The Code of Maya Kings and Queens: Encoding and Markup of Maya Hieroglyphic Writing

Martin de la Iglesia, Franziska Diehr, Uwe Sikora, Sven Gronemeyer, Maximilian Behnert-Brodhun, Christian Prager and Nikolai Grube

Electronic version
URL: https://journals.openedition.org/jtei/3336
DOI: 10.4000/jtei.3336
ISSN: 2162-5603

Publisher
TEI Consortium

Electronic reference
Martin de la Iglesia, Franziska Diehr, Uwe Sikora, Sven Gronemeyer, Maximilian Behnert-Brodhun, Christian Prager and Nikolai Grube, “The Code of Maya Kings and Queens: Encoding and Markup of Maya Hieroglyphic Writing”, Journal of the Text Encoding Initiative [Online], Issue 14 | April 2021- March 2023, Online since 01 April 2021, connection on 13 March 2023. URL: http://journals.openedition.org/jtei/3336 ; DOI: https://doi.org/10.4000/jtei.3336

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.
ABSTRACT

Maya hieroglyphic script (300 BCE–1500 CE) is a semi-deciphered logographic and syllabic autochthonous writing system from the Americas and is one of the most significant writing traditions of the ancient world. Because of its incomplete state of decipherment, complexity and variation in graphematics, and partially lost lexicon, transliterations cannot be used within the encoding. The project Text Database and Dictionary of Classic Mayan approaches this challenge with an encoding strategy relying on stand-off markup, which is enriched with additional information sources. Using different formats (RDF, XML) and standards (CIDOC CRM, TEI P5), the inscriptions are encoded in a multilevel corpus: (1) a tei_all-compliant schema defining values and rules for the encoding of the text’s topological and structural features, (2) a “Sign Catalogue”
for the classification of Maya hieroglyphs, and (3) the tool ALMAH (Annotator for the Linguistic analysis of MAYa Hieroglyphs) for linguistic analyses. In this paper, we focus on the TEI schema and highlight our strategy for encoding hieroglyphs without using linguistic transliterations and transcriptions.

INDEX

Keywords: Maya writing, hieroglyphic writing, digital epigraphy, Maya culture, Mesoamerica

1. Text Database and Dictionary of Classic Mayan

In 2014, the project Text Database and Dictionary of Classic Mayan was established at the University of Bonn by the North Rhine-Westphalian Academy of Sciences, Humanities and Arts and the Union of the German Academies of Sciences and Humanities to research the written language of the pre-Columbian Maya. The project intends to use digital methods and technologies to compile the epigraphic contents and object histories of all known hieroglyphic texts. The aim is to create a dictionary of Classic Mayan, a language whose script has not yet been completely deciphered. For this purpose, we will compile a machine-readable corpus of all known Maya texts, which are written on about ten thousand text carriers and four codices made of bark paper. To create a holistic environment that provides a solid information base for the dictionary, we have also developed additional resources and tools: a documentation of the text carriers, a classification of the Maya signs, and a tool for linguistic annotation and analysis, as well as supplementary archival materials and a bibliography. The dictionary will be a highly concentrated extract of all these information sources. However, the machine-readable corpus is the central part of the virtual environment: it is enriched by the other information resources using stand-off markup and data-linking mechanisms.

In the first two sections of the paper, we present the complex writing system of Classic Maya culture. Then we present and discuss our TEI encoding strategy for the corpus. Toward the end, we look at the data-linking mechanisms connecting other information sources to the corpus. We conclude by discussing our approach to an ontological corpus.
2. Introducing Classic Maya Writing

