

**WG2 N5163**

**L2/21-108**

***Date: 2021-06-14***

**Title: Consideration for the encoding of an extended Egyptian Hieroglyphs repertoire**

**Source: Michel Suignard**

**Status: Individual Contribution**

**Distribution: UTC, WG2**

**Replace: N5128R (L2/20-068R)**

## Executive Summary:

This document presents the current stage of research on the Egyptian Hieroglyph extensions as done by the author. Unlike previous version, it does not formally propose a set of characters for immediate encoding. Instead, it describes a large set of glyphs (total over 10,350 glyphs) that represent a maximum envelope of candidates in which to pick the appropriate abstract signs. It presents a methodology in how to start from that large set of glyphs, based on priorities, themselves based on criteria such as number of attestations, references by established references, etc.

This document also provides information on the taxonomy of glyphs, the sources of the glyphs, a description of how the priority settings for glyphs were determined (page 8) and encoding principles guidelines (pages 9-10).

Not all glyphs in the list are necessarily candidates for encoding. The list includes compound glyphs, made of smaller combined components, which could be handled as sequences. Mirrored or rotated versions of signs are also included, but these could be encoded as sequences with control characters. Note that a separate proposal for the format characters is being prepared by Mark-Jan Nederhof and Andrew Glass.

Two files are appended to this document:

The first appended file is the exploratory set of 3090 glyphs, presented in a format similar to a character proposal. It contains those characters deemed to have the highest priority (i.e., those in the category A, B, C and D, based on the number of attestations as described on page 8). The current set contains one character located in the last open spot in the existing Egyptian Hieroglyph block (U+1342F) and the remainder are in a new block Egyptian Hieroglyphs Extended-A (U+13450..U+143FF). In this draft document, annotations may contain source information, but this would not typically appear in the actual code chart. It is expected that a derived version of that set could be soon proposed as a first tranche for encoding in the Unicode standard.

The second file is a PDF of a spreadsheet containing 10,350 glyphs, the maximum set of candidates from which characters can be selected for encoding. The list is not intended to replace the full functional database, as is found in Thot Sign List (TSL). The spreadsheet does not include the sequences that could be used to represent the complex glyphs, but the author expects Egyptologists to come to consensus on how to represent the compound signs.

Other comments:

- The sources used to compile the glyph list has been expanded from earlier documents. It now includes elements from the Thot Sign List, as well as information provided by Daniel Werning's team from Thesaurus Linguae Aegyptiae (as determined by usage frequency), the Pyramid texts and Karnak.

- Stability for the list is provided by use of PUA code points on Plane 15 (U+F0000-U+FFFFD). The PUA code points are indicated in the “Index” field in the spreadsheet.
- The question of variants is discussed on page 11.
- The naming scheme has not yet been finalized (see pages 15-16). There has been some discussion amongst Egyptologists on the topic. Some Egyptologists have asked for names that reflect sources, but the source names themselves are not consistent (cf.. bottom of page 12).
- This document has adopted the terminology used by TSL and JSesh (and is described on page 3), which varies from the typical Unicode usage.

A review of the characters is currently being done by Egyptologists. Special attention is required so Ptolemaic characters get adequately reviewed.

## Introduction

Egyptian hieroglyphs have evolved from a set slightly above 700 characters during most of the classical period (from Old Kingdom, Middle Kingdom, to New Kingdom) to a much larger repertoire in the Greco-Roman period. In that era, the number of hieroglyphic characters has expanded to over 7000 – 10000 signs. Typically, that larger set is known as Ptolemaic, but this is a simplistic assumption because many of these extensions have also been found to be in use in the classical period.

The explosion in glyph diversity was also accompanied by an increase in the number of phonetic values (phonemes) that could be assigned to a single glyph, up to twenty or even more readings. Overall, deciphering late usage of the script has proven to be challenging.

This document does not pretend to describe in detail the various complexities of the extended Egyptian Hieroglyphs and its usage. There are many publications on that subject, the bibliographic references section has been extended in the recent versions of this document to covers some related documents on that subject.

Unlike the previous version of this document (WG2 N5218R, L2/20-068R) this version does not represent an encoding proposal. It describes a full set of glyphs (over 10,350) that includes the already encoded 1,071 signs in the Egyptian Hieroglyphs block, and many signs for which there is not even a glyph available in outline forms (font element or SVG). Some but not all these glyphs are candidates for sign encoding, based on their attestation and common use criteria. The document should be seen as a set of considerations, helping progress through the creation of several tranches of Egyptian hieroglyphs through multiple Unicode block extensions.

The glyph list still uses the glyph taxonomy used by the previous document based on the [IFAO], but this does not constitute a final decision on what should be the naming scheme for an eventual encoding. The sign taxonomy is aimed at providing a visual classification of these glyphs. By no mean it aims at creating a functional sign taxonomy which is beyond the scope of this document.

The list of glyphs also contains compound glyphs, made of smaller combined components. For some of them it may make more sense to encode the corresponding sign using sequences of sign components associated with format characters. Others are mirrored or rotated versions, again many of these signs could be encoded using sequences of simpler signs if control characters are available to describe these transforms.

The intent of these considerations is to set the base for the encoding of the extended sets, allowing Egyptologists to communicate data in a stable set. The stability is provided using an immutable Unicode Private User Area (PUA) code for each of these glyphs in the U+F0000-U+FFFFD block. These codes are used as indexes along with the catalog number based on the IFAO classification.

## Terminology

It may be useful to define some terms used in this document (inspired from TSL and JSesh documentation):

- Token – particular sign in situ, a photography, a fac-simile, a character in a font, or a hand-drawn character in a book. In practice, the token will always be a specific drawing, from a font or from a file, perhaps referring to an actual source.
- Sign – (abstract) defined by its linguistic values, according to the usual substitution rules. For instance, there would be a sign for the "owl", regardless of its actual representations, or one for the scribe outfit. The sign is the most abstract representation. A sign can itself be encoded as atomic unit (single Unicode code point) or a sequence of atomic units (compound sign).
- Class – subset of a sign occurrence, according to meaningful and discrete graphical feature, or selection in possible values of the related sign. The classes constitute a layer between the token and the sign.

