

**Proposed Update** Unicode® Standard Annex #57**UNICODE EGYPTIAN HIEROGLYPH DATABASE  
(UNIKEMET)**

Version	Unicode 16.0.0
Editors	Michel Suignard
Date	2025-01-06
This Version	<a href="https://www.unicode.org/reports/tr57/tr57-4.html">https://www.unicode.org/reports/tr57/tr57-4.html</a>
Previous Version	<a href="https://www.unicode.org/reports/tr57/tr57-3.html">https://www.unicode.org/reports/tr57/tr57-3.html</a>
Latest Version	<a href="https://www.unicode.org/reports/tr57/">https://www.unicode.org/reports/tr57/</a>
Latest Proposed Update	<a href="https://www.unicode.org/reports/tr57/proposed.html">https://www.unicode.org/reports/tr57/proposed.html</a>
Revision	3

**Summary**

*This document describes the organization and content of the Egyptian Hieroglyph database.*

**Status**

*This is a **draft** document which may be updated, replaced, or superseded by other documents at any time. Publication does not imply endorsement by the Unicode Consortium. This is not a stable document; it is inappropriate to cite this document as other than a work in progress.*

**A Unicode Standard Annex (UAX)** forms an integral part of the Unicode Standard, but is published online as a separate document. The Unicode Standard may require conformance to normative content in a Unicode Standard Annex, if so specified in the Conformance chapter of that version of the Unicode Standard. The version number of a UAX document corresponds to the version of the Unicode Standard of which it forms a part.

Please submit corrigenda and other comments with the online reporting form [\[Feedback\]](#). Related information that is useful in understanding this annex is found in Unicode Standard Annex #41, [“Common References for Unicode Standard Annexes.”](#) For the latest version of the Unicode Standard, see [\[Unicode\]](#). For a list of current Unicode Technical Reports, see [\[Reports\]](#). For more information about versions of the Unicode Standard, see [\[Versions\]](#). For any errata which may apply to this annex, see [\[Errata\]](#).

## Contents

- 1 [Introduction](#)
  - 2 [Mechanics](#)
    - 2.1 [Database Design](#)
    - 2.2 [Unikemet.txt](#)
  - 3 [Property Types](#)
    - 3.1 [Catalog Indexes](#)
    - 3.2 [Sources](#)
    - 3.3 [Description](#)
    - 3.4 [Function](#)
    - 3.5 [Core](#)
    - 3.6 [Mirroring and Rotation](#)
  - 4 [The Properties](#)
    - 4.1 [Alphabetical Listing](#)
    - 4.2 [Listing by Version of Addition to the Unicode Standard](#)
  - 5 [History](#)
  - [Appendix: Encoding Principles](#)
  - [References](#)
  - [Acknowledgements](#)
  - [Modifications](#)
- 

## 1 Introduction

The Unikemet database is the repository for the Unicode Consortium's collective knowledge regarding the Egyptian hieroglyphs contained in the Unicode Standard. It contains ancillary data to help implement support for the Egyptian hieroglyphs. (The term 'kemet' meant 'black land' in old Egyptian and was used as the official name of their country.)

Formally, Egyptian hieroglyphs are defined within the Unicode Standard via their names and assigned code points. However, while the first block: Egyptian Hieroglyphs (U+13000..U+1342F) has character names based on the Gardiner convention, the extended block: Egyptian Hieroglyphs Extended-A (U+13460..U+143FF) use algorithmic names of the type EGYPTIAN HIEROGLYPH-xxxxx where xxxxx is the 5-digit hexadecimal value of the code point, therefore providing little information about the identity of the character. The ancillary data provided by the database define additional information such as a detailed description of the character, various sources, catalog entries, and function. It also defines properties related to these hieroglyphs, such as belonging to a Core set, whether they rotate or not, and whether they mirror or not.

This document is a guide to that data, describing the mechanics of the Unikemet database, the nature of its contents, and the status of the various properties.

## 2 Mechanics

### 2.1 Database Design

The database consists of a number of fields containing data for each Egyptian hieroglyph in the Unicode Standard. The fields, all of which correspond to properties, have names that consist entirely of ASCII letters and digits with no spaces or other punctuation except for underscore. For historical reasons, they all start with a lowercase k.