The subject of our work is the written language of the Classic Maya culture on the Yucatán Peninsula, whose cultural area extended over the territory of the present-day states of Guatemala, Belize, and parts of Honduras and Mexico. Maya writing has been used for more than fifteen hundred years and can be found, for example, on freestanding monuments (stelae, altars), architectural elements (lintels, doorjambs), portable objects (ceramics, shells), and natural features such as caves or rocks (Grube 1994). The writing system comprises about two thousand figurative graphs exhibiting signs for words and syllables. It has survived on more than ten thousand inscription carriers from more than five hundred archaeological sites dating between 300 BCE and 1500 CE. The graphs of this glottographic writing system include representations of figurative objects from the natural environment, material culture, human and animal body parts, heads of humans and animals or portraits of supernatural beings, and abstract ideas. The high language of the hieroglyphs is called Classic Mayan and has been preserved in greater parts in the colonial and modern Ch’olan and Yucatecan languages (Wichmann 2006). A large number of texts display calendar dates for the precise dating of events and thus also provide unique data on the history of writing and language. Classic Mayan can therefore be reconstructed very accurately and compared with the findings of historical linguistics. All inscriptions originate from the vicinity of the palaces of the divine kings who ruled over independent city-states. The inscriptions contain biographical information and provide written evidence for inter- and intradynastic connections of the ruling families. Public monuments depict and describe actions such as war or royal visits, and they exhibit information about ceremonies and religious rituals in the context of accessions to the throne, ancestor worship, calendar anniversaries, inaugurations, processions, and other occasions of royal daily lives (Martin and Grube 2008).

Typologically, Maya writing is a morpho-syllabic writing system with two basic sign functions: syllabograms and morphograms. Of the latter, logograms that denote concrete words are the most common. Syllabic signs represent vowels and open syllables of the form CV (consonant and vowel). These permitted the syllabic spelling of lexical and grammatical morphemes. In addition, they were also used as pre- or postfixed phonetic complements for logograms. Thus, it was possible to write words entirely with syllabic signs or using only logograms. Usually, logograms and syllabic signs were combined, providing logo-syllabic pronunciation of words (figure 1). A high level
of calligraphic complexity was further achieved through allographic spellings and allomorphic representation of signs. This phenomenon allowed scribes to compose texts that were aesthetically sophisticated without necessarily repeating graphs. There were at least two or more different allographs for frequently used syllables (figure 2), a situation which explains the extremely high number of syllabic signs (about three hundred) within the total inventory of approximately two thousand graphemes in the Maya script (Grube 1994).
Figure 1. Different sign classes and their combination in writing words. The word bahlam in italics represents the phonemic transcription stressing the length of the first vowel. Bold words represent the graphematic transliteration. Drawings by Sven Gronemeyer.

Morphographs  Syllabographs

bahlam "jaguar"

morphemic  syllabic

BALAM  ba-la-ma

morpho-syllabic

BALAM-ma
The signs were combined with each other to build roughly quadratic blocks. One such hieroglyphic block usually corresponds to the concept of a word among the pre-Hispanic Maya. In most cases, these blocks were paired in double columns that were read from left to right and from top to bottom. Sentences were formed by the combination of hieroglyphic blocks indicating various parts of speech. Multiple sentences were combined to produce complex texts, whose syntax and textual structure are comparable to those of modern Mayan languages. The individual elements in the hieroglyphic block are traditionally subdivided into main and small graphs, with the main graphs being spatially larger and approximately square in shape, and the small elements attached to the periphery of the main characters oriented along the axis of the main graph. Within the block, the individual graphs were not only arranged side by side or on top of each other. They could also be conflated into one single outline, similar to a ligature. In addition, two or more graphs could also be partially or completely overlapped or inserted into each other (figure 3).
The altered shape of the individual graph within the block itself did not influence the spelling and meaning of the hieroglyph. The morphology of the graphs and their arrangement in a block is one of the main challenges for epigraphy in those cases where either all or individual signs are not deciphered or only tentatively deciphered and thus elude linguistic verification. The documentation of the original spelling—the actual graph variation and the intrablock arrangement using TEI XML—is therefore essential for the epigraphic work with hieroglyphic writing systems, since a simple and linear transliteration of a text no longer shows the original spelling or the placement of the glyphs within the block (Prager and Gronemeyer 2018, 145).

While these above-mentioned graphemic and graphotactic strategies only concern the graphic realization of words in written Mayan, the principle of underrepresentation of specific word endings also has an impact on the pronunciation of hieroglyphs in modern epigraphy. The omission of signs further enabled the scribes to vary individual words and texts (Zender 1999, 130–142). This scribal practice also has a high impact on the lexicography of Classic Mayan. By means of the above-mentioned graphemic and graphotactic strategies, Maya scribes were able to create
a broad variety of texts with the avoidance of graphic repetitions. They were skilled calligraphers, sought a maximum of visual splendor and designed texts and pictorial works as individual pieces of art. On the other hand textual and pictorial contents are rather formulaic.

