Meeting report on Egyptian Hieroglyphs
Meeting date: Thursday – Friday, September 12-13, 2019 Thursday: 9:30 a.m. – 5 p.m
Location: Home of Michel Suignard, 77220 Tournan en Brie, France
Attendees: Michel Suignard (host and author of L2/19-220, Editor of ISO 10646), Svenja Gülden (AKU – Altägyptische Kursivschriften), Pierre Fournier (LaBex Archimede – UPVM), and Frédéric Rouffet (Arcanae)
Author: Michel Suignard

Summary:

The following document is organized as a disposition of comments made on the agenda and various documents presented at the meeting (either as separate documents or appendix to the agenda). While most of these points were discussed during the meeting, the text below represents an augmented version of those discussions with some additional thoughts by the author. It is intended to be used as input by the Unicode Script Committee for further direction concerning the proposal. Comments are presented indented, and the dispositions and considerations are italicized.

Main topic: Review of L2/19-220 repertoire
## 1. Comments on glyphs in current code charts (Unicode block: 13000-1342F)

a) Bob Richmond

> **1. N037A Comments**: When we encoded Egyptian originally, EGYPTIAN HIEROGLYPH N037A was rectangle containing 2 strokes. Now it has 3 strokes in the latest code chart.
>
> As I recall, Gardiner used 2 strokes in his variant of N037 so we encoded this. Hieroglyphica used a different 3 stroke glyph (and did not include the Gardiner form) and it seems this error has now been propagated into Unicode docs. Segoe UI Historic and Google Noto still look ok. 2 [strokes,] not 3 is significant. See below for an example of a text that would be broken if encoded with a font built according to current code chart.
>
> If a 3 stroke version is required this must be added to the repertoire.
>
> **Recommended action:** Change current glyph for N037A from 3 to 2 strokes. Get input whether a separate 3-stroke character is needed.
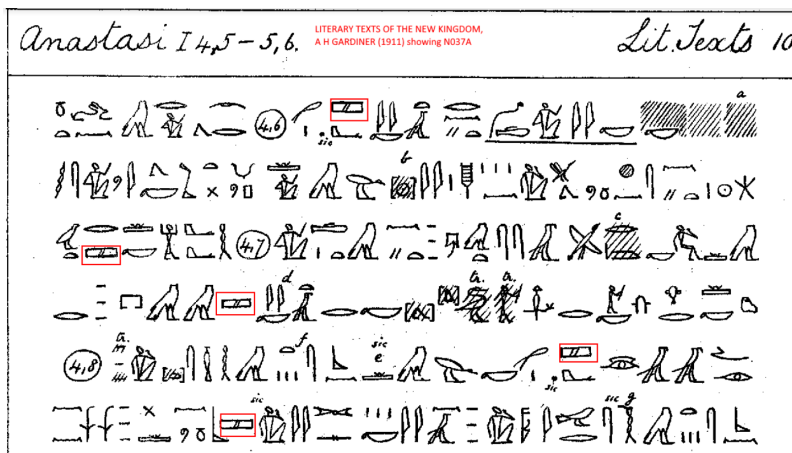
Jochen Hallof has created over 2000 attestations of the current N37A (with three strokes). The author checked that actual evidence matched it for at least one attestation (Dendara, Volume 11, 'Planches' 2 and 11, sign included in group 11) and it indeed the picture is the 3-stroke form. While the referenced text above (Gardiner 1911) is obviously using a 2-stroke form, it is also an abstraction of the original evidences.

Unlike Bob, the meeting participants did not see the number of strokes as particularly significant but mere variants (one, two or three strokes variants are found, and the strokes can also be mirrored). There was however no objection at reversing N37A to only show 2 strokes as requested by Bob. There seems to be no need to separately encode a 3 stroke forms.

2. D0101 Comments: New D010 is full quadrat height whereas Gardiner is about 2/3. I'm not sure about this yet.

Recommended action: Get input on reducing size of current code chart glyph of D0101. Current code chart:

13080   EGYPTIAN HIEROGLYPH D010

The meeting participants did not see a need to reduce the size. The code chart is not prescribing exact size.

3. D027A Comments: New D027A is full quadrat height and this was clearly not Gardiners intention or common practice (he lists it among the small signs).

Recommended action: Reduce the size of the current code chart glyph for D027A. Current code chart (with D027 shown for comparison):

13091   EGYPTIAN HIEROGLYPH D027
13092   EGYPTIAN HIEROGLYPH D027A

Gardiner 1928:14:

Again, while sizing is not prescriptive, it was still felt that the glyph size for D027A could be slightly decreased.

4. D28 / D28A Comments: Inexplicably, Hieroglyphica coded Gardiner D28 as D28A and D28 a larger variant. Probably just an error. New D028 is ok.

Recommended action: No action. Current code chart:

13093   EGYPTIAN HIEROGLYPH D028

The meeting participants concurred.

5. D031A Comment: I'm not convinced by the new glyph for D031A - needs discussion.

Recommended action: Get input on current glyph in the code chart. Current code chart:

13097   𓃗   EGYPTIAN HIEROGLYPH D031A

Earlier glyph (from Egyptian hieroglyph proposal by B Richmond and M Everson N3237):

𓃗

While the participants did not see an issue with the difference in appearance, it was agreed to make the content protruding more (in a version intermediate between the two forms).

6. D059 (/D036/D058) Comments: Gardiner clearly places D036 in front of D058 in the "size a" D59 in his Catalogue of signs but close inspection of his "size c" shows the opposite indicating an error one way or another. This "size c" form makes sense as its the reading order D036 (ꜥ) before D058 (b) so I therefore encoded the latter version and the old font is clear about this. The new D059 changes this by placing the D036 in front. Unless there is clear evidence, the previous order should be restored.