All data in the Unikemet database is stored in UTF-8 using Normalization Form C (NFC). Note, however, that the “Syntax” descriptions below, used for validation of property values, operate on Normalization Form D (NFD), primarily because that makes the regular expressions simpler.

## 2.2 Unikemet.txt

Included with the [UCD] is a file called `Unikemet.txt`. This is a snapshot of the public contents of the Unikemet database as of the release date for this version of the Unicode Standard.

The file is a single text file, in UTF-8, NFC, and using Unix line endings which contain the values for all properties in the Unikemet database. Properties are described by categories in this document but are nevertheless included in a single file (unlike, for example, the UniHan database).

In this file, blank lines may be ignored; lines beginning with # are comment lines used to provide the header and footer. Each of the remaining lines is one entry, with three, tab-separated fields: the Unicode Scalar Value, the property name, and the value for the property for the given Unicode Scalar Value. For most of the properties, if multiple values are possible, the values are separated by spaces. No hieroglyph may have more than one instance of a given property associated with it, and no empty properties are included in `Unikemet.txt`.

There is no formal limit on the lengths of any of the property values. Any Unicode characters may be used in the property values except for control characters (especially tab, newline, and carriage return). Note that unlike UniHan, double quotes are allowed but are discouraged, and will likely be removed in a future version.

The data lines are sorted by Unicode Scalar Value and property type as primary and secondary keys, respectively.

The file’s header includes a summary of the properties the file contains.

## 3 Property Types

The data in the Unikemet database serves a multitude of purposes, and the properties are most conveniently grouped into categories according to the purpose they fulfill. We provide here a general discussion of the various categories, followed by a detailed description of the individual properties, alphabetically arranged.

### 3.1 Catalog indexes

Two catalog indexes are defined: `KEH_Cat` and `KEH_UniK`. The catalog index `KEH_Cat` is defined using a sign taxonomy based on a publication by Institut Français d’Archéologie Orientale (IFAO), see [KEH\\_IFAO](#). It is written using a three-level classification: a group index, a sub-group index, and an index within that sub-group. The higher level, the group index, is a combination of the Gardiner A-Z (and Aa) classification and the IFAO chapter classification (I to XXX in Roman notation). The second level uses the IFAO sub-chapter classification already present in the IFAO publication. The third level is a new index and just orders items sequentially within each sub-group. For example, the catalog index 'A-01-001' represents the first element, designated by 001, of the sub-group 'A-01'. The element 01 in 'A-01' represents the first sub-group of the group 'A'.

Within the group level, IFAO may include a few more items, but these can be easily mapped into existing Gardiner groups. For example, the IFAO groupings 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:

Gardiner groups	IFAO (translated from French)
A. Man and his occupations	I. Men and monarchs
B. Woman and her occupation	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 edifices
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 workshop 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-XXVIII lists is identical, except for the two groups XXI and XXII (which correspond to the groups U and T, respectively).
- IFAO Chapter XXIX (Uncertain identity signs) and Chapter XXX (Conventional signs) are not used in the taxonomy because they are seldom used by other references.
- Some characters originally in the group 'AA.XXVIII Unclassified Ill-defined signs' have been moved to other groups when their identity could be confirmed. Some members originally in the IFAO group XXIX have also been reclassified.

Because this catalog number is still a work in progress, its status is provisional.

The `KEH_UniK` catalog index was originally defined exclusively for the original Unicode Egyptian Hieroglyph block and is part of the formal character name for these code points. This catalog index has been extended to cover all newly encoded signs. The code points which refer to the same Hieroglyphica and JSesh source value use the prefix HJ followed by a space and the common value between Hieroglyphica and JSesh but zero padded to 3 digits. For example, the catalog index for U+1346C is HJ A072A indicating that the code point is associated with the same Hieroglyphica and JSesh source value: A72A. New entries not common to Hieroglyphica and JSesh were given new values without a prefix. The main rationale for the catalog index is to provide a Gardiner-like notation for all Egyptian hieroglyphs, which is a feature requested by Egyptologists. A significant issue is that the name space shared among the original Gardiner notation, the Unikemet original catalog index, Hieroglyphica and JSesh values has many collisions. For example, U+1304E has A71 as sources for both Hieroglyphica and JSesh, but was assigned to A069 in the original block. In comparison, U+1346A in the extended block has A69 as sources for both Hieroglyphica and JSesh. To avoid an apparent name collision, the catalog index for this character is not HJ A069, but A069A. Therefore, the notation 'HJ' is only used for new characters when the common Hieroglyphica and JSesh source values do not collide with `KEH_UniK` values used in the original block.

