Rethinking Unicode: How to Digitally Encode Non-Linear Written Artefacts?

A Tentative Encoding of Codex Mendoza, Folio 2r

Antonio Perri, Luciano Perondi & Daniele Capo

Abstract. Unicode encoding, even after 23 years of development since the release of version 1.0, continues to reflect the limitations of analog and alphabetic typography, emphasizing a linear and analytic structure where characters are arranged in a mechanistic sequence based on the arbitrary segmentation of speech inherent in alphabetic systems. In critically assessing this aspect, at the same time we address the main problem of avoiding the bias of *homo alphabeticus*, and of course *typographicus*.

As several scholars argued, Aztec, Mixtec, and other non-Maya Mesoamerican pictographic writing systems (often considered as basically non-glottographic systems) cannot be reduced to linear sequences of isolated characters; rather, their graphic plasticity conveys semantic (and linguistic) contents through conventional patterns of units arranged in graphic space according to non-linear principles. This, in fact, explains the enduring opposition to accepting such a notation as writing "proper"; however, this account still suffers from a theoretical bias according to which only more or less systematic phoneticism defines "true writing". In contrast, any accurate consideration of the ways in which writing works in specific visual artifacts shows that writing, unlike speech, does not need to represent phonetically an utterance in order to provide a complete encoding of linguistic content.

We argue that the relevant visual values in this system are ordered in subsets (topological, spatial, and logical) according to the variables devised by Bertin. We suggest then that each unit is a "character" (in the Unicode vocabulary), but it could be coded either as a positive unit (i.e., a graphic sign) or as an "empty" spatial, topographical relation between visually expressed and coded signs. Following the digital method aimed at processing sets of basic graphic features as non-linear combinations of variables, we pointed out that Aztec pictorially agglutinated complexes (or multi-unit arrangements forming perceptually autonomous units, e.g., a toponym or a proper name) display an overall structure

Antonio Perri 0000-0002-0453-4849
Università degli Studi Suor Orsola Benincasa - Napoli, Italy
E-mail: antonio.perri@unisob.na.it

Luciano Perondi 0000-0002-7198-4030
Università Iuav di Venezia, Italy
E-mail: luciano.perondi1@gmail.com

Daniele Capo
Università degli studi Sapienza Università di Roma, Italy

Y. Haralambous (Ed.), *Graphonlinguistics in the 21st Century 2024. Proceedings*
Graphonlinguistics and Its Applications (ISSN: 2681-8566, e-ISSN: 2534-5192), Vol. 12.
Fluxus Editions, Brest, 2026, pp. 633–660. https://doi.org/10.36824/2024-graf-perr
ISBN: 978-2-487055-12-4, e-ISBN: 978-2-487055-13-1

similar to Unicode encoding of emoji since it is a combination of different layers of variables. However, as more complex and higher-ranked coded variables are needed to account for broader textual compositions, we suggested that the coding systems available so far in the Unicode standard (including the one devised for emojis) are still not flexible enough to accommodate an appropriate encoding of Aztec textual units, above all the spatial arrangement.

In this paper, we will sketch the guidelines for an exhaustive coding of the perhaps most iconic page of Codex Mendoza: the folio 2r, which relates the founding of Mexico-Tenochtitlan and has been studied extensively by Joaquín Galarza and his research group at the beginning of the 1980s.

A Speculative Hypothesis for Encoding Non-Linear Scripts in Mesoamerican Writing Systems

This article addresses the complexities involved in digitally encoding non-linear written artefacts, particularly historical documents such as the *Codex Mendoza*.

Unicode is crucial for accurately representing diverse writing systems, particularly in a multilingual context. Prior to its establishment, many writing systems were constrained by limited encoding options, typically supporting only two scripts, which created challenges for documents that included multiple languages. The introduction of Unicode (Dombrowski, Goregaokar, Joeng, and Kamran, 2024) has resolved many of these compatibility issues, allowing for a more coherent approach to text encoding that supports the display and processing of complex characters and scripts, including those that require bidirectional text rendering. While Unicode has transformed the digital representation of text by encompassing over 128,000 characters from various scripts and languages into a coherent digital framework, its application often falls short when dealing with non-linear texts that incorporate intricate layouts and visual elements. While it has succeeded in providing a foundation for representing a vast array of scripts, the system inherently inherits limitations from its origin in linear and alphabetic paradigms. These limitations are particularly evident when attempting to encode complex systems that require bidirectional text rendering or contextual markers, as is often the case with historical and non-linear writing systems like *non-Maya Mesoamerican ones*.

Non-linear texts often involve complex layouts and visual elements that standard Unicode encoding does not adequately address. As noted by Perri et al. (2024), *Aztec writing* shows a more or less systematic (but not merely phonetic) alignment between graphic elements and linguistic content, where *visual variables* and layout composition integrate a semantic content which must be expressed linguistically.

In critically assessing this aspect, at the same time we address the main problem of how to avoid the bias of *homo alphabeticus* (Brokaw,

2022)—and, of course, *typographicus*. Folio 2r of *Codex Mendoza* is a prominent example of the resources visually encoded by images-glyphs, since it is structured as an historical and geographical account based on a specific spatial arrangement of signs (Bleichmar, 2015).

As several scholars argued, Aztec, Mixtec, and other *non-Maya Mesoamerican pictographic writing systems* (often considered as basically non-glottographic systems) cannot be reduced to linear sequences of isolated characters; rather, their graphic plasticity conveys semantic (and linguistic) contents through conventional patterns of units arranged in graphic space according to non-linear principles (Zamora Corona, 2022).

When analyzing Mesoamerican writing systems, it is important to note that iconic representations are inherently two-dimensional and, in some cases, may incorporate a 'coded' third dimension. This third dimension emerges through the sequencing of visual layers or superimposed levels, achieved via relative dimensional contrasts on the inscribed surface (Fedorova and Perri, 2022).

For example, in folio 2r of *Codex Mendoza*, the upper section of the planar textual space—which at first glance appears as a diagrammatic and cartographic layout—in fact articulates a narrative of the founding of Tenochtitlan. This area is divided into four sections by streams of water, whose diagrammatic and planar arrangement conveys clear linguistic information in a visual form.

Case Study: Folio 2r of *Codex Mendoza*

We selected folio 2r of the *Codex Mendoza* (Figure 1) for its highly complex layout, including various features characteristic of *Aztec writing*; indeed, it is an ideal stress test for our provisional hypothesis of encoding non-linear Aztec and other non-linear writing systems.

The *Codex Mendoza* (Berdan and Anawalt, 1992) is a significant Aztec early colonial codex drawn around 1541, shortly after the Spanish conquest of the Aztec Empire. Commissioned by Antonio de Mendoza, the first Viceroy of New Spain, this document played—and still plays—the role of a vital record of Aztec history, culture, and social structure. It is formed by three main sections (two of them copies of pre-Hispanic documents), each providing distinct insights into Aztec history, economy, and society intended to be of use by the new Spanish administration during the early colonial period.

The first section of the *Codex Mendoza* chronicles the history of the Aztec rulers from the founding of Tenochtitlan in 1325 up to the conquest by the Spanish in 1521. It details the reigns of various rulers and the towns they conquered, illustrating the progressive expansion and consolidation of the Aztec Empire. The second section focuses on the

Fig. 1. Codex Mendoza, folio 2r. Public domain image, sourced from the Bodleian Libraries

tribute system established by the Aztecs. This economic and administrative information is closely related to the so-called *Matrícula de Tributos*, another pre-Hispanic document that also detailed tribute payments. The third section of the *Codex Mendoza*—which many scholars assume does not follow a traditional model but was conceived for use by the Conquerors while adhering to the pictographic rules of the system—provides an account of everyday life among the Aztecs.

