L2/24-093

Additional lost signs for Ancient Egyptian hieroglyphic text

Mark-Jan Nederhof

November 10, 2023

Following the proposal in L2/21-248, Unicode now has 4 characters to represent lost signs in hieroglyphic text, corresponding to a 2×2 grid. After reanalyzing the use of such characters in place of hieroglyphs of different sizes and aspect ratios, we propose to add 12 further characters for lost signs, to obtain a natural family of 16 such characters, corresponding to a 4×4 grid.

1 Introduction

As explained in L2/21-248, complete and undamaged texts are the exception in the corpus of Ancient Egyptian hieroglyphic inscriptions. This motivates introducing 'lost sign' characters that can be used in place of hieroglyphs, which implies they can interact with control characters for Ancient Egyptian, such as the horizontal joiner and the vertical joiner. They are also subject to the same rules of scaling and positioning as normal hieroglyphs (cf. L2/19-331), which means their sizes are not absolute, but may be influenced by their place inside a larger group of hieroglyphs and other lost signs.

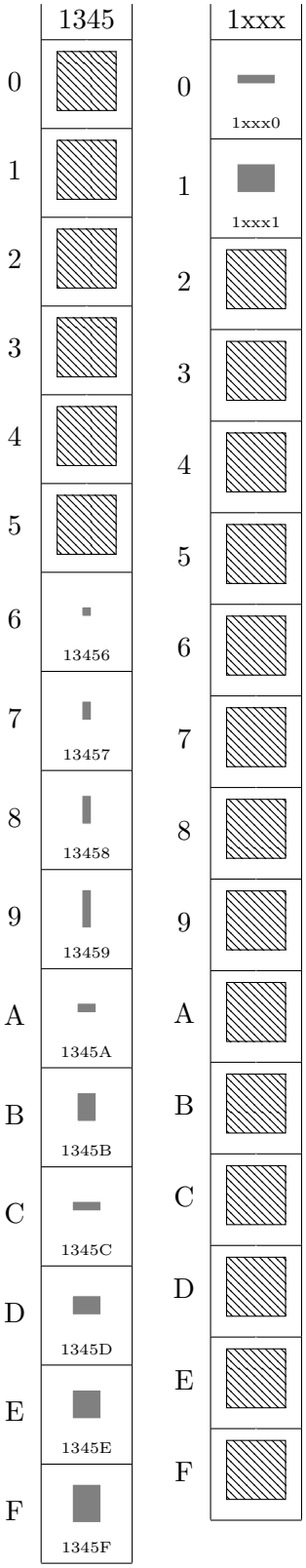
There are currently four 'lost sign' characters:

- U+13443: the LOST SIGN  of size 1 em × 1 em,
- U+13444: the HALF LOST SIGN  of size 0.5 em × 0.5 em,
- U+13445: the TALL LOST SIGN  of size 0.5 em × 1 em, and
- U+13446: the WIDE LOST SIGN  of size 1 em × 0.5 em.

Here 1 em represents the unscaled height of the common sign  .

The rationale for having multiple 'lost sign' characters is that a particular choice informs the reader about the size and aspect ratio of an unreadable part of the writing surface. This in turn suggests to the reader which sign(s) may have existed there before that part of the inscription became unreadable, which is essential to correctly interpreting the inscription. The above four 'lost sign' characters correspond to similar primitives of widely used hieroglyphic typesetting systems since the 1980s. Including these four characters in Unicode was therefore a bare minimum. This does not address the question however whether these four 'lost sign' characters are sufficient. We will argue that encoding partially damaged and incomplete hieroglyphic texts would greatly benefit from having 12 additional 'lost sign' characters, listed in Table 1. The new characters are at the granularity of multiples of 0.25 em, to naturally incorporate the four already existing 'lost sign' characters, to form a family of 4×4=16 such characters.

In Section 2 we first investigate the traditional view of the repertoire of hieroglyphs as divided into four classes depending on size and aspect ratio. Section 3 presents an empirical investigation that suggests that the traditional four classes are insufficient to account for the diversity of sizes and aspect ratios, on the basis of an hieroglyphic font. This translates to the need for a higher granularity for the 'lost signs', beyond the