### 3.2 Sources

Sources are among the normative parts of the Unikemet database, and refer to some well-known Egyptian hieroglyphs collections. These sources are defined as `KEH_HG`, the Hieroglyphica classification, `KEH_JSesh`, the JSesh index, and `KEH_IFAO`, the IFAO entries. While these values are normative, they are not immutable. Some values may be a matter of interpretation or may contain errors. Many of these sources only use glyphic evidence, don't refer to the original paleographic attestations, and don't provide a formal description of the referred sign.

Detailed descriptions of the syntax used for these sources are to be found in [Section 4.1, Alphabetical Listing](#), below.

### 3.3 Description

While the description `KEH_Desc` is only informative, it is an essential part of the identity of an Egyptian hieroglyph. Because many attestations of these signs are imprecise, due to the imperfect preservation of the original evidence, Egyptologists had to come to a rough consensus on how to describe the abstract form of these signs as precisely as possible. While this description still allows variation in the font style used for their representation, it is expected that all these variants will adhere to the description as stated by this property. Due to the complexity of some of these signs, the description can be a rather long expression.

For example, the description for U+13A6E reads as follows:

'A ram (*Ovis longipes palaeo-aegyptiacus*), standing, without a beard, with a cobra (*Naja haja*), standing up, with expanded hood (*Uraeus*)(l64) on its head, with the wings of a bird on its back, spread in a v-shape.'

Note that the description currently uses the Hieroglyphica/JSesh references in many of these descriptions to designate another sign included in the sign. The example above, 'l64' refer to U+13D79 which is itself described as 'A cobra (*Naja haja*), standing up, with

expanded hood (Uraeus)'. Because Hieroglyphica and JSesh do not always coincide, in case of differences, the JSesh reference prevails.

### 3.4 Function

The function type `KEH_Func` and its corresponding function value `KEH_FVal` are only provisional; they are still a work in progress. All signs are expected to have a function type representing either a pictogram, a logogram, a phonogram (or “phonemogram”), a classifier (or “determinative”), phono-repeater, a radicogram, or an interpretant. The function type also includes a function value with transliterated text.

The following text defines the function types:

- Pictogram – pictorial symbol. It typically has no pronunciation.
- Logogram – sign that represents a word in the Egyptian language. As such it has a pronunciation and a meaning.
- Phonogram or phonemogram – sign that represents a sound in the writing system. It does not carry a semantic value. Strictly speaking, if we make a distinction between phonetics and phonology, the term phonemogram would be preferred to denote a phonological concept, but the two terms tend to be used interchangeably.
- Classifier – sign written at the end of words that indicates the semantic category to which the respective word belongs. As such it is always mute. It is traditionally called a determinative.
- Phono-repeater – a sub-category of classifier which has a phonetic meaning.
- Radicogram – graphemes that both point to some form and some content, but are not able to refer to an autonomous lexeme alone.
- Interpretant – non-autonomous graphemes that interpret the phonemic values of other semograms or phonograms.

The function value uses the transliteration format convention that is commonly known as the Gardiner 1957 convention. This convention already appears in the names list annotations of the original Egyptian Hieroglyph block. The transliteration format uses the following letters: ʾ, ʿ, y, ʿ, w, b, p, f, m, n, r, h, ḥ, ḫ, ḏ, s, š, k, ʔ, g, t, ṯ, d, ḏ. It may also contain additional punctuation for optional part, alternative, semantic element, etc. This will be developed in future versions of this document.

A single hieroglyph may have multiple function types. At the moment, most of the hieroglyphs have a single documented type, but in reality many of them have multiple types. For example, the fact that a sign have a given documented function type and a variant a different documented function type should be interpreted as the base sign using these two function types (or more), and not as a discrepancy.