The codex combines Aztec pictograms with Spanish text and glosses and is among the most extensively studied manuscripts of the *Nahuatl* pictorial tradition. This 'bilingual' nature of the manuscript enhanced scholarly studies of the Mendoza, soon making it a vital resource for historians, linguists, and anthropologists alike.

Folio 2r of the *Codex Mendoza* (Galarza et al., 1983a,b) provides a critical glimpse into the founding of Tenochtitlan, the capital of the Aztec Empire. It specifically captures the momentous event of Tenochtitlan's founding, embodying the cultural significance of this event in Aztec history by showing (and narrating) the foundational myth and rituals crucial to the Aztecs' identity and governance. The visual text of folio 2r articulates at once geographic space, key figures, and events involved in the establishment of the city.

Graphic elements such as colors, positions, orientations, and attire of pictorial characters are not merely decorative or generic expressions of a visual storytelling; on the contrary, each unit integrates one (or more) linguistic roles. Any alteration of the arrangement of these elements, then, would result in a change of textual meaning and linguistic content. This shows that the overall linguistic content to be spoken is inherently encoded within these visual features.

The *Codex Mendoza* exemplifies a sophisticated integration of graphic and linguistic elements in a spatial organization, revealing the *synsemic* nature of its graphic structure. The codex employs a multidimensional approach where elements are arranged in ways that encode multiple layers of meaning simultaneously. This is particularly evident in the ability of writers-painters to merge topography, chronology, and sociopolitical hierarchies within a single compositional framework.

For instance, the spatial arrangement of folio 2r narrates the founding of Tenochtitlan, the central political and cultural site of the Aztec Empire. The representation of the city is sectioned into four quadrants by water channels, a configuration that acquires a linguistic value inasmuch as it integrates the *Nahuatl* phrase *in atl quinxexeloa nahui calpulli* ('the water divides the space into four districts'). This graphic division is not merely decorative but conveys critical geographic and administrative information, illustrating the early city's organization into *calpulli* (districts) governed by specific noble lineages.

Moreover, chronological information is also integrated into the visual narrative. A frame of glyphs enumerating years surrounds the cen-

tral imagery, anchoring the depicted events within a specific temporal flow. The emplacement and sequencing of these year-glyphs are devised according to the Aztec calendrical system, in which cyclical time structured both historical narration and ritual practices. Notably, the glyph marking the *xiuhmolpilli* ('binding of the years') highlights the completion of a calendric cycle, reinforcing the temporal depth of this narrative.

Hierarchical relationships among figures, such as nobles and warriors, are visually expressed in the layout through size, position, and orientation. Thus, for example, the central depiction of the eagle perched on a cactus—the latter transcribing the place name of Tenochtitlan—is surrounded by warriors and district leaders, whose spatial arrangement diagrammatically expresses their roles and status within the Aztec sociopolitical hierarchy.

The *Codex Mendoza*, thus, demonstrates how what we elsewhere called *synsemia* (Perondi, 2012), i.e., a deliberate spatial (and non-linear) articulation of elements, is devised to encode complex, multilayered meanings that transcend linear textual conventions. This specific feature underscores the need for encoding models that are able to preserve not only the visual integrity of such artifacts but also their semantic and cultural contexts.

Exploring *Synsemia* and *Entaxis*: Visual Approaches to Non-Linear Linguistic Structures in *Aztec writing*

Unlike phonetic alphabets, *non-Maya Mesoamerican writing systems* operate through a combination of smaller graphic elements (what we may call "glyphs") and spatial configurations. Both of them express systemically relevant linguistic contents.

This page of the *Codex Mendoza* can be analyzed at two levels: we will label them the *entactic* level and the *synsemic* level. The relevant framework for analyzing *Aztec writing* articulates these two distinct levels. The entactic level, as described by Vaillant (1999), involves the composition of graphic units into coherent forms, representing smaller linguistic components such as morphemes or syllables, and may also explain the structural composition of glyphs. In contrast, the synsemic level, outlined by Perondi and Romei (2010), encompasses the broader spatial and relational arrangements that convey meaning through contextual and hierarchical relationships, effectively describing the 'toposyntax' of the text.

To encode these levels effectively, this study leverages the semiology of Bertin (2011), highlighting key visual variables such as shape, color, texture, orientation, and position. These variables interact hierarchically, facilitating graphical "association" and "dissociation," which enable the construction of complex graphic structures at both the entactic and synsemic levels (Bertin, 2011; Perri, 2006; Vaillant, 1999). This

model incorporates spatial and relational markers to preserve the nonlinear semantics of *Aztec writing* in a multilayer framework. It demands considerable flexibility to account for the distinctive features of *Aztec writing* in the *Codex Mendoza*. At the entactic level, this is necessary to address the agglutinative structure of the *Nahuatl* language, and at the synsemic level, to accurately capture the spatial and hierarchical relationships between units.

A key feature of *Nahuatl*, as an agglutinative and polysynthetic language, is its direct role in shaping the intricate relationship between glyphs and language in the *Codex Mendoza*. In polysynthetic systems, words are formed incorporating two or more lexical morphemes, each carrying specific semantic or grammatical functions. This morphological characteristic, together with its omipredicativity (Launey, 1994)—, the syntactic property according to which the names could act as predicates in a given sentence, if conjugated to a given grammatical person (incidentally, for the third person of any narrative, a zero morpheme)—is aptly mirrored in the visual composition of glyphs, where multiple graphic elements are agglutinated into complex units to convey layered meanings and narratives. For example, the glyph for *Ocelopantzin* (Noble Lord of the Jaguar Banner) integrates distinct components: a banner (*pantli*), adorned with jaguar spots, and the noble suffix (*-tzin*), each corresponding to individual morphemes in *Nahuatl*. This seamless integration of linguistic and visual elements illustrates how Aztec glyphs encode both structural and semantic principles of the *Nahuatl* language. This system leverages spatial and graphic configurations to encapsulate meaning, highlighting the critical role of linguistic agglutination in any effort to digitally encode such systems (Perondi, Perri, et al., 2018; Zamora Corona, 2022).

However, current Unicode standards are inadequate for encoding the continuous variability and relational complexity of Aztec graphemics. For instance, variables such as shape demonstrate virtually infinite variability, highlighting the need for innovative encoding approaches. Since the structure of these writing systems is mapped onto the agglutinative nature of languages transcribed, Unicode encoding on a character-by-character basis would result in the generation of thousands of characters; moreover, it would not solve the problem of giving a readable linguistic value to the spatial arrangement of elements.

The proposed encoding model is built on the principles of relational encoding and hierarchical composition encoding. *Relational encoding*, related to the associative variables following the terminology introduced by Bertin (2011, pp. 42–48), involves the use of "connecting characters" and agglutination markers to link distant units and articulate their semantic relationships. This encoding reflects the use of associative variables (e.g., objects of the same color, orientation, and shape are related) and graphic connectors to establish relationships within the visual text.