3. Encoding Allography and Graphotactics

Regular transliterations in TEI encoding would not be able to represent the graphemic variability of Maya writing even if all signs had been fully deciphered. Instead, we encode a proper description of each hieroglyphic block: the actual graphic variant and the spatial relation of each graph to another. A digital Sign Catalogue modeled in RDF provides the backbone of the TEI encoding. Each sign is indexed using a unique number, each type of graphic variation using a two-letter suffix (Prager and Gronemeyer 2018). The URI of each graph is referenced in the TEI. Furthermore, such digital approaches to epigraphy can help overcome some shortcomings of predigital epigraphy, such as dealing with ambiguity.

The encoding in TEI can be divided into a formal, descriptive part (the text layout and design) and a textual part (the content and the glyphic spellings). In addition to the TEI encoding itself, we needed to create an environment that supports the entire workflow through which different materials such as texts, photographs, drawings, artifact information, historical facts, and established research results are made available. In this virtual environment, a TEI-encoded corpus is enriched by other information resources which are themselves heavily annotated with metadata. The following sections present the key features of the encoding and enrichment strategy employed.
4. Textual Elements and Structure

A Maya text consists of three units: fields, blocks, and glyphs (figure 4). Text fields are encoded using the element <tei:div> (text division) with a defined @type ("textfield"). Strictly speaking, the @type attributes would not be necessary, but they facilitate the human interpretation of generic TEI elements as representations of specific linguistic units. The same principle is used for encoding glyph blocks, for which the element <tei:ab> (anonymous block) is used. For encoding single glyphs, the element <tei:g> (character or glyph) is used with the attributes @ref and @n. The attribute @n is used to indicate the catalogue number of the sign with its graph variant according to the classification used by the project’s Sign Catalogue. Instead of pointing to a <tei:glyph>, @ref points to the URI in the Sign Catalogue where the corresponding glyph is described. This mechanism is explained in more detail below.
The textual elements are being brought into a logical structure by nesting the elements (figure 5): `<tei:div>` for the text field is the parent element to the `<tei:ab>` elements of all the glyph blocks contained in that text field, in this example from A1 to D4, and the glyph block contains all glyphs belonging to that block (D1G1 to D1G3). Every element is provided with an `@xml:id` attribute for referencing purposes. In order to identify glyphs that form a group within a block, a fourth element, `<tei:seg>` (arbitrary segment), is used. This is needed to represent the positioning of the glyphs (see section 5).

Note how the usage of TEI elements differs slightly from approaches with similar aims, such as Comic Book Markup Language (CBML). CBML is designed for the encoding of a “class of documents that tightly integrate pictorial images and text. Comic books are just one such type of complex graphic document; other examples include illuminated manuscripts; seventeenth-century alchemical manuscripts, with hand-drawn figures and graphic symbols; artists’ books; artists’ sketchbooks; illustrated poems like those of William Blake; letterpress productions like those of the Kelmscott Press; illustrated children’s books; newspaper and magazine advertisements; and even Web pages and other born-digital media” (Walsh 2012).
can easily imagine that Maya texts fit within this scope as well. The main encoding difference is the introduction of new specific elements under the CBML namespace, most notably `<cbml:panel>`.

According to Walsh (2012), “<cbml:panel> is a modification of TEI’s `<div>` element, which represents a generic subdivision of the text in the TEI model.” The `<cbml:panel>` element and other custom elements in the CBML namespace are used alongside generic TEI elements such as `<tei:div>` and `<tei:sound>`; together they constitute the markup of comic book pages or entire comic book issues. In contrast, all of the XML elements used in the Maya encoding scheme presented here are taken from the TEI namespace. No custom elements or attributes are introduced. Instead, information specific to Maya texts is encoded in the values of attributes such as `@type` (see above) or `@rend` (see below). While these attribute values are controlled by project-specific taxonomies, this approach has the disadvantage that this information is not easily made machine-readable and interchangeable. The advantage, on the other hand, is that the markup as a whole validates to `tei_all` and may thus be processed and interpreted by generic tools.

