Corruption of the basic sign list of Egyptian hieroglyphs in Unicode 16

Mark-Jan Nederhof (University of St Andrews)

Abstract

Unicode 5.2 introduced a first batch of 1071 Egyptian hieroglyphs. Unicode 16 introduced 3995 further hieroglyphs. In the process however, some of the original 1071 signs were corrupted. This document aims to identify all differences between Unicode 5.2 and Unicode 16, and focusses on the corrections needed to restore the basic list to its original intentions.

Unicode 5.2

In preparation for the list of Egyptian hieroglyphs in Unicode 5.2 (released in 2009), a database called UniKemet was created, published in 2006 as pp. 33-51 of:

https://www.unicode.org/L2/L2006/06354-n3181-egyptian.pdf

This database rigorously linked all the hieroglyphs either to publications from the Griffith Institute, mainly the grammar by Alan Gardiner, or to the Hieratische Paläographie of Georg Möller. Note that each row lists the exact source as well as the page number. The sources include Gardiner's grammar, and a number of ancillary documents, which were made publicly available:

https://www.unicode.org/L2/L2005/05313-Gardiner28-57.pdf

Gardiner's grammar provides extensive information about the signs, together with bibliographic references, and also the ancillary documents provide, albeit to varying degrees, information about many additional signs. Some signs are linked to particular inscriptions listed in the Griffith Institute Topographical Bibliography (GITB), which allows these signs to be studied in context.

The signs were given Unicode names that were similar to the names used by Gardiner. However, all numbers were padded with zeros to make 3-digit numbers; for example, Gardiner's A1 became A001. One rationale was to avoid any clash with sign names used by disparate sign lists, such as various editions of the Manuel de Codage and Hieroglyphica, and the sign list of the JSesh hieroglyphic editor. A related rationale was to avoid any suggestion that the Unicode signs stood for anything other than the signs from the publications of Gardiner and Möller, and in case of doubt about appearance or function, one would have to consult these publications and not others.

Unicode 16

For the development of the extended sign list of Unicode 16, two databases were used, one of which happens to be also named UniKemet:

https://www.unicode.org/Public/UCD/latest/ucd/Unikemet.txt

and the other will be referred to henceforth as the Database:

Surprisingly however, none of the information from the original UniKemet was included, and it is unclear whether the work was done with any awareness that that earlier database rigorously tied code points to sources. On the contrary, it seems the glyphs in the code charts were taken as mere suggestions, and new signs were arbitrarily assigned to old code points while old signs moved to new code points.

It was never announced beforehand that changes to the basic list would be made, and the changes were never discussed outside narrow circles. In particular, there was no consultation of the user community of the basic list, some of whom had spent over a decade encoding texts and developing teaching materials with that list. The only way one could find out that there had been any changes at all was to look closely at the Unicode 16 code charts.

This has thrown the user community into turmoil. Encodings produced in the past 15 years may no longer be valid, as old code points now stand for different signs, and the intended signs have moved to different code points. Hieroglyphic fonts created in the past 15 years are now inconsistent with the current code charts and recalling and updating all those fonts would be impossible.

When I spotted the first instances of corruption of the basic list, I flagged these up, and some fixes are now scheduled for Unicode 17. However, I did not realize until very recently just how many code points had been affected, after a systematic comparison of Unicode 5.2 and Unicode 16:

https://nederhof.github.io/newgardiner/unicode5to16compare0.html

Although in some instances the changes are harmless and do not require any fixes, there are many other cases where code points were given entirely different identities, without any justification whatsoever in the Database. In order to limit the harm, those code points need to be restored to their original intentions as soon as possible, ideally in Unicode 17, so that the corruption can remain limited to only one Unicode release.

This document contains a snapshot of:

https://nederhof.github.io/newgardiner/unicode5to16corruption.html

which will be updated as soon as there are new developments. Listed below are all discrepancies between Unicode 5.2 and Unicode 16 that I have been able to identify, divided into apparent errors and other noteworthy differences. I'm also listing cases where the Unicode 16 glyph is correct but the UniKemet documentation is inaccurate. Where two glyphs are given for the same code point, the first is from Unicode 5.2 and the second is from Unicode 16. The glyphs were automatically extracted from the official PDF code charts. Lines in blue starting with code point and kEH_Desc are descriptions copied verbatim from UniKemet. Comments in green indicate that the issues were resolved.

Apparent errors
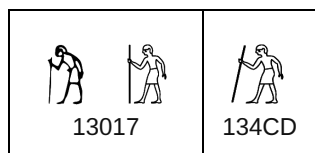
U+13017 kEH_Desc Man, standing, with a bend back, right arm forward, holding a staff/stick near the top, left arm hanging beside the body.

The staff in the original shape of U+13017 already had a slight backward angle, unlike the Unicode 16 form, where the staff is perfectly vertical. This will be solved in Unicode 17 by redrawing U+13017 to have the staff at a backward angle, like in the original shape. But note that then U+13017 and U+134CD will differ by no more than the exact degree of the angle, which seems an inappropriately small graphical distinction for the purposes of Unicode.
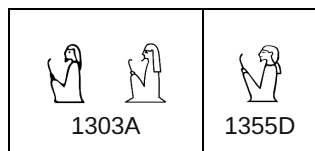
1303A   1355D

Considering the hairstyles, U+1303A was misplaced at U+1355D. This will be solved in Unicode 17 by swapping the shapes and ancillary documentation of U+1303A and U+1355D from Unicode 16.


13055

In the original shape, the hair is not long. The Database has "Nederhof example is a different sign though [...]". Different from what? Also, it is not my example, it is *the* example that was *the* justification for introducing this code point in Unicode 5.2 in the first place. The only ground truth in determining what a code point should stand for is the original UniKemet database used to compile the Unicode 5.2 basic list. See p. 34 of:

https://www.unicode.org/L2/L2006/06354-n3181-egyptian.pdf

It defines B005A as the sign in GITB 804-013-500:

http://www.griffith.ox.ac.uk/gri/8rel100.pdf

In the transcription the hair is definitely not long. Unless one argues that the mentioned tomb relief was misread, and no one has argued that so far, the hair should be short in the code charts. In fact, the Database confirms that the sign in Page-Gasser & Wiese (1997) has "short hair". One could argue that "no female features" might be a reason to adjust the glyph with regard to the original, but since there is no doubt the sign represents a female, who cares to what extent the "female features" are accentuated in the glyph.
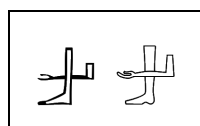
130C1

U+130C1 kEH_Desc A forearm, with the palm of the hand facing upwards (D36) written over a human foot and lower leg (D58).

