

A suggestion that builds upon the content of L2/16-177 A comprehensive system of control characters for Ancient Egyptian hieroglyphic text (preliminary version)

William Overington

L2/16-199

5 July 2016

In the document L2/16-177 A comprehensive system of control characters for Ancient Egyptian hieroglyphic text (preliminary version) there is section 11.

Section 11 begins as follows.

Because Unicode has its limitations, extended formats with additional functionality are needed for advanced purposes. It is highly desirable that Unicode and the extended formats can be converted seamlessly one to the other. In particular, converting Unicode to extended formats should be a matter of converting syntax only, and converting extended formats to Unicode should be a matter of systematically removing functionality that is unavailable in Unicode, while retaining an acceptable rendering.

End of quote

Here is a suggestion that applies a technique based upon an idea in the L2/15-145R Proposal for additional regional indicator symbol document, that idea being the use of a base character followed by a sequence of tag characters.

If an extra control character were added to the collection of suggested control characters that are in the original document, together with a default visible glyph for it, then the facility for conserving, within the Unicode character stream, of information of that functionality of extended formats that is unavailable in Unicode, could be achieved.

The extra control character could be as follows.

13459 EGYPTIAN HIEROGLYPH PARAMETER BASE CHARACTER

with the short name PARAMETER being useful.

Use would then be as follows, whenever desired, within whatever rules are decided.

A sequence of PARAMETER followed by zero or more characters from the range U+E0020..U+E007E, followed by one and only one U+E007F character.

The characters in the range U+E0020..U+E007F are tag characters. U+E007F is the CANCEL TAG character.

In practice, a meaningful parameter would have one or more characters from the range U+E0020..U+E007E, yet the possibility of having zero such characters is included in the definition so that a sequence of PARAMETER U+E007F is valid, so that such a sequence of two characters can be used as a placeholder parameter and may be validly included in a document by an end user as he or she is working on developing the content of a document.