### 3.5 Core

The provisional property `KEH_Core` determines whether an Egyptian hieroglyph is part of a 'Core' set. The 'Core' set is a curated subset of characters from the full Egyptian hieroglyph encoded set. It is the recommended set for Egyptologists and should be implemented in widely used fonts. The Core set represents the opinion of experts who reviewed the evidence that was provided to them. (The same group reviewed the full set.) This set is similar to UniHanCore2020 for CJK, which is the minimal set of required ideographs for East Asia. For a description of the selection process for the Core set by the Egyptologists involved, see the “Principles” Appendix. Characters in the Core set were verified by an image in photographs and trustworthy facsimiles. Transcription (a hand-drawn sketch of a sign) alone was not normally considered to be verified evidence. Images from hieratic texts

could be considered if the hieroglyphic nature of the sign could be easily reconstructed (cursive hieroglyphs). Possible values for this enumerated property are 'C' for Core, 'L' for Legacy, and 'N' for None. The Legacy value is used primarily for code points located in the Egyptian Hieroglyphs block (U+13000..U+1342F) to denote that these characters may be present in fonts for legacy reasons, but that their usage is discouraged. The 'None' value is used in the new Egyptian Hieroglyphs Extended-A block (U+13460..U+143FF) to denote that the code points with that property value are not fully attested, but may eventually become part of the 'Core' set.

The following are the exceptions to the requirement for verification:

- the sign appears in the Unicode 5.2 repertoire,
- the sign could not be verified and could not be constructed using an overlay or insertion mechanism.

While the property is provisional, the eventual intent is to make it normative in a future version of this document.

### 3.6 Mirroring and Rotation

The properties `KEH_NoMirror` and `KEH_NoRotate` indicate specific and rare behavior for some Egyptian hieroglyphs.

Most Egyptian hieroglyphs are expected to mirror relative to the reading direction. For example, for asymmetrical 'faces', the face is expected to face the start of the text, whether the line runs RTL or LTR. In very rare cases, the sign has a fixed orientation concerning mirroring. For example, U+130BB and U+130BD are an apparent set of mirrored walking legs. However, these two signs indicate opposite walking directions. In these rare cases, the property value `KEH_NoMirror` will be set to 'Y'.

Similarly, most Egyptian hieroglyphs can be rotated without changing their meaning. Because these rotations are a common occurrence, variation selectors should be used to represent these alternate representations. However, there are some signs where the rotation is significant and therefore, they cannot be rotated. In these rare cases, the property value `KEH_NoRotate` will be set to 'Y'.

## 4 The Properties

We now give two listings of the properties in the Unikemet database. The first is an alphabetical listing, with information on the property contents and syntax. The second is a listing of the properties by the version of the Unicode Standard in which they were first introduced.

### 4.1 Alphabetical Listing

For each property we give the following information in the alphabetical listing: its *Property* tag, its Unicode *Status*, its *Category* as defined above, the Unicode version in which it was *Introduced*, its *Delimiter*, its *Syntax*, and its *Description*.

The *Property* name is the tag used in the Unikemet database to mark instances of this property.

The Unicode *Status* is either *Normative*, *Informative*, or *Provisional*, depending on whether it is a normative part of the standard, an informative part of the standard, or neither. We may also include *Deprecated* as a Unicode Status if the property is no longer to be used.



Most of the properties which allow multiple property values have a *Delimiter* defined as “space” (U+0020 SPACE). Properties which do not have multiple property values have this defined as “N/A.” Some properties do not currently have multiple values in the data but may do so in the future.

For most properties with multiple values, the order of the values is arbitrary and has no particular significance. The most common order in such cases is alphabetical or numerical.

Because the property `kEH_Func` describing the function type may correspond to multiple types and may have also multiple values, the syntax is more complex. If there are multiple types, the types are separated by '/', but in most cases they share the same value. Multiple values are typically separated by either '/' or '|'; the "space" cannot be used because it may be part of a value field. Note that this is a work a progress, it denotes the current status among Egyptologists and may evolve over time. Note, however, that the vast majority of Egyptian hieroglyphs have a single function type and a single function value.

Validation is done as follows: The entry is split into subentries using the *Delimiter* (if defined), and each subentry converted to Normalization Form D (NFD). The value is valid if and only if each normalized subentry matches the property's *Syntax* regular expression. Note that any given property's *Syntax* is not guaranteed to be stable and may change in the future.

