**GkHebKeyboard - Development Readme**

These notes supplement comments in the source code in the hope that it will make some of the development logic clear.

Note that *all* characters are created in Unicode and the processing has been constrained by what is or is not available in the Unicode representations of Greek and Hebrew.

**Handling Greek Characters**

Standard (simple) Greek characters are found in the Unicode block 0x0384 to 0x03EC. (We have also used the digamma characters in 0x03DC and 0x03DD.) The bulk of these characters are simple letters of the alphabet (both majuscule and miniscule). We shall refer to these as “base characters”. There are a few additional characters (such as characters with diaereses). Some characters are replicated in the later block (described below) and we have generally used those later characters.

Classical Greek are generally represented with “furniture”: breathings (rough and smooth), accents (using common terms, grave, acute and circumflex), iota subscript, diaeresis (somewhat like an umlaut) and a few other adornments. These are separate, discrete characters and this fact has determined the processing of Greek text.

Let us explain. If a word begins with a single vowel, in classical Greek that vowel must have either a rough breathing or a smooth breathing. For simplicity, let’s assume that the initial vowel is an alpha (α). The user will type this and then type the key for a smooth breathing (say). This must replace the character, alpha, with an alpha plus a smooth breathing (ἀ). If the user subsequently types an acute accent, the character must be changed again to have both the smooth breathing and the accent (ἄ). All the Greek processing is to enable this kind of substitution seamlessly.

**Handling Hebrew Characters**

The situation with Hebrew is completely different. Here, the consonants for the “base characters” while vowels are added above and below these consonants. However, the Unicode consortium has provided vowels as “Unicode combining characters”. Thus, an aleph followed by a long a is two Unicode characters but are displayed as a single unit (אָ). This is also true of dagesh and accents.

Of course, remember also that Hebrew characters work from right to left and that this can cause problems when inserting them into left-to-right text.