

Gentium - Greek Encoding Details

Common ancient Greek letter+diacritic combinations have been supported in Gentium since its first release. Version 7 extends this support to less common and even very rare combinations with much more robust OpenType handling.

Most OS and application environments handle common combinations well, as the text often uses the large set of precomposed character clusters included in Unicode (e.g. ᾗ U+1F83 GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI). Complex combinations that do not have corresponding precomposed characters (e.g. alpha + macron + dasia) can cause more difficulty. These require that the characters are in a precise order that follows Unicode ordering - see section 2.11 of [The Unicode Standard](#).

Each cluster consists of a base letter followed by one or more diacritics and an optional iota, and must be in the following order. Note that this is not necessarily the order in which they are commonly written from left to right, especially in the case of capitals.

 base + [length] + [breathing|diaeresis] + [accent] + [circ] + [iota]

Groups:

- base: all alphabetic characters with no additional diacritics
- length: U+0304 macron; U+0306 vrachy (breve)
- breathing|diaeresis: U+0313 psili (smooth, comma above); U+0314 dasia (rough, reversed comma above); U+0308 dialytika (diaeresis)
 - U+0343 koronis is also in this group but is not recommended
- accent: U+0300 varia (grave); U+0301 oxia, tonos (acute)
 - U+0344 dial+tonos is also in this group but is not recommended
- circumflex: U+0342 perispomeni
- iota: U+0345 ypogegrammeni

This order also assumes that the clusters are fully decomposed (separated) into their individual characters. You cannot assume that you can add a diacritic to a precomposed cluster. For example, you cannot reliably construct U+1F04 GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA with the sequence U+1F00 GREEK SMALL LETTER ALPHA WITH PSILI + U+0301 COMBINING ACUTE ACCENT. This may work in some environments but not others.

Gentium version 7 does, however, provide specific support to hybrid precomposed+diacritic combinations that include U+0304 COMBINING MACRON. This increases the likelihood that these combinations will render properly:

 (precomposed base with length) + [breathing|diaeresis] + [accent]

Example:

U+1FD1 GREEK SMALL LETTER IOTA WITH MACRON + U+0314 COMBINING REVERSED COMMA ABOVE + U+0301 COMBINING ACUTE

We still strongly recommend using the normal decomposed sequence for this cluster:

U+13B9 GREEK SMALL LETTER IOTA + U+0304 COMBINING MACRON + U+0314 COMBINING REVERSED COMMA ABOVE + U+0301 COMBINING ACUTE

Properly encoding and rendering complex sequences depends on many factors: the software keyboard that turns keystrokes into characters, internal operating systems routines that handle Unicode composition and decomposition, OpenType rules in the font, and the rendering engine used by the application. If a sequence is not rendering well the problem could be with any one or more of these factors. Contact us if you have problems, and tell us details of the operating system, application, keyboard, and font version. Please also include the exact character sequence you're using and a screenshot of the problem.