Applying that terminology to the large glyph list mentioned above, the task is to split the list in these 3 groups. By default, all glyphs are tokens. Some glyphs may be used to represent abstract signs, and other the classes related to these signs. Some list elements can also represent compound signs. It is possible that in creating abstract signs it will be necessary to add signs required to represent compound signs when one sign component is not yet currently attested. Only signs and classes should be encoded as Unicode points. In the rest of this document, the term sign is used to represent the set constituted of the abstract sign and its related classes (unless detailed otherwise).

It should be noted that these principles are not uniformly used in Unicode encoding. Some blocks such as the Phaistos Disc block are encoding tokens, partly because the set is undeciphered. Most of the symbol blocks are mostly pictorial, and therefore are also mostly composed of tokens, even if some leeway is possible in the actual implementation in fonts. Most of the blocks representing writing systems are a mix of signs and classes.

## Sign Taxonomy

The proposed sign taxonomy uses a three-level classification. The higher level is a combination of the Gardiner A-Z (and Aa) classification and the IFAO chapter classification (I to XXX in roman notation). While IFAO has a few more items, they can be easily mapped into existing Gardiner groups (for example the Gods (Chapter III) and Goddesses (Chapter IV) can be combined in the Gardiner group C (Anthropomorphic Deities). The following is the list of the first level groups and their relationship with the IFAO groups:

<b>Gardiner groups</b>	<b>IFAO (translated from French)</b>
A. Man and his occupations	I. Men and monarchs
B. Woman and her occupations	II. Women and monarchs
C. Anthropomorphic deities	III. Gods IV. Goddesses
D. Parts of the human body	V. Human body parts
E. Mammals	VI. Mammals
F. Parts of mammals	VII. Mammal body parts
G. Birds	VIII. Birds
H. Parts of birds	IX. Bird parts
I. Amphibious animals, reptiles, etc.	X. Reptiles, amphibians
K. Fishes and parts of fishes	XI. Fishes and parts of fishes
L. Invertebrate and lesser animals	XII. Insects and arachnids
M. Trees and plants	XIII. Plants
N. Sky, earth, water	XIV. Sky, earth, water
O. Buildings, parts of buildings, etc.	XV. Edifices and parts of edifice

<b>Gardiner groups</b>	<b>IFAO (translated from French)</b>
P. Ships and part of ships	XVI. Boats and parts of boat
Q. Domestic and funerary furniture	XVII. Everyday and funeral furniture
R. Temple furniture and sacred emblems	XVIII. Temple furniture
S. Crowns, dresses, staves, etc.	XIX. Crowns XX. Jewels, clothes, staves
T. Warfare, hunting, butchery	XXII. Warfare, hunting, fishery, butchery
U. Agriculture, crafts, and professions	XXI. Agriculture and shop tools
V. Rope, fiber, baskets, bags, etc.	XXIII. Rope, baskets, bags
W. Vessels of stone and earthenware	XXIV. Vases
X. Loaves and cakes	XXV. Bread loaves
Y. Writings, games, music	XXVI. Writings, games, music
Z. Strokes, signs derived from Hieratic, geometrical figures	XXVII. Geometric shapes
AA. Unclassified	XXVIII. Ill-defined signs

Notes:

- The order of the A-Z and I to XXVIII lists is identical except for the 2 groups T-U and XXI-XXII.
- IFAO Chapter XXIX (Uncertain identity signs) and Chapter XXX (Conventional signs) are not used in the taxonomy because they have no known references in the Hieroglyphica set and are seldom used by other references.
- Many characters originally from the group 'AA.XXVIII Unclassified Ill-defined signs' have been moved to other groups when their identity could be clarified. Some members originally from the IFAO group XXIX have also been reclassified.

To facilitate the transition from the current Gardiner sign taxonomy, it seemed preferable to keep the same naming convention A to Z for the higher level, to eschew J, and to put the uncertain code points in the group 'AA'.

The second level uses the IFAO classification within each IFAO chapter to further enumerate the content of each of these groups. For example, the IFAO classification for 'A: Man and his Occupations' has the following categories (translated from French):

Men and Rulers (homme et souverains)

1. Man seated or kneeling empty handed (homme assis ou agenouillé les mains vides)
2. Man standing empty handed (homme debout les mains vides)
3. Man, head down (homme, la tête en bas)
4. Man adoring or bent (homme adorant ou penché)
5. Man on the ground or in water (homme à terre ou dans l'eau)
6. Man standing, holding a staff (homme debout, tenant un bâton)
7. Man seated or kneeling, holding something (homme assis ou agenouillé, portant quelque chose)
8. Man standing, holding something (homme debout, portant quelque chose)
9. Man seated, pouring water (homme assis, versant de l'eau)
10. Man standing, pouring water (homme debout, versant de l'eau)
11. Man hiding (homme se cachant)
12. Man working (homme au travail)
13. Shepherd and porter (berger et portier)
14. Man carrying a bundle (homme portant un baluchon)
15. Man standing, composed with a hieroglyphic sign (homme debout, en composition avec un signe hiéroglyphique)
16. Man or god holding a weapon (homme tenant une arme)
17. Soldier (soldat)
18. Prisoner and enemy (prisonnier et ennemi)

19. Dancer and acrobat (danseur et acrobates)
20. Musician (musicien)
21. Man and animal (homme et animal)
22. Man in a boat (homme dans un bateau)
23. Child (enfant)
24. Dwarf (nain)
25. Noble seated on a chair (notable assis sur un siège)
26. King or god seated without crown (Roi ou dieu assis sans couronne)
27. King or god standing without crown (Roi ou dieu debout sans couronne)
28. King or god wearing the white crown (Roi portant la couronne blanche)
29. King or god wearing the red crown (Roi portant la couronne rouge)
30. King or god wearing another crown (Roi portant une autre couronne)
31. Dead person kneeling (défunt agenouillé)
32. Mummy (momie)
33. Statue (statue)

This categorization is used in the new taxonomy, with just one additional value (34. Man, varia) added to cover signs that are not attributable to any of the previous sub-groups. Once these members are created, they are immutable, i.e. they cannot be removed or renumbered. However, new members can be created in each of the A to Z and AA first level groups if such a need arises. Finally, the glyphs from Hieroglyphica have been mapped into each of these 2<sup>nd</sup> level sub-groups to create the 3<sup>rd</sup> level content. The format use the following syntax: ((A-IK-Z){1}|AA)-(0-9){2}-(0-9){3}. Examples are A-01-001, M-21-024, AA-08-006.

