All authors publishing with Dumbarton Oaks are required to submit text that is Unicode compliant. This policy applies to the submission of any text in any language, not simply polytonic Greek. For some authors, particularly those who work only in western European languages, this poses no problem, since the Latin alphabet has been de facto Unicode compliant for decades. For those of you who use Greek, however, the prospect of creating Unicode-compliant Greek may be daunting. You may ask, why convert, especially if a particular font has suited your needs for some time. Given the frequency with which computing changes, it is sensible to wonder how long this standard will endure, and how complicated it is to configure your computer.

This guide is intended to answer these questions, by explaining the importance and benefit of this standard, and by providing instructions on how to set up your computer to be Unicode compliant. Only the most essential information about Unicode Greek is presented here. Suggestions for further reading can be found at the end of this guide.

What is Unicode Greek?

To understand the need for Unicode Greek it helps to understand something of the development of codes and the telegraph. The earliest telegraphs transmitted and received electronic pulses, which were transcribed onto long paper strips. As the voltage changed, either a pen wiggled across the paper, or a set of needles or stamps punctured or indented it, creating a visual pattern representing the message. (The audible signal we associate with telegraphs was a later development.) To interpret the message both sender and recipient needed to use the same code, listing the assignments of pulses to letters or numbers. These code tables were the prototypes of the code tables used in computers, where electrical impulses represent characters and symbols.

The initial codes consisted of a limited set of characters. International Morse Code, for instance, has only fifty-one characters: the twenty-six letters of the English alphabet (assumed to be uppercase), the ten digits, and fifteen signs of punctuation. All the earliest telegraph codes had a limited number of characters. It was more efficient to have a small character set than an expansive one, since it reduced the number of dots and dashes, and therefore the time, needed to encode, transmit, and decode a message.

As technology developed so did communication needs, and the telegraph prototypes gave way to more expansive code tables that included both upper and lowercase letters, as well as commands, such as "carriage return" or "end of transmission," meant to instruct the receiving device about the format or shape of the transmission. At first there was little uniformity in the development of new codes. The number and variety of coding systems developed over the first century of telecommunications compelled the International Organization of Standards (ISO) to develop a single standard for coding telecommunications. Thus was born in the 1960s the American Standard Code for Information Interchange (ASCII), a 128-character code that included upper and lowercase letters, the digits, standard punctuation, and commonly used command codes.

Since computers work on a principle similar to telegraphs, it was natural for the earliest computers to take advantage of these standardized character sets. Consequently IBM and Apple each made ASCII the basis for the character sets of their new computers. It was recognized, however, that the 128-character set excluded a number of other letters, symbols, and commands that could and should be represented. Both IBM and Apple developed a 256-character set, but, because the two systems were developed independently, the upper 128 characters did not correspond to each other. Thus, the initial attempts to expand the basic character set were marked by inconsistency.

The character sets computers used through the mid-1990s catered almost exclusively to the Latin alphabet. Both PC and Macintosh upper character sets (spaces 128–255) assigned some slots to Greek

letters, but these were intended to serve, not speakers of modern Greek or scholars dealing with ancient and Byzantine Greek texts, but mathematicians who needed to write equations. Thus, throughout the 1980s and 1990s anyone who wanted to use computers to work with Greek or other alphabets had to invent creative ways of getting around the Latin alphabet.

Most often, the way to circumvent the problem was simply to change the font. By all appearances, this seemed to imitate what had been done in printing in the past. In handset or letterpress printing, several languages in a single font—here, literally a set of lead blocks or molds that were inked and pressed onto the paper—could be mixed at will. In computers, these physical blocks are replaced, not by the computer's printer, as might be expected, but by computer code. This code is a description of a typeface, a way of presenting a particular form of a letter, number, or other character on the computer screen or the printed page. So the earliest method for working with Greek text on a computer was to use either the Macintosh or IBM 256-character set, but with a specially created Greek font. In this scenario, when Latin letters were typed, what appeared was a Greek letter the font designer assigned to that place in the character set. For instance, by pressing the ell key, a lambda would appear. The underlying data was still an English letter (in this case, ell), but it looked Greek (that is, like a lambda).

This technique, although still used widely today, has for several reasons been less than satisfactory.