Lost signs

13456		EGYPTIAN HIEROGLYPH LOST SIGN ONE QUARTER BY ONE QUARTER
13457		EGYPTIAN HIEROGLYPH LOST SIGN ONE QUARTER BY ONE HALF
13458		EGYPTIAN HIEROGLYPH LOST SIGN ONE QUARTER BY THREE QUARTERS
13459		EGYPTIAN HIEROGLYPH LOST SIGN ONE QUARTER BY ONE
1345A		EGYPTIAN HIEROGLYPH LOST SIGN ONE HALF BY ONE QUARTER
1345B		EGYPTIAN HIEROGLYPH LOST SIGN ONE HALF BY THREE QUARTERS
1345C		EGYPTIAN HIEROGLYPH LOST SIGN THREE QUARTERS BY ONE QUARTER
1345D		EGYPTIAN HIEROGLYPH LOST SIGN THREE QUARTERS BY ONE HALF
1345E		EGYPTIAN HIEROGLYPH LOST SIGN THREE QUARTERS BY THREE QUARTERS
1345F		EGYPTIAN HIEROGLYPH LOST SIGN THREE QUARTERS BY ONE
1xxx0		EGYPTIAN HIEROGLYPH LOST SIGN ONE BY ONE QUARTER
1xxx1		EGYPTIAN HIEROGLYPH LOST SIGN ONE BY THREE QUARTERS

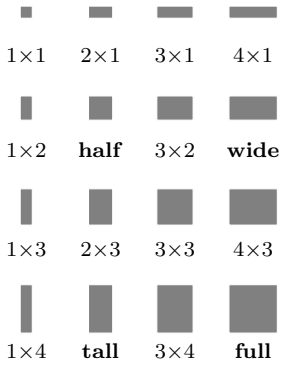


Table 1: Portion of the code chart prospectively containing the new ‘lost sign’ characters, with underspecified code points for two characters that would no longer fit in the existing Egyptian Hieroglyph Format Controls block. The diagram below the list of new characters depicts the four existing ‘lost sign’ characters (labelled **half**, **wide**, **tall**, **full**), together with the 12 newly proposed ones, which are labelled by width × height as multiples of 0.25 em.

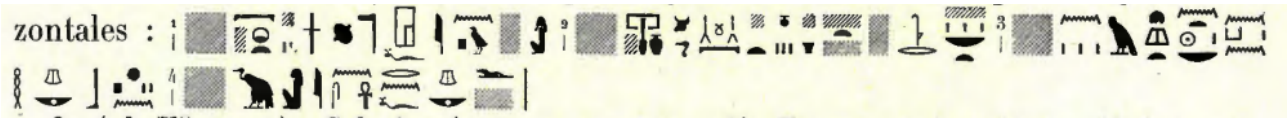


Figure 1: An hieroglyphic transcription printed well before the age of digital text processing tools [1, p. 58]. Several of the blocks of shading in this example have aspect ratios different from 1:2, 2:1 or 1:1. The second block in the first line appears to be about 2:3, and the block immediately preceding line number (3) appears to be about 4:1.

existing 2×2 grid. An obvious choice is then a 4×4 grid, such that the width and height of a 'lost sign' becomes a multiple of 0.25 em. The alternative of adding lost signs with height and width 0.33 em and 0.66 em is discussed in Section 4, and rejected on the basis of several arguments. Section 5 presents examples where the higher granularity of the 4×4 grid enhances the ability to suggest the correct interpretation of texts with lost signs, in an experimental implementation.

2 The traditional four classes

Some publications divide hieroglyphs into four classes:

- large (approximately square) signs,
- small (approximately square) signs,
- tall and narrow signs, and
- wide and low signs.

This perspective can be traced back to Gardiner's grammar [3, pp. 547–548], which provided an index to easily find a subset of the signs based on their shape, provided they are 'tall narrow signs', 'low broad signs', or 'low narrow signs', leaving an implicit fourth class of anything else. However, Gardiner makes no attempt to quantify the dimensions that determine that a sign belongs to one class rather than another. Moreover, the sizes and aspect ratios within one class can vary considerably.[1] For example, the listed 'low narrow signs' include signs such as  , which are round and thereby have a bounding box that is square, as well as  and , whose bounding boxes are clearly not square.

Both the Manuel de Codage [2] and PLOTTEXT [4, 5] include four primitives for obtaining shaded areas of the same dimensions as the four 'lost sign' characters that are now in Unicode. However, there is no reason to believe that these four primitives exactly correspond to the four classes in Gardiner's index. Moreover, these software tools were developed in a time when digital printing technology for hieroglyphic text was in its infancy, and the choice of these four primitives is more likely a technical artefact, attributable to the relative ease of their implementation in that technology, and does not reflect any inherent property of the Ancient Egyptian writing system. It can also be observed that well before the 1980s, printed hieroglyphic texts regularly used a wider range of shaded blocks; cf. Figure 1.

