most frequently occurring ligature or abbreviation of all. Indeed, this letter was so well known that a Canadian Christian missionary borrowed it to create 8̄ LATIN LETTER OU, which is used today in the Algonquin language. (This LATIN LETTER OU is also missing from the UCS, and in another document I propose its addition.) The following characters (with proposed code positions) should therefore be considered for addition to the UCS:

03D7 GREEK KAI SYMBOL		
03D8 GREEK KAI SYMBOL WITH VARIA	03DD:	
03DB GREEK SMALL LETTER STIGMA		
03DD GREEK SMALL LETTER DIGAMMA		
03DF GREEK SMALL LETTER KOPPA	03DF:	
03E1 GREEK SMALL LETTER SAMPI		
03F4 GREEK CAPITAL LETTER OU	03E1:	
03F5 GREEK SMALL LETTER OU		
	03F4:	
03D7:		
		
03D8:		
	03F5:	
03DB:		
		

A. Administrative

1. Title

Additional Greek characters for the UCS

2. Requester's name

Michael Everson, EGT (WG2 member for Ireland)

3. Requester type

Expert contribution

4. Submission date

1998-05-25

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 characters to existing block? Name?

Yes. Greek Symbols and Coptic

2. Number of characters

8

3. Proposed category

Category A

4. Proposed level of implementation and rationale

Level 1 noncombining characters.

5a. Character names included in proposal?

Yes

5b. Character names in accordance with guidelines?

Yes

5c. Character shapes reviewable?

Yes (see above)

6a. Who will provide computerized font?

Michael Everson, Everson Gunn Teoranta

6b. Font currently available?

Yes

6c. Font format?

TrueType

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes. ISO 5428:1984

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Hardcopy is provided for WG2 distribution.

<http://www.indigo.ie/egt/standards/iso10646/pdf/iso-5428.pdf>

8. Does the proposal address other aspects of character data processing?

No, there aren't any.

C. Technical -- Justification

1. Contact with the user community?

Yes: TC46/SC4, U.S. Library of Congress, and EAOT.

2. Information on the user community?

Librarians, users of ISO 5428:1984, linguists, publishers, and the general public.

3a. The context of use for the proposed characters?

Archaic and typographic Greek. Compatibility with ISO 5428:1984.

3b. Reference

See references given above.

4a. Proposed characters in current use?

Yes

4b. Where?

In Greece and elsewhere in the world.

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Contemporary use and keeping them with the other Greek characters.

6. Should characters be kept in a continuous range?

No.

7a. Can the characters be considered a presentation form of an existing character or character sequence?

GREEK SYMBOL KAI could be considered a ligature of KAPPA and ALPHA and IOTA but in fact it is not a ligature, it is an abbreviation symbol, just as & is not a ligature of E and T, even when represented as &.

7b. Where?

7c. Reference

8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?

No

8b. Where?

8c. Reference

9a. Combining characters or use of composite sequences included?

Yes. GREEK SYMBOL KAI WITH VARIA can be decomposed to GREEK SYMBOL KAI and COMBINING GRAVE ACCENT.

9b. List of composite sequences and their corresponding glyph images provided?

Yes

10. Characters with any special properties such as control function, etc. included?

No

D. SC2/WG2 Administrative

To be completed by SC2/WG2

1. Relevant SC 2/WG 2 document numbers:

2. Status (list of meeting number and corresponding action or disposition)

3. Additional contact to user communities, liaison organizations etc.

4. Assigned category and assigned priority/time frame

Other Comments