Finally, the *Description* contains not only a description of what the property contains, but also source information, known limitations, methodology used in deriving the data, and so on.

The properties covered in the table are: `kEH_Cat`, `kEH_Core`, `kEH_Desc`, `kEH_Func`, `kEH_FVal`, `kEH_HG`, `kEH_IFAO`, `kEH_JSesh`, `kEH_NoMirror`, `kEH_NoRotate`, and `kEH_UniK`.

Property	<b>kEH_Cat</b>
Status	Informative
Category	Catalog Indexes
Introduced	16.0
Delimiter	N/A
Syntax	<code>([A-IK-Z] AA)-[0-9]{2}-[0-9]{3}</code>
Default	N/A
Description	Catalog entry corresponding to the IFAO-based taxonomy

Property	<b>kEH_Core</b>
Status	Provisional
Category	Core
Introduced	16.0
Delimiter	N/A
Syntax	<code>C L N</code>
Default	N



Description	This enumerated property determines whether an Egyptian hieroglyph is part of the 'Core' set (value 'C'), Legacy (value 'L') or None (value 'N'). The Legacy value is primarily used for hieroglyphs in the original Egyptian Hieroglyphs block but which are not part of the Core Set.
-------------	---

Property	<b>kEH_Desc</b>
Status	Informative
Category	Description
Introduced	16.0
Delimiter	N/A
Syntax	[^\\t"]+
Default	N/A
Description	Detailed description of the appearance of the hieroglyph. It can be any Unicode character, except for control characters.

Property	<b>kEH_Func</b>
Status	Provisional
Category	Function
Introduced	16.0
Delimiter	/ (see description)
Syntax	[^\\t"]+
Default	N/A
Description	All signs are expected to have a function type representing a pictogram, a logogram, a phonemogram (or “phonogram”), a classifier (or “determinative”), a phono-repeater (sub-category of classifier), a radicogram or interpretant. It can be any Unicode character, except for control characters. Some types such as logogram have an English description, while others such as phonemogram typically do not. Most signs have a single type, but some have multiple types (separated by '/'). Sometimes additional context may be included in the type description, including transliterated text. This text can also use '/' to denote alternative description. Finally, while some signs are clearly attested, their type is uncertain, unknown, or undocumented as yet. That uncertainty is mentioned in the text itself.

Property	<b>kEH_FVal</b>
Status	Provisional
Category	Function
Introduced	16.0
Delimiter	/ or   (see description)
Syntax	[i y' w b p f m n r h h h h s s k k g t t d d . /   - ; ( ) s ] +
Default	N/A
Description	All signs are expected to have a function value corresponding to their function type. The value is expressed using the Gardiner 1957 convention for Egyptian hieroglyph transliteration. The delimiters '/' or ' ' are used to separate alternative values, while other punctuations may represent syntax elements, optional values, etc. The current value field represents a draft version, as work

	is still in progress and will be refined, based on feedback. Some signs still do not have a function value but are expected to be documented in the future.
--	---

Property	<b>KEH_HG</b>
Status	Normative
Category	Sources
Introduced	16.0
Delimiter	space
Syntax	([A-IK-Z] AA)[0-9]{1,3}[A-Za-z]{0,2}
Default	N/A
Description	Hieroglyphica source as specified in <i>Hieroglyphica – Sign List</i> , Nicholas Grimal, Jochen Hallof, Dirk van der Plas, 2nd edition, 2000. Multiple Hieroglyphica entries could be assigned to the same code point.

Property	<b>KEH_IFAO</b>
Status	Normative
Category	Sources
Introduced	16.0
Delimiter	space
Syntax	[0-9]{1,3},[0-9]{1,2}
Default	N/A
Description	IFAO source value defined as page number and order in that page, separated by a comma. IFAO is defined as <i>Catalogue de la fonte hiéroglyphique de l'imprimerie de l'I.F.A.O.</i> , Institut Français d'Archéologie Orientale du Caire, 1983, IF607, SEVPO, Paris, France. Multiple IFAO entries could be assigned to the same code point.