3 An hieroglyphic font

To investigate whether hieroglyphs naturally fall into a number of classes according to their size and aspect ratio, Figure 2 plots the distribution of different combinations of width and height of the 1072 hieroglyphic

[1] By the aspect ratio of a sign we mean the ratio between the width and the height of its bounding box.

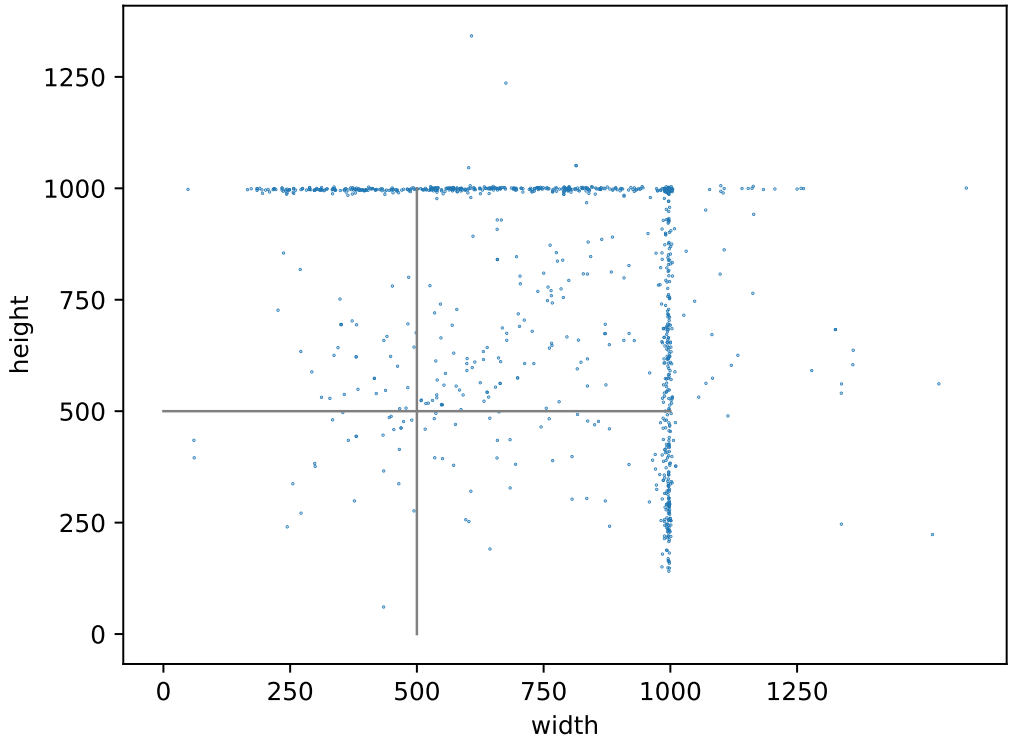


Figure 2: Distribution of width and height of hieroglyphic signs in a font, where 1000 stands for 1 em.

signs in the NewGardiner font.[2] It is clear that a significant portion of signs are either 1 em wide or 1 em high, or both. However, there is no obvious dense cluster of signs around 0.5 em × 0.5 em, and if we restrict our attention for now to square signs, then the need emerges for 0.25 em × 0.25 em and 0.75 em × 0.75 em 'lost sign' characters, to achieve better coverage of the range of sizes.