First, the assignment of the various letters of the Latin alphabet to that of the Greek has been capricious. For example, some fonts assign to chi the x, others, the c. Some fonts use the left parenthesis for a rough breathing, others, the J. Some fonts assign precomposed combinations of vowels and their diacritical marks to arbitrary, hard-to-remember places in the upper character set. Other fonts split vowels and their individual accents. Other examples are legion. In the end, every font of this sort requires the user to learn a new keyboard configuration.

Second, like other types of intellectual property, fonts are copyrighted and cannot be shared without the designer's approval. Many of the best fonts are expensive, a reflection of how much time and effort went into creating the font. Anyone who wishes to share a Greek text with someone else requires the recipient to have the same font. The recipient, then, is placed in the awkward position of having either to break the law or buy the font.

Third, it is difficult or impossible, because of accents and breathings, to search through text using conventional word processors or internet browsers. If you are searching for ἤδη (ēdē) do you enter the key combination hdh, hjvdh, hvjdh, h[dh, ≥dh, or some other combination? You must know exactly how that particular font's architecture works, and hope that the word was keyed in correctly. That same search routine will probably not work on other fonts.

Fourth, even the most beautifully designed Greek font might produce shoddy typography. For instance, a rough breathing may be correctly centered over a lowercase eta, but that same breathing over an iota will be off center. An iota subscript properly centered under an omega will be incorrectly positioned under an eta, which takes the iota subscript under its left leg. For many scholars such finesse has not been a concern. But for publishers who value their craft, this has contributed to erosion of standards in typography and the art of bookmaking.

Fifth, some fonts are incomplete in their repertoire of extra nonalphabetic characters. For instance, a font may lack either an obelisk or the Greek numeral for six. This poses a problem for many scholars who deal with Greek texts. If the primary font does not have certain characters, then the editor is forced to use two or more Greek fonts, producing the potential for confusion, and awkward or ugly typography.

Sixth, exchanging information between Apple or Macintosh and IBM-compatible PC systems proves an insoluble problem. Even if both users have copies of the same font, one for the PC and the other for the Macintosh, this does not guarantee seamless transmission. If the font assigns important glyphs to the upper 128 characters, then those glyphs will be lost in transmission, since PC and Macintosh have completely different assignments to the upper character set.

These are very serious problems. And for every challenge Greek poses, there are dozens more in other alphabets and languages, such as Chinese, Arabic, and Tibetan.

Acknowledging all these problems, computer developers at Xerox and Apple in the late 1980s began to work on solutions, beginning with efforts to standardize Han Chinese. Other corporations, such as Sun, Adobe, Microsoft, and IBM, joined the effort to develop a common standard and in 1991 founded the Unicode Consortium, the non-profit organization that supports and develops the Unicode Standard. Their intent was and is simple, to develop a character code that addresses every need in every language, a code that is universal, uniform, and unique (hence the name).

Unicode is universal in that it addresses the computing needs of all the world's languages. The initial versions of Unicode had a character map consisting of 65,536 (2^{16}; hence called 16 bit) code points, but this was later enlarged to 1,114,112 (2^{32}; i.e., 32-bit code) points, to accommodate all possible scripts, living and historical. Under this plan each and every character in the world's writing systems can be assigned a unique, unambiguous code point.

Furthermore, Unicode is efficient, in that it has built into its structure a set of equivalencies. That is, it establishes rules that tell the computer, for instance, that the two keystrokes alpha and rough breathing is the same as a single keystroke containing the precomposed glyph ἁ. There is some flexibility in the way Greek text is entered into the computer; the result is the same. Thus, any software that takes full advantage of Unicode will allow you to search for Greek text more easily and accurately than older software did with non-Unicode Greek fonts.

There are a number of other advantages and caveats to Unicode that require a more technical explanation, beyond the immediate scope of this presentation. In sum, Unicode makes possible for Greek texts beautiful typography, a complete set of characters, accurate searching, ease of typing, and a seamless exchange between users without any loss of information.

Will Unicode become obsolete any time soon? No. This question is asked mainly because it seems that the computing industry is abandoning one standard for another. But this is not the case. Unicode has not rendered the older system, ASCII, obsolete. ASCII is still an industry standard, but it is now a proper subset of Unicode. What is being rendered obsolete is the incorrect use of ASCII. Scholars have been trying to use ASCII to do things for which it was never designed. Unicode has been designed specifically to address the needs of those who have had to make ASCII do what it was not intended to do.