These are arguably only insignificant graphical variants of the same thing, but unless one can argue that one variant is real and the other variant is not real (never occurs), one should revert to the original Unicode 5.2 variant, and change description and image accordingly. Incidentally, both forms exist in GEG, and are in fact found in the same line (top line of p. 458). In 2006-2009, there were discussions about which of the two was the most appropriate form for Unicode, and in consultation with Egyptologists it was decided to opt for the form in Unicode 5.2. Whether or not one agrees now this was the right decision at the time should be irrelevant. Code charts should not be changed without a compelling reason.
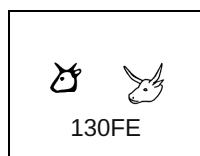

130FE

U+130FE kEH_Desc The head of a bovid (ox) with outwards curving horns.

The description suggests outwards curving horns, unlike the Unicode 5.2 and Unicode 16 shapes. Some of the tokens in the TSL have outwards curving horns and some have inwards curving horns:

https://thotsignlist.org/mysign?id=2871

Does it make sense to specify the curvature in the description when the curvature is so variable in actual inscriptions? I propose to change the description to just have "with horns"?


130F9   13ABE
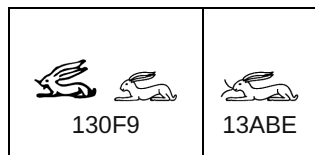
U+130F9 kEH_Desc A desert hare, lying down.
U+13ABE kEH_Desc A desert hare, lying down, with whiskers.

The original shape of U+130F9 unmistakably had whiskers. This will be solved in Unicode 17 by swapping the shapes and ancillary documentation of U+130F9 and U+13ABE from Unicode 16. (There is no need though to make the length of the whiskers as exaggerated as they are currently in U+13ABE.)


130FA

In the original shape, the hare has whiskers. Moreover, the difference between U+130FA and U+130F9 should be in the height, while the width should be roughly the same, i.e. U+130FA is flatter than U+130F9. See Gardiner (1957) for the typographical motivation. Both issues will be solved in Unicode 17.
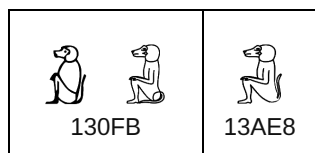
U+130FB kEH_Desc A hamadryas baboon (Papio hamadryas), seated, hands on knees, tail folded under the rear.
U+13AE8 kEH_Desc A hamadryas baboon (Papio hamadryas), seated, hands on knees, tail upwards.

The original shape (tail up) was misplaced at U+13AE8. This will be solved in Unicode 17 by swapping the shapes and ancillary documentation of U+130FB and U+13AE8 from Unicode 16.
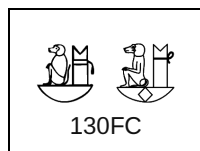
U+130FC kEH_Desc A hamadryas baboon (Papio hamadryas), seated, hands on knees, tail folded under the rear (E35), in front of a sheath or receptacle with V shaped indentation at the top, bound at the top with a loop a the backside (V36G); on top of an alabaster basin with a diamond shaped inner marking (W3).

The original shape had the tail up. This will be solved in Unicode 17 by restoring the orientation of the tail.
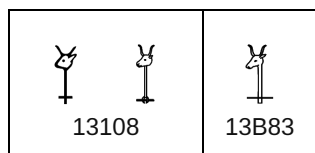

U+13108 kEH_Desc The head and neck of a long-necked horned animal, with a forked end to the neck with a horizontal line over the forked end.
U+13B83 kEH_Desc The head and neck of a long-necked horned animal, with a horizontal line written over the neck, near the bottom.

The original shape was misplaced at U+13B83. The only reasonable solution is to swap shapes and ancillary documentation from Unicode 16.

In the Database there is not a shred of an acknowledgement that the shape of U+13108 changed from Unicode 5.2 to Unicode 16, let alone a justification for the change.
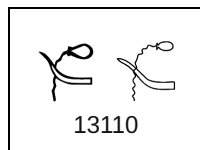
U+13110 kEH_Desc A vase on its side, with liquid issuing from it (W54), written over the horn of a bovid (F16).

Even though these are insignificant graphical variants, code charts should not be changed without a good reason. Unless a compelling case can be made that the horn over the liquid cannot occur in real inscriptions, revert to the Unicode 5.2 form.
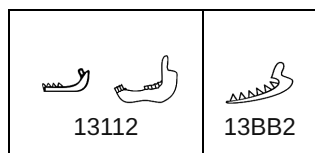
The appearance of U+13112 has changed considerably and one could even argue that U+13BB2 is closer to the original shape than to the new shape; both the original shape and U+13BB2 have a single row of pointed teeth while the new shape of U+13112 has two isolated rows of less pointed teeth. Gardiner (1957) characterizes the sign as "lower jaw-bone of ox", and as far as I know, oxen are not particularly known for having pointed teeth. But does this necessarily mean that the original shape of U+13112 and U+13BB2 differ in what they are meant to represent? Unless there is solid evidence that the original shape of U+13112 was wrong or uncharacteristic, it may be better to revert the shape to something closer to Unicode 5.2. It is then an open question whether the current shape of U+13BB2 deserves its own (core) code point, or whether it is just a duplicate of U+13112.

I don't see the original shape of U+13112 anywhere in the Database, and there doesn't seem to be even a shred of an acknowledgement that the shape changed from Unicode 5.2 to Unicode 16, let alone a justification for the change.
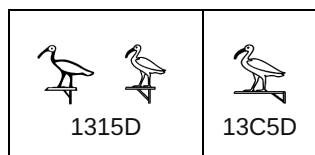
The description of U+1315D does not match the images. The original shape has a standard closer to R92A. So if a graphical variant is needed with R12, then it should be at a new code point, so here that would be U+13C5D. In other words, flip the ancillary documentation and give U+13C5D the R12 shape. The draft Unicode 17 code charts have this the wrong way around. I hope this can still be corrected.
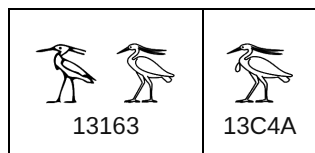
It is by no more than a few black pixels, but it is unmistakable that the original shape was meant to exhibit a lappet. In other words, the shape of U+13163 was misplaced at U+13C4A. The only reasonable solution is to swap shapes and ancillary documentation from Unicode 16.
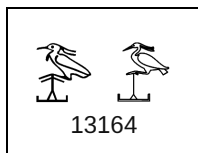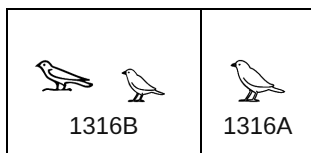
Give the shape a lappet to revert to Unicode 5.2.
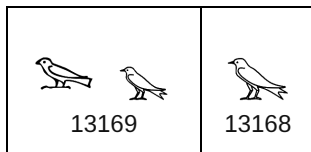
The width of U+13169 should be the same as that of U+13168, but the height should be smaller. See Gardiner (1957) for the typographical motivation. This will be corrected in Unicode 17.