5. Glyph Positioning in Glyph Block

Figure 6. Partial example showing the encoding of the relative positions of glyphs in a glyph block using `@rend` with predefined values. Furthermore, `@corresp` indicates the position of each glyph or glyph group in relation to its corresponding glyph or glyph group. Monument of unknown provenance, Bonampak area. Drawing by Linda Schele © David Schele (SD-6000). Photo courtesy Ancient Americas at LACMA (http://ancientamericas.org/collection/aa010566).

```xml
<tei:ab xml:id="D1">
  <tei:g xml:id="D1G1" tei:rend="left_beside" tei:corresp="#D1S2"/>
  <tei:seg xml:id="D1S1" tei:rend="right_beside" tei:corresp="#D1G1">
    <tei:g xml:id="D1G2" tei:rend="above" tei:corresp="#D1G3"/>
    <tei:g xml:id="D1G3" tei:rend="beneath" tei:corresp="#D1G2"/>
  </tei:seg>
</tei:ab>
```
As explained above, the arrangement of glyphs within a block does not indicate a reading order. When encoding the glyphs in the TEI document, only their arrangement is made explicit. The correct reading order is established later by the linguistic analysis. For this reason, it would be misleading to describe the relative positions of glyphs using the @next and @prev attributes. This is especially true when keeping in mind that Maya writing is not yet fully deciphered and there may be multiple possible reading orders. The advantage of keeping the encoding of arrangement and reading separate is that in the process of analyzing a text, one can still be improved without having to change the other. Therefore, instead of @next and @prev the @rend attribute is used with predefined values. This attribute, combined with the @corresp attribute, can indicate a glyph’s position in relation to neighboring glyphs, logically creating a statement like “The current glyph is arranged in a specific manner (@rend with respect to another glyph (@corresp).” The order in which the XML elements are encoded in the document, such as which <tei:g> element appears first inside <tei:seg>, is purely arbitrary and is not meant to imply any reading or writing order.

As the example in figure 6 shows, the <tei:seg> element is used to describe the positions more accurately. The first glyph, D1G1, is positioned on the left-hand side of (“left_beside”) the corresponding segment, D1S1, which contains two glyphs that are stacked over each other, their positions indicated by “above” and “beneath.”
Figure 7. Predefined values for @rend describing glyph positions in Maya writing. In transliteration practice, operators like ".", ":", and "+" are used to describe positions. In contrast to that, using TEI encoding is much more precise because each glyph is described on its own. Drawings by Christian Prager.

| Transliteration operator | value of @rend               | value of @rend               |
|--------------------------|-----------------------------|-----------------------------|
| .                        | left_beside                 | right_beside                |
| :                        | above                       | beneath                     |
| °                        | infixed_in                  | encloses                    |
| +                        | conflated                   | conflated                   |
| g                        | attached_at_top_left        | is_adjoined_to              |

There are other positioning options, of which we shall mention only a few here: for instance, glyphs can be “conflated,” that is, merged so that they share a mutual outline. As in the example on the right of the upper right side in figure 7, the hand-shaped glyph is used as an outline for the head-
variant glyph. Glyphs can also be “infixed” by being completely enclosed by another glyph. In the example on the left of the upper right side in figure 7, the hand-shaped glyph is enclosed by the head-variant glyph.

In a way, these scribal phenomena are reminiscent of medieval European handwriting which, when encoded in TEI, requires special attention to be paid to special characters, abbreviation marks, punctuation, diacritical marks, and scribal sigla. Some of these medieval scribal phenomena may also be encoded using custom @rend values (Fredell, Borchers, and Ilgen 2013). The main difference between medieval European and Maya texts, however, is that the former are still based on a standard Latin alphabet and use XML tags only to supplement a linear succession of Unicode characters, whereas Maya glyphs cannot, to date, be sufficiently represented by Unicode characters.1

To represent Maya glyphs in a block-like structure, the Ideographic Description Sequences (IDS) as presented by Unicode seem promising. The IDS are used to describe characters that are not present in the Unicode standard. Twelve special characters (ranging from U+2FF0 to U+2FFB) are used as operators, prefixing other characters to indicate their arrangement (Unicode Consortium 2010). However, many Maya glyph blocks, like the one in figure 8, cannot be fully encoded using the IDS because only some of the glyph positions in this block are covered by the IDS characters. Our approach can deal with a variety of different arrangements and therefore overcomes the insufficiency of IDS for describing Maya glyph block arrangement.
Figure 8. Example of encoding conflation and interference in Maya hieroglyphic signs in the context of the term *ch’ahoom anaab* (“sacrificer and artist”) found on Sculptured Stone 5 from Bonampak. Drawings by Christian Prager.
6. Reading Direction and Text Arrangement

Figure 9. To encode the reading directions of text fields, a @style attribute is used with project-specific values in a <tei:div> element in <tei:body>. Drawing by Alexandre Safronov. Courtesy Wayeb Drawing Archive.

