

What a Biblical Scholar Should Know about Unicode

Rodney J. Decker, Th.D.
Baptist Bible Seminary, Clarks Summit, Pennsylvania
v. 1.0.¹ Copyright 2003. All rights reserved.
<<http://www.NTResources.com>>

The continual changes and developments in computer-related technology require that the biblical scholar stay up to date in a number of areas related to contemporary scholarship. If you want to be able to read material on the web that is related to the biblical languages or exchange such data with others, you need to get up to speed on current developments in the font world. The new standard for fonts is what is known as Unicode. This system has been in development for a number of years and there is now adequate support in operating systems, software, and fonts to make it both practical and very useful in biblical studies. This document describes these developments and explains how to implement and use Unicode fonts for polytonic Greek and for classical/biblical Hebrew.²

1. Terminology

Like any specialized study, there is a technical vocabulary that you must understand to make sense of the discussion. Here are the basics from the font world as well as a few key Greek definitions.

1.1. Character

In an alphabetic language, a character is a consonant, vowel, digit, punctuation mark, or diacritic used to represent a sound, symbol, sign, or a relationship between any of these elements.

1.2. Glyph

A glyph is an individualization of a character (i.e., a specific way to draw/write a character). The same character may be represented by a wide variety of glyphs without changing the meaning of the character represented. E.g., each of the following glyphs represent the character 'lower case a': **a α ᾶ ᾷ ἄ ἅ ἆ ἇ Ἀ Ἁ ἰ**

¹ This is an expanded version of an earlier paper titled "What a NT Student Should Know about Unicode." The bulk of the material here is still NT-oriented, but I have tried to expand the material to be more useful to biblical studies in general so as to include Hebrew. See also my web page devoted to this same topic: <<http://faculty.bbc.edu/rdecker/unicode.htm>>.

² The following information is provided "as is" and carries no support services and makes no guarantees. It reflects my understanding of the issues involved and describes what works for me. Beware that I am a NT scholar, not a programmer, though I have also worked as an amateur in font design. *Caveat emptor!* Thanks to Peter Kirk who provided some clarifications and helped me refine some of the terminology at several points.

1.3. Font

A font is a computer program which describes a collection of characters having glyphs of a similar style/design with instructions enabling the computer to either display these characters on a monitor or print them to a printer in a variety of sizes. In an alphabetic font this is normally a collection of characters adequate to display or print written text in at least one language. (Some specialized fonts may contain only a partial set of unique glyphs intended to supplement another complete font, but most fonts include a full range of characters.)

1.4. Keyboard

1.4.1. Physical keyboard

A physical keyboard is a device used to enter instructions into a computer, typically through a set of slightly more than 100 mechanical switches (keys).

1.4.2. Logical keyboard

A logical keyboard is a computer program which assigns specified characters to each key on a physical keyboard. Specifying a different logical keyboard results in a different set of characters being sent to the computer when the keys on the physical keyboard are pressed.

E.g., the “standard” logical keyboard on a U.S. computer system is the QWERTY keyboard in which the character values of each key is the same as the character imprinted on that key. But a Dvorak (logical) keyboard can be specified (using the same physical keyboard) which assigns characters to the keys in a different order.³



Dvorak Keyboard Layout

In the layout shown just above the character assignments of the Dvorak keyboard are shown in large type; the standard QWERTY keyboard assignments appear in smaller type.

³ The Dvorak keyboard was designed to facilitate typing speed. By contrast, the traditional QWERTY keyboard was deliberately designed to slow typists down on the early typewriters so that the mechanical linkages did not jam.

1.5. Monotonic and Polytonic Greek

There are several forms of written Greek that are relevant to understanding the issues described here. The Greek of the Bible (either NT or LXX) can be described as koine Greek (or Hellenistic Greek). That refers to a diachronic description of one historical era of the Greek language.⁴ This language can also be described in terms of its orthography as polytonic Greek. By contrast, modern Greek is written in a monotonic form. Monotonic is a fairly recent (1982) development⁵; all previous Greek was written in polytonic form.

These “forms” refer to the diacritical markings of the language. “Tonic” refers to the accents. Monotonic Greek is a system that uses a single (“mono,” μόνος) accent—the tonos (τόνος). The monotonic system is a deliberate simplification of polytonic orthography imposed by legislation in modern Greece. The polytonic system is that which uses the full range of traditional diacritical marks with which the NT student is familiar: acute, grave, and circumflex accents, smooth and rough breathing marks, dieresis, etc. This distinction is important to keep in mind when reading anything about “doing Greek” on a computer. Often modern discussions assume monotonic Greek, but that is of little value for the biblical scholar for whom polytonic Greek is essential. The material in this paper which deals with Greek addresses only polytonic Greek.

2. How a computer knows what glyph to print

When a user presses a key on the (physical) keyboard, the computer uses the logical keyboard to look up the character value.

If the lower left key on the physical keyboard is pressed and the standard US/Western European (logical) keyboard is being used (i.e., a QWERTY keyboard, the default on all U.S. computer systems—and the only one ever used on most systems), the computer assigns the value of ‘z.’ But if a Dvorak keyboard is being used, the computer assigns the character value of ‘semicolon’—even if the physical keyboard has the character ‘z’ imprinted on the physical key that was pressed.

The computer then queries the font program for instructions on how to “draw” that character on the screen. Depending on the font being used and other character formats that may be assigned (size, style, etc.), the computer instructs the font rasterizer (another computer program) to determine which “dots” (pixels) on the screen to “turn

⁴ Koine Greek is roughly dated from the late 4th C. BC to the early 4th C. AD. The earlier stage of Greek is known as classical Greek and the following one Byzantine Greek.

⁵ For a more complete discussion of the move to the monotonic system of Greek, see “Greek & Greek Extended” <http://www.tlg.uci.edu/~opoudjis/unicode/unicode_gkbkgd.html>, part of “Greek Unicode Issues” <<http://www.tlg.uci.edu/~opoudjis/unicode/unicode.html>>, by Nick Nicholas (URLs accessed 7/5/03).

on” to display a representation of this character in the appropriate glyph size and style being used. (A similar process is used to print a character.)

