

**To: UTC**

**From Debbie Anderson, SEI, UC Berkeley**

**Subject: Zoom Meeting on Egyptian Hieroglyphs (April 8, 2020)**

**Date: April 10, 2020**

**Participants:** Michel Suignard, Daniel Werning, Bob Richmond, Frédéric Rouffet, Jorke Grotenhuis, and Debbie Anderson

The discussion focused on different issues related to the draft Unicode proposal by Michel Suignard, “Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, all Groups (A to Z and AA)” ([L2/20-068r](#)) and its [related database](#). This revised proposal includes many more characters than in previous versions, for a total of 6,826 new characters, and is a union of characters from Hieroglyphica, JSesh, Dendara (Jochen Hallof), requests from Barbara Richter of UC Berkeley, and IFAO entries in *Valeurs Phonétiques*. (Via email, Serge Rosmorduc didn’t recommend relying on *VPhon*, since it was a “working index” that drew on publications, which may have been limited by fonts at the time or the particular editor’s interpretation. Michel reported he used it only as one input to his database.)

### **Atomic Encoding**

The general approach in the proposal by Michel Suignard was to use atomic encoding when characters that contain complicated elements cannot be easily created by fonts on the fly. This is especially true with “container” glyphs, due to the need to have multiple sizes.

Daniel Werning (speaking both for himself and colleagues) was very concerned about relying on this approach: he said experts were not as concerned with having a nice typeset product, but wanted the characters to be searchable and indexable. While he agreed he could create programs to compose or decompose characters, he wondered whether low-level searches by Google could find the elements in a character, if it were atomically encoded. By having some atomic and some handled by sequences, wouldn’t this present two different ways of encoding signs? He voiced strong support for decomposed characters (=sequences). This option would allow one to create the various combinations that are encountered.

Michel stressed that Egyptian Hieroglyphs will not be part of Unicode normalization.

### **Differences between font sets (JSesh, Hieroglyphica, etc.)**

Frédéric Rouffet has a list of differences between the main font sets (i.e., cases where the same code point has different signs). Frédéric agreed to share the list, which could be compared with similar comparisons by others (i.e., Daniel also has such a comparison).

### **Character Repertoire and Variants**

Discussion took place on which sign differences are meaningful and to whom.

- Daniel showed an example in which the only difference was the direction of the knife:



In this case, the knife position is probably a variant (and not important), but in other cases how would students know to pick the “right” one? Are the variants semantically different or not? What should be the default case? Michel mentioned that users could define a subset of

characters they want students to use, but Daniel voiced concern about different localities/projects defining different subsets, and the need for a knowledgeable IT person to be involved. Perhaps create a Unicode-based font with a selected subset of characters?

- Jorke Grotenhuis pointed out the description fields in the [Thot Sign List](#) (i.e., A3 is “Man, seated on heel”). The sign entries also reference characters by the “codes” used by Gardiner, Hieroglyphica, JSesh, and Unicode (if encoded). The descriptions may provide a hint to what is deemed important, though there was discussion on whether this information is always a reliable gauge on important features in a sign.
- The experts on the call were concerned that few – if any – experts have the time to review the entire set of proposed characters, and may not have the background to judge whether particular details (such as a particular hand or foot position, the object in the hand, etc.) are important or not for all periods.
- Daniel reported that the [Thesaurus Linguae Aegyptiae project](#) (TLA) has produced a list of signs and their frequency. He suggests those signs appearing at least 2x be added to Michel’s database and the weight be re-calculated based on frequency. Daniel suggested such frequent characters be prioritized, since they reflect signs Egyptologists are using.
- Daniel and colleagues were very concerned about the large number of Ptolemaic variants from IFAO, and it is premature to adding these signs at this stage. They pollute the repertoire with so many variants, causing problems for searching. (Note by Debbie: One impetus for Michel’s work was because Ptolemaic signs were missing.)
- Usage of Variation Sequences was discussed as one way to handle the wide range of variants in plain text, though Michel reported the Unicode group did not recommend this approach: the font needs to support them and layout engines and apps need to support them. (Variation Sequences are default ignorable, so they get ignored if not supported. See §23.4 of the *Unicode Standard*.)
- Jorke strongly believes the characters should be based either on an attested original source or 4-5 modern sources. He agreed most of Hallof’s attestations from Dendara are ok. His main concern are students using characters in Unicode that may be incorrect or don’t really exist, though he admitted if they appear in secondary sources they actually do exist. He also didn’t advocate including cursive signs that are based on hieratic.
- Should just a subset be encoded, or the entire set? Michel voiced concern about extending the proposal process out over too long a time period: work on a full Unicode proposal is very time-consuming.
- Michel pointed out the Unihan database, which is a publicly available database allowing users to search on properties.

**To do:**

- Frédéric R. to share his comparison of the different fonts.
- Daniel W. to share frequency counts with Michel in incorporate in his database.
- Jorke G. to ask Stephane about sharing 7K tokens from the Thot Sign List project.
- Michel will make his database available, first removing the “dbCode” field (which is not for Unicode code points but only used internally for his database) and will instead assign Plane 14 code points. He also needs to get permission to share the font. Note: Plane 14 characters will be a catalog including all characters, including those that are not proposed. Experts are invited to

add phonemes, etc., and share any comments on what variants (such as foot position) are meaningful, and which are not.

- Debbie to circulate the summary.