The following table shows an example for the sub-group A31 made of 13 characters: (1<sup>st</sup> column shows the PUA index, 2<sup>nd</sup>: catalog number, 3<sup>rd</sup>: representative glyph; 4<sup>th</sup>: UCS code point (for already encoded sign); 5<sup>th</sup>: IFAO index. The glyph which is shown as a black silhouette is not part of a font. It was created using a graphic snapshot of the IFAO documentation.

**Table for A31: Dead person kneeling (only showing IFAO source)**

index	Catalog	glyph	UCS	IFAO
F0640	A-31-001		1303D	46;11
F0641	A-31-002			
F0642	A-31-003			
F0643	A-31-004			
F0644	A-31-005			
F0645	A-31-006			46;13
F0646	A-31-007			46;14
F0647	A-31-008			46;15
F0648	A-31-009			47;1
F0649	A-31-010			47;2
F064A	A-31-011			47;3
F064B	A-31-012			
F064C	A-31-013			

## List content

Based on the classification above, a list of glyphs was created using the union of the following collections:

- Characters encoded in the Unicode Egyptian Hieroglyphs block (U+13000..U+1342F)
- Hieroglyphica elements (including addition made by Jochen Hallof in reporting Dendara and Esna attestation)
- Hieroglyphica extensions based on various demands
- JSesh elements (version 7.5.5)
- Thot Sign List elements (online)
- Thesaurus Linguae Aegyptiae (TLA) elements as determined by usage frequency
- Pyramid Text publication of Sethe occurrences
- Karnak sign occurrences
- IFAO entries

Hieroglyphica is by far the biggest contributor to the list but shares approximately 6,800 glyphs with JSesh; this represents the largest intersection between two sources of this list. The list currently contains approximately 10,350 elements. All these elements have an immutable PUA code point. To facilitate extension, room has been left between each sub-group so that related glyphs can be added while staying close to existing ‘similar’ glyphs. At some point the catalog values may also become fixed, but this is not an immediate goal. However, the intent is to keep them as stable as possible.

Once a list was established, it became necessary to triage the list according to various criteria:

- Attestation of use (either as an atomic character or as part of a compound character)
- Priority setting
- Encoding principles

## Attestation of use

In addition to the collections used to create the list, additional references were used to determine the scope of attestation.

The following index and sources were analyzed:

- Gardiner, 1953
- Hieroglyphica index (including addition made by Jochen Hallof in reporting Dendara vol X to XIV attestation)
- Hieroglyphica extension (various sources)
- Douros extensions (Aegyptus)
- Thot Sign List (TSL) accessed online at [http://thotsignlist.org/sign\\_gen](http://thotsignlist.org/sign_gen)
- Thesaurus Linguae Aegyptiae entries, as provided by Daniel Werning
- Pyramid Text entries, as provided by Daniel Werning
- Karnak entries, as provided by Daniel Werning
- JSesh word processor, version 7.5.5
- Unikemet (as presented in existing Unicode Egyptian Hieroglyphs block)
- Dendara, Le fonds hiéroglyphique au temps de Cléopâtre, Sylvie Cauville, 2001
- Catalogue de la fonte hiéroglyphique de l'imprimerie IFAO, 1983
- Christian Leitz – Einführungen und Quellentexte zur Ägyptologie
- Kurth Einführung ins Ptolemaische, 2009
- Hornung & Schenkel (2007, last modified in 2015)
- Möller extensions (as conveyed in L2/16-250)
- Richmond extensions (as conveyed in L2/16-303), see <https://github.com/HieroglyphsEverywhere/Docs/blob/master/Archive/AnalysisOfSomeMdCCodedTexts.pdf>
- Jochen Hallof attestation, referencing Dendara volumes X to XV, Les chapelles Osiriennes, Sylvie Cauville, 1997-2012
- Jochen Hallof attestation, referencing Le Temple d'Esna Volume 7, Serge Sauneron, 2009
- Valeurs Phonétiques, 1988, entries with at least one phonetic entry count for one attestation

NOTE: Other sources such as older publications for Dendara vol 1 to 9 could also be considered because they are referenced in Valeurs Phonétiques. The situation is similar for publications related to the Esna Temple and the Edfou Temple (see the References section , other sources) .

Based on source attestation a weight index was created to denote the number of sources found. In some cases, where multiple sources described a character not found in an existing font, the glyph was created as an extension to the Hieroglyphica set. The following table shows weight values going from 3 to 6, indicating significant usage for the group B06.

**Table for B06: Queen or goddess, seated, wearing a crown (showing all sources with at least a member for this group)**

index	catalog	glyph	GID	TLA	Cauv.	IFAO	Kurth	H&S	JH-D	VP	W
F07A0	B-06-001		B38	2		52;10	1;54c			B;115	4
F07A1	B-06-002		B38A	7	31;4	52;11	1;54c		2	B;111	6
F07A2	B-06-003		B39A	3		52;12				B;109	3
F07A3	B-06-004		B39	41		52;13		B;3401		B;113	4

Legend:

- Index: PUA code point,
- Catalog value
- Glyph: representative glyph,
- GID: Glyph index value (mostly identical to Hieroglyphica index)
- Thesaurus Linguae Aegyptiae (TLA) frequency count,

- Cauv.: Dendara by Sylvie Cauville, reference,
- IFAO: reference,
- Kurt: reference,
- H&S: Hornung & Schenkel reference,
- JH-D: Jochen Hallof Dendara attestation count
- VP: Valeurs Phonétiques reference,
- W: Weight value

Because some signs have no attestation beyond Hieroglyphica and JSesh, their weight value is blank. The list has currently about 1000 items with blank weight values (leaving more than 9350 with some attestation).

