

# Guidelines and Suggested Amendments to the Greek Unicode Tables

Yannis Haralambous  
Département Informatique  
École Nationale Supérieure des Télécommunications de Bretagne  
B.P. 832, 29285 Brest, France  
<http://omega.enstb.org/yannis>  
<mailto:yannis.haralambous@enst-bretagne.fr>

## ABSTRACT

In this paper we present guidelines for using the part of Unicode related to the Greek writing system, namely the tables of “basic Greek” and “extended Greek.” These guidelines are followed by a series of suggestions of amendments to the Unicode standard v3.2, as well as a certain number of characters we submit for insertion into the standard. The aims of the three parts of this paper are to explain, correct and complete the Greek part of Unicode.

## ΠΕΡΙΛΗΨΗ

Τὸ ἄρθρο αὐτὸ περιέχει ὁδηγίες χρήσης τοῦ ἑλληνικοῦ μέρους τῆς κωδικοσελίδας Unicode, δηλαδὴ τῶν πινάκων «βασικῶν» καὶ «ἐκτεταμένων» ἑλληνικῶν. Τὶς ὁδηγίες χρήσης ἀκολουθεῖ μία σειρά ἀπὸ ἀλλαγές καὶ διορθώσεις ποὺ προτείνουμε ὅπως καὶ σωρεῖα χαρακτήρων ποὺ ὑποβάλλουμε γιὰ ἔνταξη στὴν κωδικοσελίδα. Σκοπὸς τοῦ τρίπτυχου αὐτοῦ ἄρθρου εἶναι νὰ ἐξηγήσει, νὰ διορθώσει καὶ νὰ ὀλοκληρώσει τὴν ἑλληνικὴ πλευρὰ τῆς τυποποίησης Unicode.

## 1 Guidelines

The Greek Unicode tables have three major inconvenients:

1. the fact that there are *two* Greek tables in Unicode (pages 372 and 489-90 of the Unicode book) with conflicting contents. In particular, some letter+accent combinations appear twice in the Unicode standard, and therefore users should agree to which occurrence to use;
2. the fact that the Greek alphabet is used both for Greek language *and* for mathematical notation and that Greek Unicode tables mix these two aspects;
3. the confusing and redundant way diacritics have been encoded:
  - (a) there are combining diacritics for accents and breathings, but no combining accent+breathing combinations;
  - (b) there are spacing versions of accents, breathings and their combinations;

- (e) there are even spacing versions of accents, dieresis, and their combinations, which never appear in front of vowels and hence are illogical.

For these reasons we would like to suggest a certain number of rules for using the Greek Unicode tables. Most of these rules are prohibitive: the reason for this is that, because of historical reasons<sup>1</sup> there are characters which either are illogical, or which appear twice in the encoding, and we consider that people should refrain from using them. Of course we supply also the necessary rationale for these rules.

### 1.1 Rule 1: Lowercase Letters with Acute Accent

**RULE:** *Do not use characters* U+1F71 GREEK SMALL LETTER ALPHA WITH OXIA, U+1F73 GREEK SMALL LETTER EPSILON WITH OXIA, U+1F75 GREEK SMALL LETTER ETA WITH OXIA, U+1F77 GREEK SMALL LETTER IOTA WITH OXIA, U+1F79 GREEK SMALL LETTER OMICRON WITH OXIA, U+1F7B GREEK SMALL LETTER UPSILON WITH OXIA, U+1F7D GREEK SMALL LETTER OMEGA WITH OXIA, U+1FD3 GREEK SMALL LETTER IOTA WITH DIALYTIKA AND OXIA, U+1FE3 GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND OXIA.

**RATIONALE:** In the “basic” Greek table all vowels with acute accent are already provided: U+03AC GREEK SMALL LETTER ALPHA WITH TONOS, U+03AD GREEK SMALL LETTER EPSILON WITH TONOS, U+03AE GREEK SMALL LETTER ETA WITH TONOS, U+03AF GREEK SMALL LETTER IOTA WITH TONOS, U+03CC GREEK SMALL LETTER OMICRON WITH TONOS, U+03CD GREEK SMALL LETTER UPSILON WITH TONOS, U+03CE GREEK SMALL LETTER OMEGA WITH TONOS, as well as vowels with dieresis and acute accent U+0390 GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS, U+03B0 GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS.

Since we have to choose between two occurrences of exactly the same letter, our suggestion is to systematically use the first occurrence, namely the one in the “basic” Greek table.

### 1.2 Rule 2: Uppercase Letters with Accent Only

**RULE:** *Do not use characters* U+1FBA GREEK CAPITAL LETTER ALPHA WITH VARIA, U+1FBB GREEK CAPITAL LETTER ALPHA WITH OXIA, U+1FC8 GREEK CAPITAL LETTER EPSILON WITH VARIA, U+1FC9 GREEK CAPITAL LETTER EPSILON WITH OXIA, U+1FCA GREEK CAPITAL LETTER ETA WITH VARIA, U+1FCB GREEK CAPITAL LETTER ETA WITH OXIA, U+1FDA GREEK CAPITAL LETTER IOTA WITH VARIA, U+1FDB GREEK CAPITAL LETTER IOTA WITH OXIA, U+1FEA GREEK CAPITAL LETTER UPSILON WITH VARIA, U+1FEB GREEK CAPITAL LETTER UPSILON WITH OXIA, U+1FF8 GREEK CAPITAL LETTER OMICRON WITH VARIA, U+1FF9 GREEK CAPITAL LETTER OMICRON WITH OXIA, U+1FFA GREEK CAPITAL LETTER OMEGA WITH VARIA, U+1FFB GREEK CAPITAL LETTER OMEGA WITH OXIA.

**RATIONALE:** An uppercase letter carries an explicit accent *only* if it is the initial letter of the word, and hence necessarily carries also a breathing. These characters do not carry breathings but only accents, hence they are illogical (note that the “monotonic” combination of uppercase vowels with acute accent is covered in the “basic” Greek table by characters U+0386 GREEK CAPITAL LETTER ALPHA WITH TONOS, U+0388 GREEK CAPITAL LETTER EPSILON WITH TONOS, U+0389 GREEK CAPITAL LETTER ETA WITH TONOS, U+038A GREEK CAPITAL LETTER IOTA

<sup>1</sup>This is the non-offensive way of referring to human errors...

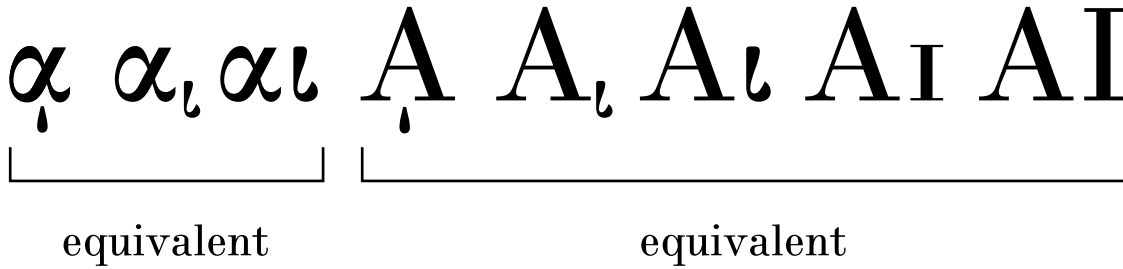


Figure 1: Equivalent representations of the mute *iota* grammatical phenomenon (characters U+1FB3 GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI and U+1FBC GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI, which we suggest to rename into U+1FB3 GREEK SMALL LETTER ALPHA WITH MUTE IOTA and U+1FBC GREEK CAPITAL LETTER ALPHA WITH MUTE IOTA).

WITH TONOS, U+038C GREEK CAPITAL LETTER OMICRON WITH TONOS, U+038E GREEK CAPITAL LETTER UPSILON WITH TONOS, U+038F GREEK CAPITAL LETTER OMEGA WITH TONOS).

### 1.3 Rule 3: Mute Iota

**RULE:** *If you wish to encode a letter with mute iota, do not pay attention to the fact whether the Unicode encoding defines it as having a subscript or an adscript iota. Leave the choice of mute iota representation to the rendering engine.*

**RATIONALE:** In Unicode, lowercase letters carry a “subscript” iota and uppercase letters an “adscript” one. This distinction is *meaningless*: the choice between these two representations is merely stylistic and should be left to the rendering engine. Note also that lowercase letters can very well carry adscript iotas and uppercase letters subscript ones: all combinations are possible and none carries any additional semantics<sup>2</sup> (see fig. 1).