This approach also enables the representation of elements such as paths or sequences, exemplified by the glyph *othli* for "road," which visually connects starting and ending points. *Hierarchical composition encoding*, related to the dissociative variables (Bertin, 2011), employs differences in size and position to signify priority and establish hierarchy within texts, akin to the layering commonly used in graphical interfaces.

The proposed encoding model not only might enable the digital preservation of the non-linear nature of *Aztec writing* but also provides valuable insights for encoding other non-alphabetic systems, such as modern visual artifacts like infographics and flowcharts. This approach aligns with the theories of Harris (Harris, 1986; 2000) and Klinkenberg and Polis (Klinkenberg and Polis, 2018), which conceptualize writing as a semiotic system distinct from speech.

Hypothesis for a Descriptive Metalanguage

To implement the encoding principles outlined in the previous chapter, it was necessary to incorporate features that are currently absent (or only partially present) in the Unicode standard. Specifically, these include:

- a counter,
- seven distinct layers for the graphic attributes of the elements (covering color, orientation, value, texture, relative position, number, and shape),
- three synsemic composition layers addressing absolute orientation, absolute position, and absolute size.

The system is structured into two primary levels: entactic and synsemic. The entactic level is designed to encode agglutinated units, while the synsemic level focuses on the overall organization of these units. To facilitate this, two pairs of character markers are incorporated: one pair for marking the beginning and end of agglutination, and another for marking the beginning and end of *synsemia*. Furthermore, character-counters are proposed to act as pointers to specific sections of the encoded text.

The encoding of simple units relies on characters representing seven graphic attributes: Shape, Color, Texture, Amount, Position, Orientation, and Value. This mechanism exhibits similarities to the Unicode approach for encoding colors in emoji characters (Consortium, 2024), in particular Emoji Modifiers (1.4.4) and Emoji Sequences (1.4.5). Since this is a speculative hypothesis, we did not fully adhere to Unicode notation, but we assume it is possible to at least adapt the notation for *entaxis*.

To represent the encoding, we devised a metalanguage where the markers are represented between two | and the units carry the infor-

mation about the graphic attributes that apply to the unit. The syntax for each unit takes this form:

```
Unit A (NAME, name in Nahuatl)
    [layer a.1 -- Shape (SHAPE, name in Nahuatl),
    layer a.2 -- Color (COLOR, name in Nahuatl),
    layer a.3 -- Texture (TEXTURE, name in Nahuatl),
    layer a.4 -- Amount (NUMBER, number in Nahuatl),
    layer a.5 -- Position (AROUND, position in Nahuatl),
    layer a.6 -- Orientation (AROUND, position in Nahuatl),
    layer a.7 -- Value (PERCENTAGE)];
```

The counter `A` identifies the element itself inside the text encoding. `NAME` is the Unicode name of the character, followed by the name to be read in *Nahuatl*. After the unit, in brackets, the seven graphic attributes are listed as layers with their names always in *Nahuatl*.

The marker of agglutination takes this form:

```
| initial marker of agglutination [A #counter of agglutinations#]
(NAME OF THE UNIT A+NAME OF THE UNIT 1+…+NAME OF THE UNIT M,
name in Nahuatl)|
```

The counter between brackets identifies the agglutination itself inside the text encoding. After the counter, in parentheses, the units enclosed by the agglutination are rendered explicitly with their Unicode name and joined with the addition sign. After this, the name in Nahuatl of the whole agglutination is also given.

The syntax for the synsemic marker takes this form:

```
| initial marker of synsemia [N #counter of synsemia#]
[M #number of agglutinations#]
    [Layer b.1 Orientation (-) - Layer b.2 Size (x; y) - Layer b.3
    Position (#main object#; #verse#; #relative position of other
    objects#)]
```

The counter `N` refers to the synsemic group itself, while the counter `M` refers to the number of agglutinated units contained inside the synsemic group. Synsemic layers (like unit layers) are listed between brackets. The syntax of `Layer b.3 (Position)` identifies the position of elements (agglutinations and units) enclosed by the synsemic group. The first attribute refers to the position of the main element inside the synsemic group, the second attribute identifies the verse in which the elements are meant to be displayed. Finally, the third attribute gives the position of the other elements enclosed relative to the main object.

An example of the use of this metalanguage is given below:

```
| initial marker of synsemia [N #counter of synsemia#]
  [M #number of agglutinations#]
  [Layer b.1 Orientation (-) - Layer b.2 Size (x; y) -
   Layer b.3 Position (#main object#; #verse#;
   #relative position of other objects#)]
```

```
| initial marker of agglutination [A #counter of agglutinations#]
  (NAME OF THE UNIT A+NAME OF THE UNIT 1+…+NAME OF THE UNIT M,
   name in Nahuatl)|
Unit A (NAME, name in Nahuatl)
[layer a.1 -- Shape (SHAPE, name in Nahuatl),
 layer a.2 -- Color (COLOR, name in Nahuatl),
 layer a.3 -- Texture (TEXTURE, name in Nahuatl),
 layer a.4 -- Amount (NUMBER, number in Nahuatl),
 layer a.5 -- Position (AROUND, position in Nahuatl),
 layer a.6 -- Orientation (AROUND, position in Nahuatl),
 layer a.7 -- Value (PERCENTAGE)];
| end marker of agglutination [A] |
| end marker of synsemia [N] |
```

For brevity, the full code is not included; instead, […] is used to indicate sections that are repetitive and have not been explicitly detailed.

Hypothesis for Segmenting and Encoding Folio 2r

Folio 2r of *Codex Mendoza* provides an overview where the warriors are positioned inside a specific portion of the landscape, which is sectioned by water. In turn, the warriors surround Tenochtzin, with all of them focused on the depiction of the Tenochca war at the center. The hypothesis suggests that both an entactic level and a synsemic level could be utilized in the encoding process. While this dual-level approach might be implemented in editing software, it has been coded here as a speculative hypothesis in order to code also the linguistic dimension of synsemic composition.

The Reckoning of Years

On the page, the chronological framework is meant to be read first (Figure 2), either by fully phonetizing the fifty-one years of the count or by identifying the year at the beginning (*Ome Calli xihuitl*, 'Two House year') and the year at the end of the count (*Matlactli Omeyi Acatl*, 'Thirteen Reed year'). Additionally, the year Two Reed is marked to signify the conclusion of an Aztec calendric century of fifty-two years, known as the *xiuhmolpilli* ('the binding of the year').

> *Ome acatl xihuitl, ipan in yancuican oqu-ilpique in inxiub in huehuetque Mexica ye intoca Tenochca in oncan atlitic Tenochtitlan.*
>
> 'Year Two Reed, for the first time the Old Mexicans called Tenochca did the binding of their years in the water [the lagoon] of Tenochtitlan.'