## Priority setting

The next step was to use the various sources and attestation to determine an encoding priority value, based on the following criteria:

Priority value	Explanation
A	<ul style="list-style-type: none"> <li>• combined attestation equal or larger than 100</li> </ul>
B	<ul style="list-style-type: none"> <li>• combined attestation equal or larger than 10</li> </ul>
C	<ul style="list-style-type: none"> <li>• TSL entries, all corresponding to well-documented entries in the TSL database collecting description, functional analysis of a given sign</li> <li>• Combined attestation between 4 and 9</li> </ul>
D	<ul style="list-style-type: none"> <li>• Triple attestation</li> <li>• Provide at least one element for a sub-group</li> <li>• Explicit request for encoding for various other reasons</li> </ul>
E	<ul style="list-style-type: none"> <li>• Dual attestation</li> </ul>
F	<ul style="list-style-type: none"> <li>• Single attestation (hapax)</li> </ul>
(blank)	<ul style="list-style-type: none"> <li>• Other entries</li> </ul>

Examples in the group A1:

index	catalog	glyph	GID	Priority	TLA	TSL	JH-D	JH-E	VP	Explanation
F0001	A-01-002		A1F	C		1_82_01				TSL entry
F0002	A-01-003		A1D	D			2		1	Attestation =3
F003D	A-01-005		ES120	D				3		Attestation =3
F0006	A-01-007		A71A	B			89		1	Attestation >=10
F0007	A-01-009		A426B	A	1		106	2		Attestation >=100
F000A	A-01-012		A68	B	2		1	6	1	Attestation >=5
F000B	A-01-013		A72	B				7		Attestation >=5
F003C	A-01-014		ES008	F					1	Attestation =1
F0022	A-01-38		A426C	E	2					Attestation =2

Signs that are already encoded in Unicode such as A1 (U+13000), A3 (U+13002), etc. which have for most of them high usages were not given priority values to simplify the process. At this point, the recommendation is to consider signs with Priority 'A', 'B', 'C', 'D' for the first tranche of encoding.

## Encoding principles

Once the set of candidates has been established through the priority list, encoding principles must be created. The following list enumerates some of them:

- Clear description of the character; the description should list all relevant and distinct features of the sign
- Distinctive traits for signs should be clearly determined (for example: beard types for human, horn for snakes, etc...)
- Any sign which is clearly a quadrat compound sign should not be encoded as an atomic unit
- Any signs which can be encoded by a sequence using existing signs and format characters should be encoded using that sequence as much as possible
- If to create such sequences, a base character is needed, it should be encoded, even if there are no isolated attestation/reference for the base character
- Mirrored signs should not be encoded unless they may occur in the same context as the unmirrored sign with a different function/meaning.
- Rotated signs should typically not be encoded as separate atomic unit, it is expected that additional format characters will be created to allow mirroring/rotation of base characters.

However, these principles are just guidelines, it is always possible to deviate from them, if there are existing practice or preferences to encode sign in other ways. For example, it may be preferable to encode some characters as atomic characters while they could be encoded as sequences. Considering the following signs:

index	catalog	glyph	GID	UCS	TLA	TSL	JH-D	Possible sequence
F25C0	M-12-001		M13A			1_3805_01	2	
F25C1	M-12-002		M14A		1	1_3824_01		
F25C5	M-12-006		M13	131C5		1_3805_00	599	
F25D2	M-12-019		M14	131C6	324	1_3824_00	387	
F25D4	M-12-021		M221					
F25D5	M-12-022		M121		1	1_3794	8	
F25D6	M-12-023		DE352				1	

In the 7 signs above, 5 have TSL references which may be sufficient to justify encoding. Note that the sign corresponding to is not prioritized for encoding (it exists in the database as GID I10D located at index value F21C5). Unless encoded, it would be necessary to use the sequence to represent the signs corresponding to GID M121 and DE352. It

may be simpler to encode the atomic sequences. Note that M14  was encoded as an atomic sequence at U+131C6 in the existing Egyptian Hieroglyphs block.

The other type of signs to be probably encoded as sequences are signs contained into another (there is no inclusion operator yet, so it is replaced by the notation 'inc' in the table below. Some examples:

index	catalog	glyph	GID	UCS	TLA	TSL	JH	Possible sequence
F108D	D-18-046		D177A		7			 inc 
F108A	D-18-043		D177		26		96	 inc (rotated 90deg  )
F2BC0	O-03-050		O92C		7	1_6834		 inc  or  
F2BE0	O-03-083		O369			1_6831		 inc  or    or   *  * 

The last two examples present some interesting issues.

The rectangular enclosure ( *hwt*) is often used to enclose other signs and that usage is extremely productive. Therefore, it is not realistic to try to encode atomically all combinations, because the content itself can be itself made of several combined signs. There are however two possible mechanisms to do so:

- 1) Using a standard base container (full enclosure), followed by the contained object (can itself be a sequence). Implementation of such sequences would probably require the resulting rendered sign to be created as a ligature in the font because the size of the container has to be adjusted to actual content.
- 2) Using a cartouche system with a start enclosure and an end enclosure. While requiring longer sequences, it would allow a completely productive rendering because the container outline is created on the fly. There is no need to create glyphs in the font to represent ligatures. The font simply uses a slightly smaller size of the contained element(s).

The last example shows a common variant of R26 () , identified as R26B in both Hieroglyphica and JSesh () which is the sign  (lung and windpipe) surrounded by lotus stems  (south) and  (north); the phonogram in Valeurs Phonétiques is *sm3-t3wy*. The whole construct corresponds to O369 with a Thot Sign List identifier value of TSL\_1\_6831 and is described as the logogram *hw.t-sm3-t3.wy* (mansion uniting two lands). The encoding of this sequence can be done atomically with a single code point, or as an enclosed content, itself either encoded atomically or again as a sequence.

This document does not present sequences proposal for these signs. It is expected that review by Egyptologists will create a form of consensus for all these compound signs.

## Variant issues

A common criticism of repertoires like Hieroglyphica is that they represent a modern abstraction of what is inherently a much larger repertoire of original classic attestation as found in stones, wood carving, etc. Every publication has struggled with this issue by creating their own abstract reference, using a single sign to instantiate slightly different paleographic evidence (token).