[Please note that this is an oversimplified explanation, but it is adequate for the present purpose—which is not that of computer engineering or programming!]

3. How do you make a computer display Greek or Hebrew instead of English?

3.1. *The “old way”*

By using an “incorrect” font the computer can be “tricked” into displaying a Greek or Hebrew glyph instead of an English glyph for the character specified.⁶ That is, the computer receives instructions from the physical keyboard, passed through and interpreted by the logical keyboard, to display, e.g., the character ‘lower case g.’ But if a font is being used that has the “wrong” glyphs assigned (i.e., Greek glyphs have been inserted into the font in English character positions), the character ‘g’ is displayed in a form that looks like a Greek ‘gamma.’ The computer still thinks that it is working with the character ‘g’ even though it displays (and prints) the Greek glyph ‘gamma.’

3.2. *The problem?*

There is no problem as long as you always use the same font on the same computer. That is, it “works.” But if you send your document to someone else with a different Greek font (or without a Greek font at all), your document will be gibberish because the data encoded in the file consists of English characters, even though it displays with Greek glyphs on your own computer. There is no standard for which Greek glyphs to place in which English character slots, and in any case, the computer still thinks that it’s working with English. Exchanging data is very difficult.

An illustration: you can make the computer display the Greek word $\theta\lambda\iota\psi\iota\varsigma$, in this instance by using the “incorrect/doctored” font Galilee, but the computer still “thinks” of this word as ‘qli:yiV’ (i.e., a character string composed of ASCII values 113 108 105 058 121 105 086, which equals the hex values 71 6C 69 3A 79 69 56—which is actually stored and processed by the operating system as a binary string: 1110001 1101100 1101001 111010 1111001 1101001 1010110!). There is nothing here which tells the computer that this is Greek. And then if one changes the font from Galilee ($\theta\lambda\iota\psi\iota\varsigma$) to another Greek font, e.g., Porson Greek, you now have $\rho\mu\iota:\alpha\iota X$; with GreekYY, you get $\theta\lambda\iota:\psi\iota$, if Grec de Roi, $\theta\lambda\iota\cdot\psi\iota\eta$, and the Mounce font yields $\theta\lambda\iota\psi\iota\varsigma$. To the computer they are all the same: ‘qli:yiV.’ But the user will find it impossible to decipher text treated in this manner.

⁶ The illustrations given here apply equally to Greek and Hebrew even though the specific examples are Greek.

3.3. The limitations of older operating systems using ASCII

Computers have, until recently, worked with only a limited number of character assignments. The reasons for this date back to the origins of personal computers and operating systems for them. To conserve space in the original systems, only 7 or 8 bits of data were used to store character information. Using a hexadecimal system, this allowed a maximum of 128 (7 bit) or 256 (8 bit) character positions ($16 \times 16 = 256$).⁷ They were assigned like this:

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1																
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8	€		,	f	„	...	†	‡	^	‰	Š	‹	Œ		Ž	
9		‘	’	“	”	•	–	—	~	™	š	›	œ		ž	ÿ
A		¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	-	®	-
B	°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

[The chart above shows the Windows Western encoding. Read the position number first from the left column, then the top row, thus ‘upper case J’ = 4A, ‘lower case b’ is 62 (all in hex notation).]

The top two rows are invisible control codes (e.g., carriage return, etc.). Some positions are reserved or not assigned. 20 is the space character; A0 is a “no-break space.” Rows 0–7 are “lower ASCII” positions used by 7-bit systems. The second half of the table

⁷ Hexadecimal has 16 characters (0 1 2 3 4 5 6 7 8 9 A B C D E F), thus the 16×16 grid.

consists of the “upper ASCII” positions that are added on 8-bit systems. All systems use the same lower positions, but there is a great diversity in the upper range: Mac uses one encoding and Windows uses another, and both of these are different from the older DOS encoding—and these vary depending on whether a US or “foreign” encoding is being used. Other languages also use this layout, but assign a different encoding (roughly = “logical keyboard”). That is, in a “standard” US system on the Windows operating system (tech. = Windows codepage 1252, Western encoding) position F0 is the character ð (eth). But when using a “foreign” codepage, e.g., Vietnamese (Windows codepage 1258), the same position F0 is the character đ.

The problems this system poses have long been obvious. Only one language can be displayed. English works just fine, and most languages that use the Latin alphabet can be squeezed in using all available positions in the 256 character grid, but there is not room for additional languages. And some languages don’t fit at all. Ethiopic, e.g., contains 345 characters, and Chinese consists of over 80,000!

3.5. *The Unicode solution*

Unicode has been designed to solve this problem by expanding the number of character assignments that can be specified. Initially the 8-bit encoding was extended to 16-bit allowing 65,536 characters (256 × 256), and the present version of Unicode is essentially a 32-bit encoding that enables 1,114,111 characters—enough for all the languages of the world to have unique character positions. These million plus character slots have been divided into discrete language ranges and the characters of each language have been assigned unique values.⁸ Thus not only can a million plus characters be assigned a unique identification, but the operating system (and any computer program, such as a word processor which “understands” Unicode) can now also distinguish what language is being displayed since each language has a discrete range. The Latin alphabet is included in the “CO Controls and Basic Latin” range <U+0000–U+007F>, Greek is in the “Greek and Coptic” range (= Greek alphabet and the modern monotonic Greek system <U+0370–U+03FF>) and in the “Greek Extended” range (= polytonic, ancient Greek <U+1F00–U+1FFF>), and Hebrew is found in the “Hebrew” <U+0590–U+05FF> and “Alphabetic Presentation Forms” <U+FB00–U+FB4F> ranges.⁹

⁸ See the list of language ranges at <<http://www.unicode.org/charts/>>.