In the future Unicode will expand (it is currently at version 4.1), but it will preserve all the standards that have already been established. This growth involves the inclusion of new linguistic blocks, as the Unicode Consortium develops Unicode to serve the needs of the entire world. It will be important for some Byzantinists to keep track of these developments, since new symbols are occasionally introduced for inclusion in Unicode. (Dumbarton Oaks is currently developing just such a proposal for numismatics and sigillography. For more information about this project contact the publications office.)

Preparing your system to handle Unicode Greek

Unicode is merely a code, a conceptual infrastructure. To take advantage of it, you must ensure that the hardware and software you are running is Unicode compliant. Hardware normally is not a problem.

Any Macintosh or IBM-compatible computer that runs Unicode-compliant software is itself Unicode compliant. The software is the most critical aspect. There are four different areas that must be attended to: the operating system, the fonts, the keyboard driver, and the word processor.

Operating system. The operating system is the software your computer initially installs, allowing you to run other software. Macintosh users must be running OS X. Previous versions (OS 9 and below) will not work. PC users must have Windows CE, Windows NT 4.0, Windows 2000, or Windows XP. Other operating systems (such as Windows 98) will not work. If you have an old computer that cannot run this software, then it is time for you to upgrade to a new system.

Fonts. As already mentioned, fonts are sets of instructions for the proper display or printout of a particular typeface. Different fonts are designed with different numbers of characters. Unicode makes it possible (but not necessary) for a font to have both the Latin alphabet and all the polytonic Greek characters. If you are running Windows XP/2000 or Macintosh's OS X, you already have Unicode-compliant fonts that reflect this capability: Lucida Sans, Palatino Linotype, Tahoma, and Arial Unicode MS (all PC) as well as Lucida Grande (Mac). The most recent version of Macintosh's operating system, Tiger (OS 10.4), includes two new preinstalled fonts with polytonic Greek characters: Helvetica and Times.

There are many other Unicode-compliant Greek fonts available, many for free. One we have traditionally used for our publications, Kadmos (now renamed KadmosU), is available free for download through the generosity of the original proprietor, Marc Cogan: http://socrates.berkeley.edu/~pinax/greekkeys/GreekKeys.html At the end of this document are links that will show you how to obtain many other Unicode Greek fonts.

In the past, Macintosh and PC fonts could not be interchanged. With the latest operating systems this difficulty is gone. Any font that works on a Mac (OS X) will now work on a PC (Windows XP), and vice versa, thanks to a new format of font called OpenType, which is completely Unicode compliant. If you are working with an OpenType font you are using Unicode. If you are using a TrueType font or PostScript font, you might not be working with a Unicode-compliant font. If you are uncertain if the font you are using is Unicode compliant and want to check, contact the publications department.

Keyboard driver. Although Unicode has expanded its character set to over one million characters, our keyboards still have little more than a hundred keys. In the days of the 256-character set, this was convenient, since every character was only one or two keystrokes away. (For instance, a capital A needs two keys: the shift and the a.) In the age of Unicode, where dozens of different languages have their own text blocks, and some languages, such as Han Chinese, have tens of thousands of individual characters, the use of the keyboard is not as straightforward.

Every computer uses a keyboard driver to interpret the keys being typed. Keyboard drivers are built into the operating system, and are selected by the user to tell the computer what language block should be typed, and even what physical keyboard is being used. (German keyboards, for instance, switch the y and the z; other national keyboards, such as Russian, bear little resemblance to the Qwerty arrangement familiar to many computer users.)

Macintosh OS X includes dozens of possible keyboard drivers as a standard part of the operating system. Unfortunately, prior to OS 10.4 (Tiger), polytonic Greek was not one of the options. For those not yet running Tiger (and those who run it, but are unhappy with the configuration of its polytonic Greek keyboard), third-party software is required to tell the computer that you want to type in polytonic Greek. This same software provides its own method of what keystrokes correspond to what letters, accents, or breathing marks. We recommend SophoKeys, a free keyboard driver. The instructions

on how to download and install this file are provided here: http://www.apple.com/downloads/macosx/system_disk_utilities/sophokeys.html.