**Recommended action**: Change code chart glyph so D036 is in front of D058 in D059. Current code chart:

NO.	a	b	c	d	e
36					
58					
59					

Current code chart:

1309D   EGYPTIAN HIEROGLYPH D036
130C0   EGYPTIAN HIEROGLYPH D058
         • transliterated as b
130C1   EGYPTIAN HIEROGLYPH D059

Earlier glyph (from Egyptian hieroglyph proposal by B Richmond and M Everson N3237):
1309D EGYPTIAN HIEROGLYPH D036
130C0 EGYPTIAN HIEROGLYPH D058
       = b
130C1 EGYPTIAN HIEROGLYPH D059

The participants did not feel that the z ordering of the components was implying a reading order but instead indicates just a convention when a leg is combined with another mark. In the vast majority the leg is 'behind'. It also seems that the recommended action is contradicting the comment. In conclusion, the participants preferred to keep the current shape of D059.

(other similar comments made in
https://github.com/HieroglyphsEverywhere/Docs/blob/master/UnicodeRepertoire/ExtendingUnicodeEgyptianHieroglyphicRepertoire20190911.pdf)

Cartouche begin/end signs V011A, V011B and V011C have shrunk in current code chart (important: these are supposed to be full height).

13377   EGYPTIAN HIEROGLYPH V010
13378   EGYPTIAN HIEROGLYPH V011
13379   EGYPTIAN HIEROGLYPH V011A
      • beginning of cartouche
1337A   EGYPTIAN HIEROGLYPH V011B
      • end of cartouche
1337B   EGYPTIAN HIEROGLYPH V011C
      • end of knotless cartouche

It was agreed to modify them to be full height.

V012A and HG H-27 look suspiciously alike. So, is the glyph a rope (Gardiner) or part of a bird (Hieroglyphica)?...

1337D   EGYPTIAN HIEROGLYPH V012A

V012A is sourced to Gardiner (V12) from Gardiner 1953 (see WG2 N3182 and N3237). Although not exposed in the current tranches (A to F) of L2/19-220, the character V012A has been reclassified as a bird feather according to its Hieroglyphica source. As stated later in this document, this is just a classification decision, and it can still be used as a rope sign if desired, but it should be only encoded once. Overall, it cannot be avoided that some characters in the original Unicode Egyptian Hieroglyph block will be classified differently in the new taxonomy.*

Some orientation and other variants given by Gardiner didn't make it to the Unicode release so these should be checked all present in Unikemet2 for inclusion in the next expansion.

Not actionable because no details were provided.

b) Jorke Grotenhuis

a. Incorrect image: U131AC should have legs, as it is the click beetle (I currently only know its existence from the Pyramid texts, it is in the article by Meeks about insects). Current code chart:

131AC   EGYPTIAN HIEROGLYPH L008

The suggested source image by Jorke Grotenhuis is:

In previous feedback provided to the author of L2/19-220, it was suggested that L008 was not at all an insect and in fact its source may not have been attested correctly. Based on this, it seems unwise to completely change the identity of L008. It would seem more optimal to add a new sign based on the source image provided. IFAO has in its 'scarabée' category a mix of beetles, not just dung beetles (another less glamorous name for scarab), and we could probably reference the click beetle as IFAO 229,2 which looks reasonably close to the picture above (which means we should probably rename the

group from 'scarabs' to 'beetles') allowing both a more diverse group of beetles to be encoded in that group.

b. Not technically smaller, but just slightly broader: U1310B - U1310C (F13)
1310B would do fine on by itself, I do not see the benefit of having two slightly variations. (Note that in Jsesh and Hieroglyphica F13A is not 1310C!) if you can mark one of the two as theoretically non-functional, it would already work. Current code chart:

| 1310B | ⎵ | EGYPTIAN HIEROGLYPH F013 |
| 1310C | ⎵ | EGYPTIAN HIEROGLYPH F013A |

The fact that the name space used in the current encoded Egyptian Hieroglyph block is different from the name space used in JSeh and/or Hieroglyphica is well known. It is in fact one of the reasons that a new taxonomy is proposed in L2/19-220. It is also clear that for many experts the character encoded as U+1310C is not necessary. However, because it is already encoded it cannot be removed, it could possibly be annotated as being not recommended (if such consensus arises).

c. Seems to me a meaningless angle: U13319 - U1331A (T14)
Current code chart:

| 13319 | ⎞ | EGYPTIAN HIEROGLYPH T014 |
| 1331A | ⎞ | EGYPTIAN HIEROGLYPH T015 |

While the utility of the variant shown in U+1331A is debatable it is now encoded and cannot be removed. See remark on previous case.

d. Minor change in form, not meaningful: U13214 - U13215 (N34); U1337D - U1337E (F20A)
I see no reason to keep both (I know they cannot be removed, but if they can be marked as duplicates of each other, it would already work. Every case of transcription into a standardized font is already an adaptation, in which case the extreme minor form variants as shown here would not really matter. If you wish to mark a 'primary' variant, use 13214 and 1337D.
Current code chart:

13214	⌓	EGYPTIAN HIEROGLYPH N034
13215	⌓	EGYPTIAN HIEROGLYPH N034A
1337D	⌒	EGYPTIAN HIEROGLYPH V012A
1337E	⌒	EGYPTIAN HIEROGLYPH V012B

Same remark as above, an annotation could be added in the code chart.