⁹ Note that you should not use the accent or accented characters from the basic Greek range. These are monotonic forms. In NT studies we use only the “unadorned” alphabetic characters from this range. All other characters for polytonic Greek must come from the Greek Extended range. The glyphs may look the same, but they are different and should not be interchanged. For example, the character U+03AC in the basic Greek range looks like a lower case alpha with an acute accent (ά), but it is not; it is a lower case alpha with the (monotonic) Greek accent tonos. The correct character for NT work is found in the Greek Extended range at U+1F71: ά (known in Unicode terminology as “GREEK SMALL LETTER ALPHA WITH OXIA”). The

Although no font will probably ever be created that uses all these character positions, fonts can now include a selected combination of languages that is appropriate for the font's intended use. Thus a biblical studies font can contain not only the Latin character set but also Greek and Hebrew, etc.¹⁰

4.0 How do I use Unicode?

4.1. What you need

4.1.1. A modern operating system

A modern operating system such as Windows NT4, 2000, or XP, Mac OS X, or Unix is needed to use Unicode. It is possible to use Unicode text to a limited extent under Windows 98 and ME, but only in MS Word. It is not possible to use Unicode on any of the "classic Mac OS" versions, and application support is very limited even under OS X.¹¹ The following discussion will assume Windows XP since that is the most commonly used Unicode-capable system.

4.1.2. A Unicode font

There are not a great number of Unicode fonts presently available which support polytonic Greek, though their number is beginning to increase significantly.¹² All are in TrueType or OpenType format. The "free" Unicode fonts that contain both a complete Latin character set and a complete character set for polytonic Greek include Gentium, Palatino Linotype, Tahoma, Cardo, and Arial Unicode.¹³ (There are also some commercial fonts available such as Minion Pro Greek.) Some of these also provide Hebrew support.

Keyman entry system described in this paper assigns the correct character automatically so these are not issues you need to worry about—unless you use the Insert Symbol command to enter text manually.

¹⁰ There are few Unicode fonts presently available that include both polytonic Greek and classical Hebrew. Two that do are Cardo and Tahoma. The Galilee font, when available in Unicode format [hopefully not later than winter 2004], will also have both Greek and Hebrew support.

¹¹ Apple began including some Unicode support in the Mac OS as of v. 8.6, but application support from vendors of productivity applications never materialized. Even today with OS X (now a variant of Unix) utilizing Unicode natively for text, none of the major word processing programs (MS Word for Mac X, NisusWriter, AppleWorks, etc.) support Unicode. Browser support is nearly as poor; only OmniWeb handles Unicode well at this point. Mozilla versions for OS X after 1 September 2002 have slightly buggy, but readable support of Unicode polytonic Greek. Neither the Mac version of Internet Explorer nor Netscape (through v. 7.0) provide usable support Unicode polytonic Greek in OS X.

¹² There are perhaps more options available for biblical Hebrew than there are for polytonic Greek.

¹³ There are some others as well, including Code2000, Titus Cyberbit Basic (both of which are large, specialized fonts), etc. I suspect that other standard Microsoft fonts will add polytonic support as they are updated along with newer software or OS versions. It should also be noted that the Logos Libronix system as well as Silver Mountain's Workplace for TLG also come with Unicode fonts.

Gentium is one of the best of the fonts which includes good polytonic Greek characters as well as the Latin set. It was designed by Victor Gaultney as part of a masters thesis project at the Univ. of Reading (England) and is available from the SIL web site.¹⁴ It includes both a regular and italic face (bold and bold italic faces are in development). Gentium is well suited to serve as a standard text/book face for materials that include both English and Greek text. (This paper is formatted using Gentium.)

abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ

Sample of continuous Greek text set in Gentium: Mark 1:1–6, Ἀρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ [υἱοῦ θεοῦ]. 2 Καθὼς γέγραπται ἐν τῷ Ἡσαΐα τῷ προφήτῃ, Ἴδου ἀποστέλλω τὸν ἄγγελόν μου πρὸ προσώπου σου, ὃς κατασκευάσει τὴν ὁδὸν σου· 3 φωνὴ βοῶντος ἐν τῇ ἐρήμῳ, Ἐτοιμάσατε τὴν ὁδὸν κυρίου, εὐθείας ποιεῖτε τὰς τρίβους αὐτοῦ –4 ἐγένετο Ἰωάννης βαπτίζων ἐν τῇ ἐρήμῳ καὶ κηρύσσω βᾶπτισμα μετανοίας εἰς ἄφεσιν ἁμαρτιῶν. 5 καὶ ἐξεπορεύετο πρὸς αὐτὸν πᾶσα ἡ Ἰουδαία χώρα καὶ οἱ Ἱεροσολυμίται πάντες, καὶ ἐβαπτίζοντο ὑπ’ αὐτοῦ ἐν τῷ Ἰορδάνῃ ποταμῷ ἔξομολογούμενοι τὰς ἁμαρτίας αὐτῶν. 6 καὶ ἦν ὁ Ἰωάννης ἐνδεδυμένος τρίχας καμήλου καὶ ζώνην δερματίνην περὶ τὴν ὀσφύν αὐτοῦ, καὶ ἐσθίων ἀκρίδας καὶ μέλι ἄγριον.

Palatino Linotype (not to be confused with the older font named just “Palatino”) comes with some current Microsoft products (e.g., MS Publisher XP, and perhaps others).

abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ

Sample of continuous Greek text set in Palatino Linotype: Mark 1:1–6, Ἀρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ [υἱοῦ θεοῦ]. 2 Καθὼς γέγραπται ἐν τῷ Ἡσαΐα τῷ προφήτῃ, Ἴδου ἀποστέλλω τὸν ἄγγελόν μου πρὸ προσώπου σου, ὃς κατασκευάσει τὴν ὁδὸν σου· 3 φωνὴ βοῶντος ἐν τῇ ἐρήμῳ, Ἐτοιμάσατε τὴν ὁδὸν κυρίου, εὐθείας ποιεῖτε τὰς τρίβους αὐτοῦ –4 ἐγένετο Ἰωάννης βαπτίζων ἐν τῇ ἐρήμῳ καὶ κηρύσσω βᾶπτισμα μετανοίας εἰς ἄφεσιν ἁμαρτιῶν. 5 καὶ ἐξεπορεύετο πρὸς αὐτὸν πᾶσα ἡ Ἰουδαία χώρα καὶ οἱ Ἱεροσολυμίται πάντες, καὶ ἐβαπτίζοντο ὑπ’ αὐτοῦ ἐν τῷ Ἰορδάνῃ ποταμῷ ἔξομολογούμενοι τὰς ἁμαρτίας αὐτῶν. 6 καὶ ἦν ὁ

¹⁴ <http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=Gentium>.