### 1.4 Rule 4: Variant Letterforms and Greek Symbols

**RULE:** *Use U+03D0 GREEK BETA SYMBOL, U+03D1 GREEK THETA SYMBOL, U+03D2 GREEK UPSILON WITH HOOK SYMBOL, U+03D5 GREEK PHI SYMBOL, U+03F0 GREEK KAPPA SYMBOL, U+03F1 GREEK RHO SYMBOL, U+03F4 GREEK CAPITAL THETA SYMBOL, U+03F5 GREEK LUNATE EPSILON SYMBOL, U+03F6 GREEK REVERSED LUNATE EPSILON SYMBOL only in mathematical formulas, and never in Greek text. Do not use characters U+03D3 GREEK UPSILON WITH ACUTE AND HOOK SYMBOL, U+03D4 GREEK UPSILON WITH DIAERESIS AND HOOK SYMBOL.*

**RATIONALE:** Character U+03D0 GREEK BETA SYMBOL is a medial/final beta. The distinction between initial and medial/final should be left to the rendering engine (as this is the case for the Arabic alphabet).

<sup>2</sup>One of the amendments we propose is precisely to change names of Unicode characters with mute *iota*, see §2.2.



Figure 2: Some letterforms in various typefaces: here we haven’t kept all glyphs included in each typeface, but rather those that are the most natural and most frequently used.

Characters U+03D1 GREEK THETA SYMBOL and U+03D5 GREEK PHI SYMBOL, also known as “open theta” and “closed phi” are available in all fonts, but their use is not always very natural (in italic fonts one uses rather the open theta and in straight fonts, the closed one, but this rule is by no means universal; on the other hand, the closed phi is rarely used in Greek text).

Characters U+03F0 GREEK KAPPA SYMBOL and U+03F1 GREEK RHO SYMBOL are simply the same as the usual letters, but in different fonts (it is like if a Garamond ‘a’ is considered as a distinct Unicode character than a Times ‘a’...).

However, the presence of these characters in the Unicode standard is necessary, since it conforms to the principle of *completeness*. These variant forms of Greek letters have been used in mathematical typesetting systems (to start with  $\TeX$ , where they are produced by macros called `\varthetaeta`, `\varphi`, `\varkappa`, etc.), and express different semantics in mathematical formulas than the “plain” forms. Hence, our suggestion is to use these characters strictly in a mathematical context *only*.

Characters U+03D3 GREEK UPSILON WITH ACUTE AND HOOK SYMBOL and U+03D4 GREEK UPSILON WITH DIAERESIS AND HOOK SYMBOL are a mixture of variant forms used in mathematical forms with diacritics from Greek language: in other words, they are illogical, and should not be used.

If you happen to like the variant letterforms more than the standard ones, the solution is simple: use a font containing those letterforms (for example U+03F0 GREEK KAPPA SYMBOL is the letter *kappa* of the Linotype Times Ten font, while U+03BA GREEK SMALL LETTER KAPPA is the same letter in, for example, the Monotype Times Greek font (both are “Times” fonts)). See also fig. 2 illustrating these letterforms.

Note that we did not include U+03D6 GREEK PI SYMBOL in the rule because its situation is not very clear: it *may* be used in Greek text, even if the typographical conventions and the semantics it may carry are unclear.

On the other hand, it is perfectly clear that U+03F2 GREEK LUNATE SIGMA SYMBOL and its uppercase version (which we propose for insertion in the Unicode standard, see §3.1) are variant versions of U+03C2 GREEK SMALL LETTER FINAL SIGMA, U+03C3 GREEK SMALL LETTER SIGMA and U+03A3 GREEK CAPITAL LETTER SIGMA with *special semantics*, namely the ones of context-free *sigma* letters, both uppercase and lowercase.

Figure 3: On the left, number 1996 in lowercase and uppercase form. On the right, two words written using alphabetic *koppa* and *digamma*.

### 1.5 Rule 5: Archaic Letters

**RULE:** Use U+03DA GREEK LETTER STIGMA, U+03DB GREEK SMALL LETTER STIGMA, U+03DE GREEK LETTER KOPPA, U+03DF GREEK SMALL LETTER KOPPA, U+03E0 GREEK LETTER SAMPI, U+03E1 GREEK SMALL LETTER SAMPI as numerals only, and U+03D8 GREEK LETTER ARCHAIC KOPPA, U+03D9 GREEK SMALL LETTER ARCHAIC KOPPA, U+03DC GREEK LETTER DIGAMMA, U+03DD GREEK SMALL LETTER DIGAMMA as letters only.

**RATIONALE:** This is current typographical practice. (See fig. 3 for an example.)

### 1.6 Rule 6: Diacritics

**RULE:** To represent letters with accents, breathings and dieresis, use either combining diacritics or pre-accented characters.

The same rule, rephrased: Do not use spacing diacritics to represent uppercase letters with accents and breathings.

**EXAMPLE:** to encode a capital *alpha* with acute accent and rough breathing, write either U+1F0D GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA OR U+0391 GREEK CAPITAL LETTER ALPHA followed by U+0314 COMBINING REVERSED COMMA ABOVE (Dasia) followed by U+0301 COMBINING ACUTE ACCENT (Oxia, Tonos), but *not* U+1FDE GREEK DASIA AND OXIA followed by U+0391 GREEK CAPITAL LETTER ALPHA.

**RATIONALE:** Spacing diacritics are against the Unicode principle of post-positive notation. Also, in Greek typography, a diacritic can have zero width or not, depending on the context.

### 1.7 Suggestion 6': Pre-accented characters vs. combining diacritics

**SUGGESTION:** Use pre-accented characters rather than combining diacritics.

**RATIONALE:** diacritics are placed on top of letters, but are not necessarily centered. In fact, finding the exact position of a given diacritic over a given letter is a difficult process and is best done by the font designer. Therefore, if you use pre-accented letters you have more chances of having diacritics optimally placed on letters (while a rendering engine compatible with combining diacritics will rather center all diacritics upon letters).

This is rather a “suggestion” than a “rule” since it is based purely on esthetic criteria. From a logical point of view, pre-accented characters and letters followed by combining diacritics are equivalent.

ὁ κυρ' Ἀλέξανδρος, κλπ.  
 l'apostrophe, l'effet, etc.

Figure 4: Compare apostrophe, smooth breathing and comma on the first line (Greek) and the apostrophe and comma on the second line (French): Greek apostrophe is similar in shape to the smooth breathing, while “Latin” apostrophe is similar in shape to the comma.

«Καὶ εἶπε “ὄχι!”»

Figure 5: Greek first order and second/higher order quotation marks, as in font Monotype Greek 90.

## 1.8 Rule 7: Apostrophe vs. Smooth Breathing

**RULE:** Use U+2019 RIGHT SINGLE QUOTATION MARK rather than U+1FBF GREEK PSILI as an apostrophe.

**RATIONALE:** the reader may ask “why on earth would anyone use U+1FBF GREEK PSILI as an apostrophe in the first place?” Well, the fact is that the shape of the apostrophe in Greek typography is *exactly* the one of the smooth breathing U+1FBF GREEK PSILI. The “Latin” apostrophe U+2019 RIGHT SINGLE QUOTATION MARK is a bit longer, and looks more like a comma than like a smooth breathing. This difference seems small, but is quite unpleasant when you have both an apostrophe and a smooth breathing close to each other, and this happens quite often in Greek. Therefore one could be tempted to use U+1FBF GREEK PSILI (which after all, is a spacing diacritic) as an apostrophe. (See also fig. 4.)

This would be against Unicode principles: in fact, it is the rendering engine which should change the shape of the apostrophe in a Greek context.

*En passant*, note that Greek is the only language in the world where a blank space is systematically inserted after an apostrophe.

## 1.9 Rule 8: Quotation marks

**RULE:** Use U+00AB LEFT-POINTING DOUBLE ANGLE QUOTATION MARK and U+00BB RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK as first order quotation marks, and U+201F DOUBLE HIGH-REVERSED-9 QUOTATION MARK and U+201D RIGHT DOUBLE QUOTATION MARK as second and higher order quotation marks.