e. Reversed position without reflection in Gardiner: U13413 - U13414 (Aa7)
As before, I would prefer control characters that allows us to reverse the direction of signs, instead of doubling the signs, considering for these that my initial research showed no meaningful functional variation between the orientation. (There are more signs like this in the list, that I did not hit on yet, but might be better under the same point.) Note that this all depends on the possibility to change the orientation and position of the Unicode sign.
Current code chart:

| 13413 | ⟋ | EGYPTIAN HIEROGLYPH AA007 |
| 13414 | ⟍ | EGYPTIAN HIEROGLYPH AA007A |

Gardiner 1928: 36

Gardiner has a separate entry (7*) for the reversed form which is probably the rationale for that encoding. A higher protocol can always change the orientation of a Unicode character. Typically, characters should only be duplicated in another orientation if they can appear both in the same plain text (or if both orientations can be proven to have different properties in term of sound, semantic, etc.). For historic scripts such as Egyptian hieroglyphs limited attestation is always an issue to determine such need. Note again that these 2 characters are already encoded and therefore 13414 cannot be removed.

f. Slightly smaller: U13168 - U13169 (G36); U13209 - U1320B (N25); U1339F - U133A0 (V30); U133B1 - U133B2 (W3); U13213 - U133E7 (N33A)

… I want this to be shown, and the second smaller ones to be marked as 'not functional', as I do not see the rational of making smaller versions of some signs, but not all, or vise-versa. I get that the current set cannot be changed, but my main point is that there are issues in the current list…I would prefer not to have repeated in any addition (except for my existing objection to the inclusion of a long new list as well, partly due to the issues I spotted here). …we should try to limit these issues, and we should make it clear why we consider some Unicode signs less useful than others (and at a place where it is easy to spot). True errors are difficult to discuss, as I have not seen every hieroglyph in existence, but that is why I work on the side of safety, working from the principle that an ancient source needs to show me the existence of a sign before I consider that it exists, instead of just believing that there were no mistakes, or misinterpretations in the list.

Another point I left out concerns U13155, (G19) [see below] which is a Gardiner sign, which I could not even locate at all, as Gardiner gives no sources for it, and except for the sign list of Hornung & Schenkel, which directly refer back to Gardiner, there is no other sign list I know of which even supports the existence of this sign, including Kurth and Cauville (Ptolemaic sources). I do not know what Gardiner based this on, but I guess it is from Hieratic, in which case the reading is dubious, as it could have been transcribed as U13155 as well.
Current code chart:

13168		EGYPTIAN HIEROGLYPH G036
13169		EGYPTIAN HIEROGLYPH G036A
13209		EGYPTIAN HIEROGLYPH N025
1320A		EGYPTIAN HIEROGLYPH N025A
1339F		EGYPTIAN HIEROGLYPH V030
133A0		EGYPTIAN HIEROGLYPH V030A
133B1		EGYPTIAN HIEROGLYPH W003
133B2		EGYPTIAN HIEROGLYPH W003A
13213		EGYPTIAN HIEROGLYPH N033A
133E7		EGYPTIAN HIEROGLYPH Z002B
13155		EGYPTIAN HIEROGLYPH G019

Note that there are a few ovals and circles as well which are repeated, but as they technically are variants of signs without inner detail, and therefore could be considered different (as we made them classes in the TSL). From these points, I find the V30 and W3 variant most problematic, if you want to point out two.

Postscript: If needed, I can provide you as well with a few signs that I had to rework, as I found that the Gardiner sign is actually not the majority, or exists (to my current knowledge). Further we had a few cases where the image of Gardiner and Jsesh versus Hieroglyphica and Unicode does not line up (for example, A43, where Gardiner and Jsesh do not have a beard, but Hieroglyphica and Unicode does), which gave some interesting effects. Usually this is really minor, as both options exist, and I don't want to go into who is right, but it is another feature that needs to be taken into regard when copying only one font for Unicode, and basing it only on modern sources, instead of the ancient ones.

… I was actually able to locate both version in hieroglyphs, so in this case both versions exist, but I cannot say this is not always the case.

To restate my point in general, although I do not agree with the current move forward, if it happens, I would just want to ask that more care is taken with the details and precision. As most people know that the extended library of Hieroglyphica is really poorly sourced, I feel really uncomfortable by just uploading that, without taking the other sources in regard. If it needs to be done by modern sources, please make sure more than one list is checked, preferably Kurth and Cauville, although these have real issues in them as well (although more on the function side than the actual shape). I will take a look today at the proposal, to see if I can spot any obvious problems, but I have to say that I currently do not have the time for a really deep look.

There are several points made in the text above and these are answered separately:
1. *The concern about encoding various sizes cannot be fully accommodated for existing code points because of the Unicode stability principle. However, it is in general agreed that additional encoding of that nature should not be pursued in the extension. In addition, annotation can always be added to discourage the usage of some characters.*
2. *The concern about lack of evidence for already encoded character (like G019 above or L008 in a previous comment) is real but cannot justify the removal of a character. Reasonable care should be exercised for any extensions, but it is impossible to totally avoid the issue on such a large repertoire.*
3. *There are some imprecisions on even the base Gardiner list. For A43 (and A44) which are representation of a king with the white crown, you can find pictures with or without the beard. It does not seem wise to encode both versions as most experts would consider them interchangeable. At the end it becomes a matter of preference. The new version of the Unicode chart uses a bearded version for both A43 and A44. But other fonts may select to use versions without beard.*
4. *Concerning multiple source references for the categories A-F, it should be noted that a single group of sources (Dendara, Cauville) already covers about a third of the Hieroglyphica set. The ValPhon index (Valeurs Phonétiques) covers over half all of the entries with again a third including outside source attestations. Kurth source is also extensively used. In conclusion, the author of L2/19-220 has already used an extensive list of sources to justify the encoding of the extension.*

2. Comments on L2/19-220
a) Jorke Grotenhuis's comments

Threshold for deciding whether to add characters
First, the regard if a sign should be added or not (I might have misinterpreted his point here though). The value of 1 or more is sufficient to add this? I strongly disagree, as when one finds a hieroglyph somewhere in a modern source, it would already be at 1, so already past the threshold.