Ἰωάννης ἐνδεδυμένος τρίχας καμήλου καὶ ζώνην δερματίνην περὶ τὴν ὄσφυν αὐτοῦ, καὶ ἐσθίων ἀκρίδας καὶ μέλι ἄγριον.

Cardo includes not only English and Greek, but also Hebrew. It was designed by a classics teacher, David Perry; available at <<http://scholarsfonts.net/cardofnt.html>>.

abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ

Sample of continuous Greek text set in Cardo: Mark 1:1–6, Ἀρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ [υἱοῦ θεοῦ]. 2 Καθὼς γέγραπται ἐν τῷ Ἑσαΐα τῷ προφήτῃ, Ἴδου ἀποστέλλω τὸν ἄγγελόν μου πρὸ προσώπου σου, ὃς κατασκευάσει τὴν ὁδὸν σου· 3 φωνὴ βοῶντος ἐν τῇ ἐρήμῳ, Ἐτοιμάσατε τὴν ὁδὸν κυρίου, εὐθείας ποιεῖτε τὰς τρίβους αὐτοῦ —4 ἐγένετο Ἰωάννης βαπτίζων ἐν τῇ ἐρήμῳ καὶ κηρύσσων βάπτισμα μετανοίας εἰς ἄφεσιν ἁμαρτιῶν. 5 καὶ ἐξεπορεύετο πρὸς αὐτὸν πᾶσα ἡ Ἰουδαία χώρα καὶ οἱ Ἱεροσολυμίται πάντες, καὶ ἐβαπτίζοντο ὑπ’ αὐτοῦ ἐν τῷ Ἰορδάνῃ ποταμῷ ἐξομολογούμενοι τὰς ἁμαρτίας αὐτῶν. 6 καὶ ἦν ὁ Ἰωάννης ἐνδεδυμένος τρίχας καμήλου καὶ ζώνην δερματίνην περὶ τὴν ὄσφυν αὐτοῦ, καὶ ἐσθίων ἀκρίδας καὶ μέλι ἄγριον.

Hebrew glyphs from Cardo: אבגדהוזחטיכּלמנספצקרשת

¹Gen בְּרֵאשִׁית בְּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת הָאָרֶץ² וְהָאָרֶץ הָיְתָה
 תְּהוֹ וְבָהוּ וְחֹשֶׁךְ עַל־פְּנֵי תְהוֹם וְרוּחַ אֱלֹהִים מְרַחֶפֶת עַל־פְּנֵי הַמַּיִם:

Tahoma is one of Microsoft’s standard sans serif fonts and v. 3.05 now contains both polytonic Greek and Hebrew support. (You could have an older version that does not have this support.)

abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ

Sample of continuous Greek text set in Tahoma: Mark 1:1–6, Ἀρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ [υἱοῦ θεοῦ]. 2 Καθὼς γέγραπται ἐν τῷ Ἑσαΐα τῷ προφήτῃ, Ἴδου ἀποστέλλω τὸν ἄγγελόν μου πρὸ προσώπου σου, ὃς κατασκευάσει τὴν ὁδὸν σου· 3 φωνὴ βοῶντος ἐν τῇ ἐρήμῳ, Ἐτοιμάσατε τὴν ὁδὸν κυρίου, εὐθείας ποιεῖτε τὰς τρίβους αὐτοῦ —4 ἐγένετο Ἰωάννης βαπτίζων ἐν τῇ ἐρήμῳ καὶ κηρύσσων βάπτισμα μετανοίας εἰς ἄφεσιν ἁμαρτιῶν. 5 καὶ ἐξεπορεύετο πρὸς αὐτὸν πᾶσα ἡ Ἰουδαία χώρα καὶ οἱ Ἱεροσολυμίται πάντες, καὶ ἐβαπτίζοντο ὑπ’ αὐτοῦ ἐν τῷ Ἰορδάνῃ ποταμῷ ἐξομολογούμενοι τὰς

ἀμαρτίας αὐτῶν. 6 καὶ ἦν ὁ Ἰωάννης ἐνδεδυσμένος τρίχας καμήλου καὶ ζώνην
 δερματίνην περὶ τὴν ὄσφυν αὐτοῦ, καὶ ἐσθίων ἀκρίδας καὶ μέλι ἄγριον.

Hebrew glyphs from Tahoma: אבגדהוזשחטיכלמנסעפצקשרשת

תִּהְיוּ נֹבְחִים וְחֹשְׁשׁוּ עַל־פְּנֵי תְהוֹמוֹ וְרוּחַ אֱלֹהִים אֶתְשֶׁה מִיָּמַי וְאֶתְהַרְצֶה עַל־פְּנֵי הַיְתֵהוּ² 1 Genesis

Arial Unicode is rather unwieldy (it's a very large font [22.73MB!] designed mostly to showcase the capabilities of Unicode¹⁵) and, I think, quite ugly (especially for extended Greek text). But it does contain polytonic Greek as well as Hebrew.¹⁶

abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ
 α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ ς τ υ φ χ ψ ω
 Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Σ Τ Υ Φ Χ Ψ Ω

Hebrew glyphs from Arial Unicode: אבגדהוזשחטיכלמנסעפצקשרשת

נֹבְחִים וְחֹשְׁשׁוּ עַל־פְּנֵי תְהוֹמוֹ וְרוּחַ אֱלֹהִים אֶתְשֶׁה מִיָּמַי וְאֶתְהַרְצֶה עַל־פְּנֵי הַיְתֵהוּ תִהְיוּ² 11Gen

SIL Unicode fonts provide both polytonic Greek and biblical Hebrew, though in two separate fonts.¹⁷ The Greek font is named “Galatia SIL” and the Hebrew is “Ezra SIL.”¹⁸

αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ
 αβγδεζηθικλμνξοπρσςτυφχψω ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΣΤΥΦΧΨΩ

Sample of continuous Greek text set in Galatia SIL: Mark 1:1–6, Ἀρχὴ τοῦ
 εὐαγγελίου Ἰησοῦ Χριστοῦ [υἱοῦ θεοῦ]. 2 Καθὼς γέγραπται ἐν τῷ Ἑσαΐα τῷ
 προφήτῃ, Ἴδου ἀποστέλλω τὸν ἄγγελόν μου πρὸ προσώπου σου, ὃς
 κατασκευάσει τὴν ὁδόν σου· 3 φωνὴ βοῶντος ἐν τῇ ἐρήμῳ, Ἐτοιμάσατε τὴν
 ὁδὸν κυρίου, εὐθείας ποιεῖτε τὰς τρίβους αὐτοῦ –4 ἐγένετο Ἰωάννης
 βαπτίζων ἐν τῇ ἐρήμῳ καὶ κηρύσσων βάπτισμα μετανοίας εἰς ἄφεσιν
 ἁμαρτιῶν. 5 καὶ ἐξεπορεύετο πρὸς αὐτὸν πᾶσα ἡ Ἰουδαία χώρα καὶ οἱ
 Ἱεροσολυμίται πάντες, καὶ ἐβαπτίζοντο ὑπ’ αὐτοῦ ἐν τῷ Ἰορδάνῃ ποταμῷ

¹⁵ The Arial Unicode font contains all 51,180 glyphs that were defined in v. 2 of the Unicode standard.

¹⁶ The Hebrew glyphs in Arial Unicode are nearly identical to Tahoma. The outline is the same, but Arial’s are a bit smaller and have a slightly heavier stroke.

¹⁷ Neither font is intended as a general purpose font in that Latin glyphs are present simply for the necessities of font design and the punctuation glyphs of the Hebrew font are not in a Latin style. In other words, if you use these fonts for Greek and Hebrew, you should be using a different font for your English text.

¹⁸ Note that these are not the same as earlier, non-Unicode versions named “SIL Galatia” and “SIL Ezra”!

ἔξομολογούμενοι τὰς ἁμαρτίας αὐτῶν. 6 καὶ ἦν ὁ Ἰωάννης ἐνδεδυσμένος
 τρίχας καμήλου καὶ ζώνην δερματίνην περὶ τὴν ὄσφυν αὐτοῦ, καὶ ἐσθίων
 ἀκρίδας καὶ μέλι ἄγριον.

Hebrew glyphs from Ezra SIL: אבגדהוזשחיכלמנסעפצקרשת

1¹ Gen בְּרֵאשִׁית בָּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת הָאָרֶץ: ² וְהָאָרֶץ הָיְתָה תֹהוֹ
 וּבְהוֹי וְחֹשֶׁךְ עַל־פְּנֵי תְהוֹם וְרוּחַ אֱלֹהִים מְרַחֶפֶת עַל־פְּנֵי הַמַּיִם:

4.1.3. Some means of entering Unicode text

Modern operating systems have logical keyboards that can be used to switch back and forth from one language to another, but these are not practical for entering text which consists of a mixture of English, Hebrew, and Greek.¹⁹ The solution for such varied textual needs is a keyboard manager program that allows you to stay in English mode, but easily pass Unicode values from the Greek and Hebrew ranges to the computer. The only such program for Windows of which I am presently aware is Tavultesoft's Keyman.²⁰ This program is available free for personal use. In addition to this, you also need keyboard maps for each language you want to use.²¹ A free polytonic Greek keyboard map for Keyman is available from Manuel Lopez.²² A keyboard map for classical/biblical Hebrew is included with the Ezra SIL font. These programs let you set up a hot key to shift to Greek or Hebrew, which you then enter using the appropriate keymapping. It's a very slick system that is easy to learn and makes entering even large quantities of Greek and Hebrew text efficient.

¹⁹ Other clumsy ways to do this include using the "Insert Symbol" command in Word, or the "Character Map" utility. But these are only of value for inserting a very few characters. Although they are very useful for finding specialized or rarely used characters not included in the Keyman keyboards described above, they are very tedious for entering much text—though it can be done.

²⁰ Keyman is available at <<http://www.tavultesoft.com/keyman/>>. I know of nothing exactly equivalent for the Mac OS, though there are some ways to do the same thing (see the info on such things linked on my Unicode page: <<http://faculty.bbc.edu/rdecker/unicode.htm#Links>>). I've not been able to experiment with any of these Mac options.

²¹ There may be other such keyboards available, but the two discussed here appear to be the most efficient. For polytonic Greek there are two other choices. David Perry has provided a Greek keyboard that is designed to accompany his Cardo font, and Galatia SIL also comes with a Greek keyboard. The Cardo font also has a Hebrew keyboard, but it is not yet working correctly under XP. Any such keyboard can be used with any Unicode font that has the necessary characters. Thus the Classical Greek keyboard can be used with the Cardo or Palatino Linotype fonts, and the Cardo keyboard can be used with Gentium, etc.

²² The Classical Greek keyboard is available at: <<http://members.aol.com/AtticGreek/>>.

4.1.4. A word processor that understands Unicode

Since one of the primary uses of Unicode for biblical scholars is for text entry, one must have a word processor that can process and display Unicode correctly. On Windows, MS Word (2000 and XP editions) handles Unicode very well (though it still can't hyphenate Unicode text! perhaps the MS Office 2003 release in October 2003 will fix this problem).²³ I do not have any knowledge of other Windows word processors (WordPerfect, etc.) as to how (or if) they handle Unicode. If they do handle Unicode text, then the Keyman system described here should work fine there as well.

4.1.5. Handy extra

A complete Greek NT and a LXX is available in a MS Word document formatted in Unicode text. With this you can copy and paste blocks of text a lot faster than you can type.²⁴ There are also Hebrew OT files available in Unicode format, though not in Word documents.²⁵ Other options include Bible study software programs that use Unicode.²⁶

4.2. How do I install it?

First, download the Unicode font/s you intend to use. Install it using the Fonts Control Panel.