The width of U+1316B should be the same as that of U+1316A, but the height should be smaller. See Gardiner (1957) for the typographical motivation. This will be corrected in Unicode 17.
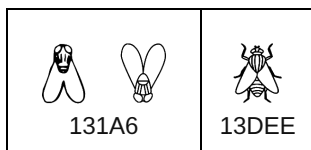
In 2006-2009, there were protracted discussions about what the most appropriate orientation of U+131A6 was for Unicode, since both orientations occur in GEG (p. 477 versus p. 545), and since the sign was taken from iconography and is not attested in actual running text. In consultation with Egyptologists it was decided to opt for the form in Unicode 5.2, fixed formally by letting UniKemet link the code point to the occurrence on p. 477 as opposed to the occurrence on p. 545. This was motivated by similar signs (like the newly introduced U+13DEE) that have the head up as well. For more information, see pp. 8 and 39 of:

https://www.unicode.org/L2/L2006/06354-n3181-egyptian.pdf

Whether or not one now agrees this was the right decision at the time, having studied (or not) the arguments laid out by Egyptologists in 2006-2009, all this should be irrelevant. The glyph should

not change on a whim relative to what it has been for the past 15 years since Unicode 5.2. If anyone feels the need for the other orientation, a variation selector for rotation can be used.

Apropos, the Database has the fly with the head up as well, so it is not even clear to me any of the philologists actually wanted to have this sign rotated relative to the original orientation. In the end, the rotation may be due to clerical error.
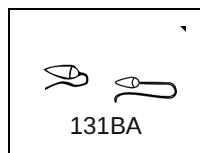
U+131BA kEH_Desc A lotus bud with a long stalk, folded under itself.

I couldn't care less about the length of the stalk in the glyph. This is a palaeographic detail and unimportant at the level of Unicode. I do object against the Database not even acknowledging that anything was changed, despite the fact that the description above explicitly invalidates the character identity from Unicode 5.2 when the stalk was definitely not long. In particular, there is nothing in the Database that suggests that a short stalk was wrong or atypical and therefore needed to be changed into a long stalk. The original form with the short stalk reappears in a different entry with the comment "Does not seem to be a meaningful variant". Ok, but if the difference between the shapes was not significant, why was there any need for the shape to be changed?

My proposal is to delete "long" from the description. Ideally one would also make the stalk a little shorter than the Unicode 16 form but it could be a little longer than the Unicode 5.2 form.
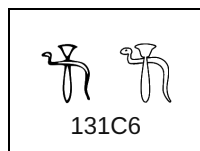
U+131C6 kEH_Desc A cobra in repose (Naja haja, I10), written over a stem of papyrus with a bud (M13).

Revert to the shape from Unicode 5.2. The description in fact already correctly reflects the original shape.
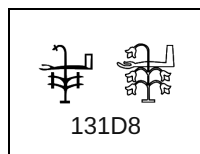

U+131D8 kEH_Desc A forearm, with the palm of the hand facing upwards (D36), written over a desert plant, with four branches, with flowers on every branch, on a horizontal base. (M26).

It is unclear which sign is on top in Unicode 5.2, so fonts have the liberty to interpret this as desired, but at least make sure the shape is consistent with the description, which is currently not the case. The current shape has the plant over the arm, not the arm over the plant.

Because the original shape did not fix which sign was on top, I would prefer for the description to leave this underspecified, by using a phrase other than "written over". Perhaps "overlaid with"?

| 131F4 | 131F6 | 13EDE | 13EDF | 13EE0 |
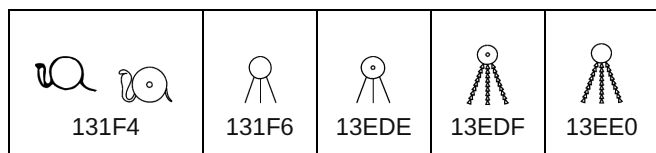
U+131F4 kEH_Desc The sun within a halo, encircled by a cobra (Naja haja), standing up, with expanded hood (Uraeus).
U+131F6 kEH_Desc The sun, with thee beams of sunlight coming from it.
U+13EDE kEH_Desc The sun, within a halo, with three beams of sunlight coming from it.
U+13EDF kEH_Desc The sun, within a halo, with three beams of sunlight coming from it, resemling lines of triangles.
U+13EE0 kEH_Desc The sun, with three beams of sunlight coming from it, resemling lines of triangles.

There is no inner circle in the original shape of U+131F4. One may argue this is an insignificant detail. But if one then introduces multiple code points U+131F6 and U+13EDE for the same sign with and without inner circle, and similarly U+13EDF and U+13EE0, then one simultaneously suggests that the inner circle is important and that the inner circle is not important. See also Point 5 of the Encoding Principles, which more often than not are divorced from what was actually done:

https://www.unicode.org/reports/tr57/tr57-4.html#EncodingPrinciples

In the Database there is not a shred of an acknowledgement that the shape of U+131FA changed from Unicode 5.2 to Unicode 16. The Database does have the shape without inner circle, but in a separate entry, with the comment "Technically the more common form". Then what could possibly be the justification for changing the glyph if the Unicode 16 form is less common than the Unicode 5.2 form?
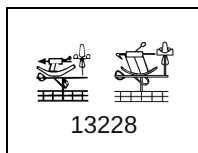

| 13227 | 14256 |
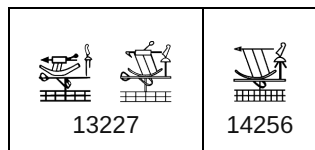
U+13227 kEH_Desc An one-barbed harpoon, with handle, written horizontally (T21), on top of a crescent moon shape, connected by three lines, in front of a feather (H6), angled forwards on top of a standard with a round top, with an short vertical line beside the main pole (R14); on top of a standard used for carrying religious symbols (R12), written on top of a parcel of land with irrigation ditches (N24).
U+14256 kEH_Desc An spear, arrow without fletching or a harpoon without handle, written horizontally, on top of a crescent moon shape, connected by four lines, in front of a feather (H6), angled forwards on top of a standard with a round top, with outwards angled lines coming from the tip of the pole (R14B); on top of a standard used for carrying religious symbols (R12), written on top of a parcel of land with irrigation ditches (N24).

The differences of U+13227 may be insignificant, but it appears to me that most of the description matches the original shape better than the Unicode 16 shape, in particular the orientation of the feather and the "short vertical line". It may be appropriate to revise the glyph, and consult an expert whether the shape is drawn correctly.

It also seems to me that U+14256 is closer to the new shape of U+13227 than the new shape of U+13227 is to the original shape of U+13227. If U+14256 is so fundamentally different that it warranted a new code point, then how can one justify treating the old and new shapes of U+13227 as interchangeable? There seems to be no consistency in any of this.

U+13228 kEH_Desc An one-barbed harpoon, with handle, written horizontally (T21), on top of a crescent moon shape, connected by three lines, in front of a spear made into a standard, with a circle on either side of the speartip, with a loop over the standard (Jsesh/Gardiner R15); on top of a standard used for carrying religious symbols (R12), written on top of a parcel of land with irrigation ditches (N24).

I think the description matches the spear better in the original shape than in the Unicode 16 shape. It may be appropriate to revise the glyph, and consult an expert whether the shape is drawn correctly.