As mentioned above, the correct reading order of a glyph block is established by linguistic analysis, but we can also identify a reading direction on the level of a text field. To describe the writing mode, “vendor-specific” CSS definitions are used within the @style attribute in the <tei:div> element. In the example pictured, text field 1 is written left-to-right, text field 2 right-to-left, and text field 6 top-to-bottom.

At first glance, these @style values seem similar to some of the standard values of the CSS writing-mode property, namely horizontal-tb, vertical-rl, and maybe also vertical-lr (W3C 2019). However, on closer inspection, it turns out that their semantics are slightly different. Consider, for instance, text field 6 from the example above. Whereas in reading glyphs in a horizontal band, the order would be from left to right (texts spelled from right to left are rare),
text field 6 consists of glyphs stacked one upon the other in a single vertical line, intended to be read from top to bottom. Standard CSS forces us to choose between vertical-rl and vertical-lr, meaning that any following vertical lines would be placed on the left or right, respectively, of the line in question. This makes no sense in the context of Mayan inscriptions, as there are never two or more consecutive vertical lines. In Maya inscriptions hieroglyphs are usually arranged in parallel columns, which are to be read two columns at a time, beginning with the uppermost glyph in the left-hand column, and then from left to right and top to bottom. Then the next two columns are read in the same order, and so on. Thus, to define any reading direction for sequences of vertical lines would be misleading here. Furthermore, one would have to additionally specify the orientation of the glyphs: without the "text-orientation" property, CSS assumes that “[t]ypographic character units from vertical scripts are typeset with their intrinsic orientation” (W3C 2019). “Vertical scripts” include Chinese, Japanese, and Korean. For Maya hieroglyphic writing, however, there is no universally accepted specification that defines whether or not it counts among those vertical scripts, which makes it necessary to explicitly indicate the character orientation as text-orientation:upright. Otherwise one might think that the glyphs would be rotated 90° clockwise (as are characters from so-called “horizontal-only” scripts, such as Latin, when typeset as “vertical-rl”). Therefore, in place of cumbersome and ultimately inappropriate standard CSS expressions such as writing-mode:vertical-lr; text-orientation:upright, the vendor-specific property -idiom-writing-mode:tb is a more concise way of expressing the same rendering while at the same time being a more meaningful classification of a Mayan writing mode.
Figure 10. Capturing the arrangement of the text by relating `<tei:surface>` and `<tei:zone>` in `<tei:sourceDoc>` with the corresponding `<tei:div>` in `<tei:body>`. Copan, Stela J. Drawing by Linda Schele © David Schele (SD-1014). Photo courtesy Ancient Americas at LACMA (http://ancientamericas.org/collection/aa010014).

When arranging text on a carrier, Maya scribes were very creative. Figure 10 shows Stela J from the archaeological site of Copán, Honduras. This stela has a very elaborate design: the sections of the text are braided like ribbons. The text arrangement is described within the `<tei:sourceDoc>` element using the elements `<tei:surface>` and `<tei:zone>`. With the `@rotate` attribute, we can indicate the relative orientation of `<tei:zone>`. This one is rotated 45° clockwise from its upright orientation. To relate the `<tei:zone>` elements to the logical text structure described in the `<tei:body>`, we use `@xml:id` attributes.
Figure 11. Using stand-off markup to connect image annotations stored in a separate XML file containing SVG objects to the TEI document. Copan, Stela J. Drawing by Linda Schele © David Schele (SD-1014). Photo courtesy Ancient Americas at LACMA (http://ancientamericas.org/collection/aa010014).