FIG. 2. The reckoning of years, section of the folio 2r

```
| initial marker of synsemia [1-52 (list of agglutinations)]
[…]
| initial marker of agglutination [1] (TWO+HOUSE+YEAR, ome calli xihuitl)|
[…]
| end marker of agglutination [1] |
[…]
| end marker of synsemia [1-52 (list of agglutinations)]
```

Year Two House (Figure 3)

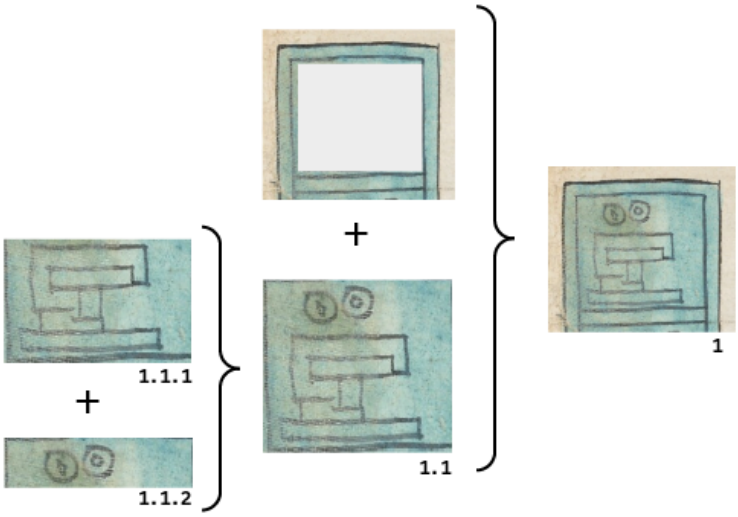

FIG. 3. Year Two House, glyph decomposition, section of the folio 2r

```
| initial marker of synsemia [52 (list of agglutinations)]
[Layer b.1 Orientation (UPRIGHT) - Layer b.2 Size (1/12; 1/18) -
 Layer b.3 Position (AROUND; FROM UPPER LEFT; COUNTERCLOCKWISE)]

| initial marker of agglutination [1] (TWO+HOUSE+YEAR, ome calli xihuitl)|
Unit 1.1 (YEAR, xihuitl)
    [layer a.1 -- Shape (SQUARE, xihuitl),
    layer a.2 -- Color (TURQUOISE, xihuitl),
    layer a.3 -- Texture (-),
    layer a.4 -- Amount (1),
    layer a.5 -- Position (AROUND),
    layer a.6 -- Orientation (-),
    layer a.7 -- Value (100%)];
```

```
| initial marker of agglutination [1.2] (TWO+HOUSE, ome calli)|
Unit 1.2.1 (HOUSE, calli)
    [layer a.1 -- Shape (A HOUSE, calli),
    layer a.2 -- Color (TURQUOISE),
    layer a.3 -- Texture (-),
    layer a.4 -- Amount (1),
    layer a.5 -- Position (BELOW),
    layer a.6 -- Orientation (-),
    layer a.7 -- Value (100%) ];
Unit 1.2.2 (TWO, ome)
    [layer a.1 -- Shape (CIRCLE),
    layer a.2 -- Color (TURQUOISE),
    layer a.3 -- Texture (-),
    layer a.4 -- Amount (2, ome),
    layer a.5 -- Position (ABOVE),
    layer a.6 -- Orientation (-),
    layer a.7 -- Value (100%)]
| end marker of agglutination [1.2] |
| end marker of agglutination [1] |
[…]
| end marker of synsemia [1-52 (list of agglutinations)]
```

Year Two Reed (Figure 4)

```
| initial marker of synsemia [1-52 (list of agglutinations)]
[…]

| initial marker of agglutination [27] ({inin}+YEAR+REED
+TWO CIRCLES+BINDING, Ome Acatl xihuitl)|
[…];

| initial marker of agglutination [27.1] |
Unit 27.1.1 (REED, acatl)
[…];

Unit 27.1.2 (TWO, ome)
[…];

Unit 27.1.3 (BINDING, ipan tlaxiuh molpilli)
        […];
| end marker of agglutination [27.1] |

Unit 27.2 (...)
        [layer a.1 - Shape (STROKE, AZTEC CONNECTING CHARACTER)
        …
        layer a.5 - Position (INBETWEEN)];

| initial marker of agglutination [27.3] (PRECIOUS+ARROW+SMOKES
+ONTO THE HEAT OF THE YEAR TO COME, xiuhtlacochtli) |
```

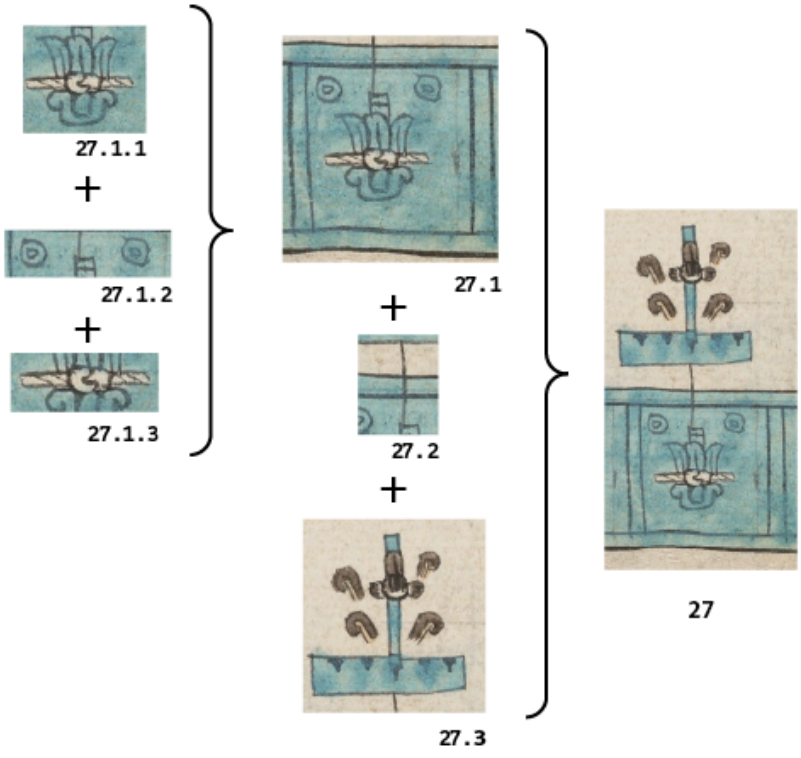

FIG. 4. Year Two Reed, section of the folio 2r

```
Unit 27.3.1 (ARROW, tlacochtli)
[…];

Unit 27.3.2 (PRECIOUS, xihuitl)
[…];

Unit 27.3.3 (SMOKES, popoca);

Unit 27.3.4 (ONTO THE HEAT OF THE YEAR TO COME, ipan in xiuhtlatlaliztli);

| end marker of agglutination [27.3] |
| end marker of agglutination [27] |
```

FIG. 5. Divided in four by water, section of the folio 2r

Divided in Four by Water (Figure 5)

```
| initial marker of synsemia [2] [1 (list of agglutinations)]
    [Layer b.1 Orientation (-) -
     Layer b.2 Size (2/3; 7/8) -
     Layer b.3 Position (INSIDE UP; FROM OUTSIDE; ONE INSIDE THE OTHER)]

| initial marker of agglutination [1] (DIVIDED+IN FOUR+BY WATER,
in atl quinxexeloa nahui calpulli)|
Unit 1 (SURROUNDED BY WATER, inten atl apano)
    [layer a.1 -- Shape (SQUARE, nohuiampa ixquich),
     layer a.2 -- Color (TURQUOISE, xihuitl),
     layer a.3 -- Texture (WAVING EDGE, apano #flowing water#),
     layer a.4 -- Amount (1),
     layer a.5 -- Position (AROUND),
     layer a.6 -- Orientation (-),
     layer a.7 -- Value (100%)];

| initial marker of agglutination [1.1] (INTO FOUR PARTS, -)|
Unit 1 (INTO TWO PARTS, -)
    [layer a.1 -- Shape (A CANAL, apantli),
     layer a.2 -- Color (TURQUOISE, xihuitl),
     layer a.3 -- Texture (WAVING EDGE, apano),
     layer a.4 -- Amount (1),
     layer a.5 -- Position (BELOW),
     layer a.6 -- Orientation (FROM UPPER RIGHT TO BOTTOM LEFT),
     layer a.7 -- Value (100%)];
Unit 2 (INTO TWO PARTS, -)
    […];
| end marker of agglutination [1.1] |
    […]
| end marker of agglutination [1] |
| end marker of synsemia [2] |
```