Apropos, the Database refers to:

http://sith.huma-num.fr/karnak/1585

The spear there is like in the Unicode 5.2 form.
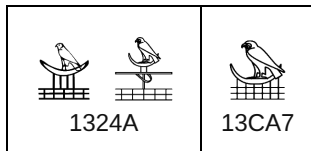
U+13CA7 kEH_Desc A falcon (G5), on top of a moon sickle shaped boat, on top of a sledge that resembles a parcel of land with irrigation ditches (N24).

The original shape of U+1324A was misplaced at U+13CA7. The original and current shapes of U+1324A are not even remotely the same sign.

The Database writes "No token of this exact grapheme [the Unicode 16 glyph of U+1324A] could yet be located". This means there was no reason for the sign to even be in Unicode and one should have kept the Unicode 5.2 shape for U+1324A, because that is attested, for example here:

http://sith.huma-num.fr/karnak/1128

(Incidentally, this has a token closer to the original Unicode 5.2 form of U+1324A than to the Unicode 16 form of U+13CA7.)

The only reasonable solution is to swap shapes and ancillary documentation from Unicode 16. I would then make U+13CA7 non-core, as it will then get to harbor an unattested shape.
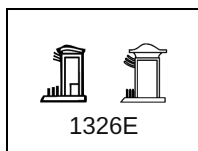
U+1326E kEH_Desc An archaic shrine, seen from the side, with a downwards sloping roof, with upwards curving lines coming from the roof, and short vertical lines in front of the shrine.

The description of the shape of the roof matches the original shape but not the new shape. In the new shape the roof is symmetric, rather than sloping downwards towards the back, as I would understand it.
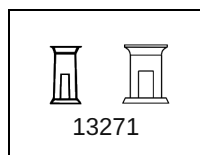

U+13271 kEH_Desc A façade of a shrine with a flat roof, with vertical sides, with a small doorway.

In the original, the sides were more oblique than vertical. In the TSL:

https://thotsignlist.org/mysign?id=4552

both the normalized glyph and some (but admittedly not all) of the tokens have oblique sides. If the TSL does not provide convincing evidence to motivate changing the sides from oblique to vertical, then I don't think anything else will.
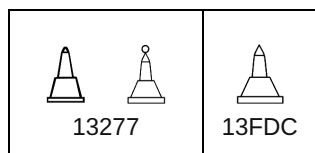
U+13277 kEH_Desc An obelisk of an Old Kingdom sun temple, with an circle (sun disk) on top of the obelisk.
U+13FDC kEH_Desc An obelisk of an Old Kingdom sun temple.

The original shape (without circle) was misplaced at U+13FDC. The only reasonable solution is to swap shapes and ancillary documentation from Unicode 16.

The Database has a link to TSL_1_4521_01, where one finds both the shape with circle and the shape without circle. There are two tokens without circle and one token with circle. The only thing this tells us is that the sign occurs both without circle and with circle. It does not in any way tell us that the originally intended shape of U+13277 in Unicode 5.2 should have a circle. The only ground truth in this respect is the original UniKemet, which defines O025A to be the shape that occurs on p. 246 of Gardiner (1931) and on p. 4 of Gardiner (1953) and that very clearly has no circle. Cf. p. 42 of:

https://www.unicode.org/L2/L2006/06354-n3181-egyptian.pdf

and pp. 44 and 49 of:

https://www.unicode.org/L2/L2005/05313-Gardiner28-57.pdf

Gardiner (1931) gives us some indication about the function of the sign: "Det. names of sun-obelisks". But this is consistent with the function "sun temple" listed by the TSL to cover both the circled and the uncircled forms, and so also this offers no justification for changing the shape of U+13277.

1329F
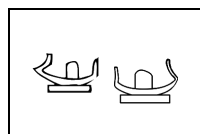
The new shape is mirrored. Since there is a mirroring control, there is no reasonable justification for this, even if the sign normally occurs the other way around.

132BC 14107
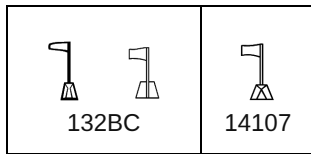
U+132BC kEH_Desc A cloth wound on a pole, an emblem of divinity (R8), written over a butchers block, with the pole at the bottom of the block.
U+14107 kEH_Desc A cloth wound on a pole, an emblem of divinity (R8), written on top of a butchers block (T28).

The original shape of U+132BC was misplaced at U+14107. The only reasonable solution is to swap shapes and ancillary documentation from Unicode 16.

There is no justification in the Database for changing the shape of U+132BC other than the comment that Hieroglyphica/JSesh happens to have another shape. There even seems to be a photo in the Database that confirms that the original Unicode 5.2 shape is attested.
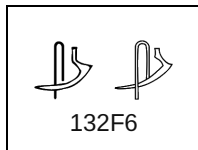
132F6

U+132F6 kEH_Desc A sickle (U1), written over a folded piece of cloth (S29).

The description matches the original shape but not the new shape. In the new shape the piece of cloth is over the sickle.
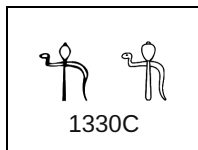
1330C

U+1330C kEH_Desc A cobra in repose (Naja haja) (I10), written over a mace with a pear-shaped head, written vertically (T3).

The original shape underspecified which of the signs was on top, and it may be safest to leave this underspecified in the description. In any case, the glyph should be consistent with the description. In the new shape, the cobra is *under* the mace.
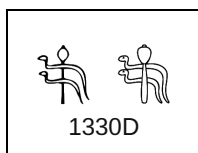
1330D

U+1330D kEH_Desc Two cobras in repose (Naja haja) (I10), arranged vertically, written over a mace with a pear-shaped head, written vertically (T3).

The description matches the original shape but not the new shape. In the new shape the mace is over the cobras.
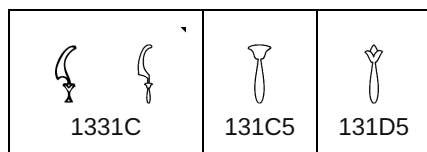

```
U+1331C kEH_Desc A scimitar (khopesh), written vertically, blade curving forwards, with a
handle resembling a stem of papyrus with a bud (M13).
U+131C5 kEH_Desc A stem of papyrus with a bud.
U+131D5 kEH_Desc A stem of papyrus with a flowering bud.
```

In the original code charts, the handle most certainly does not resemble M13 (U+131C5), If anything, it looks closer to U+131D5. I do not mind much if the glyph in the code charts currently has that shape, as the shape of the handle seems a minor detail. But by putting "with a handle resembling ..." in the description, one is suggesting that M13 is an essential part of the character identity of U+1331C, whereas (as far as I know) scimitars do not generally have a handle resembling M13. The Database does not motivate departing from the original shape. The TSL has one single token of T016A (without citation!), which is rather blurry and does not seem to suggest M13 at all:

https://thotsignlist.org/mysign?id=5850

Various examples taken from iconography are here:

https://www.reddit.com/r/Cowofgold_Essays/comments/zcfv4z/the_khopesh_of_ancient_egypt/