**RATIONALE:** this is current typographical practice. As in rule 7, quotation marks take special shapes in Greek: U+00AB LEFT-POINTING DOUBLE ANGLE QUOTATION MARK and U+00BB

RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK are—in some fonts—round, rather than angular, and U+201F DOUBLE HIGH-REVERSED-9 QUOTATION MARK and U+201D RIGHT DOUBLE QUOTATION MARK have exactly the shape and height of “double breathings” (in fact, the former is a “double rough breathing” and the latter, a “double smooth breathing”). (See also fig. 5.)

COROLLARY TO RULE 8: *Do not use double breathings U+1FFE GREEK DASIA or double U+1FBF GREEK PSILI as second or higher order quotation marks.*

## 2 Amendments to Unicode v3.2

In this section we present five amendments (two major and three minor ones) to Unicode v3.2.

The two major amendments concern: the controversial issue of accenting system (“monotonic” versus regular (or “polytonic”) Greek, §2.1), and the issue of the *mute iota* grammatical phenomenon (§2.2).

The three minor amendments concern the shape of the *perispomeni* accent (§2.3), the side-by-side placement of accents (§2.4) and the numeric/alphabetic usage of *stigma* and *digamma* (§2.5).

The amendments proposed in this paper agree with Unicode technical reports TR#15 Normalization (in particular form NFC), TR#25 (Unicode and math), TR#27 (Unicode 3.1), TR#28 (Unicode 3.2), TR#20 (Unicode in markup languages), TR#21 (Case mapping) and TR#24 (Script names). The author would like to thank Asmus Freytag for his valuable advice.

### 2.1 Unicode and the Greek Accent Controversy

*I have had quite a few conversations with our Greek subsidiary. I am well aware that polytonic Greek is making a modern revival in the past few years.*

Paul Nelson (Lead developer of the Microsoft Typography division)

#### 2.1.1 The historical facts

On February 11, 1982, after midnight, the Greek Parliament (that is, the 30 parliamentaries that were present at that late hour) voted law 1228/1982, deprecating the use of breathings, grave and circumflex accent and mute iota. The way that law was presented at the Parliament is remarkably obscure: the title of the law is *Κύρωση τῆς ἀπὸ 11.11.1981 πράξης τοῦ Προέδρου τῆς Δημοκρατίας περὶ ἐγγραφῆς μαθητῶν στὰ Λύκεια τῆς Γενικῆς καὶ Τεχνικῆς καὶ Ἐπαγγελματικῆς Ἐκπαίδευσεως (Validation of the Presidential Act of 1981-11-11 about the inscription of pupils to Schools of General, Technical and Professional Education)* [3]. And indeed, the first article of law 1228/1982 is the validation of a Presidential Act concerning pupils and schools. This article is 79 lines long.

*But*, after that first—completely innocent—article, comes a second article, 9 lines long, saying:

Μετὰ τῆ δημοσίευση τοῦ παρόντος νόμου, ὁ τονισμὸς τοῦ γραπτοῦ ἑλληνικοῦ λόγου γίνεται σύμφωνα μὲ τὸ μονοτονικὸ σύστημα.  
Μὲ Προεδρικὰ Διατάγματα ποὺ θὰ προταθοῦν ἀπὸ τοὺς Ὑπουργοὺς Ἐθνικῆς Παιδείας καὶ Θρησκευμάτων καὶ Προεδρίας τῆς Κυβέρνησης, θὰ καθορισθοῦν



τὸ εἶδος τοῦ μονοτονικοῦ, οἱ κανόνες του καθῶς καὶ οἱ λεπτομέρειες γιὰ τὴν ἐφαρμογὴ τοῦ παρόντος νόμου στὴν Ἐκπαίδευση καὶ στὴ Διοίκηση.

In English:

After publication of this law, the accenting of the written Greek language will be done accordingly to the monotonic system.

Presidential Decisions<sup>3</sup> established by the minister of Education and Religious Matters, and the minister of Presidency will determine the nature of monotonic system, its rules and the details of its application to Education and Administration.

Well understood, this article has no relation whatsoever with the title of law 1228/1982, and obviously is simply “hidden” in that law, to remain unnoticed.

That’s politics. On the other hand, the major argument of supporters of the “monotonic” system was that, contrarily to French and German, accents and breathings had no incidence anymore to the *phonetics* of Greek language.

This argument is irrelevant since these diacritics may have little direct impact on the *phonetics* of the language but are important on the *morphological*, *syntactical* and *etymological* level<sup>4</sup>.

The real reason for deprecating diacritics was the pressure by the printing industry, and especially by daily newspapers: the monotonic system reduced their production costs by 40% (without of course lowering the price of journals) and allowed them to computerize the printing process more easily, since at that time there were indeed technical problems involved in typesetting regular Greek.<sup>5</sup>

<sup>3</sup>Indeed, a presidential decision [4] was voted on April 22, 1982, providing the rules of “monotonic” system.

<sup>4</sup>For example, the circumflex accent allows disambiguation of words: ὠραία is the feminine adjective “beautiful,” while ὠραῖα is the adverb “beautifully,” or the plural neutral form of the adjective “beautiful.” This is a morphological difference between words.

The grave accent plays a syntactical/semantical rôle: γὰρ has the meaning of “because” while γὰτί has the meaning of “why.” Here we see that the grave accent *has indeed* an impact on phonetics of Greek language, since γὰτί will be pronounced with a raising voice (and *oxía* means precisely that), while γὰρ will be pronounced with a neutral or lowering voice (and that’s what *varía* means).

The rough breathing plays an etymological rôle: ὄρος means “term” (of an agreement), while ὄρος means “mountain.” When words are combined, the rough breathing changes the consonant preceding it: for example: ἐπὶ + ὄρος = ἐφορος (the *pi* has been transformed into a *phi*), which wouldn’t be the case if the second word meant “mountain” instead of “term,” etc.

The rough breathing has the same origin as the initial ‘h’ found in many European languages (for example, words starting with *homo* or *hetero* derive from Greek words ὅμοῦ (together) and ἕτερος (other)). In English and German this ‘h’ is still pronounced, while in French, for example, it is not pronounced, but kept for etymological reasons. In other words, French keeps a cultural heritage of Greek origin, which the Greek government itself has rejected...

<sup>5</sup>Computers are not able to typeset your language? You have two solutions: either enhance computers, or you can change your language, all depends on your degree of civilization and respect of cultural heritage...

### 2.1.2 The author's opinion

By that reform, the Greek government committed a crime against Greek language...

The reform is now part of Greek history. But the fact is that during all those years the Academy of Athens and many publishers have resisted and are still publishing their books in regular Greek. Indeed there are more and more books in regular Greek, even though accents and breathings are not anymore taught in school and most younger Greeks do not know how to use them.

And, well understood, the community of scholars (of ancient, medieval or modern Greek) outside Greece has never adopted the “monotonic” system. This is not a minor issue: the Greek language has been studied in the West for centuries, since Greek culture is part of the foundations of Western civilization. Like all landmarks of civilization, one can say that Greek language belongs to the entire world, and not only to a small, *profit-droven* group of people called “the Greek government.”

But how does Unicode get involved with these issues?

As Unicode aims to be used by all kinds of people, it should be “politically correct,” and avoid explicit or implicit acception or refusal of one of the two systems (regular vs. “monotonic” Greek).

Is this currently the case?

Unfortunately no. For example, on page 167 of the Unicode book, “Polytonic Greek” is presented as an exceptional case of Greek, *used for ancient Greek only*. This is totally wrong and implicitly pro-“monotonic,” and therefore unacceptable.

[On the other hand, page 71 of the Unicode book shows a very beautiful close-up picture of some words in Greek, and these words happen to be in *polytonic Modern Greek!* (the word καφές = “coffee,” can hardly be considered ancient or byzantine...)]

The suggestions of amendments that follow aim to indicate such pitfalls and supply alternatives.

### 2.1.3 Amendment 1a

AMENDMENT of p. 167 of the Unicode book:

The Greek script is written in linear sequence from left to right with ~~occasional~~ frequent use of non-spacing marks. For pre-composed combinations of letters and nonspacing marks, see the following subsection, “Greek Extended: U+1F00-U+1FFF.” Greek letters come in upper- and lowercase pairs.