Tenochtitlan (Figure 6)

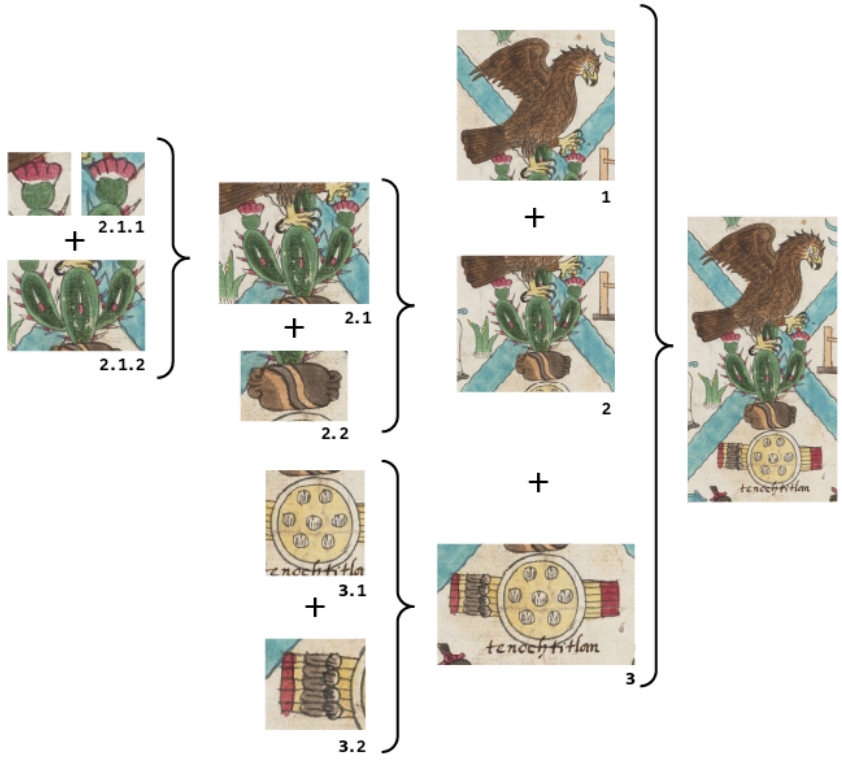

FIG. 6. Tenochtitlan, glyph decomposition, section of the folio 2r

In huey quauhtli motlalia ipan in tenochtli, motlalia Tenochtitlan in xochilê nopalli, auh in huey Tenochca chimalli ipan in mitl, in Tenochca yaoyotl

'The big eagle rests onto a stone cactus, rests in the midst of a Stone Cactus with many cactus flowers, at Tenochtitlan, and the big Tenochca shield is on the arrows, there is the war of the Tenochca.'

```
| initial marker of synsemia [3] [1-3 (list of agglutinations)]
    [Layer b.1 Orientation (UPRIGHT) -
     Layer b.2 Size (1/5; 3/8) -
     Layer b.3 Position (IN THE CENTER; FROM UP; ONE BELOW THE OTHER)]

| initial marker of agglutination [1] |
Unit 1.1 (EAGLE, quauhtli)
```

```
    […];
| end marker of agglutination [1] |

| initial marker of agglutination [2] (FLOWERS+CACTUS+STONE,
Tenochtitlan in xochilê nopalli)|
| initial marker of agglutination [2.1] (FLOWERS+CACTUS, xochilé nopalli) |
    Unit 2.1.1 (FLOWERS, xochilê;
    […];

    Unit 2.1.2 (CACTUS, nopalli);
    […]
| end marker of agglutination [2.1] |

Unit 2.2 (STONE, tetl)
    […];

| end marker of agglutination [2] |

| initial marker of agglutination [3] |
Unit 3.1 (TENOCHCA SHIELD, Tenochca chimalli)
    […];

Unit 3.2 (ARROWS, mitli)
    […];

| end marker of agglutination [3] |
| end marker of synsemia |
```

Tenochtzin Speaks Preciously (Figures 7 and 8)

Ipan in mictlampa calpopan nepantla apantli, tollantzalan, xiuhacatzalan, tlanahuatia Teochtzin in huey teopixqui auh huey tlatoani; yhuan Tenochtzin motlalia inquax in eyi pilli teyaochihuanimê: Acacitlitzin, Mexcalmetzitzin, Xocoyotzin […] quittâ in Tenochitlan yaoyotl

'In the northern subsection surrounded by water canal, in the middle of reeds, in the middle of canes, Tenochtzin rules, the high priest and high nobleman; furthermore Tenochtzin is seated in front of three noblemen warriors: the revered Reed Rabbit, the revered Little Maguey, the revered Foot Bell […] they look at the war of Tenochtitlan.'

```
| initial marker of synsemia [4] [1-4 (list of agglutinations)]
    [Layer b.1 Orientation (TOWARD SYNSEMIA [3] (they look)) -
     Layer b.2 Size (2/3; 5/16) -
     Layer b.3 Position (ON THE LEFT/NORTH SUBSECTION [2], mictlampa;
     SCATTERED; SURROUNDING)]

| initial marker of agglutination [1] (NOBLE PRIEST STONE CACTUS
SPEAKS PRECIOUSLY, Tenochtzin xiuhtlatoani)|
[…]
```