I don't see the shape of M13 in any of them. I would simply omit "with a handle resembling a stem of papyrus with a bud (M13)" as overspecific.
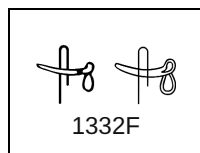
By plausible deniability one could argue the new shape underspecifies which sign is on top. But in Unicode 16, the shapes of overlays generally commit to one or the other. It would be better therefore to revert to the Unicode 5.2 shape.
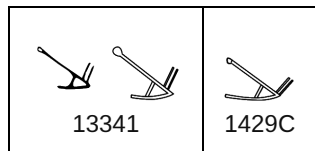

```
U+13341 kEH_Desc A plough, with a cross-bar and two vertical lines on the end of the long
beam.
U+1429C kEH_Desc A plough, with a cross-bar and two vertical lines on the long beam,
without a circle at the front of the long beam.
```

It is unclear to me whether the original shape of U+13341 is closer to its new shape, or whether it rather corresponds to U+1429C. The original shape had a pronounced circle, like U+1429C, while the new shape has more of a bulb. Does it even make sense to have two code points here?

1334C	1334B	142B7	142B8

U+1334C kEH_Desc A hair-pin or burin, without the horizontal line on the top.
U+1334B kEH_Desc A hair-pin or burin, with a horizontal line above and below the oval of the pin.
U+142B7 kEH_Desc A hair-pin or burin, with a horizontal line above and below the circle of the pin.
U+142B8 kEH_Desc A hair-pin or burin, with a horizontal line under the oval of the pin, and a triangular shape on top of the oval.

The original shape of U+1334C has a horizontal line on the top, so the new shape is wrong. By the original UniKemet from 2006, U+1334C stands for "a later form of U23" (Gardiner 1931) as found on the 26th Dyn. sarcophagus of Ankhnesneferibra (BM EA32). See p. 45 of:

https://www.unicode.org/L2/L2005/05313-Gardiner28-57.pdf

See the bottom-left corner of the image in the link below, where it reads at the end of the column "n mr=s n [...]".

https://www.britishmuseum.org/collection/object/Y_EA32?selectedImageId=1495613001

There is most definitely a line on the top.

The Database provides no evidence of having even consulted the original UniKemet. Instead, the Database refers to the Pyramid of Unas, which is off by 2 millennia so this is unlikely to be relevant to U+1334C. It is unclear to me how other newly introduced code points for hair-pins are supposed to relate to the original shape of U+1334C, or even how to characterize the difference between U+1334C and usual hairpin U+1334B. It becomes an exercise in futility if new code points are introduced for barely distinguishable shapes, as if those are meaningful for Unicode, while exact specifications of existing code points are ignored to reuse these code points for totally different things. One either takes the character identities of code points seriously, or one does not, in which case why do we even bother with Unicode?
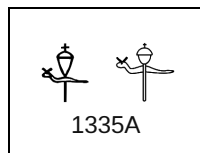

1335A

U+1335A kEH_Desc A horned desert viper (Cerastes cerastes) (I9), written over a spindle (U34).

The description matches the original shape but not the new shape. In the new shape the spindle is over the viper.
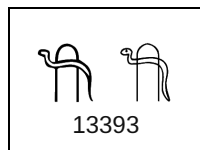

13393

U+13393 kEH_Desc A cobra in repose (Naja haja) (I10), written over a hobble for cattle without a crossbar (V20).

Apart from plausible deniability (see above), the description matches the original shape but not the new shape. It would be better to revert to the Unicode 5.2 shape.
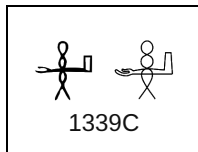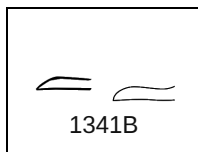

1339C

U+1339C kEH_Desc A forearm, with the palm of the hand facing upwards (D36), written over a wick of twisted flax, consisting of three loops (V28).

Apart from plausible deniability (see above), the new shape, unlike the description, suggests the wick is over the arm. Make the new shape consistent with the description. I would prefer for the description to leave underspecified which sign is on top, as it is not clear from the original shape either.
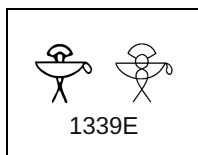

1339E

U+1339E kEH_Desc A swab made from a hank of fibre with two loops (V29), written over a wickerwork basket with a handle, orientated with the handle to the back (V31).

Even though these are insignificant graphical variants, code charts should not be changed without a good reason. Unless a compelling case can be made that the basket over the hank of fibre cannot occur in real inscriptions, revert to the Unicode 5.2 form and change the description accordingly.


1341B

U+1341B kEH_Desc A representation of two ribs, top line curving downwards.

The description matches the original shape but not the new shape. Correct the new shape to have the top line curve downwards.
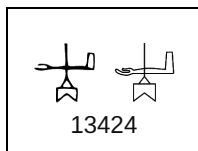
13424

U+13424 kEH_Desc A forearm, with the palm of the hand facing upwards (D36), written over an object consisting of a horizontal rectangle with a triangular indentation at the bottom, with a vertical line on top of it, connected through a triangle (Aa21).

Apart from plausible deniability (see above), the new shape, unlike the description, suggests the object is over the arm. Make the new shape consistent with the description. I would prefer for the description to leave underspecified which sign is on top, as it is not clear from the original shape either.
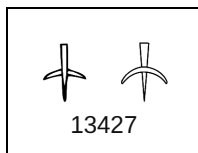
U+13427 kEH_Desc A crescent moon shape written over a thin triangle, point downwards.

Even though these are insignificant graphical variants, code charts should not be changed without a good reason. Unless a compelling case can be made that the "thin triangle" over the moon cannot occur in real inscriptions, revert to the Unicode 5.2 form and change the description accordingly.

Other noteworthy differences (maybe not errors?)

U+13043 kEH_Desc Man, standing, back bend forward, both arms toward the front, holding a hoe with a rope connecting the two pieces (U6), with the forward piece at the level of the feet.

Originally, there was a line representing the ground, and there was no connecting rope. Now there is no line for the ground and there is a rope connecting the two pieces of the hoe. In this instance, the Database seems to offer adequate justification for a change to a glyph. Both Urk IV 57,3, supposedly the source of the sign according to Gardiner (1928), and the earlier publication by Bouriant it was based on (in Recueil de traveaux 12, 1890) have the Unicode 16 form. For the latter, see p. 107 of:

https://digi.ub.uni-heidelberg.de/diglit/rectrav1892/0117/image,info

U+1304F kEH_Desc Man, seated, right knee raised, raised arms at either side of the body, hands held vertically, with the handpalms inwards.