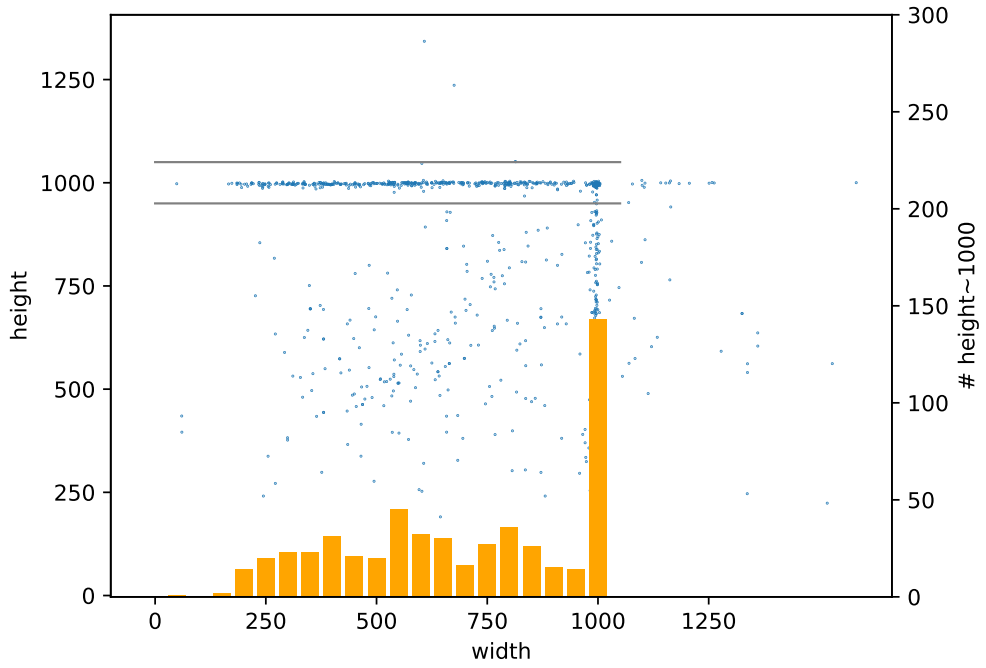
Similarly there are no obvious clusters around 0.5 em × 1 em and 1 em × 0.5 em. To make this clearer, Figure 3(a) superimposes the numbers of signs of approximately 1 em high, of varying widths in buckets of 0.05 em, and Figure 3(b) shows the height distribution for signs of approximately 1 em wide.

Figure 3(a) shows that there is a fair portion of tall signs that are a little wider than 0.5 em, but one could also point to many tall signs around 0.75 em wide, and a fair portion of the tall signs is close to 0.25 em wide. This would suggest the need for, among others, 0.25 em × 1 em and 0.75 em × 1 em 'lost sign' characters.

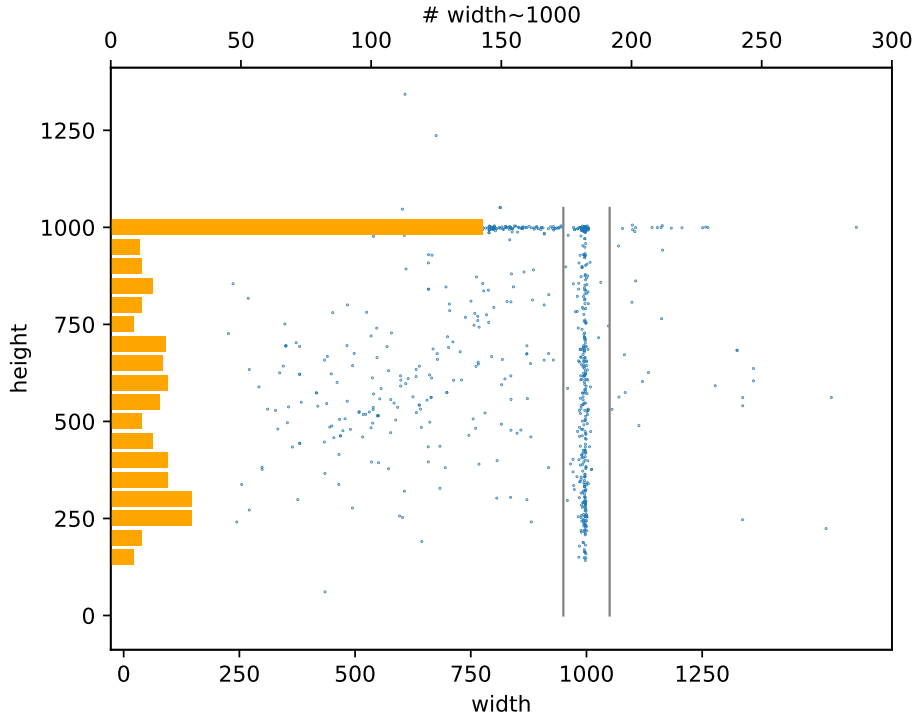
Figure 3(b) shows that there is no obvious cluster of wide signs that are around 0.5 em high. In particular, there are many wide signs that are around 0.25 em high, and there are many wide signs that are somewhere between 0.5 em and 1 em high. This would suggest the need for, among others, 1 em × 0.25 em and 1 em × 0.75 em 'lost sign' characters.

More generally, Figure 2 would suggest 'lost sign' characters for all combinations of widths and heights that are multiples of 0.25 em up to 1 em, in order to achieve better coverage over the wide spread of dimensions. There are signs in the font with width and height strictly greater than 1 em, but these are typically numerals, such as  , which are compositional. To represent such an unusually tall or wide sign that is lost, one could use multiple 'lost sign' characters, each for one occurrence of the constituent

[2] https://mjn.host.cs.st-andrews.ac.uk/egyptian/fonts/newgardiner.html



(a)



(b)

Figure 3: (a) Distribution of widths of signs of approximately 1 em high. (b) Distribution of heights of signs of approximately 1 em wide.