²³ Unfortunately the Mac version of Word/Office is still oblivious to Unicode.

²⁴ You can get these files at: <<http://www.users.dircon.co.uk/~hancock/antioch.htm>>. (The link is toward the bottom of the page under "Free Extras.") After you download them, I recommend that you change the Normal style definition to whatever font you intend to use as your standard for Unicode Greek text. I believe that these documents are also defined as italic text—which you can also change by changing the Normal style definition.

²⁵ Unicode Hebrew OT files of which I am aware include the following: On a web site (not downloadable as a complete file): with vowels only or with full cantillation marks:

Vowels only: <<http://www.anastesontai.com/b-vocalisee/anglais.htm>>

Cantillation: <<http://www.anastesontai.com/b-cantilee/en-cant.asp>>

This is on a French site, though there is an "English" interface (obviously written by a non-native speaker). You can view a chapter at a time on screen.

For downloadable files, the Jewish site Mechon Mamre offers the complete Hebrew OT in UTF-8 Unicode .html files: <<http://mechon-mamre.org/>>. (These files are entirely Hebrew text based on the Aleppo Codex; no English help here at all.) Follow the download links listed as "ZIP files downloadable from Simtel Net in either Hebrew with cantillation marks, [or] Hebrew with vowels only." They also have a parallel, Hebrew-English edition. The Westminster Hebrew Institute also offers a downloadable copy of the Leningrad Codex in XML, though it is still in alpha version:

<http://whi.wts.edu/WHI/Members/klowery/eL/index_html>. (Change the extension on the download file from '.uni' [= Unicode?] to '.xml' to facilitate use. Keep the included CSS file together with the main text file; you can also edit this style sheet to control display of the text.)

²⁶ If you use the new Libronix Digital Library System series X from Logos Research Systems, the biblical Greek (NA27) and Hebrew (BHS) texts there are all Unicode based and can be copied out as needed.

Second, download Tavultesoft Keyman and install it. (It comes as an .exe installer file.) Be sure that you get the most recent version (6.0.160 as of this writing).

Third, download and install the keyboard files you are going to use. As part of the install process you can sometimes specify a hotkey to use. The keyboard file usually suggests one, but I changed it to a more mnemonic one: Alt-G for Greek and Alt-H for Hebrew. You can also change these settings from the “Keyman Configuration...” dialog when the program is loaded (accessed with a right click on the icon in the task bar).

I suggest that you do *not* configure Keyman to load automatically when Windows starts (that caused me problems when I tried it). Keep a shortcut icon for it on your desktop or in the Task Bar and load it when you need it. You can start Keyman in the midst of a Word session and you do *not* need to restart Word. Once Keyman is loaded, the hot key settings toggle Unicode Greek or Hebrew text entry on and off.

And that’s it.²⁷

4.3. How do I use it?

It’s quite simple. When you are typing in Word (or any other program that is Unicode-aware) and want to enter Greek text, load the Tavultesoft Keyman program (if it isn’t already), return to Word, then press the hot key you assigned (e.g., Alt-G or Alt-H). From now on *anything* you type, either in the document or a dialog box, will be translated to Unicode Greek (or Hebrew) using the “translation” keys appended to this paper. When you are finished typing Greek or Hebrew and want to return to normal English text, press the hot key again. That’s all there is to it.

4.3.1. Translation Chart for the Classical Greek Keyboard

The Classical Greek keyboard uses the basics of what is called “beta code”—the same transliteration scheme used by TLG.²⁸ Most key equivalents are obvious: alpha = a, beta = b, etc. Here are some that are not obvious (accents/diacritics are typed *after* the letter over/under which you want it to appear; multiple diacritics will “stack”):

²⁷ You might also want to load the right-to-left text support in Windows, but Keyman seems to handle such things adequately on its own even without RTL. Installing RTL will also install about a number of Unicode fonts, including more than a half dozen Hebrew fonts. (How to install RTL is beyond the scope of this paper.) If you use the full installer that comes with the Classical Greek keyboard check to make sure that it does not replace a newer version of Keyman. That installer did not include the newest version of Keyman the last time that I checked and it also does not install all the support files. You can always install the newest version of Keyman from the Tavultesoft website over the top of an existing installation and it will update the program, keeping the keyboards and your settings intact.

²⁸ TLG, *Thesaurus Lingua Graeca*, the database of all known Greek literature: <http://www.tlg.uci.edu/>. The beta code system is documented at <<http://www.tlg.uci.edu/BetaCode.html>>.

eta, η, h	acute, ῀, /	dieresis, ï, + [plus sign]
theta, θ, q	grave, ῀, \	breve/vrachy, ı̇, - [dash]
xi, ξ, c	circumflex, ῀, =	macron, ī, _ [underscore]
chi, χ, x	smooth breathing, ῀,]	digamma, F, alt-v (not all Unicode fonts have a digamma)
iota subscript, η,	rough breathing, ῀, [

(These are illustrations only; see the more complete chart in the appendix.)

4.3.2. Translation Chart for the SIL Hebrew Keyboard

Most key equivalents are obvious: כּ = b, גּ = g, דּ = d, etc. Here are some that are not obvious (points/accents, etc. are typed *after* the letter over/under which you want it to appear; multiple diacritics will “stack”):

⌘, >	⌘, c	dagesh, =	segol, e	holam, o
⌘, x	⌘, f	sheva, ;	tsere, E	maqaf, -
⌘, v	⌘, j	patah, a	hiriq, i	sof pasuq, :
⌘, <	⌘, S	qamats, A	qubuts, u	meteg, alt-1

(These are illustrations only; see the more complete chart in the appendix.²⁹)

4.3.3. Normalization

You need to realize that a Unicode font does not work like the traditional “work around” for adding accents and diacritics for polytonic Greek. The recommended way to do polytonic Greek is through the use of precomposed characters (technically, “normalization form C”). In the old way of doing things, the alphabetic characters and the diacritics are two separate characters—and they remain that after you type them. The diacritic normally “backspaces” and appears over the preceding character.³⁰ But in Unicode, Keyman converts these separate key entries to composite characters.³¹ Thus when you type an alpha followed by a smooth breathing mark and an acute accent (= α ῀ ῀ / when using Keyman as described here), the result is not three separate characters (α ῀ ῀) that are displayed on top of each other, but a different, composite character (ᾗ,

²⁹ The spellings for Hebrew points, etc. given above are those used in Unicode; the English spelling differs from those forms more traditionally used in biblical studies.