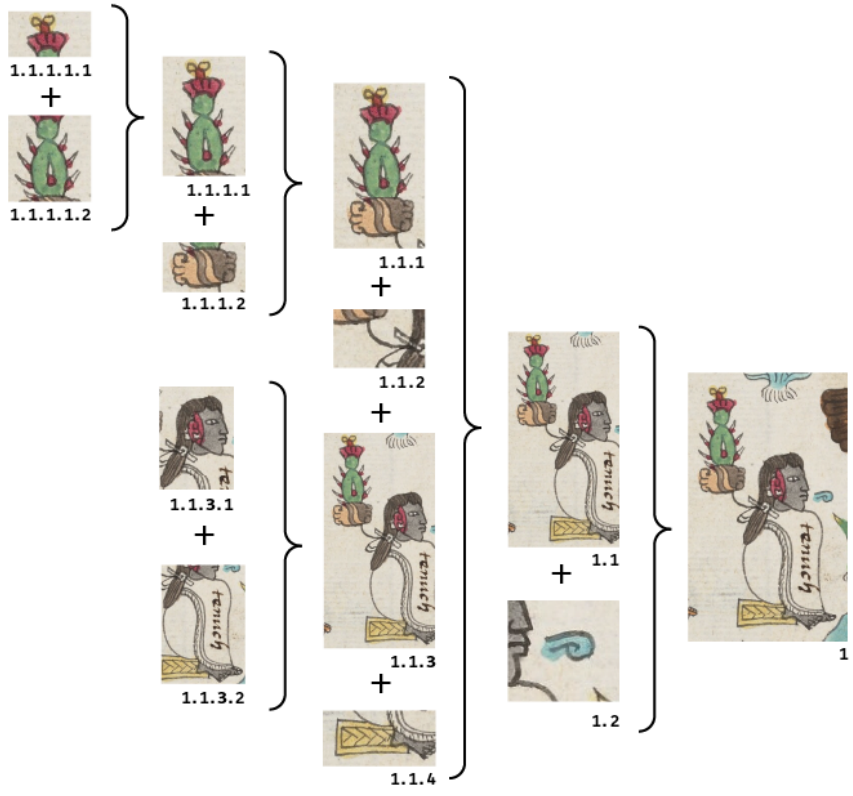

FIG. 7. Tenochtzin speaks preciously, glyph decomposition, section of the folio 2r

```
| end marker of agglutination [1] |

| initial marker of agglutination [2] (REEDS, acatl)|
Unit 2.1 (REED, acatl)
    [layer a.1 -- Shape (REED, acatl),
     layer a.2 -- Color (-),
     layer a.3 -- Texture (-),
     layer a.4 -- Amount (1),
     layer a.5 -- Position (SCATTERED),
     layer a.6 -- Orientation (-),
     layer a.7 -- Value (100%)];
| end marker of agglutination [2] |

| initial marker of agglutination [3] (CANES, tollan)|
Unit 3.1 (CANES, tollan)
[…];
```

```
| end marker of agglutination [3] |

| initial marker of agglutination [4] (NOBLE WARRIOR REED RABBIT,
Acatochtzin)|
[…];
| end marker of agglutination [4] |

| initial marker of agglutination [4] (NOBLE WARRIOR LITTLE MAGUEY,
Mexcalmetzintzin)|
[…];
| end marker of agglutination [4] |

| initial marker of agglutination [5] (NOBLE WARRIOR FOOT BELL,
Xocoyotzin)|
[…];
| end marker of agglutination [5] |

| end marker of synsemia [4] |
```

Tenochtzin (Figures 7 and 8)

Tenochtzin, xochilê nopaltzin, iitzon hueyac in iztacmolpilli, inacaz, in tlatlauhqui auh iztac, iicxi in tlilli, iizti in tlilli auh caccopinquê, in huey teopixqui, in iztactilmatini, in motlalia petlaicpalpan, in huey pilli, auh in xiuhtlatoa, in huey tlatoani

'The revered Stone Cactus, the revered Cactus with flowers, his hair is long and bent with a white knot, his ear is red and white adorned, his face is dark, his feet are dark and bare, the high priest, the one who wears a white cloak, who is seated onto a mat of reeds, is a high nobleman, and he is the one who speaks preciously, the Great Lord.'

```
| initial marker of agglutination [1] (NOBLE PRIEST+STONE+CACTUS
SPEAKS PRECIOUSLY, teopixqui Tenochtzin xiuhtlatoani?)|
…

| initial marker of agglutination [1.1] (NOBLE PRIEST+STONE+CACTUS,
teopixqui Tenochtzin)|
| initial marker of agglutination [1.1.1] (FLOWERS+CACTUS+STONE,
Tenochtitlan in xochilê nopalli)|

Unit 1.1.1.1 (STONE CACTUS, tenochtli)
[…];
| initial marker of agglutination [1.1.1.1] |
    Unit 1.1.1.1.1 (FLOWERS, xochilê)
    [layer a.1 Shape (FLOWERS, xochilê),
    …];
    Unit 1.1.1.1.2 (CACTUS, nopalli)
    [layer a.1 Shape (CACTUS, nopalli),
    …];
| end marker of agglutination [1.1.1.1] |
```

Fig. 8. Tenochtzin, section of the folio 2r

```
Unit 1.1.1.2 (STONE, tetl)
[layer a.1 Shape (STONE, tetl),
…];
| end marker of agglutination [1.1.1] |

Unit 1.1.2 (NAME-OF-PERSON CONNECTOR)
[…];

| initial marker of agglutination [1.1.3] |
    Unit 1.1.3.1 (PRIEST teopixqui (hair))
    [layer a.1 Shape (A PRIEST (hairstyle)),
    …];
    Unit 1.1.3.2 (ARISTOCRAT, -tzin)
    [layer a.1 Shape (WITH NAKED FEET, iicxi -tzin),
     layer a.2 -- Color (WHITE, iztactilmatli, -tzin),
    …];
| end marker of agglutination [1.1.3] |

Unit 1.1.4 (MAT OF WOVEN RUSH STRAW (rules) cozticpetlatl)
[layer a.2 -- Color (YELLOW),
…];
| end marker of agglutination [1.1] |

Unit 1.2 (SPEAKING PRECIOUSLY, xihutlaoa)
[layer a.2 -- Color (Turquoise, xihuitl),
…];
| end marker of agglutination [1] |
```

Ocelopantzin (Figure 9)

```
| initial marker of agglutination [1] (NOBLE+WARRIOR+OCELOT FLAG,
Ocelopantzin)|
Unit 1.1 (FLAG)
[layer a.1 -- Shape (A FLAG, pantli),
 layer a.2 -- Color (YELLOW),
 layer a.3 -- Texture (OCELOT-SKIN, ocelo-),
 layer a.4 -- Amount (1),
 layer a.5 -- Position (ON THE SIDE-LEFT),
 layer a.6 -- Orientation (-),
 layer a.7 -- Value (100%)];

Unit 1.2 (NAME-OF-PERSON CONNECTOR)
[layer a.1 -- Shape (STROKE, AZTEC CONNECTING CHARACTER),
 layer a.2 -- Color (-),
 layer a.3 -- Texture (.),
 layer a.4 -- Amount (-),
 layer a.5 -- Position (INBETWEEN),
 layer a.6 -- Orientation (-),
 layer a.7 -- Value (100%)];
```