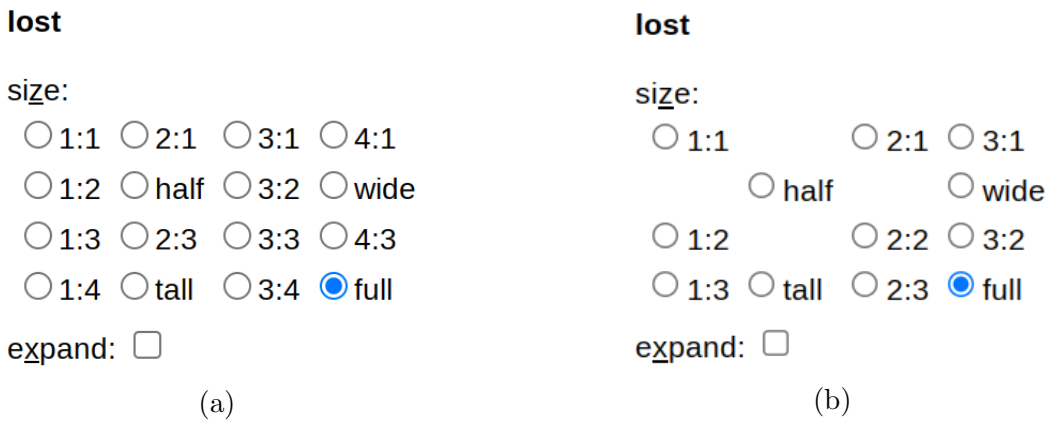


Figure 4: (a) Part of the user interface of the HieroJax editor for choosing the dimensions of a 'lost sign', when there are 16 possibilities. (b) The same but now including 'lost signs' with widths or heights 0.33 em or 0.66 em.

element, in this case  , possibly in combination with the 'expanding' variation selector **VS1** (cf. pp. 12–13 of L2/21-248) to join the 'lost sign' characters together.

4 An alternative system

In original hieroglyphic inscriptions, top-level groups of hieroglyphs are often not square.[3] In particular, top-level groups in horizontal text tend to be less wide than they are high, often with a ratio of around 2:3. Therefore, we received the suggestion to explore 'lost sign' characters of which the width or height was 0.33 em or 0.66 em. This would result in eight further characters next to the existing four, effectively superimposing a 3×3 grid on top of the existing 2×2 grid. The dimensions of the new characters would then be 0.33 em × 0.33 em, 0.33 em × 0.66 em, 0.33 em × 0.1 em, 0.66 em × 0.33 em, 0.66 em × 0.66 em, 0.66 em × 0.1 em, 1 em × 0.33 em, 1 em × 0.66 em.

There are a number of problems with this system of in total 12 'lost sign' characters. First, the fact that some top-level groups have a 2:3 aspect ratio does not necessarily mean that lost signs *within* groups have a 1:3 or 2:3 (or 3:1 or 3:2) aspect ratio. Moreover, our investigation from Section 3 does not support the idea that 1:3 or 2:3 aspect ratios have a privileged status in the repertoire of hieroglyphs.

Furthermore, integrating 'lost sign' characters with width or height 0.33 em or 0.66 em together with the existing four 'lost sign' characters leads to an awkward system of 12 characters that are hard to explain to users. Cf. Figure 4(b), where the **half** 'lost sign' character is positioned askew between four of the new characters.

Lastly, we would note that the differences between 0.33 and 0.5 and between 0.5 and 0.66 are smaller than the differences between 0.25 and 0.5 and between 0.5 and 0.75. This means that each new 'lost sign' character with a width or height 0.33 em or 0.66 em would contribute less to covering the full range of dimensions of hieroglyphs, relative to the proposed new 'lost sign' characters whose widths and heights are multiples of 0.25 em. Related to this, where a 'lost sign' character should suggest a hieroglyph of rough dimensions 0.66 em × 1 em, it is highly likely that a 'lost sign' character of dimensions 0.75 em × 1 em would be just as suitable; possibly the human eye would not even be able to distinguish between blocks of shading of these similar dimensions, as part of ordinary printed text.

[3] Despite of this, the misnomer 'quadrat' regrettably continues to be used by some to refer to a top-level group of hieroglyphs.