Hands were oriented outwards, now inwards. The supposed source is GITB 801-626-810, but this does not have the sign:

http://www.griffith.ox.ac.uk/gri/s14.pdf

This leaves us without any reliable information about the intended character identity of this code point. It does not seem entirely far-fetched though to assume the intended sign is the same as:

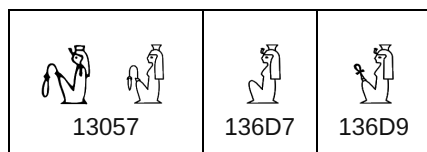https://thotsignlist.org/mysign?id=883

in which case, the hands are inwards, provided the blurry photos are interpreted correctly. Another occurrence is on p. 134, l. 14 and pl. DCCCLXXXIV of:

https://www.ifao.egnet.net/uploads/publications/enligne/Temples-Dendara009.pdf

U+13057 kEH_Desc Woman, seated, both knees up, with covered legs and arms, with long hair, wearing a diadem, holding a flower which curves forward, with the flower orientated downwards.
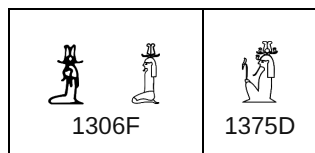U+136D7 kEH_Desc Woman, seated, both knees up, with covered legs and arms, with long hair, wearing a diadem and ureaus.
U+136D9 kEH_Desc Woman, seated, both knees up, with covered legs and arms, with long hair, wearing a diadem and ureaus, holding a tie or strap, used with sandals (ankh-sign, S34), which angles forward.

For U+13057, there was a uraeus in the original, now there is not. The Database motivates this by asserting that the uraeus was a mistake because it is not in Gardiner's source (Ann 42, 479). However, the font used therein is extremely coarse:

https://archive.org/details/ASAE-42-1943/page/n251/mode/2up

My hunch is that if the uraeus had existed, it would not have been visible in the transcription anyway.

Considering U+136D7 and U+136D9, which do have uraei, is it not plausible that the original shape might have existed? If so far no token of U+13057 was located with uraeus, then what if such a token will eventually be found? Will a new code point be introduced for it? This is disastrous for the stability of Unicode. Would it not have been safer to introduce a new code point for the shape without uraeus? It is not like new code points were used sparingly; a staggering number of new code points were introduced for barely distinguishable variants of the heron alone.


U+1306F kEH_Desc God, seated, both knees down, with a long curved beard and long hair/wig, wearing a headdress consisting of two feathers on rams horns (S77).
U+1375D kEH_Desc God, seated, both knees up, with covered legs and arms, with a long curved beard and long hair/wig, with an uraeus on the forehead, wearing a headdress consisting of two feathers and a sun-disk on rams horns, with an uraeus at either side of the feathers; holding a sceptre with a straight shaft, topped with the head of the Seth animal vertically.

For U+1306F, the sun-disk was removed. This is fine if the sun-disk in the original shape seems to have been due to a misinterpretation of known tokens. But note that there are headdresses in similar signs like U+1375D with sun-disk.

In the Database I find "redraw of C18G without sun" but without justification.
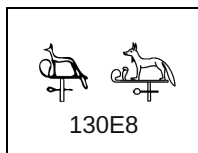
130E8

U+130E8 kEH_Desc A jackal, standing, tail down (E17), on top of a standard used for the carrying of religious symbols, with an uraeus and SdSd-pretuberance at the front of the standard, with a mace (T3) written horizontally over the vertical pole of the standard.

The Unicode 5.2 form did not have the uraeus. The Database justifies adding the uraeus by the observation that many occurrences have a uraeus, calling this "correcting" of the glyph. I don't think the word "correcting" is appropriate if occurrences without uraeus exist, and then the glyph without uraeus would not be incorrect per se. Nonetheless this change to the glyph may cause relatively little disruption, provided no one will now consider introducing a new code point for the original form without uraeus.

This identifies a major shortcoming of the kEH_Desc descriptions: a description describes a single 'representative' token that the philologists happened to find, down to minute detail. The new UniKemet also typically lists only the one function of that one token. But this fails to capture the type. For example, if in the future the intention is that U+130E8 can be used for the sign with uraeus and the sign without uraeus, then currently no documentation exists that expresses this intention. In this sense, this is a step back from the documentation in GEG, which, by being less overspecific about appearance of signs, made it clear to encoders that signs could be used for a range of graphical variants.
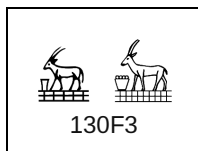
130F3

U+130F3 kEH_Desc A basket with four pieces of grain or fruit (M39) in front of an oryx, standing (E28), on top of a parcel of land with irrigation ditches (N24).

Unicode 5.2 form did not have the four pieces of grain or fruit. This seems an insignificant detail and I have no objection to the embellishment of the glyph with the four pieces of grain or fruit. I think though that one should simplify the description to just mention that there is a basket, without (over)specifying what is in it.


13146 13C7C
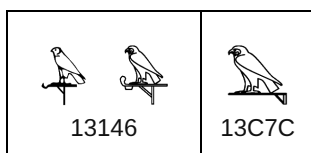
U+13146 kEH_Desc A falcon (G5), on top of a standard used for carrying religious symbols (R12).
U+13C7C kEH_Desc A falcon (G5), on top of a standard used for the carrying of religious symbols with the vertical stick at the far side (R92A).

It is unclear to me what kind of standard was intended with the original shape of U+13146, and it seems an insignificant detail. But if one starts introducing multiple code points depending on the choice of standard (cf. U+13C7C), then it is not clear to me that the change of shape of U+13146 is within the normal bounds of allowable graphical variation for a given code point, and perhaps a fresh code point would have been warranted.

The Database does not contain a shred of reflection on the fact that the shape of U+13146 has changed.
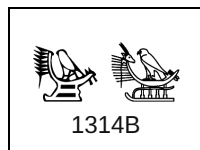

U+1314B kEH_Desc A falcon (G5), on top of a moon-sickle shaped boat with an higher prow than stern, with a rudder, with the head of an antilope (F81) on top of the prow, facing inwards, with a downwards line with multiple horizontal lines written over it coming from the head, on top of a sledge (P60B).

The head of antelope was not previously discernible. If I correctly interpret the very fuzzy images in the TSL:

https://thotsignlist.org/mysign?id=3196

then the sign may occur with and without head of antelope. Nonetheless, it would have been nice to see some acknowledgement in the Database that the glyph has changed. Again, there is the problem that the description specifies the token, not the type.

Are *two* rudders visible, in both glyphs? In that case the description and the glyph are inconsistent.
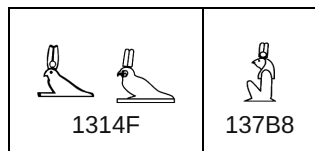
U+1314F kEH_Desc An archaic image of a falcon/cult image of a falcon, wearing a headdress consisting F1E34of two plumes (S9).
U+137B8 kEH_Desc God, seated, knees up, with covered legs and arms, with the head of a falcon, wearing a headdress consisting of two plumes with a sun disk (S63A/S70).

For U+1314F, the sun-disk was removed. This is fine if the sun-disk in the original shape seems to have been due to a misinterpretation of known tokens. But note that there are headdresses in similar signs like U+137B8 that do have a sun-disk.