Property	<b>KEH_JSesh</b>
Status	Normative
Category	Sources
Introduced	16.0
Delimiter	space
Syntax	([A-IK-Z] Aa NL NU Ff)[0-9]{1,3}[A-Za-z]{0,5}  (US1 US22 US248 US685)([A-IK-Z] Aa NL NU)[0-9]{1,3}[A-Za-z]{0,5}
Default	N/A
Description	JSesh source as specified in Rosmorduc, Serge. (2014). JSesh Documentation. [Online, version 7.5.5] Available at: <a href="http://jseshdoc.qenherkhopeshef.org">http://jseshdoc.qenherkhopeshef.org</a> [Accessed Feb 23rd 2021]. Current version is 7.6.1 as of October 4th 2023, and sources values may have to be updated accordingly. Multiple JSesh entries could be assigned to the same code point.

Property	<b>KEH_NoMirror</b>
Status	Normative
Category	Mirroring and Rotation
Introduced	16.0

Delimiter	N/A
Syntax	Y N
Default	N
Description	It determines whether an Egyptian hieroglyph does not mirror. Note the reverse property because by default, most hieroglyphs can be mirrored depending on the reading direction.

Property	<b>kEH_NoRotate</b>
Status	Normative
Category	Mirroring and Rotation
Introduced	16.0
Delimiter	N/A
Syntax	Y N
Default	N
Description	It determines whether an Egyptian hieroglyph does not rotate. Note the reverse property because by default, most hieroglyphs can be rotated without affecting their meaning.

Property	<b>kEH_UniK</b>
Status	Provisional
Category	Catalog Indexes
Introduced	16.0
Delimiter	N/A
Syntax	$([A-IK-Z] AA NL NU)[0-9]\{3\}[A-Z]\{0,2\}$ $  HJ ([A-IK-Z] AA)[0-9]\{3\}[A-Z]\{0,2\}$
Default	N/A
Description	Original Unikemet catalog index used by the Egyptian Hieroglyph block, augmented for the extended blocks. Note that this is a work in progress with some issues.

## 4.2 Listing by Version of Addition to the Unicode Standard

The table below lists the properties of the Unikemet database by the version of the Unicode Standard in which they were first added.

Version	Properties Added	Properties Removed
16.0	<a href="#">kEH_Cat</a> , <a href="#">kEH_Core</a> , <a href="#">kEH_Desc</a> , <a href="#">kEH_Func</a> , <a href="#">kEH_FVal</a> , <a href="#">kEH_HG</a> , <a href="#">kEH_IFAO</a> , <a href="#">kEH_JSesh</a> , <a href="#">kEH_NoMirror</a> , <a href="#">kEH_NoRotate</a> , <a href="#">kEH_UniK</a>	














## 5 History















The Unikemet database originated as a concept proposed by the original Egyptian Hieroglyph proposal (ISO/IEC JTC1/SC2/WG2 N3237 =L2/07-097) as an appendix to that document but never materialized as a true dataset. It contained original source references which have been partly superseded by this version. It should also be noted that N3237 is

not 100% identical to what was eventually adopted by ISO and Unicode and was not updated to reflect the final code point values.