Second, this system ignores the fact that often the modern sources refer to one another, which might be hidden if you do not look closely to the notes. Kurth often states see VP, IFAO, or Cauville.

Third, as my point has always been, to me a modern source is nothing, I can live with it if the value one is given to a list when a ancient source is mentioned, I get that he does not have the time to look at everything, but I do not think the value of 1 should already be given when there is no source at all for the form, or if it only refers to another source (For example, when the value is 2, VP and Kurth, but Kurth refers to VP, the value should be considered 1, not 2).

So, in my opinion the value of the threshold for encoding proposal should be at least 3 or more, with a minimal of 1 reference to an ancient source (which should be noted with the proposed sign, so that it can be checked by Egyptologists). If the only sources for a sign are modern sign lists etc., we really need to push the threshold higher.

The concept of modern versus ancient source is somewhat ambiguous without an exact definition of these terms ('ancient source' and 'modern source'). Most sources are already abstract representation of what the author determined to be a common set of features attributable to a set of original pictorial evidences. Should the work done by Cauville in her various publication concerning the Dendara temples considered modern or ancient?

It is true that the weight is somewhat inflated by circular references, however they still denote the fact that multiple experts expressed the opinion that a given character was useful.

Concerning the minimum threshold for encoding, this is also open to debate. For example, many of the {Dendara, Cauville} entries have thousands of attestations, although they only count as a weight of 1 in the present system. Similarly, the 'Valeurs Phonétiques' entries commonly refer to multiple attestation. The author of L2/19-220 is open to provide the files containing details of the [Dendara, Cauville] attestations.

Case of Hieratic

As a final point on this issue, there is one feature that is easy to overlook if one only looks at the modern sources. There are signs in existence, that are mere hieroglyphic representations of Hieratic signs, and do not actually occur as hieroglyphs. For example from the Gardiner list, we have U13086 (D16). This sign does not actually occur in hieroglyphic texts (as hieroglyphic texts rarely go over 1/16th if they have fractions at all). Although I understand the logic of the sign, as

every part of the Udjat eye is a different fraction, in the Hieratic script, the sign actually looks identical to U133F6 (Z11).

13086	⟨	EGYPTIAN HIEROGLYPH D016
		• 1/64
133F6	⊹	EGYPTIAN HIEROGLYPH Z011

So, the sign is actually a modern creation, and although the assumption seems reasonable, the existence of the sign is conjecture, and it is possible the Ancient Egyptians did not understand it as we now think. So the including of the specific form used is dubious at best, and I am certain this is far from the only sign where this is the case. (especially with men and the positions of arms and hands).

These characters are already encoded, therefore the concern while valid cannot be fully satisfied. However, while ancient Egyptians may not have used it, there are now common objects of communication between modern Egyptologists and as such they probably deserve such an encoding. The same rationale can be made for many more symbols. Even debating dubious but commonly referenced characters may justify their encoding to facilitate the study of Egyptian hieroglyphs.

Modern vs. Ancient Sources

To specify my point of modern sources versus ancient sources, I would refer to an issue with U13105 vs. U13106.

13105	🦅	EGYPTIAN HIEROGLYPH F007
13106	🦅	EGYPTIAN HIEROGLYPH F008

Now I do have to state that both these signs exist, but there were cases where I had transcriptions, for one stela a transcription by Lacau, and one by Sethe (Urkunden) so we have the same text, but one used U13105, where the other used U13106. If I did not have the image, I would not have known which one to trust (it turned out that Lacau transcribed correctly, and Sethe did not). Often these are minor details, but it is a reason why we have been pushing so hard to actually check modern sources with the original, and we do not blindly follow modern sources only.

These 2 characters are already encoded, but the issue is indeed common. It is a choice between creating a single abstract character or disunifying it, accepting that the same entity can be encoded with two or more abstract characters. It is also another reason to associate a database with any encoding to facilitate correlation between characters. Overall, it is typically better to disunify because an evidence of contrasted use can always be found later. For example, in the case of F007 and F008, the {Dendara, Cauville} references provide attestations for both.

Positional Variation

Second, for the list itself, for example in the column 1344. The first signs are a group of hieroglyphs with variations in had position. Here I have to wonder if it was taken into regard where the limits between variation and same are placed, how extreme must the hand position be in opposition to the other glyph to decide if should be encoded separately, or if it can be caught under one standard sign. As far as I know the Unicode works under the principles of standardisation, not epigraphical precision, so we might need to take a look if the sign variation group matters or not, before adding a theoretical sign of a man with his hand in front of him on a 80 degree angle and one with the hand on a 110 degree angle for example.

The proposed encoding for these 4 characters is based on contrasted use in the {Dendara, Cauville} attestations (2 attestations for proposed +13440, 286 for U+13441, 89 for U+13442, and 106 for U+13443

System of control characters to combine signs

Third, I spotted many signs (especially column 13BC-13C0, where I would prefer to see a system that allows us to use signs, and combine them using control characters, instead of adding every possible variant, which will allow users to create signs that are currently not covered by the list as well. I think the focus should lie on the development of these control characters before adding a lot of extra signs.

There are several different cases in the example above, some are simple insert (one object contained in a container), some are multiple objects contained in a container (e.g. U+13BE0 or U+13C02), or mix of overlay and contained (U+13BE2). It just shows that the exact layout can be complicated and cannot be created productively in current font technology. In most cases, an exact glyph representing the sequences would have to exist in the font, and therefore some system of known sequences would have to be registered and maintained. At that point it is not clear whether it is not just simpler to encode atomic characters, at least for the complex cases.