The Database does not contain a shred of reflection on the fact that the shape of U+1314F has changed. It does comment that "Jsesh glyph has a circle in crown". What evidence justified departing from both the original Unicode 5.2 shape and the JSesh shape? If no such evidence exists, then one should have kept the original shape, rather than sacrifice the stability of Unicode on a whim for no apparent reason.
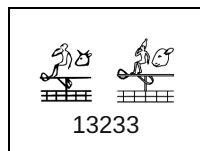
Child was given white crown. Bovid lost its horns.

The Database says no token could be found. The default action should then be to keep the shape as it is, rather than change the glyph randomly, and sacrifice the stability of Unicode on a whim for

no apparent reason.

13234
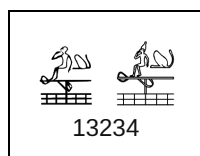
Child was given white crown.

The Database says no token could be found. The default action should then be to keep the shape as it is, rather than change the glyph randomly, and sacrifice the stability of Unicode on a whim for no apparent reason.

13235
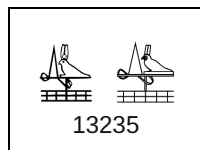
Headdress lost the sun-disk.

The Database says "Could not yet locate an example." There cannot therefore have been a justification for changing the shape.
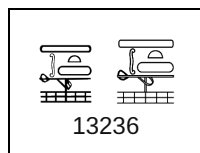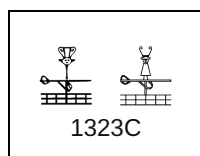

13236

U+13236 kEH_Desc A strip of land (N17), above a cluster of signs consisting of a bow, of an archaic type (Aa23), in front of a half round loaf of bread (X1), over a strip of land (N17); on top of a standard used for carrying religious symbols (R12), written on top of a parcel of land with irrigation ditches (N24).

The bow was mirrored. The Database aims to motivate the change on the basis of available tokens, but I'm far from convinced. It says "I wonder if the bow is ever facing outwards". Roughly the same group above the standard (but without standard) is at the extreme right in the photo of:

http://sith.huma-num.fr/karnak/1128

Here the bow is outward, seemingly contradicting the above. Just because one could not find an attestation of the exact Unicode 5.2 glyph within the available time, that doesn't mean the glyph was wrong and that there was a compelling reason to change it.
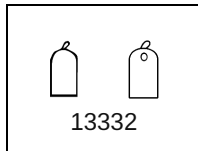

1323C

U+1323C kEH_Desc An emblem with a human face with cow ears, with two upwards line which curl inwards (R129), on top of a standard used for carrying religious symbols (R12), written on top of a parcel of land with irrigation ditches (N24).

https://www.unicode.org/L2/L2005/05313-Gardiner28-57.pdf

It is hard to escape the conclusion that the shape was incorrectly copied from Gardiner (1953) to the Unicode 5.2 code charts at the time. The Unicode 16 shape therefore seems to be an improvement.


13332

U+13332 kEH_Desc A shield with a rounded top, with a loop on top, with a boss in the center of the shield.

The original shape did not have a boss. Here it seems acceptable, though far from necessary, to refine the appearance, since Gardiner refers to an occurrence at Medinet Habu where the sign appears to have a boss; see Plate 45, column 37 of:

https://isac.uchicago.edu/sites/default/files/uploads/shared/docs/oip8.pdf

But "in the center" in the description is not accurate. I would write "near the top of the shield".
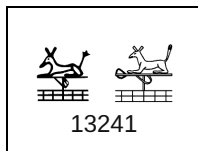
13241

U+13241 kEH_Desc The animal of Seth, lying down, tail up (E21), on top of a standard used for the carrying of religious symbols (R12) on top of a parcel of land with irrigation ditches (N24).

In the original, the standard was simplified. In the Database, I see no reflection on this. The Database has a reference to:
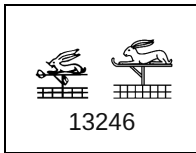
http://sith.huma-num.fr/karnak/1202

which has yet another form of the standard, with diagonal bar rather than "a loop under the horizontal beam, running over the vertical pole". It would be fine to normalize glyphs when it comes to the type of standard, and to use a default form of R12 for any signs with a standard that is similar to R12. But this is not what was done elsewhere, especially for the extended list. There appears to have been no consistent strategy for the treatment of standards.
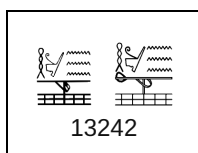
13242

U+13242 kEH_Desc A wick of twisted flax, consisting of 3 loops (V28), in front of the red crown (S3), in front of three ripples of water, vertically aligned above one another

In the original, the standard was simplified. The Database has "Transcription only. Inclusion in core based on principle, not verification with image". How then can a change to the glyph be justified?


13246

The standards are different, although both standards are often referred to as instances of R12. Also, in the original, the hare had whiskers.

The Database lists both the original shape and the new shape, in different entries, without acknowledging that the standard in U+13246 changed between Unicode 5.2 and Unicode 16. According to the Database the two forms are "practically identical". I don't necessarily disagree with this, but if the two forms are practically identical, then why was a change needed in the first place? Because an attestation could be found for one form and not for the other? Will the shape of U+13246 change back once the original form is attested?

I detect an eagerness to change glyphs in the code charts for the sake of making changes, not because there is any compelling need. To destroy and remake in one's image is not in the interests of the user community.

By the way, there is a token where the standard is somewhere in between the standards above. See 9th column from the left in the photo of:
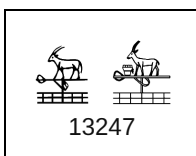
http://sith.huma-num.fr/karnak/1128

13247

There was no basket in the original shape. As the Database points out, the Unicode 5.2 form differs from what, according to the original UniKemet, should have been the source, namely Gardiner (1953). See p. 52 of:

https://www.unicode.org/L2/L2005/05313-Gardiner28-57.pdf

It is hard to escape the conclusion that the shape was incorrectly copied from Gardiner (1953) to the Unicode 5.2 code charts at the time. The Unicode 16 shape therefore seems to be an improvement.

However, the "four pieces of grain or fruit" seem overly specific and can certainly not be discerned in Gardiner (1953). I would omit this phrase from the description.


1324F
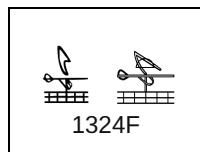
The object on the standard changed to a kind of overlay. The Database notes that a supposed token of this sign has been identified that has "X1 behind what looks like a U8". This could be in the second column in the photo of:

http://sith.huma-num.fr/karnak/1128

where X1 is to the left of what looks like U8, so "behind" in terms of reading order. (See also the first column, where there is the same combination of X1 and what looks like U8, but without standard.) I suspect the new glyph for Unicode 16 was created with the misunderstanding that "behind" meant "overlay".