### Pool sign (N37A)

In the Gardiner 1953 Supplement, it is typically represented by the following glyph: 

In Hieroglyphica, the same sign is represented by the following glyph: 

But in looking at Dendara pictorial evidence, it is possible to observe that the abstraction represents both forms. But, looking at Valeurs Phonétiques page 469, the two glyphs are given different values:

623			<i>mr</i> <i>n</i> <i>ntr</i> <i>h</i> <i>br</i> <i>h</i> <i>š</i>	Loret, Dümichen, Devéria, Junker, Junker, Sauneron, Lepsius,	<i>Manuel</i> , n° 502. <i>ZAS</i> 3, 1865, p. 58. <i>BE V</i> , 1897, p. 71. <i>Sprachliche</i> , p. 15. <i>Pylon</i> , p. 52,6, n.2. <i>Ésna VIII</i> , p. 164. <i>Königsbuch</i> , p. 175.
627	2007		<i>mr</i> <i>n</i> <i>nt</i> <i>h</i> <i>s</i> <i>š</i> <i>šw</i>	Fairman, Drioton, Drioton, Fairman, Fairman, Alliot, Daumas,	<i>BIFAO</i> 43, 1945, p. 71, n.1. <i>A.S.A.E.</i> 44, 1944, 119 C. <i>A.S.A.E.</i> 44, 1944, p. 119. <i>A.S.A.E.</i> 43, 1943, p. 238. <i>BIFAO</i> 43, 1945, p. 77. <i>Culte I</i> , p. 248. <i>Mammisis</i> , p. 305, n.4.

The Thot Sign list also gives them two class values: TSL\_1\_4265\_03 and TSL\_1\_4265\_04 with different descriptions. Based on this, it may be prudent to encode them separately to maintain the distinction in various contexts.

### Knife sign

The knife sign is currently encoded at U+1332A with a round handle and looks very similar to the knife sharpener sign at U+1332B. The very common knife variant with a straight handle is not encoded. In addition, there are some evidence of rotated and or mirrored versions as shown below (the knife pointing down is both mirrored and rotated).

index	catalog	glyph	Priority	GID	UCS	TLA	TSL	JH-D	Explanation
F3C70	T-18-001		A	T30A		374	1_5917	377	Knife with a straight handle
F3C71	T-18-002		C	T30B		2			Horizontal version of above
F3C72	T-18-003								More oblique version of above
F3C73	T-18-004								Pointing down version of above
F3C91	T-18-034		na	T31	1332B	156	1_5919		Knife sharpener
F3C92	T-18-035		C	T31A		1			Knife sharpener (rounded variant)
F3C94	T-18-037		na	T30	1332A	825	1_5916	1	Knife with a round handle

Two signs have no font representation (T-18-003 and T-18-004), and two signs are already encoded (T30 and T31). The others have priority values of A and C and should be candidate for encoding.

## Exploratory Project

While it is not possible to offer an encoding proposal until more issues are settled, it is still useful to present a set of glyphs as an exploratory project enumerating a list containing prioritized glyphs (for example priority category 'A', 'B', 'C', and 'D'). This would allow experts to solve variant issues as well as atomic versus sequence representation.

While the format is similar to an encoding proposal and use the code points reserved for the Egyptian Hieroglyph extension, it is not an encoding proposal. It would have been possible to use another Private Use Area (PUA) location to avoid the appearance of encoding in the UCS code space, but this would be even more confusing, because the glyph list already uses PUA code points as immutable references.

The current set was created using the methodology explained above and is proposed as follows:

- 1342F: 1 code point to complete the existing Egyptian Hieroglyphs block (U+13000..U+1342F): EGYPTIAN HIEROGLYPH AA030A (catalog number: O-23-002)
- 13450..14xxx: xxxx code points in a new block Egyptian Hieroglyphs Extended-A (U+13450..U+143FF) covering the following all sections A to Z and AA:

While the code chart uses a format similar to an encoding proposal, it presents annotation with some source references such as TSL, Hieroglyphica, and JSesh that would typically not be present in an actual code chart. These annotations are included to facilitate further study before actual encoding.

Catalog entries which have not been categorized as priority A, B, C, or D are not listed in the exploratory list and have no reserved code points. For example, considering the set A31 previously described, the current code chart is not reserving code points for A-31-002, A-31-003, A-31-005, and A-31-007 to A-31-013:

### A31. Dead person kneeling

*A-31-001 referenced as A052*

- 1303D  egyptian hieroglyph a052
- 13580  EGYPTIAN HIEROGLYPH A-31-004  
• HG/JSesh: A52J
- 13581  EGYPTIAN HIEROGLYPH A-31-006  
• HG/JSesh: A52A

The following example for the group I-06 Viper shows references for TSL, Hieroglyphica (HG), and JSesh, including cases where they are different:

### I06. Viper

- 13A36  EGYPTIAN HIEROGLYPH I-06-001  
• TSL: 1\_63 HG/JSesh: I9A
- 13A37  EGYPTIAN HIEROGLYPH I-06-004  
• TSL: 1\_58\_02 HG: I25C JSesh: I25A
- 13A38  EGYPTIAN HIEROGLYPH I-06-005  
• TSL: 1\_58\_00 HG: I25A JSesh: I25C
- 13A39  EGYPTIAN HIEROGLYPH I-06-006  
• TSL: 1\_57\_01 HG/JSesh: I24A
- 13A3A  EGYPTIAN HIEROGLYPH I-06-008  
• TSL: 1\_57\_03 HG/JSesh: I126
- 13A3B  EGYPTIAN HIEROGLYPH I-06-009  
• HG/JSesh: I9C
- 13A3C  EGYPTIAN HIEROGLYPH I-06-010  
• TSL: 1\_6946 HG/JSesh: I114

A full code chart for the existing block (Egyptian Hieroglyphs) and the explanatory list of 3090 signs shown in one new block (Egyptian Hieroglyphs Extended-A) is appended to this document. Note that the code charts for the existing block (U+13000..U+1342F) contains highlighted content (in blue) to show recent glyph changes introduced in Unicode 14.0.

## Glyph database

One of the essential tools of this Egyptian hieroglyphs inventory is a glyph database containing all listed entries, encoded or not, collecting all ancillary data concerning each of these entries. It is a work in progress as new elements are added for each of these entries. Currently, these elements concern source references but could be extended to cover additional information. The reference [Database] is a PDF representation of that database. It is available as a separate file associated to this document.