Even in the simpler case of simple insert, for a productive system, the font has still to maintain a smaller size glyph for all characters that are candidate for containment and multiple width of the container to accommodate various widths of the contained characters. The alternative is to create premade glyphs for all possible combinations, a concept which becomes not realistic for a repertoire made of several thousand characters.

Another idea worth pursuing is the development of character description sequences (a bit like what was done for CJK ideographs) to describe in abstract term the content of a given hieroglyph. Because the description sequences do not describe a formal decomposition or do not aim at enforcing a layout system, they are less constrained.

Duplication

> The final issue I spotted with the current list that I am still afraid duplication is occurring (as I know it to happen in Hieroglyphica as well, I sadly currently have no clear examples). For example, U14026 U14027 might be duplicated as well in category Aa (as they are variants of Aa2 (U1340F)), but due to development in Egyptology, and how we understand signs, they might be moved or duplicated.

> 14026   EGYPTIAN HIEROGLYPH F-30-004
> 14027   EGYPTIAN HIEROGLYPH F-30-006
>   F-30-008 referenced as AA003
>   → 1340F   egyptian hieroglyph aa003
> 1340F   EGYPTIAN HIEROGLYPH AA003

> For example, the sign list of Hornung & Schenkel made sure to get rid of category Aa, but by doing so, it is possible that if one uses them as the source of the proposal, the sign might show up in F, but in Aa as well, duplicating the sign. (If one does not know this happened, and bases the list on H&S and Kurth for example.) Thus, based on this, and the fact that this is only a section of the signs, I am quite certain duplication will happen again.

The opinion of the meeting attendees was to move all these characters back to the AA category to avoid the duplication concern. Apparently, there is no consensus that these signs belong to the 'excrements' category.

> For one sign I just spotted, U13F9D, it is another variant of the 1337D/1337E group I discussed from the original Unicode list.

> 13F9D   EGYPTIAN HIEROGLYPH F-16-008
>   F-16-009 referenced as V012B
>   → 1337E   egyptian hieroglyph v012b
> 1337D   EGYPTIAN HIEROGLYPH V012A
> 1337E   EGYPTIAN HIEROGLYPH V012B

It is the intent of the author of L2/19-220 to not duplicate entries. This may result in entries set in the wrong part of the taxonomy. But it is always possible through annotation to describe multiple or alternate use for a given character.

In the cases above, currently V012A is part of the 'tongue' category, while V012B is part of the 'feather' category. In the current Unicode encoding they are categorized as part of the category V (Rope, fiber, baskets, etc.).

Another I spotted is U1400F, which is the reverse of U1341D.

1400F ⌐ EGYPTIAN HIEROGLYPH F-26-011
1341D ⌐ EGYPTIAN HIEROGLYPH AA015

This was based on ValPhon categorization, but the meeting attendees thought it was better to keep the 'AA' group for these signs.

Summary

These are my major first points that I saw based on a quick look, but I would like to stress that I think that the principle of proposal needs to be changed, and made more difficult to pass the threshold, and the inclusion of ancient sources in this method.

Covered by answers above.

Second, I would prefer that some focus lies on the creation or finishing of tool that allows for the placement and interaction of existing Unicode signs, before moving forward with adding a lot of signs which could be created by the user himself.

It is important to determine the meaning of tools. There is a consensus that Egyptian control characters should be implemented by text layout engines in common platforms. But requiring tools beyond that is an impediment on universal access because any tool restricts usage in some ways. Having a solution that emphasize 'plain text' support is typically the more universal solution.

Third, one has to look at the actual tokens of the modern sources, including the tiny details, to make sure that they are actually the same, even if the difference is minor and is most likely meaningless.

See above.

b) Stéphane Polis' comments

Atomically encoded characters vs. characters created with format controls

Since some control characters have already been adopted, I do not understand why we need to encode atomically groups that can be produced with such control (overlays, corner insertions, etc.). There are still a lot of them in the current charts. I know that some simple groups, like 130B0, 130B1, 130B2, 130B3, etc. (and the numbers in general), were encoded atomically in the first version, but one should certainly stop encoding groups atomically now that we have controls (same remarks for the center insertions, which would significantly reduce the number of code points needed if this control is added).

130B0 𓂰 EGYPTIAN HIEROGLYPH D050C
130B1 𓂱 EGYPTIAN HIEROGLYPH D050D

This issue is covered several times in this document, but there is clearly a majority view that existing controls should be used to avoid unnecessary encoding. However, there are some limits to productive use based on current font technology. If it results in the need of precomposed glyphs in the font, the usage of character sequences is less desirable.

Repertoire includes modern variants of same grapheme / threshold for adding characters

I still maintain that many signs in the proposal are just *modern* variants of the same grapheme in ancient Egyptian, for which you would have a hard time finding an actual example in an ancient texts (some scholars chose to retain some iconic features when copying a sign, some others, and they copy each others, of course). The thresholds based on modern sign-list are simply not relevant, unless hieroglyphs in Unicode are meant to re-print old sign-lists. As for some other variants, if they can indeed be found in ancient texts, they are very often not meaningful. It feels to me like you would want to create code points for the <a> as written by Chaucer, the <a> as written by Balzac, and the <a> as written by Calvino. Is it really what Unicode is about?

Determining whether a variant is meaningful is hard, especially in the Ptolemaic era.

A reasonable threshold, imho, would be: at least 1 *ancient source* for a given sign. And even then, this would call for some serious thinking: take the hippo with one harpoon standing [13DF3], with two harpoons kneeling [13DF6], why not with 3, 4, 5 or 6 harpoons since all these cases are attested?

13DF3   EGYPTIAN HIEROGLYPH E-14-008
13DF6   EGYPTIAN HIEROGLYPH E-14-011

Current list is based on IFAO (page 134).