## Appendix: Encoding Principles

### General Principles

1. Hieroglyphic signs almost always have a number of different shapes that provide a visual contextualization of the text in which the sign is used. They inform the ancient (and modern) reader about the intentions and the mindset of the author. As such, the shapes are culturally relevant (chronological, geographical, social, technological, botanical, biological, religious) and can also be linguistically relevant. Different shapes may even be used within a single text. Included in such a Full List are the discrete differences that are clearly identifiable and allow to differentiate one shape from another. The Full List (extended repertoire) may grow as more evidences are identified. The Full List contains Core characters, Legacy characters (signs encoded in Unicode v5.2 which are not Core characters), and other non Core characters encoded in the extended block(s).
2. If shapes build a “continuum” or a cluster of sign variants and cannot be clearly differentiated, only a limited number of shapes will be included (e.g. a man bowing down/inclining his back will be limited to partial inclination U+13013 = A016 ) and a deep 90° bow ( U+134AA = HJ A087). “Chubby” or “slender” signs are part of this continuum and will not be differentiated.
3. Signs with a large number of repetitive elements (e.g. the waves of , the number of loops of a snake tail , besides , , , ..., ) will be limited to few selected examples based on philologists' judgment. At this moment, among these cobras, only the 2 coils, 3 coils, and the 4 coils versions are encoded, respectively as U+1319A (Core), U+13DB7 (Core), and U+13DC6 (non Core).
4. Seemingly identical signs which refer to different realities and to different functions/words will be differentiated and included (e.g. circular signs like the pupil of the eye, the pellet of sand, the geometrical circle, the tambourine, the ring, the hole in the ground).
5. Signs that exist with and without inner detailing will be included with the inner detailing (albeit sometimes reduced to “essential” details). This inner detailing is considered a relevant part of the sign and helps in avoiding confusion (e.g. the sun disk with and without an inner circle will always get the inner circle: U+131F3 = N005 ; the Sed festival chapel U+133B3 = W004  will always have the rhomb/diamond/lozenge at the bottom). If a shape is only attested in a source that provides only an outline of the hieroglyph, necessary detailing may be added as far as appropriate.
6. Differences only based on color (e.g. a yellow, red, green sun disk in some royal tombs from the Valley of the Kings) will not be included.
7. Not every detail is relevant (e.g. armlets and anklets on a person are usually irrelevant, whereas the type of clothing might be relevant [simple loincloth vs. loincloth with triangular protrusion; short vs. long loincloth, dress of the vizier, Ramesside courtier dress]; the number of arrows the soldier U+1300E = A012  is holding is irrelevant, beside the variation between a single arrow or plurality of arrows). Philologists need to judge in which cases details might be relevant (e.g. a quiver in A012 might be relevant).
8. The proposed characters are not intended to reflect fine paleographic detail, e.g.  vs.  (signs of desert hares only different in sizes). In such cases, users should instead rely on images or facsimile.






9. Signs are not eligible for separate inclusion in the proposed repertoire if they can be constructed by a base sign and overlay and/or insertion of another sign(s), or mirroring and/or rotation of a sign already in the repertoire. For description and examples of overlay, insertion etc., please see [Unicode Core Spec, Egyptian Hieroglyphs, Format Controls](#). This set includes common variants like the mirrored wickerwork basket  versus the regular , or  which is three stacked pieces of flesh , etc. (Note that these two “legacy” variants were encoded before these principles were adopted and are not included in the Core list).
10. Also excluded from the full list are signs that can be constructed with common signs such as *ḥw.t* (rectangular enclosure) , *srḥ* (palace façade) , etc. However, exceptions can be made for signs which are widely used, such as the logogram for the Hathor divinity U+13261 O10 .
11. Sequences are added as atomic characters when the functions of the component parts do not correspond to the function of a sequence. For example,  (U+131C6 phonemogram *wḏ*) is a functionally explainable overlap, as the components correspond to  (U+13193 cobra, phonemogram *wḏ*) and  (U+131C5 papyrus stem with a bud, phonemogram *ḏ*). This sign was encoded before these principles were adopted and consequently is not included in the Core list. On the other hand,  U+130BF, a classifier for “damage, injury” (*nkn*), is not a functionally explainable overlap, as the components correspond to  classifier *rd/leg* and  classifier *ds/cutting*. So, the latter is included as an atomic character. Note that Graeco-Roman standards were applied to determine attestations of the components function.
12. If a sequence can be graphically created, but there is evidence that the sign is considered a single entity in any of the Ancient Egyptian scripts (outside ligatures in cursive scripts), such as  U+13217, a monogram for *mw*, the sequence is added atomically.

### Additions to the Core list from the characters currently encoded in Unicode since v5.2

A sign from the Unicode v5.2 repertoire is added to the Core list if there is no conflict created by adding the sign, even though it might not be verified. A conflict would e.g. arise if the sign could be created as an overlay or an insertion. (An example of a conflict would be  U+13138 which should have been encoded as   = U+132F4 U+13436 U+13132 if following these principles.) Consequently, that character is not in the Core list.

### Additions to the Core list from the Full list

To be added to the Core list from the Full list, a sign needs to be verified with an image. Images (photographs and trustworthy facsimiles) originate from carved, painted or written hieroglyphic texts, and cursive hieroglyphic texts. Images from hieratic texts will only be considered if the hieroglyphic nature of the sign can easily be reconstructed (cursive hieroglyphs).