How can the positions of these text sections on the stela be indicated? Despite some recent improvements to the TEI Guidelines in the transcr (Representation of Primary Sources) module, the XML language Scalable Vector Graphics (SVG) is still an attractive alternative to TEI because of its flexibility and expressiveness in the description of geometric shapes. The measurements (i.e., the values of the SVG @points attributes) give an idea of the relative size and position of each text section. Actually, however, the numbers refer to pixel coordinates on a specific digital image of the stela. By linking the SVG data to the image file, the text section positions can be indicated as shapes in an overlay displayed on top of the image (figure 11).

In practice, we use the virtual research environment TextGrid Lab (https://textgrid.de/ www.textgrid.de) to create these data. TextGrid Lab comes with a tool called Text Image Link Editor that enables the user to mark areas on a picture and stores that positioning information in a separate XML file—the TEI/SVG hybrid document pictured above. This allows us not only to
relate a certain text passage encoded in the `<tei:body>` to the source text represented by the `<tei:sourceDoc>`, but also to enrich it with actual positions on the image. This means that it is possible to indicate the appearance of every single glyph block or glyph.

By using stand-off markup, we create a richly and deeply annotated text; by using the Text Image Link Editor, we refer to the appearance of the source text. By referencing the glyph in the Sign Catalogue, we also refer to the graph variants that are used.

7. TEI... and Beyond: Sign Catalogue

Figure 12. Hieroglyphs in the TEI-encoded corpus referring to the Sign Catalogue, where the glyph is assigned to a transliteration value that is used for linguistic analysis.

Within the TEI document we encode the glyphs with the referring URI of the `<idiom:graph>` recorded in our Sign Catalogue using a `@ref` attribute. The term graph denotes an abstract, typed form of an individually realized sign. The graph of /ja/ in the variation “bipartite right” (br) recorded in the catalog represents a type which prototypically represents all individual writing variants and thus all actual occurrences for /ja/ (figure 12).
The Sign Catalogue is modeled as an ontology and stores information about the classification of Maya signs and graphs in RDF (Diehr et al. 2018). The class `<idiom:Graph>` describes characteristics of the Maya glyphs belonging to the realm of graphematics, like the type of variation a “graph” can have. Every `<idiom:Graph>` is connected to an `<idiom:Sign>`. One `<idiom:Sign>` can have more than one `<idiom:Graph>`, but an `<idiom:Graph>` is only ever connected to one `<idiom:Sign>`. As Maya signs can have both a “logographic reading” and a “syllabic reading,” we have to distinguish those first, before finally assigning a transliteration value to the `<idiom:Sign>`. As the script of Classic Mayan is not yet fully deciphered, there are multiple hypotheses for the readings of various signs. Therefore a sign can be assigned multiple transliteration values. The Sign Catalogue also provides a mechanism for dealing with contradictory reading proposals (Diehr, Gronemeyer, Wagner, et al. 2019; Diehr, Gronemeyer, Prager, et al. 2019).

The TEI corpus and the Sign Catalogue provide information about the text as a linguistic and epigraphic subject. To bring that information together, the TEI document and the Sign Catalogue need to be processed further. To generate a transliterated text that can be analyzed, we add another component to the virtual research environment of the project: ALMAH, a tool for linguistic analysis of Classic Mayan, which has also been developed in course of the project (Grube et al. 2018, 5–7). ALMAH will provide mechanisms for multivariant text analysis, based on the TEI corpus and the linguistic information stored in the Sign Catalogue. The text structure will be extracted from the TEI document, and the (multiple) transliteration values are to be assigned from the Sign Catalogue. The assembled text will inherit linguistic, epigraphic, and structural text information. By using referencing mechanisms and stand-off markup strategies, the text will always be traceable to the original writing.
8. TEI... and Beyond: Artifact Documentation

We provide a rich documentation of the text carrier. In the `<tei:teiHeader>`, we refer to the artifact that is documented by using the Artefact Ontology of the project (Figure 13). The ontology describes not only text carrier characteristics such as material and technique, but also the finding context of the object and information regarding its scholarly documentation (e.g., measurements, designations). This Artefact Ontology consists of components from pre-existing vocabularies such as CIDOC CRM and Dublin Core Terms, as well as custom components in the idiom namespace. The information described by the Artefact Ontology is stored as RDF data. See section 10, “The Linked Corpus” below for details about the different components of our virtual environment.