5 Experimental implementation

To be able to experiment with the proposed characters, two variants were created of HieroJax[4], one for the proposed system of 16 'lost sign' characters altogether, and the other for the alternative system of 12 characters from Section 4. Both variants of HieroJax come with dedicated graphical editors, which in particular allow dimensions of lost signs to be changed, to one choice out of 16 or out of 12, respectively. For a number of fragments of hieroglyphic, two encodings were compiled and rendered, one assuming the text was fully undamaged, and the other assuming that some signs were lost.

Figure 5 is a screenshot of a web page rendered using the variant of HieroJax with a 4×4 grid. In example 1, there are two examples of tall narrow signs,  and , that were each replaced by a 0.25 em × 1 em 'lost sign' character. A 0.5 em × 1 em character instead would have suggested much wider hieroglyphs. In example 2, there is an example of a 1 em × 0.25 em 'lost sign' character for . There is also a 0.75 em × 1 em 'lost sign' for  ; here 0.5 em × 1 em would have been too narrow and 1 em × 1 em would have been too wide. Both example 3 and example 4 demonstrate the added value of having 'lost sign' characters of height 0.75 em, spanning less than the height of the line, but more than half the height of the line. For example 3, we also see the use of a 0.25 em × 0.5 em 'lost sign' character for a vertical stroke. In example 5, use is made of a 0.25 em × 0.25 em 'lost sign' character for °. In example 6, a 0.5 em × 0.5 em and a 0.75 em × 0.5 em 'lost sign' character are combined to suggest  .

Figure 6 shows the same examples, with now with the alternative system from Section 4. Inevitably, there will be individual cases where the hieroglyphs that were lost happen to have width or height around 0.33 em or 0.66 em, and where the 3×3 grid allows for slightly more desirable dimensions. Overall however, the mixed 2×2 and 3×3 system covers the wide range of dimensions less well. For example 3, where earlier we used 0.5 em × 0.75 em for  and 0.25 em × 0.5 em for the two vertical strokes, now the best fitting character for all three would be 0.33 em × 0.66 em, which is a little too narrow for and a little too tall for the vertical strokes; note the impact on the sizes and positions of the remaining five vertical strokes. In example 6, the rightmost two 'lost sign' characters are also less suggestive of than in the case of the 4×4 grid.

6 Properties

The properties of the new 'lost sign' characters will be like those of the existing ones, including the possible combination with the variation selector U+FE00 for 'expansion'. For reference, the properties of the existing characters are listed in Tables 2 and 3.

Acknowledgements

Gratefully acknowledged are discussions with Peter Dils, Andrew Glass, Jorke Grotenhuis, Stéphane Polis, and Daniel Werning.

References

[1] B. Bruyère. *Deir el Médineh – Année 1926*. L'Institut Français d'Archéologie Orientale, Le Caire, 1952.