Sometimes, it's simply funny/// Take 13D47 vs. 13D48 for instance: does anyone think that there has ever been any difference between the dagger facing left and the dagger facing right in the back of an animal? Maybe that the scribe was right vs. left-handed???

13D47   EGYPTIAN HIEROGLYPH E-02-041
13D48   EGYPTIAN HIEROGLYPH E-02-042

I do not want to be over-ironic, but please consider that if adding such signs, you should be ready for several dozens of variants for every single sign of animate, just based on the number of harpoons/daggers/other weaponry that they can receive in funerary contexts.

The current list is again on IFAO, but there is no contrastive use of both signs, therefore it seems reasonable to remove one of them (U+13D47 is not attested outside of IFAO).

Cursive forms

Why would we add cursive forms such as 1379A, 137AA, etc. which are not standardized hieroglyphs?

1379A   EGYPTIAN HIEROGLYPH A-34-002
137AA   EGYPTIAN HIEROGLYPH A-34-01

Most meeting attendees were as well in support of removal; however, they are present in Hieroglyphica and JSesh.

INSERT CENTER control character

> I strongly support the INSERT_CENTER control, already suggested in Nederhof et al. (and now considered by Andrew Glass).

Worth considering, but it may be complex to implement as noted above.

c) Bob Richmond (from
https://github.com/HieroglyphsEverywhere/Docs/blob/master/UnicodeRepertoire/ExtendingUnicodeEgyptianHieroglyphicRepertoire20190911.pdf

The Road to Repertoire Extension

> …More than one formal proposal can be in the pipeline at the same time, so it is possible to expand the repertoire in parts without waiting 2/3 years between parts. However, there needs to be a clear rationale for a multi-part strategy both for administrative and logistical reasons as well as consideration of how staging might benefit the hieroglyphic user base.
>
> The current version of the **Unikemet2** database is a good starting point in my opinion and provides a coherent basis for looking at staged expansions. Additions to the database can be made as new hieroglyphs are established as needed. Hieroglyphica as a major source links the database to the practical aspects of working with hieroglyphs as practiced by the user base.
>
> However, as noted previously, Hieroglyphica expands on Gardiner over all periods, from Early Dynastic though to Greco-Roman. Modern technology does not fully invalidate Gardiners 1928 comments on the propriety of a one size fits all font. Development of the current extension work began with a focus on Ptolemaic but the extensions being discussed now are at least as much about expansion of the Middle Egyptian treatment in Unicode, Late Egyptian and earlier periods as the Greco-Roman era.
>
> It is important for non-specialists to understand that almost all hieroglyphic from all periods is dominated by signs already encoded in Unicode. This includes the Dendera Ptolemaic Temple reliefs as published by IFAO, the encoding of which was a major contributor to the many Ptolemaic entries in the Hieroglyphica font catalogue. Furthermore, ultimately hieroglyphic is a handwritten script and there will always be applications where plain text encoding is not enough, and higher-level protocols or facsimile will be necessary to capture an ancient text to the degree deemed satisfactory to the modern writer for a given purpose…

Meeting attendees objected to the use of 'Middle Egyptian', 'Late Egyptian', and such, the use of 'Middle Kingdom', 'Late Kingdom' was preferred. In addition, the assertion that 'ultimately hieroglyphic is a handwritten script' was disputed, the term 'carved' would be preferable. Unikemet2 is the term used by Bob to describe the database associated with L2/19-220.

While the staged approach was considered, meeting attendees would prefer to see the whole repertoire: A to Z and Aa) proposed at once, even if additions are done later to the sub-groups.

Some prerequisites for extensions
Action: Detail use of EGYPTIAN HIEROGLYPH FORMAT CONTROLS in relation to Unikemet2 database

So far, I have only considered two-character combinations of hieroglyphs using the EGYPTIAN HIEROGLYPH OVERLAY MIDDLE format control. Not any of the more exotic overlays from font catalogues. See the draft https://github.com/HieroglyphsEverywhere/Docs/blob/master/UnicodeRepertoire/EgyptianHieroglyphOverlaysHieroglyphica20190911.pdf.

There are over 200 such combinations in Unikemet2 at present that are overlays of signs already encoded. Over 100 additional combinations are available if currently unencoded components are included in a repertoire expansion.

Traditional font catalogues (including Lepsius-Theinhardt, Gardiner, IFAO and Hieroglyphica) treat combinations as atomic, including overlays, so combined forms should remain listed in Unikemet2 but marked as implemented using format controls and excluded from consideration for repertoire extension.

It is important that a list of attested overlays (and maybe other combinations) is maintained outside the Unicode standard itself (overlaying random hieroglyphs is rarely useful or pretty and should not be done using font-specific quirks). That way font and software developers can ensure they do a good job with overlays that are used. By keeping the list separate, new items can be added to the attested list when identified without need to go through the standardization process. It is desirable a consensus is achieved here among interested parties so such a list (with documentation) can be actioned and consequences factored into Unikemet2. For the sake of ongoing work on implementing format controls it is highly desirable this is produced before end of the year. Discuss.

While most of the attendees were in favor of that solution, it was not clear why the list of attested overlays should be maintained outside the Unicode standard. Unicode maintain list of named sequences http://www.unicode.org/reports/tr34/ which seems to be the right vehicle for such application.

Other format controls could be added such as the mooted INSERT CENTRE that could be used for arrangements such as one or more hieroglyphs enclosed in a large D028 (there are 58 D058 enclosures in HG though their sources are undocumented so it's hard to take a view until their significance is well-understood). There would need to be a clear cost/benefit analysis that show any such addition is worthwhile.

Noted, see discussion above.