~~**Polytonic Greek.** Polytonic Greek, used for ancient Greek (classical and Byzantine), may be encoded using either combining character sequences or precomposed base plus diacritic combinations. For the latter, see the following subsection, “Greek Extended: U+1F00-U+1FFF.”~~

**Monotonic Greek.** In 1982, the government of the Hellenic Republic has deprecated by law the use of breathings, grave and circumflex accent, and mute iota. This version of the Greek writing system is called “monotonic.” Standards ISO 8859-7 and ELOT 928 have been designed for “monotonic” Greek. Even today, many Greek documents are not conformant to that reform.

RATIONALE:

- 1) As explained earlier, in regular Greek practically each word has at least one accent (U+0300 COMBINING GRAVE ACCENT (Varia), U+0301 COMBINING ACUTE ACCENT (Oxia, Tonos) OR

U+0342 COMBINING GREEK PERISPOMENI), and every word starting with a vowel has also a breathing (U+0313 COMBINING COMMA ABOVE (Psili), U+0314 COMBINING REVERSED COMMA ABOVE (Dasia)). Searching in the TLG, we found that *22,3 percent of Greek letters have at least one nonspacing mark*<sup>6</sup>; this is why their use should rather be classified as “frequent” than “occasional.”

2) We have explained above both the historical facts (§2.1.1) and our personal views (§2.1.2) about the 1982 reform. As “monotonic” Greek is a quite *local* phenomenon, both historically (only 20 years of existence, compared to eighteen centuries of regular Greek) and geographically (it is used only in Greece and Cyprus, and not by the worldwide community of scholars), it should be the one presented in the text as a special case, rather than “polytonic” Greek, which we would rather call “regular” or “ordinary” Greek, since that’s what it has always been.

#### 2.1.4 Amendment 1b

AMENDMENT of p. 167 of the Unicode book:

~~The basic Greek accent written in modern Greek is called *tonos*. It is represented by an acute accent (U+0304). The shape that the acute and grave accents ~~takes~~ take over Greek letters is generally steeper than that shown over Latin letters in Western European typographic traditions. ~~and in~~ In “monotonic” Greek the only remaining accent is the acute accent (*oxia*), called *tonos*. In earlier editions of this standard it was mistakenly shown as a vertical line over the vowel.~~

RATIONALE:

The fact that acute and grave accents are steeper than their Latin alphabet equivalents is statistically true (actually, even in Latin alphabet text, the angle of an acute or grave accent depends on the typeface used: for example, Bodoni accents are steeper than Times ones); however, this fact is by no means related to the “monotonic” system.

On the other hand, the error of earlier editions of the Unicode standard<sup>7</sup> affected only the “basic” Greek table, probably because people with little (or no?) knowledge of Greek have imagined that an accent which derives from both the acute and the grave accent should be symmetric.

As stated in [1, p. 14], in the early days of the 1982 reform, some documents used triangular or circular accents to represent the unique accent that survived the reform. This practice diminished progressively after ministerial grammar books stated clearly that the “unique accent has the shape of an acute accent.”

<sup>6</sup>In fact, texts contained in the TLG contain 177,3 million consonants, 107 million vowels without nonspacing marks and 82,4 million vowels or *rho* with nonspacing marks.

<sup>7</sup>Which could, if not corrected, have terrifying consequences for Greek typography, already in a miserable state due to technical limitations of DTP software.

## 2.2 Unicode and the Mute Iota Issue

### 2.2.1 The facts

Both in ancient and in modern (*katharevousa*) Greek one encounters the grammatical phenomenon of *mute iota*, where a *iota* is written but not pronounced (for example, τῶι Θεῶι is pronounced *tô theô* and not *tôi theôî*).

The *mute iota* can be—and has been—represented in many different ways: by a regular *iota*, a subscript *iota* (in Greek, *ypogegrammeni*), an adscript *iota* (in Greek, *proseggrammeni*), a small capital *iota*, etc. The way of representing the *mute iota* is purely stylistic and hence grammatically irrelevant: Oxford Editions will rather use a regular size *iota*<sup>8</sup> while Association Budé<sup>9</sup> or editions printed in Greece<sup>10</sup> will rather use subscript *iotas*, etc.

### 2.2.2 The author's opinion

By explicitly separating subscript and adscript *iota*, Unicode creates unnecessary confusion: all of these should be called *mute iota* (for example, U+1FF3 GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI should be called U+1FF3 GREEK SMALL LETTER OMEGA WITH MUTE IOTA instead). Rendering engines should then choose the way of explicitly representing letters with mute *iota* (this representation may very well also depend on the context and may vary inside the same document: in a completely uppercased sentence, the *mute iota* can be represented by a small capital *iota*, while in a mixed uppercase/lowercase context, a subscript *iota* can be used for capital letters with mute *iota*. Compare, for example, Ο ΑΙΔΗΣ and ὁ Ἄδης, which can both be encountered in the same document but neither one of them is conformant to Unicode v3.2.

### 2.2.3 Amendment 2a

AMENDMENT of p. 168 of the Unicode book:

~~The nonspacing mark *ypogegrammeni* (also known as *iota subscript* in English) can be applied to the vowels *alpha*, *eta*, and *omega* to represent historic diphthongs. This mark appears as a small *iota* below the vowel. When applied to a single uppercase vowel, the *iota* does not appear as a subscript, but is instead normally rendered as a regular lowercase *iota* to the right of the uppercase vowel. This form of the *iota* is called *proseggrammeni* (also known as *iota adscript* in English). In completely uppercased words, the *iota* subscript should be replaced by a capital *iota*. See *SpecialCasing.txt* on the CD-ROM. Archaic representations of Greek words (which did not have lowercase or accents) use the Greek capital letter *iota* following the vowel for these diphthongs. Such archaic representations require special case mapping.~~

The grammatical phenomenon of the *mute iota* applied to lowercase letters *alpha*, *eta*, and *omega* can be represented by a nonspacing mark *ypogegrammeni* (also known as *iota subscript* in English)

<sup>8</sup>Hart's *Rules for Compositors and Readers at the University Press Oxford* [2, p. 112] says: "The vowels α η ω may occur with the letter iota beneath them (subscript) α η ω; some authors prefer to write α η ω, in which case accents and breathings should be set on the first vowel."

<sup>9</sup>Quotation to be added later.

<sup>10</sup>Quotation to be added later.

or by a regular lowercase *iota*. The same grammatical phenomenon applied to uppercase letters can be represented by a *subscript iota*, or by a regular lowercase *iota*, or as a downsized lowercase *iota* placed to the right of the vowel (and called *prosgegrammeni*, or *iota adscript* in English), or as a small capital *iota*, or as a regular capital *iota*.

Rendering engines should choose between different representations of the *mute iota* phenomenon, based either on user's choice or on the context.

## 2.2.4 Amendment 2b

On pages 494-497 of the Unicode book, informative notes should be added to the following characters:

- U+1F80 GREEK SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI
  - lowercase alpha with smooth breathing and mute iota
- U+1F81 GREEK SMALL LETTER ALPHA WITH DASIA AND YPOGEGRAMMENI
  - lowercase alpha with rough breathing and mute iota
- U+1F82 GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA AND YPOGEGRAMMENI
  - lowercase alpha with smooth breathing and grave accent and mute iota
- U+1F83 GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI
  - lowercase alpha with rough breathing and grave accent and mute iota
- U+1F84 GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA AND YPOGEGRAMMENI
  - lowercase alpha with smooth breathing and acute accent and mute iota
- U+1F85 GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA AND YPOGEGRAMMENI
  - lowercase alpha with rough breathing and acute accent and mute iota
- U+1F86 GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
  - lowercase alpha with smooth breathing and circumflex accent and mute iota
- U+1F87 GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
  - lowercase alpha with rough breathing and circumflex accent and mute iota
- U+1F88 GREEK CAPITAL LETTER ALPHA WITH PSILI AND PROSGEGRAMMENI
  - uppercase alpha with smooth breathing and mute iota
- U+1F89 GREEK CAPITAL LETTER ALPHA WITH DASIA AND PROSGEGRAMMENI
  - uppercase alpha with rough breathing and mute iota
- U+1F8A GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA AND PROSGEGRAMMENI
  - uppercase alpha with smooth breathing and grave accent and mute iota
- U+1F8B GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA AND PROSGEGRAMMENI
  - uppercase alpha with rough breathing and grave accent and mute iota
- U+1F8C GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA AND PROSGEGRAMMENI
  - uppercase alpha with smooth breathing and acute accent and mute iota
- U+1F8D GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA AND PROSGEGRAMMENI
  - uppercase alpha with rough breathing and acute accent and mute iota
- U+1F8E GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
  - uppercase alpha with smooth breathing and circumflex accent and mute iota
- U+1F8F GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
  - uppercase alpha with rough breathing and circumflex accent and mute iota
- U+1F90 GREEK SMALL LETTER ETA WITH PSILI AND YPOGEGRAMMENI
  - lowercase eta with smooth breathing and mute iota
- U+1F91 GREEK SMALL LETTER ETA WITH DASIA AND YPOGEGRAMMENI
  - lowercase eta with rough breathing and mute iota
- U+1F92 GREEK SMALL LETTER ETA WITH PSILI AND VARIA AND YPOGEGRAMMENI
  - lowercase eta with smooth breathing and grave accent and mute iota
- U+1F93 GREEK SMALL LETTER ETA WITH DASIA AND VARIA AND YPOGEGRAMMENI
  - lowercase eta with rough breathing and grave accent and mute iota