³⁰ This makes considerable difficulty for the font designer who has to compromise the glyph designs in order to create a diacritic glyph that will look acceptable when backspaced over both wide (ω) and narrow (ι) characters.

³¹ See the code charts for Greek on the Unicode site: <<http://www.unicode.org/charts/>> and direct links to the pdf files that show the Greek ranges are at <<http://www.unicode.org/charts/PDF/U0370.pdf>> for the basic Greek range; for extended Greek range: <<http://www.unicode.org/charts/PDF/U1F00.pdf>>.

<U+1F04>) that is pre-assigned and pre-designed. Keyman watches for combinations of keystrokes as defined by the keyboard in use and converts them on the fly to the appropriate character.

Hebrew is handled differently from Greek. Hebrew consonants, vowels, and accents all remain separate characters in Unicode Hebrew rather than becoming unique composite characters (normalization form D?). This initially seems like the same old backspacing “kludge” used in pre-Unicode fonts, but it is not. Instead Hebrew Unicode fonts employ another new technology known as OpenType.³² With an OpenType font the font designer can specify precisely where and how the separate glyphs appear when used together. Both the position and the shape of the separate glyphs can be controlled so that they are always positioned correctly. This enables the various points, accents, etc. to appear in the correct position in relation to the consonant. A dagesh dot, e.g., will appear correctly positioned when used with either a wide, enclosed character such as ם, with a narrow, open character such as ן, or between the stems of ם.³³

5. Bible Study Software and Unicode

[coming... someday!]

Windows

Logos/Libronix: Unicode based

BibleWorks: *not* Unicode-aware, will move to Unicode as soon as they think the user base is adequate (Unicode capable system still under 60% of BW users) and “as soon as Microsoft gets the kinks out of their Unicode fonts for Hebrew” (email from BW support).

TLG Workplace: can be used in Unicode mode (or with a traditional, private encoding for older systems)

³² There are several other new font technologies that could be used to do the same thing, but thus far all Hebrew Unicode fonts have used OpenType. Other possibilities are Graphite (an SIL technology) and AAT (from Apple).

³³ Greek can be done the same way and the appropriate characters have been defined to enable this, but polytonic Greek has been standardized with precomposed (“composite”) characters and Hebrew with decomposed characters. This is partly due to the backwards compatibility design of Unicode which incorporated existing standards that used precomposed characters for Greek (viz., ISO 10646). The Hebrew way of doing things is the normal Unicode system for most languages that have similar features, using a base character with combining diacritics. Another reason that Hebrew has not been implemented this way is that the full set of precomposed characters would be in the thousands when all the combinations of vowels, accents, and cantillation marks are included.

Mac

Accordance, *not* Unicode based, developers are talking about a “Unicode export” option for (perhaps) next year

There’s more that could be said, but this should be sufficient to get you started.

Appendix: Keyboard charts

Classical Greek Keyman keyboard					SIL Hebrew Keyman keyboard					
α	a	acute	ή	h/	א	>		ח	x=	
β	b	grave	ῆ	h \	ב	b		ח׃	x;	
γ	g	circumflex	ῇ	h =	ג	g		חׂ	xa	
δ	d	smooth	῏	h]	ד	d		חׄ	xA	
ε	e	rough	ῑ	h[ה	h		חׅ	xe	
ζ	z	iota sub	ῒ	h	ו	w		ח׆	xi	
η	h	diacesis	ῖ	i+	ז	z		חׇ	xE	
θ	q	breve	ῗ	i-	ח	x		ח׈	xu	
ι	i	macron	Ῑ	i_	ט	v		ח׉	xo	
κ	k				י	y		ח׊	xa;	
λ	l	question?	;	;	כ	k	ך	K	ח׋	xe;
μ	m	mid-dot	·	:	ל	l			ח׌	xo;
ν	n				מ	m	ם	M		
ξ	c				נ	n	ן	N	ח׍	x-
ο	o				ס	s			:	:
π	p				ע	<				
ρ	r				פ	p	ף	P	ח׎	x alt 1

σ	s	Ϟ	c	Ϛ	C	Ⲙ	x alt 7
τ	t	ϙ	q			Ⲙ	x alt =
υ	u	ϛ	r				
φ	f	Ϝ	f	ϛ	S		
χ	x	ϝ	j				
ψ	y	Ϟ	t				
ω	w						

These two charts are available in .pdf files formatted to print on a half page each (vertical) at <http://faculty.bbc.edu/rdecker/unicode.htm#Templates>. If you print them out on card stock and keep them by the computer, they will serve as a handy reference guide as you learn to use the Keyman system for Unicode text entry.

Bibliography

Berry, John D., ed. *Language Culture Type: International Type Design in the Age of Unicode*. Surrey, UK: Association Typographique Internationale (ATypI)/New York: Graphis, 2002.

Bringhurst, Robert. *The Elements of Typographic Style*. 2d ed. Point Roberts, WA: Hartley & Marks, 1996.

Decker, Rodney J. "Biblical Language Fonts and Unicode." <http://faculty.bbc.edu/rdecker/unicode2.htm>. (This is an excerpt from a paper presented at SBL, Nov. 2001 on the use of technology in teaching Greek. The full paper is forthcoming in a volume from Sheffield edited by Stanley Porter. Also includes an "Excursus on Unicode" by Peter Constable of SIL.)

Gaultney, Victor. "Gentium: A Typeface for the Nations." http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=Gentium Accessed 10/9/03. This is one of the best English and polytonic Greek fonts now available in Unicode format.

Hudson, John. "Unicode, from Text to Type." In *Language Culture Type*, ed. J. Berry, 24–44. Surrey, UK: Association Typographique Internationale (ATypI)/New York: Graphis, 2002.

