L2/18-190

Proposal to properly support American style Dozenal notation

Eduardo Marin Silva

31/05/2018

Introduction. This completes the work started by Karl Pentzlin in L2/13-054, by supporting American style duodecimal notation in a less problematic way. In the document it is stated that there are two main ways to write duodecimal numbers (not counting hacks like using letters), one promoted by the British and another by the Americans, such proposal requested the addition of four characters, to represent the digits ten and eleven in both systems. However, the consortium only approved the digits necessary to represent the British style notation. I could not find the ad hoc report corresponding to that proposal and the minutes of the UTC (which never include rationales) sated that they should be encoded as symbols without a numeric value. Asking for a change on the general category and numeric value property may be too late, and maybe there was a very good reason to not treat these numbers the same as ASCII digits. Whatever the reason this proposal is invariant to that.

Why was only one system accepted. There are two possible reasons that come to mind, that could explain why only the British style notation was accepted.

One is that they considered both notations to be unifiable; so one should use the same set of codepoints and to change the typography as needed. This is plausible since one can observe the fact that both digits for “eleven” look like a turned number three (Ɛ), however this is not compatible with the fact that the American digit ten looks like SMALL GREEK LETTER CHI (03C7 χ) and the British one, looks like a turned digit two (Ƨ); under normal circumstances, those two glyphs would not be unifiable.

The second and more plausible explanation, is that the consortium was worried that both versions of the digit eleven were confusable and so while they were not against the American notation in principle, they would only encode one of them (preferably the one more based around the numbers) and wait and see if there was demand for atomic encoding of the other, before possibly including a difficulty for vendors that want to avoid spoofing.

Why both systems have their own merits. One advantage that the American style had over the British style, is that one could input the numbers using closely matching ASCII characters: X for ten and E for eleven; while British style notation had to get creative and use characters such as CANADIAN SYLLABICS CARRIER JU (ᘁ) for the ten, and LATIN CAPITAL LETTER OPEN E (Ɛ), both of them inaccessible to most keyboards and one of them being non-Latin in nature.

The British style however had the advantage that the glyphs were familiar to anybody who knew the respective 2 and 3 in decimal and it was perceived as more fitting than to use arbitrary letterforms.

When the decision came that only the British style would be supported, the Dozenal Society of America, made a referendum to keep their preferred glyphs or to switch over to be complicit with Unicode. Although only 86 people participated in the voting, it was only a difference of ten votes (12%) that made their decision final. Whether that was the intended effect, is irrelevant to the fact, that the partiality of the Consortium, forced a specific language community to abandon their system, in favor of a more advantaged one; that goes against the stated goals of the Consortium.

What to do now. While it is unknown if the Dozenal Society of America would reverse its decision, even if this proposal were successful, the fact remains that there are a lot of publications using the American style notation that cannot be transcribed properly into Unicode. Including up to 123 bulletins published by them:

http://www.dozenal.org/drupal/content/duodecimal-bulletin.html If nothing else, American style notation must be supported one way or another for preserving those documents to future generations.

There are two plausible models to consider, fully atomic and mixed.

In the fully atomic one, we just encode two new characters, with names such as DUODECIMAL DIGIT GREEK SMALL LETTER ENLARGED CHI and DUODECIMAL DIGIT TURNED LATIN CAPITAL LETTER EZH. The problem with the name for the second character is that most probably Dwiggins never intended a connection with the letter ezh, and in fact just used a common glyphic variant of the digit three (with a flat top), in order to avoid confusion with the letter open E, and this glyphic variant just happened to look like an ezh; in fact, it is more probable that contemporaries would have called such a glyph a “tailed z”. In that case a more fitting name would be TURNED DIGIT THREE WITH FLAT BOTTOM

The mixed option consists of just encoding the first digit atomically and considering the second digit as a glyphic variant of the TURNED DIGIT THREE (using the rationale above) and encoding an SVS to capture the difference. The SVS could take the form of:

- 218B FE00; flat bottom glyph; # TURNED DIGIT THREE

Another option that in my opinion is unacceptable is to not even code the chi like character separately and instead consider it a variation of the Greek letter proper. Not only would that introduce conflicts in parsing and casing, but the glyph is an enlarged version of the letter and therefore not confusable in any font with serifs.

The fully atomic option is convenient for anybody who wants to discuss both notations in the same document, and is the one I would prefer, however the mixed option allows to maintain the semantic link between both glyphs of both styles, if there is one like I suspect.

The mixed option encodes one character less; however, we must take into account the fact that the glyph with the flat bottom, is sufficiently distinct for it to not be confusable, even in a sans serif font, unlike the glyph for the TURNED DIGIT THREE (Ɛ), that is confusable with four other characters: EULER CONSTANT (Ɛ), LATIN CAPITAL LETTER OPEN E (Ɛ), CYRILLIC CAPITAL LETTER REVERSED ZE (Ɛ) and LATIN CAPITAL LETTER TRESILLO (Ɛ).