- U+1F94 GREEK SMALL LETTER ETA WITH PSILI AND OXIA AND YPOGEGRAMMENI
  - lowercase eta with smooth breathing and acute accent and mute iota
- U+1F95 GREEK SMALL LETTER ETA WITH DASIA AND OXIA AND YPOGEGRAMMENI
  - lowercase eta with rough breathing and acute accent and mute iota
- U+1F96 GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
  - lowercase eta with smooth breathing and circumflex accent and mute iota
- U+1F97 GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
  - lowercase eta with rough breathing and circumflex accent and mute iota
- U+1F98 GREEK CAPITAL LETTER ETA WITH PSILI AND PROSGEGRAMMENI
  - uppercase eta with smooth breathing and mute iota
- U+1F99 GREEK CAPITAL LETTER ETA WITH DASIA AND PROSGEGRAMMENI
  - uppercase eta with rough breathing and mute iota
- U+1F9A GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA AND PROSGEGRAMMENI
  - uppercase eta with smooth breathing and grave accent and mute iota
- U+1F9B GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA AND PROSGEGRAMMENI
  - uppercase eta with rough breathing and grave accent and mute iota
- U+1F9C GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA AND PROSGEGRAMMENI
  - uppercase eta with smooth breathing and acute accent and mute iota
- U+1F9D GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA AND PROSGEGRAMMENI
  - uppercase eta with rough breathing and acute accent and mute iota
- U+1F9E GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
  - uppercase eta with smooth breathing and circumflex accent and mute iota
- U+1F9F GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
  - uppercase eta with rough breathing and circumflex accent and mute iota
- U+1FA0 GREEK SMALL LETTER OMEGA WITH PSILI AND YPOGEGRAMMENI
  - lowercase omega with smooth breathing and mute iota
- U+1FA1 GREEK SMALL LETTER OMEGA WITH DASIA AND YPOGEGRAMMENI
  - lowercase omega with rough breathing and mute iota
- U+1FA2 GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA AND YPOGEGRAMMENI
  - lowercase omega with smooth breathing and grave accent and mute iota
- U+1FA3 GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA AND YPOGEGRAMMENI
  - lowercase omega with rough breathing and grave accent and mute iota
- U+1FA4 GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA AND YPOGEGRAMMENI
  - lowercase omega with smooth breathing and acute accent and mute iota
- U+1FA5 GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA AND YPOGEGRAMMENI
  - lowercase omega with rough breathing and acute accent and mute iota
- U+1FA6 GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
  - lowercase omega with smooth breathing and circumflex accent and mute iota
- U+1FA7 GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
  - lowercase omega with rough breathing and circumflex accent and mute iota
- U+1FA8 GREEK CAPITAL LETTER OMEGA WITH PSILI AND PROSGEGRAMMENI
  - uppercase omega with smooth breathing and mute iota
- U+1FA9 GREEK CAPITAL LETTER OMEGA WITH DASIA AND PROSGEGRAMMENI
  - uppercase omega with rough breathing and mute iota
- U+1FAA GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA AND PROSGEGRAMMENI
  - uppercase omega with smooth breathing and grave accent and mute iota
- U+1FAB GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA AND PROSGEGRAMMENI
  - uppercase omega with rough breathing and grave accent and mute iota
- U+1FAC GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA AND PROSGEGRAMMENI
  - uppercase omega with smooth breathing and acute accent and mute iota
- U+1FAD GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA AND PROSGEGRAMMENI
  - uppercase omega with rough breathing and acute accent and mute iota
- U+1FAE GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI

- uppercase omega with smooth breathing and circumflex accent and mute iota  
U+1FAF GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
- uppercase omega with rough breathing and circumflex accent and mute iota  
U+1FB2 GREEK SMALL LETTER ALPHA WITH VARIA AND YPOGEGRAMMENI
- lowercase alpha with grave accent and mute iota  
U+1FB3 GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI
- lowercase alpha with mute iota  
U+1FB4 GREEK SMALL LETTER ALPHA WITH OXIA AND YPOGEGRAMMENI
- lowercase alpha with acute accent and mute iota  
U+1FB7 GREEK SMALL LETTER ALPHA WITH PERISPOMENI AND YPOGEGRAMMENI
- lowercase alpha with circumflex accent and mute iota  
U+1FBC GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI
- uppercase alpha with mute iota  
U+1FC2 GREEK SMALL LETTER ETA WITH VARIA AND YPOGEGRAMMENI
- lowercase eta with grave accent and mute iota  
U+1FC3 GREEK SMALL LETTER ETA WITH YPOGEGRAMMENI
- lowercase eta with mute iota  
U+1FC4 GREEK SMALL LETTER ETA WITH OXIA AND YPOGEGRAMMENI
- lowercase eta with acute accent and mute iota  
U+1FC7 GREEK SMALL LETTER ETA WITH PERISPOMENI AND YPOGEGRAMMENI
- lowercase eta with circumflex accent and mute iota  
U+1FCC GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI
- uppercase eta with mute iota  
U+1FF2 GREEK SMALL LETTER OMEGA WITH VARIA AND YPOGEGRAMMENI
- lowercase omega with grave accent and mute iota  
U+1FF3 GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI
- lowercase omega with mute iota  
U+1FF4 GREEK SMALL LETTER OMEGA WITH OXIA AND YPOGEGRAMMENI
- lowercase omega with acute accent and mute iota  
U+1FF7 GREEK SMALL LETTER OMEGA WITH PERISPOMENI AND YPOGEGRAMMENI
- lowercase omega with circumflex accent and mute iota  
U+1FFC GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI
- uppercase omega with mute iota

## 2.3 The Shape of the Perispomeni

### 2.3.1 Amendment 3a

Table 7-1, on page 167 of the Unicode book, is as follows:

Table 7-1. Nonspacing Marks Used with Greek

Code	Name	Alternative Names
U+0300	COMBINING GRAVE ACCENT	<i>varia</i>
U+0301	COMBINING ACUTE ACCENT	<i>tonos, oxia</i>
U+0302	COMBINING CIRCUMFLEX ACCENT	
U+0303	COMBINING TILDE	
U+0304	COMBINING MACRON	
U+0306	COMBINING BREVE	
U+0308	COMBINING DIAERESIS	<i>dialytika</i>

AMENDMENT:

Table 7-1. Nonspacing Marks Used with Greek

Code	Name	Alternative Names
U+0300	COMBINING GRAVE ACCENT	<i>varia</i>
U+0301	COMBINING ACUTE ACCENT	<i>tonos, oxia</i>
U+0303	COMBINING TILDE	
U+0304	COMBINING MACRON	
U+0306	COMBINING BREVE	
U+0308	COMBINING DIAERESIS	<i>dialytika</i>
U+0311	COMBINING INVERTED BREVE	

## RATIONAL E:

There is an unfortunate confusion in the Unicode encoding between the *name* “circumflex accent,” and the *shape* of the Greek diacritic *perispomeni*, which, in Greek typography, can be either the one of a tilde, or the one of an inverted breve or the one of a macron, but *never* the one of a circumflex accent, as, for example, on Latin letter ‘ê’. As already stated in [1, p. 11], the choice between these three shapes is purely esthetic (for example, typefaces Monotype Greek 90, 91, 92, Greek Times and New Hellenic use the tilde form, while typefaces Porson and Greek Sans use the inverted breve form, and typeface Linotype Helvetica uses the macron form...).