Action: Review sources referenced in Unikemet2

Where can I find the *Hornung & Schenkel (2007, last modified in 2015), Zeichenlist* as referenced?

Unfortunately, these are copyrighted material that cannot be shared.

Question is what additional references or changes are necessary/desirable to be added to Unikemet2 before proceeding to a formal proposal for an encoding?

Note it is important that reviewers and users can potentially access the source material and we don't leave people in the future in the situation of Hieroglyphica where unpublished sign lists are stated as sources…

Each additional reference is extremely time-consuming. Many of the existing references are still copyrighted and therefore are not easy to retrieve because most of them are not published anymore. The intent of the author of L2/19-220 has been to find attestation for most of Hieroglyphica, using material provided generously by UC Berkeley. More sources and attestations are always welcome.

Action: Unikemet2 Taxonomy

The taxonomy index based on IFAO for the Unikemet2 database seems to me useful with the index using the ((A-IK-Z){1}|AA)-(0-9){2}-(0-9){3} format. This approach was taken by *Inventaire* although the categories used in the Hieroglyphica Systematical list are broader e.g. 5 subcategories of category A rather than the 34 sub-categories of L2/19-220.

I suggest dwarf be treated as a man (or woman if applicable) with no separate category and man in ship moved to varia or some broader sub-category. It's worth looking through all sub-categories to look for any similar simplifications.

N5O63 "Current Issues" lists some important point for review and other issues may be identified by specialists to add to this list.

Consensus over the full taxonomy list over all categories would be an important step in enabling possible expansions to be based on Unikemet2.

The author of L2/19-220 would rather not tinker with the IFAO classification as much as possible.

Action: Define transitional PUA system

With the fruits of expansion maybe 3 or more years away from release, it would be valuable to agree a specification of use of the Unicode Private Use Areas for codepoints during the transitional period. Such a specification will simplify work on expansions and, if managed and defined properly, deliver benefits to specialist users of hieroglyphic…

It is important that code points in such a PUA list are not changed, once specified, for stability both during the transitional period to Unicode extensions release and later when users still may want to refer to what is then obsolete material and old fonts.

I suggest the transitional PUA system be organized by taxonomy index with gaps in the code-space for wiggle room, once the Unikemet2 taxonomy is agreed. This to be deliberately defined so that the exact Unicode organization decided on for expansion is not linked to the PUA values. Software can just use a published mapping table to switch between the PUA and released standard. Font developers can use some automated process (e.g. open-source Python script) to build a release font from a PUA version.

The author of L2/19-220 has been contemplating such a scheme for a while and is willing to do the transition.

Discussion: Content of first extension

Michael Everson proposed an extension in **N4741 L2/16-250** Preliminary proposal to encode Möller's Egyptian Hieroglyphs in the SMP of the UCS. This is specifically oriented at helping with hieratic transliteration to hieroglyphic.

N5063 L2/19-220 Revised draft only details categories A to F as the first and second tranche of the Unikemet2 database but with others to follow.

If there is consensus on using Unikemet2 (with any agreed additions of changes) as the basis for future extensions, there are many ways one could proceed and stage proposals.

N5063 proposes the essential criterion for considering a hieroglyph for encoding be that the weight be 1 or greater. The question, looking towards a formal proposal, is what other criteria might be used to help with selection.

From the point of view of the user base, staging formal extensions to Unicode by category is of limited value. For instance, to release an expansion containing just Anthropomorphic signs (Categories A-C) only makes a marginal improvement to Unicode for work concerning Middle Egyptian hieratic texts. Nevertheless, from a technical perspective it may make sense nevertheless even if it means a slower uptake of Unicode among specialist users for work with hieroglyphic.

The availability of Hieroglyphica and the related Aegyptus font means font availability is assured from the onset even for 1000s of hieroglyphs. However, it is important to recognize that most fonts will not be expected to cover all parts of the extended code space. …

The author L2/19-220 notes the preference for a full repertoire instead of the tranche concept. However, it is likely that most fonts covering more than the classical period which is adequately represented by the current Unicode repertoire will include the full extended characters. As seen by the trend followed by publicly available font (such as Noto) they tend to cover in full historic repertoire, and Egyptian hieroglyphs coverage is a common ask.

Discussion: Unicode Naming Convention

The current Unicode naming convention for hieroglyphs is of the form ((A-IK-Z){1}|AA){3}{A-Z} i.e. category letter, a number padded to left by 0s to 3 digits followed by an optional upper case letter denoting graphic variant This is based on the Hieroglyphica/Gardiner system from which it is easy visually distinguished e.g. A001 vs. A1.

N5O63 proposes hieroglyph names in all extensions are based on the Unikemet2 taxonomy index, so A001 would have been named A-01-001 if it was not already named. This allows the Unicode name to communicate more of its graphical form than the current system. There is no attempt to communicate the variant concept in the name, variants are given their own code and any treatment of variants is done outside of Unicode.

While the original Hieroglyphica taxonomy (derived from Gardiner) was considered it was abandoned for the following reasons:

1. *Some of the groups become too large (over 1000 members for the A category) with similar signs spread randomly through the range.*
2. *While some member names express the concept of variants by using alphabetical prefix (such as A005A, A006A, etc…), this model is not followed consistently in Hieroglyphica and similar schemes.*
3. *Extensions to the original definition are contradicting the Unicode naming convention, and various extensions are contradicting each other.*
4. *Extension to a subgroup of similar looking sign is meaningless. For example, extension to the man sitting (typically in the A1-A9 range) can be found as A71, A75, A426, A478, A497, etc.)*

The system used in L2/19-220 is based on the IFAO taxonomy which is not perfect but has the merit of providing more details within each group. It is does not try to describe variants because variants are somewhat subjective, and it would create a major deviation from IFAO.