For Windows XP, there is a Polytonic Greek keyboard driver already built into the operating system, but you must activate it yourself. There are very detailed instructions with pictures provided by Microsoft here: http://www.microsoft.com/globaldev/wrg_redirect.asp?URL=polytonic. Unfortunately, this driver's assignment of letters and diacriticals is difficult for many people to memorize and use. You may find this free keyboard driver easier and more intuitive: http://members.aol.com/AtticGreek/. Other Windows keyboards are listed here: http://socrates.berkeley.edu/~pinax/greekkeys/WindowsFAQ.html#windowsUnicode.

One of the benefits of using some of these keyboards is that they are context sensitive. That is, the drivers can convert key combinations into single glyphs on the fly, and determine what sort of shape a character should take based on position. For instance, an alpha will change its shape as you continue to add diacriticals to it, but still consist of only one character. Also, if you end a word with a sigma, then press the space bar, the closed sigma turns into a final sigma.

Word processors. Not all word processors are Unicode-compliant, and even those that are, may not support the entire range of Unicode. There are only a few word processors that take full advantage of Unicode. WordPerfect is not at all compatible with Unicode. Fortunately, one of the most widely used word processors, Microsoft Word (on the PC, versions from 97 onwards; on the Mac, 2004 only), is fully Unicode compliant. We recommend authors use Microsoft Word or one of the programs listed here: http://www.unicode.org/onlinedat/products.html#9.

After you attend to the four components listed above your system should be able to handle polytonic, Unicode-compatible Greek. Be sure to read carefully the documentation that comes with the software you have chosen to use, so as to familiarize yourself with its conventions and requirements.

Converting text to Unicode Greek

Many authors have already committed large amounts of Greek to a font that is not Unicode compliant. Oftentimes this text needs to be submitted for publication at Dumbarton Oaks. Does this Greek need to be retyped?

Thankfully, no. Several resources help you to automatically convert your preexisting Greek into a Unicode-compliant format. You may be able to make this conversion on your own. There is a website that provides conversion of a few fonts: http://www.jiffycomp.com/smr/unicode-converter/. Greek Keys Converter (Macintosh) facilitates several kinds of conversion: http://www.lucius-hartmann.ch/programme/gkconver.php. Thessalonica offers a universal Converter (PC): http://www.thessalonica.org.ru/en/thessalonica-w97.html. Greek Transcoder, a set of Word macros (both PC and Mac), converts many fonts: http://www.greektranscoder.org/. There are several other utilities for the PC listed here: http://socrates.berkeley.edu/~pinax/greekkeys/WindowsFAQ.html#convert.

If none of these resources work, do not despair. There probably is a way to make the conversion. Contact the publications office, present a sample of your text, and we will suggest how to handle the material.

For further reading

If, at this point, you wish to learn more, or investigate any of the issues above more in depth, here are several of the many resources available on the internet:

— http://cgm.cs.mcgill.ca/~luc/greek.html Extensive list of links to Greek fonts and other resources related to Unicode Greek.

— http://faculty.bbc.edu/rdecker/unicode.htm Introduction to Unicode Greek, written for Biblical scholars; has many useful links and explanations.

— http://omega.enstb.org/yannis/ Extensive scholarly articles regarding the implementation of Greek Unicode.

— http://socrates.berkeley.edu/~pinax/greekkeys/GreekKeys.html Official site for Greekkeys. Through this site new forms of various Unicode fonts such as KadmosU, BosporosU, and AtticaU are distributed. KadmosU is noteworthy for its beauty. These fonts are especially helpful for Copticists, since they include the recently standardized Coptic text block.

— http://unicode.helmug.gr/ Discussions of Unicode issues for Macintosh users

— http://www.russellcottrell.com/greek/fonts.htm List of some of the Unicode Greek fonts that are available

— http://www.scholarsfonts.net/ Very good explanation of Greek Unicode; the free Unicode font distributed here incorporates Greek, Hebrew, and Latin.

— http://www.tlg.uci.edu/~opoudjis/unicode/unicode.html Extensive discussion about Unicode and Greek, addressing the intersection of Unicode and specific points of Greek philology, such as the two types of qoppa.

— http://www.tlg.uci.edu/help/UnicodeTest.html A long page listing available Unicode Greek fonts. See the subpage on how to configure your web browser to view Unicode Greek on the internet: http://www.tlg.uci.edu/help/Help6.html.

— http://www.unicode.org/ The official website of the Unicode Consortium, with a plethora of information about the Unicode standard. Of interest to Byzantinists will be their section on Greek: http://www.unicode.org/faq/greek.html.