[4] https://nederhof.github.io/hierojax/

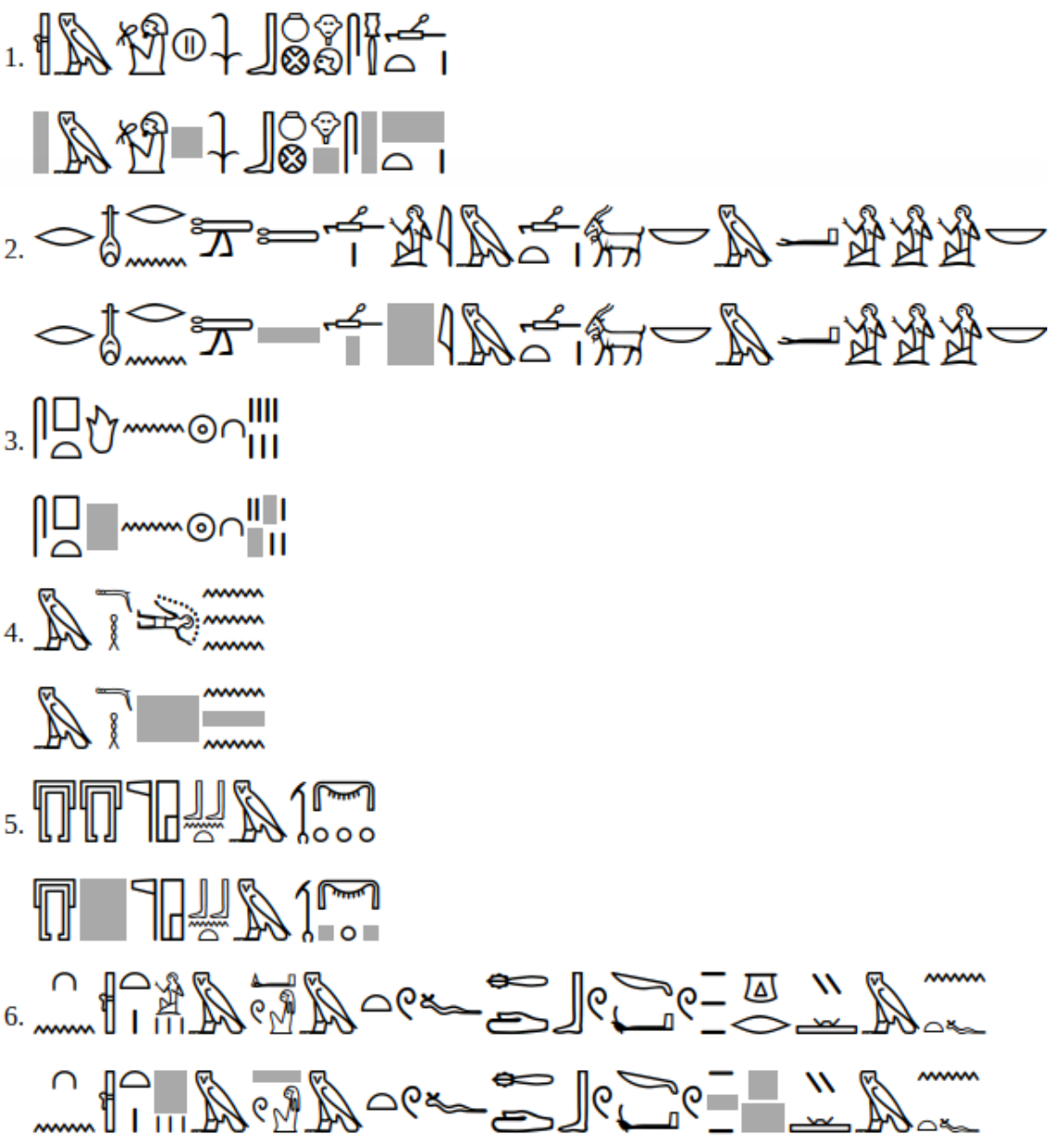


Figure 5: Examples with a 4×4 grid.