Assuming that the above text correctly identifies the sign that Gardiner had in mind (which is not at all clear to me), then a redrawing of the glyph may be in order. In any case, it should not be an overlay.
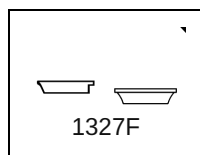

1327F

U+1327F kEH_Desc A tenoned door, written horizontally.

The original shape was asymmetric, while the new shape is perfectly symmetric. Can that be correct? I imagine a tenoned door not to be symmetric.
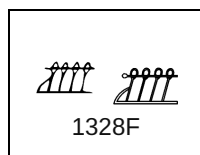

1328F

U+1328F kEH_Desc A fence, consisting of a horizontal base with four lines with bulbs at the end, angled backwards, with a curved line connecting the front of the base and the first line with a bulb, with a horizontal line with a loop at the front through the bulbs.

There was no loop at the front in the original. I don't think anyone cares that the glyph is refined. The only suggestion I have is to omit the loop from the description as overly precise.
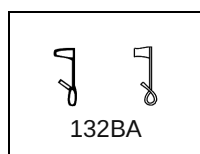

132BA

U+132BA kEH_Desc A cloth wound on a pole, an emblem of divinity (R8), on top of a bag or a sack of linen, with a tie towards the front, and a (small) loop at the back (V33).

The two shapes are similar. But I cannot relate the description to the shapes. I don't think the tie towards the front should be drawn as one with the outline of the bag. What is the "loop at the back"? May the description be overly precise?
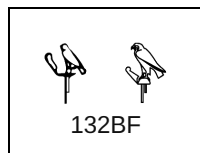

U+132BF kEH_Desc A falcon (G5) on a standard with a rounded top, with a feather (H6), angled forwards, attached to the standard in front of the falcon.

The standard has a different shape. The Database does not contain a shred of reflection on the fact that the shape has changed.
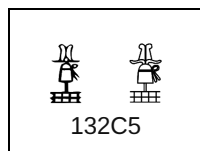
U+132C5 kEH_Desc A wig, with a fillet, with a headdress consisting of two feathers on top of the horns of a ram (S77), on top of a vertical pole, on top of a parcel of land with irrigation ditches (N24).

The headdress lost the sun disk. The Database motivates this by "more often without disk". I don't think anyone cares about the change to the glyph. But by the same token, I don't think there was ever a compelling need for any change.
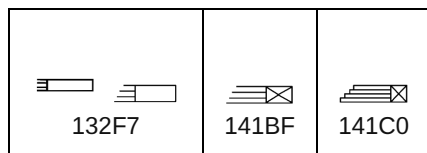
U+132F7 kEH_Desc A rectangular piece of cloth, with sloping fringes on its front short side, without internal decoration.
U+141BF kEH_Desc A rectangular piece of cloth, with sloping fringes on its front short side, with a cross shape as internal decoration.
U+141C0 kEH_Desc A square piece of cloth, with long sloping fringes on its front short side, with a cross shape as internal decoration.

In the original shape of U+132F7, the fringes were not sloping.

Erman and Grapow already listed and contrasted both the non-sloping and the sloping forms (the latter with cross). The vast majority of dictionaries and grammars use the non-sloping form of S32. One would have to have very good arguments to unceremoniously purge the traditionally used non-sloping form by replacing it with the sloping form.

I also don't see how one can argue that U+141BF and U+141C0 deserve two independent (core) code points, even though they are basically the same shape apart from aspect ratio of the crossed rectangle, while at the same time treating the non-sloping and sloping forms of S32 as the same shape to justify arbitrarily replacing one by the other in the code charts.

The Database does not contain a shred of reflection on the fact that the shape has changed.
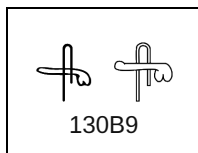
It is debatable whether the original shape has the folded piece of cloth on top, or whether the intention was to underspecify which of the two signs is on top. In the latter case, it may be adequate to describe this simply as an overlay of the two signs without specifying which of the two is on top, and leave it to the font designer how to draw this. By the same reasoning, the current glyph in Unicode 16 can also remain as it is, as one allowable graphical realization of the underlying sign.

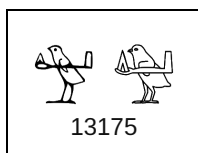There are several more such cases below.
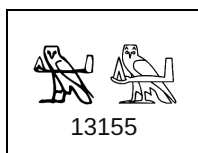
As above.
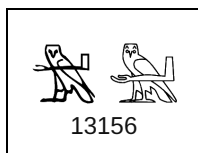
U+13156 kEH_Desc A forearm, with the palm of the hand facing upwards (D36), written over an owl (G17).
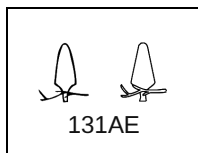
As above.
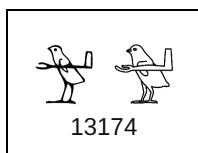
U+13174 kEH_Desc A forearm, with the palm of the hand facing upwards (D36), written over a quail chick (G43).

As above.

U+13175 kEH_Desc A forearm with the hand holding a conical loaf of bread (D37), written over a quail chick (G43).
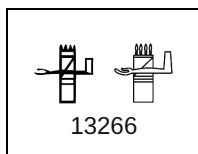
As above.

U+131AE kEH_Desc A branch, horizontally written (M3), written over a tree (M1).

As above.
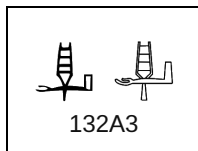

U+131B2 kEH_Desc A branch, horizontally written (M3), over an owl (G17).

As above.

U+13266 kEH_Desc A forearm, with the palm of the hand facing upwards (D36), written over a wall of the palace, with ornamental chevaux de frise on top of the wall, with internal decoration (O11).
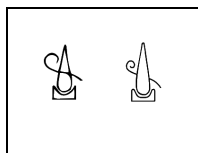
As above.

U+132A3 kEH_Desc A forearm, with the palm of the hand facing upwards (D36), written over a mast of a ship with two prongs, connected by vertical lines (P6).
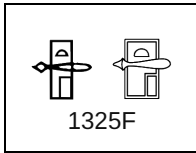
As above.
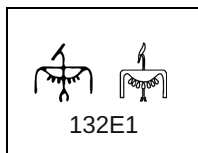
U+132E1 kEH_Desc A collar of beads (S12), written over a sceptre with a straight shaft, a forked base, topped with the head of the Seth animal (S40).

As above.

13353

U+13353 kEH_Desc A spiral, winding counter-clockwise away from its central point, ending at the right lower corner after about 1,5 turns (Z7), written behind a fire-drill in a piece of wood (U28).

As above.

1325F

U+1325F kEH_Desc A wooden column, written horizontally (O29), written over a half round loaf of bread (X1), written inside a plan of a rectangular enclosure, with an internal rectangle in the lower corner away from the reading direction (O6).

The original shape underspecified which of the signs is on top. The new shape specifies that it is the column that is on top. I have no objection to the new shape, but I would prefer for the description to leave underspecified which sign is on top, as it is not clear from the original shape either.