Hudson, John. "Windows Glyph Processing." <http://www.microsoft.com/typography/developers/opentype/>. Accessed 6/30/03.

- Leonidas, Gerry. "A Primer on Greek Type Design." In *Language Culture Type*, ed. J. Berry, 76–90. Surrey, UK: Association Typographique Internationale (ATypI)/New York: Graphis, 2002.
- Macrakis, M., ed. *Greek Letters From Tablets to Pixels*. New Castle Delaware: Oak Knoll Press, 1996.
- Perry, David. *Word Processing in Classical Languages: Latin, Germanic, Greek*. By the author. Available in pdf format at <<http://scholarsfonts.net/wordprcp.zip>>. This is a book-length discussion of relevant issues. Also see Perry's Unicode font, *Cardo*, which includes Greek and Hebrew as well as other classical character sets: <<http://scholarsfonts.net/cardofnt.html>>. Both accessed 6/30/03.
- Rourke, Patrick. "Unicode Polytonic Greek for the World Wide Web" <<http://zeno.stoa.org/~ptr/unicode/>>. Accessed 6/30/03. (An extensive site; consists of a draft copy of a book-length treatment of the subject.)
- Rusten, Jeffrey. Review of "Palatino Unicode with Polytonic Greek" (the font). <<http://www.arts.cornell.edu/classics/Faculty/Rusten/unicode/review.htm>> Accessed 6/30/03. Although this is a "review" of a beta version of the font later released as Palatino Linotype, it includes much more helpful information on Unicode for polytonic Greek in general despite the fact that it is now somewhat dated (the review appeared in 1997).
- Unicode Consortium. "The Unicode Standard, Version 4.0.0." <<http://www.unicode.org/versions/Unicode4.0.0/>>. Accessed 6/30/03. See also the Character Code Charts at <<http://www.unicode.org/charts/>>.
- Wood, Alan. "Alan Wood's Unicode Resources." <<http://www.alanwood.net/unicode/>>. Accessed 6/30/03.

Entering Specialized (Unicode) Glyphs in Word

Rodney J. Decker, Th.D.

Supplement to: “What a Biblical Scholar/Student Should Know about Unicode”

Jan. 2006

For more information, see <<http://faculty.bbc.edu/rdecker/unicode.htm>>.

I’ve learned some tricks about accessing some Unicode characters, especially those in the “Supplemental Planes” (i.e., numbered higher than the BMP¹) in MS Word. (For regular Greek text, use Keyman.) If you know the Unicode (hexadecimal) value of a character, you can use the Alt-X keyboard shortcut to enter the character directly in your document.

- Type the Unicode (hexadecimal) value of the character. (You can also use the U+# format.) Of course you have to know the hex value for the glyph you want! Some of the more important ones for NT studies are listed below.
- Press Alt-X.
- Word replaces the string to the left of the insertion point with the character you specified.

(You can also use Alt-X to display the Unicode character code for a particular character. Place the insertion point to the right of the character, and then press Alt-X. The character is replaced by its character code. Press Alt-X again to switch back to the character.)

You apparently need to switch to a font that contains the glyph that you want first (though I think this differs a bit between Word XP/2002 and Word 2003). ***In the following examples, use the Cardo font*** (the only font I know that has all these glyphs).

E.g., the NT textual criticism symbols do not appear in the Insert Symbol dialog. These were added to the 4.1 version of the Unicode Standard (spring 2005), so apparently Word doesn’t recognize them directly. (This will change in the next version of Windows: Longhorn/Vista and the subsequent version of Word.) Some are Supplemental Plane characters (others are additions to the BMP). Most of these are in the Supplemental Punctuation range (2E00 +). To insert the Nestle–Aland text critical symbols, type any of the following hexadecimal values followed by Alt-X:

¹ BMP = Basic Multilingual Plane, also known as “Plane 0.” The Supplemental Planes are additional groups of characters (technically: “code points”) in planes 1–16. Each plane contains 65,536 (10,000₁₆) contiguous Unicode code points—sufficient to encode the world’s languages and communication systems without confusion or duplication by providing a unique number for every character, regardless of language or context.

Hex Value	Resultant Glyph				
2E00	Ɔ	2E01	Ɔ	2E02	Ɔ
2E03	Ɔ	2E04	Ɔ	2E05	Ɔ
2E06	Ɔ	2E07	Ɔ	2E08	Ɔ
2E09	Ɔ	2E0A	Ɔ	2E0B	Ɔ
2E0C	Ɔ	2E0D	Ɔ		

For some other text crit symbols from the MSS (not used in NA):

2E0E	⸏	2E13	⸏	2E14	⸏
2E15	⸏	2E16	⸏	2E17	⸏

Other useful symbols (not all are text critical symbols) in the Kanji range include the following (these are not technically intended for text critical use, but the only way to get a double square bracket, etc, seems to be here):

3008	⟨	3009	⟩	300A	⟪
300B	⟩	300C	Ɔ	300D	Ɔ
300E	Ɔ	300F	Ɔ	3016	⌈
3017	⌈	301A	⌈	301B	⌈

Other symbols are as follows (in different ranges):

- Majority Text symbol, 1D510, ℞
- Papyrus symbol, 1D513, ϐ
- Septuagint, Greek Old Testament, 1D516, ⸚
- Lectionary symbol, 1D459, l

(These are part of the Mathematical Alphanumeric Symbols range, 1D400+. Other text crit. symbols using Gothic/Fraktur letters for “Egyptian text,” etc. are in the same range.) The Majority Text symbol is explicitly identified as such in the Unicode code charts. Most other Fraktur letters are not identified by their text-critical values (but 1D516, “Septuagint, Greek Old Testament” is), but only as “Mathematical Fraktur Capital B/G/K,” etc.: ℞ ™ ℞).

A handy utility program that display **all** the Unicode characters in Unicode 4.1 (including those that neither Word nor WinXP will show you in their input dialogs) is BabelMap (<<http://www.babelstone.co.uk/Software/BabelMap.html>>; be sure to get the newest version).