The current format of the database is an Excel spreadsheet with the following columns:

- Taxonomy index or PUA index. The taxonomy indexes are used as group headers for the following rows. When present, no other columns are defined. For other rows, the column contains the PUA code for the sign.
- Catalog number, using the taxonomy format described above.
- Representative glyph – may be a font glyph or a picture. The pictures use a black silhouette format, but the differences between font glyphs using vector outlines and the black filled silhouette obtained by scanning the original IFAO publication are purely a result of using different technology and are not intentional.
- Code point (UCS) – only present for value below U+13430 and represents the actual Unicode code point when a sign is already encoded. The code point U+1342F is an exception. When additional characters are approved for encoding, these rows will also acquire a UCS value.
- Priority – consisting of A, B, C, D, E, F or blank.
- Glyph index – This index is used to reference all glyphs in the repertoire. It is based on the Hieroglyphica index (next column) but it also contains values corresponding to extensions caused by various needs, such as Douros unique reference, JSesh disunification, TSL specific entries, other addition request). Some provisional values (noted as DExxx) used in Dendara referencing are also recent additions.
- Hieroglyphica (HG) index – As included in the Book 2000 edition.
- Hieroglyphica extension (HGx) – These entries corresponds to value added by Jochen Hallof after publication of the 2000 edition and includes entries to represent signs used in the latest Dendara Sylvie Cauville publication (vol X to XV).
- George Douros Aegyptus font index extension – these are mostly used to represent hieroglyphs that were not in the original Gardiner set, or in case where Hieroglyphica glyphs were different from Gardiner for the same reference. The bulk of that font index is identical to Hieroglyphica but was modified to also represent the Unicode Egyptian Hieroglyph blocks.
- JSesh sign index – based on the JSesh word processor version 7.5.5; JSesh is a word processor for ancient Egyptian hieroglyphic texts. Its sign collection uses a syntax which matches most of the Hieroglyphic indexes.
- Sum – attestation summary of the PT, TLA, and Karnak columns
- Pyramid Text (PT) frequency count, as collected by Daniel Werning's team,
- Thesaurus Linguae Aegyptiae (TLA) frequency count, also collected by Daniel Werning's team,
- Karnak frequency count, also collected by Daniel Werning's team,
- Thot Sign List – abstract signs and classes (1 prefix), the 'TSL\_' prefix is omitted

- Gardiner index based on publications from 1928 to 1953. Some indexes were collected from the 1957 Egyptian Grammar publication. All these references are already encoded in the original Egyptian Hieroglyphs block.
- Unikemet database index – based on document [L2/06-355] (same as ISO/IEC SC2/WG2 N3182) and [L2/07-097] (same as WG2 N3237).
- Dendara – Sylvie Cauville, 2001 [Cauville]. This reference uses both Hieroglyphica glyph id and IFAO index, which makes it very useful to confirm these mappings. The book references are in “page number; item number in page”.
- Catalogue de la fonte Hiéroglyphique de l’imprimerie de l’I.F.A.O., 1983 [IFAO]. This column represents IFAO entries using page number and item number within the page. Given the different styles used between Hieroglyphica glyphs and symbols used in IFAO, matching the symbols has been challenging.
- Einführungen und Quellentexte zur Ägyptologie, Louise Gestermann und Christian Leitz [Leitz]. This reference uses Hieroglyphica glyphs and Id. The book references are in “page number; item number in page”.
- Einführung ins Ptolemäische, Dieter Kurth [Kurth]. This reference uses glyphs very similar to Hieroglyphica but with additions. Some of these additions were added to the repertoire. The book references are in “page number; item number in page”.
- Hornung & Schenkel (2007, last modified in 2015). The references use the 4-digit numbers found in each A-Z and AA categories in the source in the form ‘A;0100’. Sometimes, only glyphs were provided without a clear numeric reference, in those case incomplete references using ‘X’ to denote approximation. For example, the reference ‘O;06XX’ indicates an entry associated with references found in the range O;0600 to O;0699, but not having an explicit numeric value.
- Rainer Hannig, Großes Handwörterbuch Ägyptisch-Deutsch, 1995. The reference uses a glyph and identifier similar to Hieroglyphica and JSesh. The book references are in “page number; item number in page”.
- Yvonne Bonnamy, Dictionnaire des Hiéroglyphes. The reference also uses a glyph and identifier similar to Hieroglyphica and JSesh. The book references are in “page number; item number in page”.
- Möller extensions (as conveyed in L2/16-250) <http://www.unicode.org/L2/L2016/16250-n4741-moller-egyptian.pdf>, references that could be identified in Hieroglyphica were added as referenced sources in the database. The references used for that column are the ones (slightly edited for concision) provided in annotation of the name list in L2/16-250. For example, that document proposes EGYPTIAN HIEROGLYPH A084 with annotation ‘Moeller a020bis v3’, the reference column in the database has the value ‘a20bis v3’.
- Bob Richmond extensions (as conveyed in [L2/16-303]), see <https://github.com/HieroglyphsEverywhere/Docs/blob/master/Archive/AnalysisOfSomeMdCCodedTexts.pdf>, the references used are the ‘CODE’ values in Table 1.
- Jochen Hallof, referencing: Dendara (JH-D), Les chapelles Osiriennes, Sylvie Cauville, Volumes X to XV, 1997-2012 and La Porte d’Isis, 1999. The index uses numeric values showing the number of attestations in the documents; for example ‘1’ indicates an hapax).
- Jochen Hallof, referencing: Esna VII (JH-E), Le Temple d’Esna Volume 7, Serge Sauneron, 2009. The index uses numeric values showing the number of attestations in the documents; for example ‘1’ indicates an hapax).
- Valeur Phonétiques des signes Hiéroglyphiques d’Époque Gréco-Romaine (VP), Montpellier, 1988. The index uses the categories A-Z. For the current subsets, because God and Goddess use separate the same index space in VP, the goddess value has been shifted by 1000 to create unique values for the goddess entries (C;1 -> C;1001).
- The actual value associated with a VP index, multiple values are separated by ‘;’. These values are Latin transliteration of the graphemes of these elements without determination of their functional taxonomy classification. This value does not count for the weight index calculation. But it is used to determine the priority.
- Weigh index – computed by accumulating the number of references found for a given entry. Only references belonging to Gardiner, and columns on the left are counted. The weight index can be blank for entries with no references.