Therefore we propose the replacement of U+0302 COMBINING CIRCUMFLEX ACCENT by U+0311 COMBINING INVERTED BREVE in the list.

Related to this confusion between circumflex (the name of the accent) and tilde, inverted breve and macron (its possible forms) is also the following amendment:

**2.3.2 Amendment 3b**

AMENDMENT for p. 168 of the Unicode book:

U+0342 COMBINING GREEK PERISPOMENI may appear as either a ~~circumflex or a tilde~~ or an inverted ~~breve or a macron~~. Because of this variation in form, the perispomeni was encoded distinctly from U+0303 COMBINING TILDE.

**2.4 Side-by-side Placement of Accents****2.4.1 Amendment 4**

AMENDMENT for p. 27 (paragraph 2.6) of the Unicode book:

For example, when used with the Greek script, the ~~“breathing marks”~~ smooth breathing U+0313 COMBINING COMMA ABOVE (Psili) and rough breathing U+0314 COMBINING REVERSED COMMA ABOVE (Dasia) require that, when used together with a following acute or grave accent, they be rendered side-by-side above their base letter (when the letter is lower case) or in front of the base letter (when the letter is upper case) rather than the accent marks being stacked above the breathing marks.

## RATIONAL E:

This is simply an omission in the text which we rectify.



## 2.5 Stigma and Digamma

### 2.5.1 Amendment 4

On page 40 and 41 of the Unicode Standard 3.2BETA, it is written:

03DA	Ϛ	GREEK LETTER STIGMA
03DB	ς	GREEK SMALL LETTER STIGMA
03DC	Ϝ	GREEK LETTER DIGAMMA
03DD	ϝ	GREEK SMALL LETTER DIGAMMA
		• used symbolically for numeral six
03DE	Ϙ	GREEK LETTER KOPPA
03DF	ϙ	GREEK SMALL LETTER KOPPA
		• most commonly used with a numeric value (90), as in the dating of legal documentation
03E0	Ϡ	GREEK LETTER SAMPI
03E1	ϡ	GREEK SMALL LETTER SAMPI

#### AMENDMENT:

03DA	Ϛ	GREEK LETTER STIGMA
		• used symbolically for numeral six
03DB	ς	GREEK SMALL LETTER STIGMA
		• used symbolically for numeral six
03DC	Ϝ	GREEK LETTER DIGAMMA
03DD	ϝ	GREEK SMALL LETTER DIGAMMA
03DE	Ϙ	GREEK LETTER KOPPA
		• used symbolically for numeral ninety
03DF	ϙ	GREEK SMALL LETTER KOPPA
		• used symbolically for numeral ninety
03E0	Ϡ	GREEK LETTER SAMPI
		• used symbolically for numeral nine hundred
03E1	ϡ	GREEK SMALL LETTER SAMPI
		• used symbolically for numeral nine hundred

#### RATIONALE:

There is a confusion between alphabetical and numerical archaic letters: letters *stigma* Ϛ Ϛ, *koppa* Ϙ Ϙ and *sampi* Ϡ ϡ are clearly numerical, since they are used even nowadays for numbers 6, 90 and 900. On the other hand, letters *archaic koppa* Ϙ Ϙ and *digamma* Ϝ ϝ are clearly alphabetic, and are used nowadays for the transcription of ancient inscriptions.

This is the current usage of these letters; of course an epigraphologist may always find an inscription where *archaic koppa* or *digamma* represent numbers, or inversely *koppa* and *sampi* are used alphabetically, but this would be an exception rather than the rule.

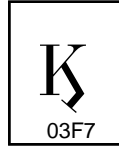
Additional confusion is generated from the fact that *stigma* has been used until two centuries ago in printed and handwritten text as a *sigma-tau* ligature (and in fact, in modern Greece, the letters στ used together represent number six (although σ by itself is 200 and τ is 300)).

It would be contrary to the principles of Unicode to use U+03DA GREEK LETTER STIGMA and U+03DB GREEK SMALL LETTER STIGMA as *sigma-tau* ligatures, that is why it is especially important to qualify them as being numerics.

### 3 Additions to Unicode v3.2

We propose the following additions to the Greek tables of Unicode v3.2. At the end of this section, the reader will find the Unicode v3.2 tables with the characters of this proposal inserted.

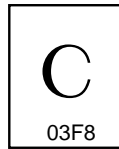
#### 3.1 Uppercase Versions of kai and lunate sigma



U+03F7 GREEK CAPITAL KAI SYMBOL

**RATIONALE:**

In an uppercased-only context the U+03D7 GREEK KAI SYMBOL character also gets uppercased. Therefore it is necessary to include its uppercase version as an additional character in the Unicode standard.

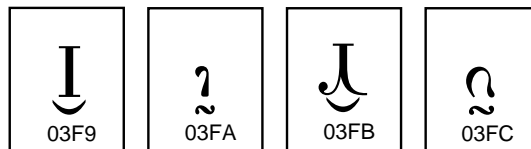


U+03F8 GREEK CAPITAL LUNATE SIGMA SYMBOL

**RATIONALE:**

The lunate sigma U+03F2 GREEK LUNATE SIGMA SYMBOL, which is actually not a “symbol” (in the sense that it is used for Greek text only and *not* for mathematics), but the traditional way of representing a *context-free sigma* (neither initial nor final), requires an uppercase version.<sup>11</sup> Therefore it is necessary to include its uppercase version as an additional character in the Unicode standard.

#### 3.2 Reversed Iota and Upsilon Vowels



<sup>11</sup>Although one may argue that there is no difference between initial and final uppercase sigma, traditionally when typesetting Greek text with lunate lowercase sigmas, one uses also a lunate sigma for uppercase.

ἦταν γῖὰ σένα πειρὸ καλὰ  
δρὸ ἦλιους ν' ἀποχτήσης

Figure 6: A sentence with some occurrences of U+03FA GREEK SMALL REVERSED LETTER IOTA WITH TILDE BELOW and U+03FC GREEK SMALL REVERSED LETTER UPSILON WITH TILDE BELOW taken from a popular theatrical play of the beginning of 20th century.

U+03F9 GREEK CAPITAL REVERSED LETTER IOTA WITH BREVE BELOW

- Used in late 19th/early 20th century for semi-consonantic vowels

≈ 0399 032E Greek capital letter iota with combining breve below

U+03FA GREEK SMALL REVERSED LETTER IOTA WITH TILDE BELOW

≈ 03B9 032E Greek small letter iota with combining breve below

U+03FB GREEK CAPITAL REVERSED LETTER UPSILON WITH BREVE BELOW

≈ 03A5 032E Greek capital letter upsilon with combining breve below

U+03FC GREEK SMALL REVERSED LETTER UPSILON WITH TILDE BELOW

≈ 03C5 032E Greek small letter upsilon with combining breve below

#### RATIONALE:

These letters have been used very abundantly at the end of the 19th and at the beginning of the 20th century to represent the semi-consonantic sound of “yi” in *dimotiki* texts. Later on, when the spelling of *dimotiki* has been standardized, they have been replaced by non-reversed *iota* and *upsilon* (so that there is no graphical distinction anymore between vowel “i” and semi-consonant “yi” (as for example in *ποιός* pronounced “pîos” and *ποιός* pronounced “pyiós”). (See also fig. 6.)

To encode documents of that period correctly, these characters are necessary to be included in the Unicode standard.

Note that the fact that the base letters are *reversed* excludes the alternative of using combining diacritics to obtain them.

Also note that semi-consonantic vowels, by their nature, never get accented and are never located at the beginning of words, hence it is *not* necessary to provide any combination of these with the usual Greek accents and breathings.

The diacritics breve or tilde used in these constructions are both called “hyphen.” Sometimes a breve is used also for the lowercase reversed *iota* and *upsilon*, but obviously it was easier for printers to take the already existing *iota* with *perispomeni* and simply revert it. Since it is a typographical tradition to reverse letters, we are keeping it in our proposal; however, as a compatibility equivalent we propose non-reversed letters with combining breve below (the latter representing the “hyphen”).