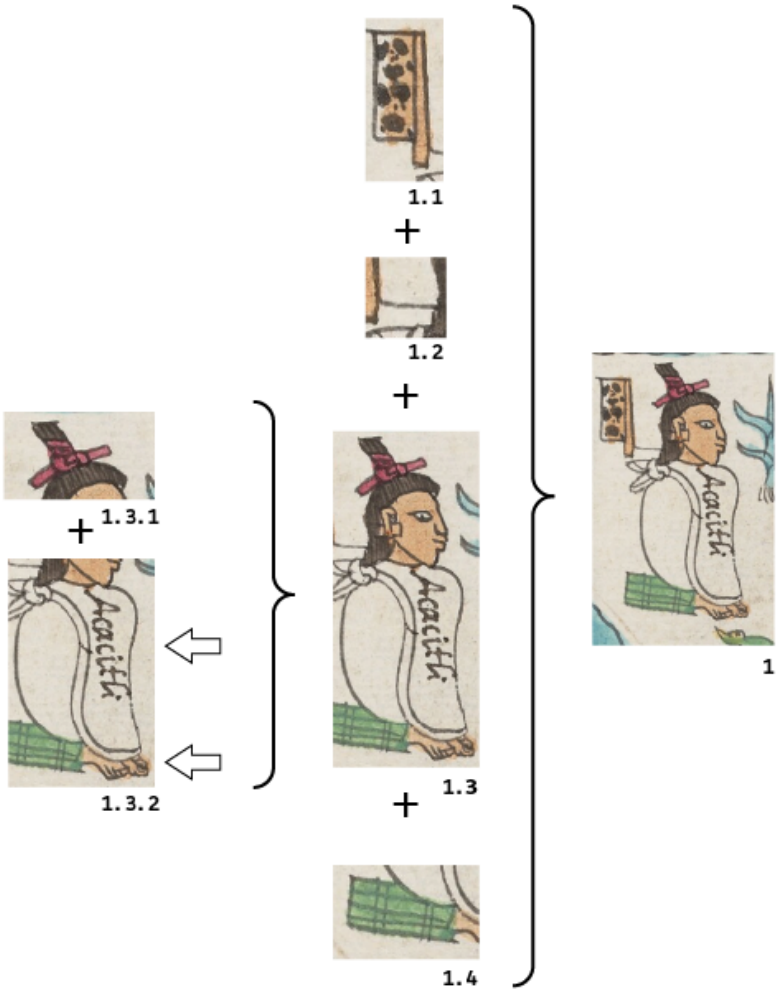

FIG. 9. Ocelopantzin, glyph decomposition, section of the folio 2r

```
| initial marker of agglutination [1.3] (NOBLE+WARRIOR) |
    Unit 1.3.1 (WARRIOR (hair))
    [layer a.1 -- Shape (A WARRIOR (hairstyle)),
     …];

    Unit 1.3.2 (NOBLE, -tzin)
    [layer a.1 -- Shape (WITH NAKED FEET, -tzin),
     layer a.2 -- Color (WHITE (dress), -tzin),
     …];
| end marker of agglutination [1.3] |

Unit 1.4 (WICKER MAT)
[layer a.1 -- Shape (A WICKER MAT),
 layer a.2 -- Color (GREEN),
 layer a.3 -- Texture (STACKED STICKS),
 layer a.4 -- Amount (1),
 layer a.5 -- Position (BELOW),
 layer a.6 -- Orientation (-),
 layer a.7 -- Value (100%)];

| end marker of agglutination [1] |
```

The Defeat of Colhuacan (Figure 10)

In tolcozticcalquaitl in teocalli mopechteca iyecapan, uetzi, in teocalli popoca, tlatla iicampa, in montemayahui, tepehualitztli Colhuacan altepetl

'The yellow roof made by reeds of the temple leans ahead, falls, the temple goes up in smoke, in flames, it is the conquest, the defeat of the village of Colhuacan.'

Notice that while the following pages of *Codex Mendoza* have extensive lists of conquest, only in this f. 2r the conquest glyph is placed on a different layer, "above" the place-name referred to. Maybe this is a visual device used to maintain the multilayered map-style of the overall page.

```
| initial marker of synsemia [8] [1-3 (list of agglutinations)]
[Layer b.1 Orientation (ONE AGAINST THE OTHER) - Layer b.2 Size
(2/5; 1/5) - Layer b.3 Position (ON THE LOW SIDE; FROM LEFT TO RIGHT;
ONE AFTER THE OTHER)]

| initial marker of agglutination [1] (NOBLE+AZTEC+WARRIORS,
Tenochca tecatl)|
Unit 1.1 (WARRIOR, tecatl)
[layer a.1 -- Shape (HUMAN FIGURE, tecatl),
 layer a.2 -- Color (RED, tecatl),
 layer a.3 -- Texture (ARMOR, tecatl),
 layer a.4 -- Amount (1),
 layer a.5 -- Position (CENTER),
 layer a.6 -- Orientation (-),
```