- Comment column – presenting remarks on the reference, note that the ‘~~~~’ notation indicates some imperfect matching.

The following table shows the ‘valeurs phonétiques’ and corresponding TSL values for the group A11:

**Table for A11: Man Hiding (showing Valeurs Phonétique values)**

index	Catalog	glyph	UCS	TSL	VP values	TSL values
F01F0	A-11-001				<i>h3p</i>	
F01F1	A-11-002		13004	1_848_01		Logogram: jmn Phonemogram: dg
F01F2	A-11-003		13005	1_848_00	<i>imn;rm</i>	Logogram: jmn Phonemogram: dg
F01F3	A-11-004					
F01F4	A-11-005				<i>imn;nm</i>	
F01F5	A-11-006				<i>thn</i>	
F01F9	A-11-007					
F01F6	A-11-008				<i>imn;thn</i>	
F01F7	A-11-009					
F01F8	A-11-010				<i>imn</i>	

(note that *imn* and *jmn* are equivalent, just using different transliteration for the yod)

Other columns may be added to the database IN THE FUTURE, such as the following:

- Description – free form description of the glyph, unless already defined by a TSL entry
- Components – list of all glyph components
- Sequences
- Variants – similar entries

## Source reference file and character name issue

A common concern is the identification of the encoded signs, and it has been mentioned several times that it would be advisable to have the character names containing some form of source information, possibly in a form close to the Gardiner syntax. The main issue with that concept is that various collections, including Hieroglyphica, JSesh, Unikemet, all diverge in their own way from the original Gardiner name space and therefore creates name collision. In addition, Unicode character names are immutable, i.e. once encoded the names cannot be changed, even if they contain mistakes, which would be an issue if they contain source information, and if that information turned out to be erroneous.

The preferred solution is to use a neutral character name (very often a combination of the block name and the code point) and push the identification to an auxiliary data file. For example, considering the following 4 characters (the names currently use the catalog numbers but could also just be EGYPTIAN HIEROGLYPH-141E7, EGYPTIAN HIEROGLYPH-141E8, etc.):

13F3F 𐩀 EGYPTIAN HIEROGLYPH U-09-077  
 • TSL: 1\_6126\_03  
 13F40 𐩁 EGYPTIAN HIEROGLYPH U-09-078  
 • HG: U25  
 13F41 𐩂 EGYPTIAN HIEROGLYPH U-09-079  
 • HG/JSesh: U25A  
 13F42 𐩃 EGYPTIAN HIEROGLYPH U-09-081  
 • TSL: 1\_6126\_02 HG/JSesh: U24C  
 13F43 𐩄 EGYPTIAN HIEROGLYPH U-09-082  
 • HG: U24H JSesh: U24G

It is possible to create a data file with following entries:

U+13F3F	kTSL	1_6126_03
U+13F40	kHG	U25
U+13F41	kHG	U25A
U+13F41	kJSesh	U25A
U+13F42	kTSL	1_6126_02
U+13F42	kHG	U24C
U+13F42	kJSesh	U24C
U+13F43	kHG	U24H
U+13F43	kJSesh	U24G

If a unique Gardiner style attribute is desired, it could be included in the same datafile as (the values shown below correspond to the GID used in the spreadsheet):

U+13F3F	kID	U25F
U+13F40	kID	U25
U+13F41	kID	U25A
U+13F42	kID	U24C
U+13F43	Kid	U24H

All these entries describe attributes related to the code points, typically described as Unicode properties, and unlike the character names, they do not have to be immutable and can be fixed if they are erroneous. In addition, because each property may have its own name space, they do not collide with each other.

## Conclusion

The work presented in this document is still preliminary and some of the information is presented in a rough format. However, the base is now a spreadsheet containing all the pertinent information which can be used to extract correlation points and can be augmented easily with ancillary data to make encoding decision easier as well as facilitate search in this vast set.

## References

### Repertoire:

[L2/06-355]: Source for the encoding of Egyptian Hieroglyphs (WG2 N3182), Michael Everson, 2006-10-29, <http://www.unicode.org/L2/L2006/06355-n3182-egyptian-sources.pdf>

[L2/07-097]: Proposal to encode Egyptian Hieroglyphs in the SMP of the UCS, Michael Everson and Bob Richmond, 2007-04-10 (the content was slightly modified for incorporation in Unicode) <http://www.unicode.org/L2/L2007/07097-n3237-egyptian.pdf>

[L2/16-079], Preliminary draft for the encoding of an extended Egyptian Hieroglyphs repertoire, Michel Suignard, 2016-04-11, <http://www.unicode.org/L2/L2016/16079-hieroglyphs.pdf>

[L2/16-257], Source analysis of an extended Egyptian Hieroglyphs repertoire (Hieroglyphica based), Michel Suignard, 2016-09-19, <http://www.unicode.org/L2/L2016/16257-n4751-hieroglyphs-new.pdf>

[L2/16-250], Preliminary proposal to encode Möller's Egyptian Hieroglyphs in the SMP (WG2 N4741), Michael Everson, 2016-09-14, <http://www.unicode.org/L2/L2016/16250-n4741-moller-egyptian.pdf>

[L2/16-307], Towards an Expansion of the Unicode Hieroglyph repertoire, Bob Richmond, 2016-10-28 <http://www.unicode.org/L2/L2016/16307-hieroglyph-repertoire.pdf> , and linked document, Analysis of Unicode Egyptian hieroglyphs in a collection of Mdc-coded transcriptions: <https://github.com/HieroglyphsEverywhere/Docs/blob/master/Archive/AnalysisOfSomeMdCCodedTexts.pdf> .