### 3.3 Letters Epsilon And Omicron With Perispomeni Accent

ϵ̅̅̅	ϵ̅̅̅̅	ο̅̅̅	ο̅̅̅̅	Ε̅̅̅	Ε̅̅̅̅	Ο̅̅̅	Ο̅̅̅̅	ϵ̅̅̅̅̅	ο̅̅̅̅̅
1F16	1F17	1F46	1F47	1F1E	1F1F	1F4E	1F4F	1FB5	1FC5

- U+1F16 GREEK SMALL LETTER EPSILON WITH PSILI AND PERISPOMENI  
 ≡ 03B5 0313 0342
- U+1F17 GREEK SMALL LETTER EPSILON WITH DASIA AND PERISPOMENI  
 ≡ 03B5 0314 0342
- U+1F1E GREEK CAPITAL LETTER EPSILON WITH PSILI AND PERISPOMENI  
 ≡ 0395 0313 0342
- U+1F1F GREEK CAPITAL LETTER EPSILON WITH DASIA AND PERISPOMENI  
 ≡ 0395 0314 0342
- U+1F46 GREEK SMALL LETTER OMICRON WITH PSILI AND PERISPOMENI  
 ≡ 03BF 0313 0342
- U+1F47 GREEK SMALL LETTER OMICRON WITH DASIA AND PERISPOMENI  
 ≡ 03BF 0314 0342
- U+1F4E GREEK CAPITAL LETTER OMICRON WITH PSILI AND PERISPOMENI  
 ≡ 039F 0313 0342
- U+1F4F GREEK CAPITAL LETTER OMICRON WITH DASIA AND PERISPOMENI  
 ≡ 039F 0314 0342
- U+1FB5 GREEK SMALL LETTER EPSILON WITH PERISPOMENI  
 ≡ 03B5 0342
- U+1FC5 GREEK SMALL LETTER OMICRON WITH PERISPOMENI  
 ≡ 03BF 0342

#### RATIONALE:

Grammar rules for modern and standardized ancient Greek do not allow the *perispomeni* accent to be placed on *epsilon* and *omicron* letters (because these letters are *short* vowels, contrarily to *eta* and *omega*, which are *long* ones). This explains why the corresponding Unicode positions have not been used.

In fact, letters *eta* and *omega* have been introduced in Athens in 403 BC (see [1, p. 3]) as long counterparts of *epsilon* and *omicron*. Before the introduction of *eta* and *omega*, both the short and the long versions of *epsilon* and *omicron* were written as *epsilon* and *omicron*.

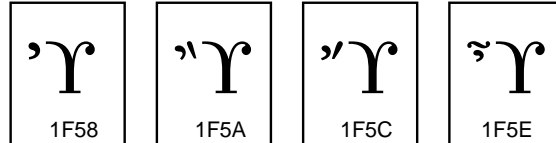
When transcribing epigraphical texts which contain long *epsilon* and *omicron* letters, scholars very often use the *perispomeni* accent upon these letters (as if they were *eta* or *omega*). For that reason, it becomes necessary to introduce pre-composed versions of *epsilon* and *omicron* with all combinations of *perispomeni* and breathings.

One may argue that these combinations of vowels and diacritics can be obtained by *combining diacritics*. This is true, *but* since the other combinations of vowels, accents and breathings all exist in pre-composed form, software designers usually do not implement

composition of Greek vowels and diacritics (since, as they argue, there is no explicit need for it).

Under these conditions it is impossible for scholars to access the “forbidden” combinations (*epsilon* and *omicron* with *perispomeni* and eventually breathings). Therefore, and also for reasons of consistency with the existing pre-composed characters, it is necessary to explicitly add these pre-composed characters to the table, at the positions which have been left empty in prevision of such an addition.

### 3.4 Letters Capital Upsilon With Smooth Breathing



U+1F58 GREEK CAPITAL LETTER UPSILON WITH PSILI

≡ 03A5 0313

U+1F5A GREEK CAPITAL LETTER UPSILON WITH PSILI AND VARIA

≡ 03A5 0313 0300

U+1F5C GREEK CAPITAL LETTER UPSILON WITH PSILI AND OXIA

≡ 03A5 0313 0301

U+1F5E GREEK CAPITAL LETTER UPSILON WITH PSILI AND PERISPOMENI

≡ 03A5 0313 0342

#### RATIONALE:

This case is quite similar to the preceding one: grammar rules normally prohibit a smooth breathing on an initial letter *upsilon*, and hence it is normally impossible to obtain a smooth breathing on a capital letter *upsilon* (which is necessarily initial, otherwise it would take no breathing in the first place).

Once again, rules of grammar do not always apply: in ancient Greek dialects (see for example [5]) one encounters quite often initial letters *upsilon* with smooth breathings.

The same argument as in the previous subsection applies in favour of adding pre-composed versions of these characters instead of relying on rendering engines to do composition of vowels and combining diacritics. Therefore, it is necessary to explicitly add these pre-composed characters to the table, at the positions which have been left empty in prevision of such an addition.

## References

- [1] Haralambous, Yannis, *From Unicode to Typography, a Case Study: the Greek Script*, Proceedings of the 14th International Unicode Conference, Boston, MA, 1999.

- [2] Hart, Horace, *Hart's Rules for Compositors and Readers at the University Press Oxford*, 39th ed., Oxford University Press, 1996.
- [3] Ἐφημερὶς τῆς Κυβερνήσεως, 11 Φεβρουαρίου 1982, τεύχος πρῶτον, ἀριθμὸς φύλλου 15, σελ. 112-113, Νόμος 1228/1982: *Κύρωση τῆς ἀπὸ 11.11.1981 πράξης τοῦ Προέδρου τῆς Δημοκρατίας περὶ ἐγγραφῆς μαθητῶν στὰ Λύκεια τῆς Γενικῆς καὶ Τεχνικῆς καὶ Ἐπαγγελματικῆς Ἐκπαιδεύσεως.*
- [4] Ἐφημερὶς τῆς Κυβερνήσεως, 29 Ἀπριλίου 1982, τεύχος πρῶτον, ἀριθμὸς φύλλου 52, σελ. 429-430, Διάταγμα 297: *Ἐφαρμογὴ τοῦ μονοτονικοῦ συστήματος στὴν Ἐκπαίδευση καὶ στὴ Διοίκηση.*
- [5] Buck, Carl Darling, *The Greek dialects; grammar, selected inscriptions, glossary*, University of Chicago Press, 1955.

0370

Greek and Coptic



03FF

	037	038	039	03A	03B	03C	03D	03E	03F
0			ι 0390	Π 03A0	ϛ 03B0	π 03C0	ϛ 03D0	Λ 03E0	κ 03F0
1			Α 0391	Ρ 03A1	α 03B1	ρ 03C1	ϛ 03D1	ϛ 03E1	ϛ 03F1
2			Β 0392		β 03B2	ς 03C2	Υ 03D2	ϛ 03E2	ϛ 03F2
3			Γ 0393	Σ 03A3	γ 03B3	σ 03C3	Υ 03D3	ϛ 03E3	ϛ 03F3
4	' 0374	' 0384	Δ 0394	Τ 03A4	δ 03B4	τ 03C4	Υ 03D4	ϛ 03E4	Θ 03F4
5	' 0375	' 0385	Ε 0395	Υ 03A5	ε 03B5	υ 03C5	φ 03D5	ϛ 03E5	Ε 03F5
6		'Α 0386	Ζ 0396	Φ 03A6	ζ 03B6	φ 03C6	ϛ 03D6	ϛ 03E6	ϛ 03F6
7		' 0387	Η 0397	Χ 03A7	η 03B7	χ 03C7	ϛ 03D7	ϛ 03E7	Κ 03F7
8		'Ε 0388	Θ 0398	Ψ 03A8	θ 03B8	ψ 03C8	ϛ 03D8	ϛ 03E8	ϛ 03F8
9		'Η 0389	Ι 0399	Ω 03A9	ι 03B9	ω 03C9	ϛ 03D9	ϛ 03E9	ϛ 03F9
A	' 037A	'Ι 038A	Κ 039A	Ϊ 03AA	κ 03BA	ϊ 03CA	ϛ 03DA	ϛ 03EA	ϛ 03FA
B			Λ 039B	Υ 03AB	λ 03BB	ϛ 03CB	ς 03DB	ϛ 03EB	ϛ 03FB
C		'Ο 038C	Μ 039C	ά 03AC	μ 03BC	ό 03CC	ϛ 03DC	ϛ 03EC	ϛ 03FC
D			Ν 039D	έ 03AD	ν 03BD	ύ 03CD	ϛ 03DD	ϛ 03ED	
E	; 037E	Υ 038E	Ξ 039E	ή 03AE	ξ 03BE	ώ 03CE	ϛ 03DE	ϛ 03EE	
F		'Ω 038F	Ο 039F	ί 03AF	ο 03BF		ϛ 03DF	ϛ 03EF	