> For most technical purposes the name is not terribly important. Software works with Unicode code points and can map these into whatever name space e.g. catalogue number as required e.g. map U+13000 to A001, A-1-001, HG.A1 (Hieroglyphica) or LT.A89 (Lepsius-Theinhardt) or whatever. Hyphen '-' is not allowed in an Adobe Feature file glyph name but a simple substitution like A-1-001 -> A.1.001 could be used. Hyphen is not allowed in MdC glyph names but some workaround could likewise be used or more likely MdC dialects would continue with Hieroglyphica style aliases.

As noted above, each application can substitute its own separator for the '-'. The Unicode naming convention does not allow any symbols in name except for the hyphen.

> Names are the main user-facing feature of Unicode so it is instructive to ask what publications like Wikipedia might look like. Users are used to the Hieroglyphica form of sign names in modern Grammars and other uses of hieroglyphs over the last 30 years. The taxonomy index as name would make sign lists look more complex so it seems unlikely these names will prove popular for less technical publications and with the users referring to them.

> There is a reasonable case to be made for an extension which encodes variants of existing signs using names based on the current system. For instance the Late Egyptian Hieroglyph variant HG.G17A would be encoded as G017A. This avoids introducing unnecessary unfamiliarity into the name space. One might also consider other factors.

> Analysis of Unikemet2 would help characterize the possibilities and enable informed discussion on this topic.

The author has provided above the reasons for not following Hieroglyphica naming scheme. In addition, there are many Unicode blocks where the names are totally meaningless because they are derived by algorithm from their code point (such as CJK Ideographs, Tangut, etc.). The current scheme is aiming at providing a reasonable middle ground where extension to an existing sub-group can be easily identified.

3. Combined signs (especially as relate to L2/19-220)

Question: What are the principles for atomically encoding characters vs. use of format control characters (L2/17-112), especially for stacking, characters with overlays? (Bob Richmond document: https://github.com/HieroglyphsEverywhere/Docs/blob/master/UnicodeRepertoire/EgyptianHieroglyph OverlaysHieroglyphica20190911.pdf
(introduction provided)

> What follows is a list of hieroglyphs from Hieroglyphica (2nd Edition, 2000) that are possible candidates for encoding using the EGYPTIAN HIEROGLYPH OVERLAY MIDDLE format control. This list is not intended to be complete, checked or definitive in any sense.
>
> I'm making this rough draft available to stimulate discussion but the eventual goal is to link check data like this to the Unikemet2 database as well as provide guidance and test data for font developers working on format control implementation.
>
> In general, the candidates for consideration should be 2-character combinations of characters already encoded in Unicode and whose overlay can be considered as "middle". This notion needs to be clarified. I've included some examples of some hieroglyphs that aren't suitable for reasons such as one or both components are not currently in Unicode.
>
> It is important to note that sequence order is crucial for EGYPTIAN HIEROGLYPH OVERLAY MIDDLE. A+B is different to B+A. The data here will contain errors, especially where the order is not clear from the pictures of signs in Hieroglyphica.

The meeting attendees in general liked the concept of using this mechanism for the 2-character combination. It should not be an issue to add components which are not yet encoded to make all these sequences possible.

The sequence order is an interesting issue. It is not clear at all that there is a direct connection between the z order and the reading. Currently ValPhon attributes both reading orders to the same glyph. The z order seems to be a convention for some of the pair (like the leg is mostly behind, independently of the reading order). For example, for D59 (U+130C1 ⸝), Valphon has the values: ꜥb ꜥb3 b b3 bꜥ. In consequence, most meeting attendees disagreed that the sequence order was crucial, or that the reading order could be simply derived from the original carving (in many cases it is ambiguous). Alternatively, if it is crucial, it would mean that both sequences should be created, which would create disunification issues.

4. Corrections to Möller field in L2/19-220 spreadsheet (Svenja Gülden/AKU)

Svenja started to provide these orally during the meeting but it was agreed that she would produce a document and sent it to Michel.

5. IFAO

Questions: Should all IFAO characters be included (and in cases where no attestation has been found, should space be reserved for them)? If added, how to correlate IFAO with Hieroglyphica?

There was no consensus to do this work. No one seems interested to do a digitized version of IFAO.

6. Proposed timeline for submission of Gardiner sets A-F to Unicode Technical Committee

Note: Once submitted to the Unicode Technical Committee, the proposal will be reviewed and will need to be approved. The proposal will also go to the ISO Subcommittee on Coded Character Sets (SC2) and its Working Group (WG2) and will eventually go onto a series of ballots for ISO 10646. The overall approval process takes about 2 years. During the approval process, additional comments can be made.

This was not discussed at the meeting. Bob Richmond has expressed a preference to submit a full repertoire, not tranches.

7. Other topics:

a) Views on additional insert control and shading control characters? (Suggestions from Andrew Glass) Insert Control character: The insert character under consideration is discussed on page 11 of L2/17-112: Cf. L2/16-210R's suggested approach with INSERT_CENTER character (section 6) (Mark-Jan Nederhof): Stéphane Polis is in support of this character.

This is covered to some length in previous part of this document but was not discussed at the meeting.

Shading character: The original proposal for Egyptian Hieroglyphs (N1944) contained shading control characters but there may be objections within the Unicode Technical Committee. One UTC member suggested using building a hack on top of the current syntax, using an existing geometric shade character.

This was not discussed at length at the meeting. Defining a syntax using the overlay control character and an existing geometric shade character is possible, but it is not clear it would be widely implemented (not trivial with current font technology).

8. Next Steps / Wrap-up

This document will be submitted to Unicode Script Committee for further feedback and review.

Acknowledgements: Many thanks to Michel Suignard (and family) for hosting the meeting!