[L2/17-073], also WG2 N4788, New draft for the encoding of an extended Egyptian Hieroglyphs repertoire, Michel Suignard, 2017-03-19, <http://www.unicode.org/L2/L2017/17073-n4788-hieroglyphs.pdf>

[L2/17-415], also WG2 N4924, Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, Michel Suignard, 2017-11-27, <https://www.unicode.org/L2/L2017/17415-n4924-hieroglyphs.pdf>

[L2/18-165], also WG2 N4944, Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, Michel Suignard, 2018-04-17, <http://www.unicode.org/L2/L2018/18165-n4944-hieroglyphs.pdf>

[L2/19-165], also WG2 N5063, Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, / First Tranche: Human, God and Goddess / Second Tranche: Human parts, Mammals, Mammal parts, Michel Suignard, 2019-06-20, <http://www.unicode.org/L2/L2018/18165-n4944-hieroglyphs.pdf>

[L2/20-068R], also WG2 N5128R, Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, all groups (A to Z and AA), Michel Suignard, 2020-03-12, <https://www.unicode.org/L2/L2020/20068r-n5128r-ext-hieroglyph.pdf>

### Structure

[L2/16-210R]: A system of control characters for Ancient Egyptian hieroglyphic text (updated version), Mark-Jan Nederhof et al, 201-01-25 <http://www.unicode.org/L2/L2016/16210r-egyptian-control.pdf>

[L2/16-232]: Preliminary analysis of Egyptian Hieroglyph quadrat type, Andrew Glass, 2016-08-04, <http://www.unicode.org/L2/L2016/16232-quadrat-types.pdf>

[L2/17-112R]: A method for encoding Egyptian quadrats in Unicode, Andrew Glass, et al, 2017-05-12, <http://www.unicode.org/L2/L2017/17112r-quadrat-encoding.pdf>

**Other sources** (beyond those mentioned in repertoire references above)

[Leitz] Einführungen und Quellentexte zur Ägyptologie, Louise Gestermann und Christian Leitz, Band 2, 2004, LIT, ISBN 3-8253 7340-4

[Kurth] Einführung ins Ptolemäische, Eine Grammatik mit Zeichenliste und Übungsstücken, von Dieter Kurth, Teil 1, Backe-Verlag, 2009

Hornung & Schenkel (2007, last modified in 2015), Zeichenliste, Prof. Dr. E. Hornung, edited by Stéphane Polis.

[Hannig] Großes Handwörterbuch Ägyptisch-Deutsch, Rainer Hannig, Marburger Edition, Philippe von Zabern, 1995, ISBN-13: 978-3805317719

[Hieroglyphica] Hieroglyphica – Sign List, Nicholas Grimal, Jochen Hallof, Dirk van der Plas, 2<sup>nd</sup> edition, 2000

[Dictionnaire] Dictionnaire des Hiéroglyphes, Yvonne Bonnamy, Troisième Edition. 2019. Actes Sud. ISBN -13 : 978-2-330-12518-9

[JSesh] Rosmorduc, Serge. (2014). *JSesh Documentation*. [Online, current version 7.5.5] Available at: <http://jseshdoc.genherkhopeshef.org> [Accessed Feb 23rd 2021]

[IFAO] IFAO, Catalogue de la fonte hiéroglyphique de l'imprimerie de l'I.F.A.O., Institut Français d'Archéologie Orientale du Caire, 1983, IF607, SEVPO, Paris, France

[Cauville] Dendara, Le fonds hiéroglyphique au temps de Cléopâtre, Sylvie Cauville, 2001, Cybèle, ISBN 2-9516758-0-1

[Dendara X to XV] Le Temple de Dendara, Les chapelles Osiriennes, Sylvie Cauville, Institut Français d'Archéologie Orientale, Tomes X to XV, Le Caire, 1997-2012 (composition hiéroglyphique Jochen HALLOF et Hans VAN DEN BERG). X: ISBN 2-7247-0199-2, XI: ISBN 2-7247-0279-4, XII: ISBN 978-2-7247-0460-0. Online versions available at: <https://www.ifao.egnet.net/publications/catalogue/>

[Dendara PI] Le Temple de Dendara, La Porte d'Isis, Sylvie Cauville, Institut Français d'Archéologie Orientale. Le Caire, 1999 (composition hiéroglyphique Jochen HALLOF et Hans VAN DEN BERG). Online versions available at: <https://www.ifao.egnet.net/publications/catalogue/>

[Edfou 1 to 15] Le Temple d'Edfou, Émile Chassinat, Institut Français d'Archéologie Orientale. Le Caire, 1897-1934 (some correction done in 1960 and later). Online versions available at: <https://www.ifao.egnet.net/publications/catalogue/>

[Esna I to VIII] Le Temple d'Esna, Serge Sauneron, Institut Français d'Archéologie Orientale. Le Caire, 1934-2009 (composition hiéroglyphique Jochen HALLOF for vol VII). Online versions available at: <https://www.ifao.egnet.net/publications/catalogue/>

[ValPhon] Valeurs Phonétiques des signes hiéroglyphiques d'époque Gréco-Romaine, Préface de Jean-Claude Goyon, 1988, Université de Montpellier.

[TLA] Thesaurus Linguae Aegyptiae, link: <https://aew.bbaw.de/tla/servlet/S05?d=d001&h=h001>

[TSL] Thot Sign List, Stéphane Polis, Serge Rosmorduc, et al. University of Liège and Berlin-Brandenburgische Akademie der Wissenschaften, general link: <http://thotsignlist.org/> , for a given TSL sign TSL\_1\_ID, use <<http://thotsignlist.org/mysign?id=ID>>

## Acknowledgements

The author wants to thank the following persons for their contribution to the effort: Debbie Anderson and Ken Whistler for their kind feedback to the previous drafts, Jochen Hallof and Barbara Richter for their contribution to the determination of the mapping to the existing encoded Egyptian hieroglyphs as well as the additional extensions beyond the original Hieroglyphica set. In addition, Barbara Richter and Debbie Anderson provided many documents helping the creation of this document. Important other resources were the original Egyptian Hieroglyphs proposal created by Michael Everson, new material created by Bob Richmond, as well as the Aegyptus font created by George Douros.

The proposal also benefited from discussions and ideas exchanged with many other Egyptologists during meetings held in Cambridge, UK in July 2016, London, UK in June 2018, Tournan, France in June 2019, and more recent discussions among the same group of experts. In particular, Daniel Werning (Humboldt-Universität zu Berlin) has contributed significantly in improving the glyphs of glyphs used for the representation of already encoded signs and providing the information concerning the Thesaurus Linguae Aegyptiae (TLA) attestation; Stéphane Polis (Université de Liège) provided mapping between Hieroglyphica/JSesh and the Thot Sign List IDs.

This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.