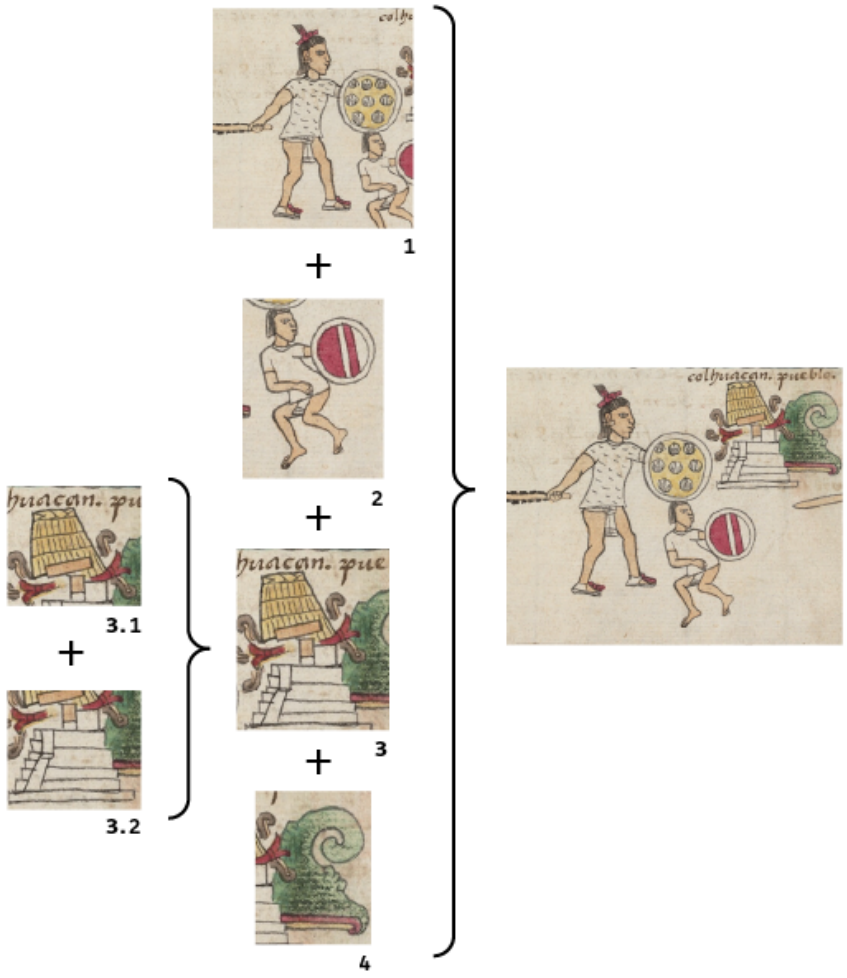

Fig. 10. The defeat of Colhuacan, glyph decomposition, section of the folio 2r

```
 layer a.7 -- Value (100%)];
| end marker of agglutination [1] |

| initial marker of agglutination [2] (NOBLE+COLHUACAN+WARRIORS,
Colhuacan tecatl)|
Unit 2.1 (WARRIOR, tecatl)
[layer a.1 -- Shape (HUMAN FIGURE, tecatl),
 layer a.2 -- Color (BLUE, tecatl),
 layer a.3 -- Texture (ARMOR, tecatl),
 layer a.4 -- Amount (1),
 layer a.5 -- Position (LEFT),
 layer a.6 -- Orientation (-),
 layer a.7 -- Value (100%)];
| end marker of agglutination [2] |

| initial marker of agglutination [3] (YELLOW ROOF+OF THE TEMPLE
+MADE BY REED+LEANS AHEAD, tolcotziccalquaitl)|
Unit 3.1 (YELLOW ROOF OF THE TEMPLE, tolcozticcalquaitl)
[layer a.1 -- Shape (TEMPLE WITH ROOF, teocalli tolcozticcalquaitl),
 layer a.2 -- Color (YELLOW, coxtic),
 layer a.3 -- Texture (MADE BY REED),
 layer a.4 -- Amount (1),
 layer a.5 -- Position (AFTER UNIT [2]),
 layer a.6 -- Orientation (LEAN AHEAD),
 layer a.7 -- Value (100%)];

| initial marker of agglutination [3.2] (THE TEMPLE GOES UP IN SMOKE,
IN FLAMES, popoca, tlatla)|
Unit 3.2.1 (SMOKE, popoca)
[layer a.1 -- Shape (SMOKE, popoca),
 layer a.2 -- Color (GRAY, popoca),
 layer a.3 -- Texture (CURLING),
 layer a.4 -- Amount (1),
 layer a.5 -- Position (UPWARD),
 layer a.6 -- Orientation (-),
 layer a.7 -- Value (100%)];
Unit 3.2.2 (FLAMES, tlatla)
[layer a.1 -- Shape (FLAME, tlatla),
 layer a.2 -- Color (RED, tlatla),
 layer a.3 -- Texture (BLAZING),
 layer a.4 -- Amount (1),
 layer a.5 -- Position (UPWARD),
 layer a.6 -- Orientation (-),
 layer a.7 -- Value (100%)];
| end marker of agglutination [3.2] |
| end marker of agglutination [3] |

Unit 4 (CURVED MOUNTAIN, Colhuacan)
[layer a.1 -- Shape (MOUNTAIN, Colhuacan),
 layer a.2 -- Color (BROWN, Colhuacan),
 layer a.3 -- Texture (ROCKY),
 layer a.4 -- Amount (1),
```

```
 layer a.5 -- Position (BELOW),
 layer a.6 -- Orientation (-),
 layer a.7 -- Value (100%)];

| end marker of synsemia [8] |
```

Conclusion

This paper suggests a speculative approach to encoding non-linear written artifacts, using folio 2r of the *Codex Mendoza* as a case study. The limitations of existing Unicode standards, which are rooted in the constraints derived from alphabetic and linear typographic traditions (Lussu, 2007), are critically assessed in light of the complex visual and semantic demands of *Aztec writing*.

The tentative system we devised would enable an encoding of writing by overcoming the main issue that a relevant part of the text is written through graphic composition, which cannot be fully represented with a mere linear sequence of characters-codes.

Our tentative model provides a multi-layered framework that integrates relational encoding, hierarchical composition, and visual variables. This approach allows for the representation of both entactic and synsemic levels, thus ensuring that spatial arrangements, relational markers, and hierarchical relationships intrinsic to Aztec graphemics are preserved. By addressing the agglutinative nature of *Nahuatl* and the non-linear, multidimensional layout of Aztec texts, the model provides a novel method for encoding the linguistic and visual complexity of these artifacts.

Folio 2r, as a critical and emblematic page, highlights the potential of this model. Its intricate spatial organization, chronological framework, and rich interplay of visual and linguistic content make it an ideal stress test for encoding non-linear texts. The method shows how graphic elements such as color, position, orientation, and texture are not merely decorative, but fundamental to conveying semantic and linguistic content.

This work underscores the importance of rethinking text encoding to accommodate non-linear systems, bridging the gap between the needs of historical artifact preservation and modern digital technologies. Beyond its application to *Aztec writing*, the principles and techniques proposed here may inform the development of encoding systems for other complex, non-alphabetic, and visual notations (including sign languages), or even a non-linear use of conventional linear script, contributing to a more inclusive and comprehensive digital representation of human linguistic diversity.

References

Berdan, Frances F. and Patricia R. Anawalt (1992). *Codex Mendoza: Four-Volume Set*. University of California Press.

Bertin, Jacques (2011). *Semiology of Graphics: Diagrams, Networks, Maps*. Redlands, CA: Esri Press.

Bleichmar, Daniela (2015). "History in pictures: translating the Codex Mendoza". In: *Art History* 38.4, pp. 682–701.

Brokaw, Galen (2022). "Homo Alphabeticus, Glottographic Exceptionalism, and the Ethnocentric Definition of Writing". In: *PMLA* 137.1, pp. 163–170.

Consortium, Unicode (2024). *Unicode® Technical Standard #51: Unicode Emoji*. URL: https://unicode.org/reports/tr51/ (visited on 01/06/2025).

Dombrowski, Quinn et al. (2024). "Encoding Multilingualism: Technical Affordances of Multilingual Publication from Manuscripts to Unicode and OpenType". In: *The Journal of Electronic Publishing* 27.1.

Fedorova, Liudmila L (2009). "The emblematic script of the Aztec codices as a particular semiotic type of writing system". In: *Written Language & Literacy* 12.2, pp. 258–275.

Fedorova, Liudmila L. and Antonio Perri (2022). "Emblematic techniques as textual strategies in non-linear and linear scripts". In: *Proceedings of Grapholinguistics in the 21st Century, 2022*. Ed. by Yannis Haralambous. Vol. 9. Grapholinguistics and Its Applications. Brest: Fluxus Editions, pp. 75–99. DOI: https://doi.org/10.36824/2022-graf-perr.

Galarza, Joaquín et al. (1983a). *Codex Mendoza, 1re page, f. 2r. Lecture nahuatl*. Paris.

——— (1983b). *Codex Mendoza, 1re page, f. 2r. Méthode d'analyse*. Paris.

Harris, Roy (1986). *The origin of writing*. Chicago and La Salle, IL: Open Court.

——— (2000). *Rethinking Writing*. London: Athlone Press.

Klinkenberg, Jean-Marie and Stéphane Polis (2018). "On scripturology". In: *Signata. Annales des sémiotiques/Annals of Semiotics* 9, pp. 57–102.

Launey, Michel (1994). *Une grammaire omniprédicative: essai sur la morphosyntaxe du nahuatl classique*. Collection sciences du langage. CNRS Éditions.

Lussu, Giovanni (2007). "The Form of Language". In: *Typography Papers* 7, pp. 139–147.

Perondi, Luciano (2012). *Sinsemie. Scrittura nello spazio [Synsemiae. writing in the space]*. Viterbo: Nuovi Equilibri.

Perondi, Luciano, Antonio Perri, et al. (2018). "Configurare lo spazio nella scrittura azteca: il Codice Mendoza come modello di trasposizione e oltre [Framing space in Aztec writing: the Codex Mendoza as a model of transposition and beyond]". In: *XY* 3.5, pp. 40–53.

Perondi, Luciano and Leonardo Romei (2010). "Letture da recuperare [Readings to retrieve]". In: *Il sole 24 ore* 245, pp. 12–13.

Perri, Antonio (2006). “Dire o mostrare? L’immagine-testo ei paradossi della pittografia-come-notazione [To Say or to Show? The Image-Text and the Paradoxes of Pictography-as-Notation]”. In: *Centro internazionale di Semiotica e linguistica-Università di Urbino-Documenti e prepubblicazioni* 350, pp. 51–75.

Perri, Antonio et al. (2024). “Alternative graphemics. Aztec writing system as a case study towards an integrated, digitalised model of non-typographic graphemics”. In: *Ocula* 24.30, pp. 43–64.

Vaillant, Pascal (1999). *Sémiotique des langages d’icônes*. Bibliothèque de grammaire et de linguistique. Paris: Éditions Honoré Champion.

Zamora Corona, Alonso Rodrigo (2022). “Towards a Complex Theory of Writing: The Case of Aztec and Mixtec Codices”. In: *Signata* 13. DOI: 10.4000/signata.3866. (Visited on 01/13/2025).