1F00

Greek Extended

1F7F

	1F0	1F1	1F2	1F3	1F4	1F5	1F6	1F7
0	ἄ 1F00	ε̇ 1F10	ῆ 1F20	ι̇ 1F30	ο̇ 1F40	υ̇ 1F50	ώ 1F60	ὰ 1F70
1	ἄ 1F01	ε̇ 1F11	ῆ 1F21	ι̇ 1F31	ο̇ 1F41	υ̇ 1F51	ώ 1F61	ά 1F71
2	ἄ 1F02	ε̇ 1F12	ῆ 1F22	ι̇ 1F32	ο̇ 1F42	υ̇ 1F52	ώ 1F62	ὲ 1F72
3	ἄ 1F03	ε̇ 1F13	ῆ 1F23	ι̇ 1F33	ο̇ 1F43	υ̇ 1F53	ώ 1F63	έ 1F73
4	ἄ 1F04	ε̇ 1F14	ῆ 1F24	ι̇ 1F34	ο̇ 1F44	υ̇ 1F54	ώ 1F64	ῆ 1F74
5	ἄ 1F05	ε̇ 1F15	ῆ 1F25	ι̇ 1F35	ο̇ 1F45	υ̇ 1F55	ώ 1F65	ῆ 1F75
6	ἄ 1F06	ε̇ 1F16	ῆ 1F26	ι̇ 1F36	ο̇ 1F46	υ̇ 1F56	ώ 1F66	ι̇ 1F76
7	ἄ 1F07	ε̇ 1F17	ῆ 1F27	ι̇ 1F37	ο̇ 1F47	υ̇ 1F57	ώ 1F67	ι̇ 1F77
8	᾿Α 1F08	᾿Ε 1F18	᾿Η 1F28	᾿Ι 1F38	᾿Ο 1F48	᾿Υ 1F58	᾿Ω 1F68	ὀ 1F78
9	᾿Α 1F09	᾿Ε 1F19	᾿Η 1F29	᾿Ι 1F39	᾿Ο 1F49	᾿Υ 1F59	᾿Ω 1F69	ό 1F79
A	᾿Α 1F0A	᾿Ε 1F1A	᾿Η 1F2A	᾿Ι 1F3A	᾿Ο 1F4A	᾿Υ 1F5A	᾿Ω 1F6A	ὐ 1F7A
B	᾿Α 1F0B	᾿Ε 1F1B	᾿Η 1F2B	᾿Ι 1F3B	᾿Ο 1F4B	᾿Υ 1F5B	᾿Ω 1F6B	ύ 1F7B
C	᾿Α 1F0C	᾿Ε 1F1C	᾿Η 1F2C	᾿Ι 1F3C	᾿Ο 1F4C	᾿Υ 1F5C	᾿Ω 1F6C	ώ 1F7C
D	᾿Α 1F0D	᾿Ε 1F1D	᾿Η 1F2D	᾿Ι 1F3D	᾿Ο 1F4D	᾿Υ 1F5D	᾿Ω 1F6D	ὠ 1F7D
E	᾿Α 1F0E	᾿Ε 1F1E	᾿Η 1F2E	᾿Ι 1F3E	᾿Ο 1F4E	᾿Υ 1F5E	᾿Ω 1F6E	
F	᾿Α 1F0F	᾿Ε 1F1F	᾿Η 1F2F	᾿Ι 1F3F	᾿Ο 1F4F	᾿Υ 1F5F	᾿Ω 1F6F	



1F80

Greek Extended

1FFF

	1F8	1F9	1FA	1FB	1FC	1FD	1FE	1FF
0	ἄ 1F80	ἥ 1F90	ὦ 1FA0	ἄ̃ 1FB0	~ 1FC0	ϊ̃ 1FD0	ϋ̃ 1FE0	
1	ἄ̄ 1F81	ἥ̄ 1F91	ὦ̄ 1FA1	ἄ̄̃ 1FB1	~̄ 1FC1	ϊ̄̃ 1FD1	ϋ̄̃ 1FE1	
2	ἄ̅ 1F82	ἥ̅ 1F92	ὦ̅ 1FA2	ἄ̅̃ 1FB2	ἥ̅̃ 1FC2	ϊ̅̃ 1FD2	ϋ̅̃ 1FE2	ὦ̅̃ 1FF2
3	ἄ̆ 1F83	ἥ̆ 1F93	ὦ̆ 1FA3	ἄ̆̃ 1FB3	ἥ̆̃ 1FC3	ϊ̆̃ 1FD3	ϋ̆̃ 1FE3	ὦ̆̃ 1FF3
4	ἄ̇ 1F84	ἥ̇ 1F94	ὦ̇ 1FA4	ἄ̇̃ 1FB4	ἥ̇̃ 1FC4		ρ̇̃ 1FE4	ὦ̇̃ 1FF4
5	ἄ̈ 1F85	ἥ̈ 1F95	ὦ̈ 1FA5	ἄ̈̃ 1FB5	ἥ̈̃ 1FC5		ρ̈̃ 1FE5	
6	ἄ̉ 1F86	ἥ̉ 1F96	ὦ̉ 1FA6	ἄ̉̃ 1FB6	ἥ̉̃ 1FC6	ϊ̉̃ 1FD6	ϋ̉̃ 1FE6	ὦ̉̃ 1FF6
7	ἄ̊ 1F87	ἥ̊ 1F97	ὦ̊ 1FA7	ἄ̊̃ 1FB7	ἥ̊̃ 1FC7	ϊ̊̃ 1FD7	ϋ̊̃ 1FE7	ὦ̊̃ 1FF7
8	Ἄ̇ 1F88	Ἡ̇ 1F98	Ὠ̇ 1FA8	Ἄ̇̃ 1FB8	Ἐ̇̃ 1FC8	Ἰ̇̃ 1FD8	Υ̇̃ 1FE8	Ὢ̇̃ 1FF8
9	Ἄ̈ 1F89	Ἡ̈ 1F99	Ὠ̈ 1FA9	Ἄ̈̃ 1FB9	Ἐ̈̃ 1FC9	Ἰ̈̃ 1FD9	Ϋ̃ 1FE9	Ὢ̈̃ 1FF9
A	Ἄ̉ 1F8A	Ἡ̉ 1F9A	Ὠ̉ 1FAA	Ἄ̉̃ 1FBA	Ἐ̉̃ 1FCA	Ἰ̉̃ 1FDA	Υ̉̃ 1FEA	Ὢ̉̃ 1FFA
B	Ἄ̊ 1F8B	Ἡ̊ 1F9B	Ὠ̊ 1FAB	Ἄ̊̃ 1FBB	Ἐ̊̃ 1FCB	Ἰ̊̃ 1FDB	Υ̊̃ 1FEB	Ὢ̊̃ 1FFB
C	Ἄ̋ 1F8C	Ἡ̋ 1F9C	Ὠ̋ 1FAC	Ἄ̋̃ 1FBC	Ἐ̋̃ 1FCC		ἲ̋̃ 1FEC	Ὢ̋̃ 1FFC
D	Ἄ̌ 1F8D	Ἡ̌ 1F9D	Ὠ̌ 1FAD	Ἄ̌̃ 1FBD	Ἐ̌̃ 1FCD	Ἰ̌̃ 1FDD	Υ̌̃ 1FED	Ὢ̌̃ 1FFD
E	Ἄ̍ 1F8E	Ἡ̍ 1F9E	Ὠ̍ 1FAE	Ἄ̍̃ 1FBE	Ἐ̍̃ 1FCE	Ἰ̍̃ 1FDE	Υ̍̃ 1FEE	Ὢ̍̃ 1FFE
F	Ἄ̎ 1F8F	Ἡ̎ 1F9F	Ὠ̎ 1FAF	Ἄ̎̃ 1FBF	Ἐ̎̃ 1FCF	Ἰ̎̃ 1FDF	Υ̎̃ 1FEF	