If verified, the sign is included in both the Core and Full list, no matter the priority, or origin. Exceptions are made for a sign that appears in verified sequences that require the addition of one of the parts of the sequence into the Core list, even though there is not yet an attestation of the sign on its own, e.g. U+137CB  Logogram (god) *ntr* in order to compose  can then be encoded as   : U+137CB U+13436 U+13257. (A sign within the sequence counts for the existence of the sign.)

Signs with a lot of variant forms need to be checked as a group (*šzp*, *ḥr.t-ntr* etc.) by philologists/specialists. Certain variants of such signs were excluded from the Core list

based on manual review.

For the sake of standardization (for example: seated god with sun-disk and seated god with sun-disk with uraeus), an unverified sign might be required to be added to the Core list.

### Inclusion in the Full List

To be automatically included in the Full list, the sign should have three or more attestations in printed or manuscript sign lists, in published computer-generated books like Dendara X, Esna VII and Athribis, in *Thot Sign List*, *Ramses*, *TLA*, *Karnak text database*, *AKU-PAL*, and the sign has a *JSesh* entry and/or Hieroglyphica number. Verification with an actual image is not a requirement.

Manual additions to the full list can be made based on common use by Egyptologists (see Leitz, Kurth or H&S sign lists), using a manual override.

Transcription (e.g. a hand drawn sketch of a sign) alone is not considered to be evidence for verification for the Core list, but if the other principles are covered, the sign might still be added to the Full list.

### Missing signs

The work to include all signs that are needed and eligible is not yet finished; only those signs which adhere to the principles cited above have been included. In many cases, no evidence has yet been found for signs that have been listed or mentioned in existing sign lists (i.e., images are lacking or the quality of the images at our disposal was not sufficient to decide). If users find a meaningful variant that is not in the list and they can provide evidence of the sign (photo or facsimile of an ancient source with precise identification of the monument), please make a submission to Thot Sign List (thotsignlist@gmail.com), including the evidence and full bibliographic information so it can be reviewed. Eligible signs will be added to a future Unicode proposal and refer to the contributors.

## References

For references for this annex, see Unicode Standard Annex #41, “[Common References for Unicode Standard Annexes](#).”

## Acknowledgements

This new database is the result of a collective work by many Egyptologists and is still a work in progress.

## Modifications

The following summarizes modifications from the previous revision of this annex.

### Revision 38

- **Proposed Update** for Unicode 17.0.0.
- Added New Appendix: Encoding Principles.

### Revision 3

- **First approved published version.**

## Revision 2

- **Draft of the first version of UAX #57 for Unicode 16.0.0.**
- **Fixed various typos in the document, added an IFAO reference and made other small edits.**
- **Fixed various regex expressions based on the actual data set.**
- **The keHCore property was changed from a binary property (Y/N) to an enumerated property (C|L|N). Description updated accordingly.**
- **Emphasis that a hieroglyph may have multiple function type values.**

## Revision 1

- **Proposed Draft of the first version of UAX #57 for Unicode 16.0.0.**
- **Split original keH\_Func into a description/type still called keH\_Func and a new property: keH\_FVal expressing the function transliterated value.**
- **Fixed various typos in the keH\_Func description.**

Previous revisions **will can** be accessed with the “Previous Version” link in the header **when appropriate.**

---

© 2024–2025 Unicode, Inc. This publication is protected by copyright, and permission must be obtained from Unicode, Inc. prior to any reproduction, modification, or other use not permitted by the [Terms of Use](#). Specifically, you may make copies of this publication and may annotate and translate it solely for personal or internal business purposes and not for public distribution, provided that any such permitted copies and modifications fully reproduce all copyright and other legal notices contained in the original. You may not make copies of or modifications to this publication for public distribution, or incorporate it in whole or in part into any product or publication without the express written permission of Unicode.

Use of all Unicode Products, including this publication, is governed by the Unicode [Terms of Use](#). The authors, contributors, and publishers have taken care in the preparation of this publication, but make no express or implied representation or warranty of any kind and assume no responsibility or liability for errors or omissions or for consequential or incidental damages that may arise therefrom. This publication is provided “AS-IS” without charge as a convenience to users.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc., in the United States and other countries.