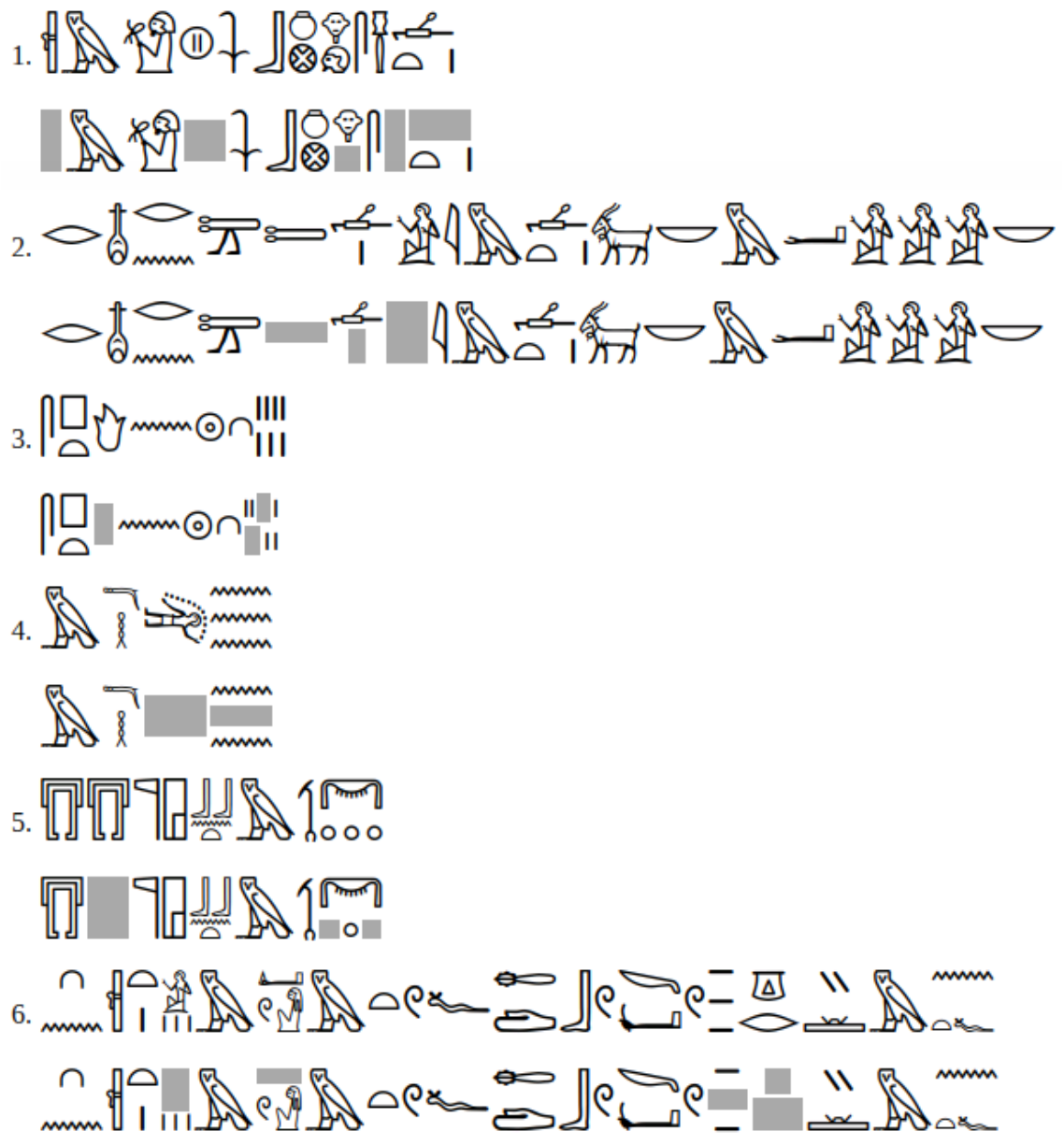


Figure 6: Examples with a 3×3 grid added to the existing 2×2 grid.

[2] J. Buurman, N. Grimal, M. Hainsworth, J. Hallof, and D. van der Plas. *Inventaire des signes hiéroglyphiques en vue de leur saisie informatique – Informatique et Égyptologie 2.* Institut de France, Paris, 3rd edition, 1988.

[3] A. Gardiner. *Egyptian Grammar.* Griffith Institute, Ashmolean Museum, Oxford, 1957.

[4] N. Stief. Hieroglyphen, Koptisch, Umschrift, u.a. – ein Textausgabesystem. *Göttinger Miszellen*, 86:37–44, 1985.

[5] N. Stief. PLOTTEXT – ein Programmsystem zur Ausgabe von Texten. Version 4.09, Regionales Hochschulrechenzentrum (RHRZ) der Universität Bonn, 2001.

```
# ================================================

# SentenceBreakProperty.txt

13441..13446  ; OLetter # Lo   [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN

# ================================================

# WordBreakProperty.txt

13441..13446  ; ALetter # Lo   [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN

# ================================================

# Blocks.txt

13430..1345F; Egyptian Hieroglyph Format Controls

# ================================================

# EastAsianWidth.txt

13441..13446; N # Lo     [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN

# ================================================

# LineBreak.txt

13441..13446 ; AL # Lo     [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN

# ================================================

# NamesList.txt

@ Blank and lost signs
13443 EGYPTIAN HIEROGLYPH LOST SIGN
        ~ 13443 FE00 expanded
13444 EGYPTIAN HIEROGLYPH HALF LOST SIGN
        ~ 13444 FE00 expanded
13445 EGYPTIAN HIEROGLYPH TALL LOST SIGN
        ~ 13445 FE00 expanded
13446 EGYPTIAN HIEROGLYPH WIDE LOST SIGN
        ~ 13446 FE00 expanded

# ================================================

# Scripts.txt

13441..13446  ; Egyptian_Hieroglyphs # Lo    [6] EGYPTIAN HIEROGLYPH FULL BLANK..
                                     #               EGYPTIAN HIEROGLYPH WIDE LOST SIGN
```

Table 2: Properties of the existing ‘lost sign’ characters (1/2).

```
# ================================================

# UnicodeData.txt

13443;EGYPTIAN HIEROGLYPH LOST SIGN;Lo;0;L;;;;;N;;;;;
13444;EGYPTIAN HIEROGLYPH HALF LOST SIGN;Lo;0;L;;;;;N;;;;;
13445;EGYPTIAN HIEROGLYPH TALL LOST SIGN;Lo;0;L;;;;;N;;;;;
13446;EGYPTIAN HIEROGLYPH WIDE LOST SIGN;Lo;0;L;;;;;N;;;;;

# ================================================

# VerticalOrientation.txt

13441..13446   ; U  # Lo     [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN
```

Table 3: Properties of the existing ‘lost sign’ characters (2/2).