It should be noted that this approach deliberately differs from ways of encoding text carrier information directly in TEI, as, for instance, discussed by Nelson (2017). The encoding of artifact-related data in TEI has been further facilitated by the recent introduction of `<tei:object>` and related elements in the TEI P5 Guidelines (see TEI Consortium 2020, 13.3.5: “Objects,” https://tei-c.org/Vault/P5/4.1.0/doc/tei-p5-doc/en/html/ND.html#NDOBJ). However, the TEI-based encoding of information about objects is best suited for contexts in which the object information itself exists as text that is part of the document(s) to be encoded, as in the example of
the historical inventories from Nelson's aforementioned article. In the case of the Text Database and Dictionary of Classic Mayan, in contrast, the texts (which are encoded in TEI) and the text carrier information (RDF) are derived from separate sources and thus do not benefit from homogeneous encoding. Additional details are given by Grube et al. (2016) and Prager et al. (2018).

9. Encoding Helper: “Classic Mayan TEI Generator”

Encoding the texts according to our schema may prove time-consuming: assessing identifiers for elements, searching the URI of each graph in the Sign Catalogue as well as the URI of the text carrier(s), correctly encoding the glyph positions, and so on. In order to establish a basic structure of the TEI document including <tei:header>, <tei:sourceDoc>, and <tei:body> (containing the glyph blocks and glyphs), we developed a parser called Classic Mayan TEI Generator that transforms an alphanumeric transliteration (i.e., sign number plus variant suffix; see figure 8) alongside some metadata (e.g., editor of the document, information on the project) into a TEI document. The description of graphotactics and the relation of graphs to each other is realized by project-specific operators (see figure 7).

The Classic Mayan TEI Generator is specifically not designed as a fully functioning GUI for editing purposes. Once the parser has created a TEI document, it cannot be used to edit the document any further. In particular, encodings like editorial notes on lost and/or reconstructed glyphs are still made within the TEI code itself. While these cannot be generated from a transliteration, we believe that through working within the TEI XML structure, the editors get a better understanding of what and how they encode. Combining automatic and manual procedures ensures a high quality of encoding.

The parser functions as a helper for providing the editors with a basic document structure, already containing all identifiers, and a text structure. Even with its reduced functionality, the parser is a welcome help for the editors since there are about ten thousand texts to encode.
10. The Linked Corpus

Figure 14. The virtual research environment of the project uses multiple information sources. With the TEI document as the central part, a kind of linked corpus is assembled.

In this paper we have laid out our strategy for encoding the hieroglyphic texts of Classic Mayan. The complexity of the writing system presents researchers with manifold obstacles and challenges. By using stand-off markup and other mechanisms to link data, a corpus is established that joins different information resources together in a virtual environment (figure 14). The TEI document forms the central part of this linked structure: the TEI encoding is used to describe the formal structure of the text, its appearance, and its layout. The encoding is enriched by multiple other resources to support specific functions and workflows for annotation, documentation, and analysis: a Sign Catalogue for the classification of signs and graphs as well as their variants; an ontology for the documentation of the text carriers; the tool ALMAH for linguistic annotation and analysis; a project bibliography (using Zotero); and documentation and organization of archival material (which is managed by the DARIAH-DE service ConedaKOR). Bringing all those information sources together provides a holistic research environment for analyzing and deciphering the script of Classic Mayan.

The data thus generated will be successively made available on the future project portal https://www.classicmayan.org/. Furthermore, the corpus data will also be made freely accessible in the TextGrid repository. All schemata created in the project (ODD, OWL) can be viewed in the public area of our Git repository and can be used under a CC BY 4.0 license: https://projects.gwdg.de/projects/documentations/repository.
Ontology-based, linked open data in a virtual research environment will for the first time provide a single point of reference in Maya studies. A corpus-based dictionary not only facilitates the research on pending questions of historical linguistics and enables a better adjustment of theories with epigraphic data. As the meaning of a word is affected by its context, the database can also support efforts to identify metaphors or stylistic devices, or to reconstruct a vocabulary that was culturally lost. The encoding as applied by our project can also be modified to describe the graphotactics of other nonlinear scripts or for other (partially) undeciphered scripts, such as the rongorongo script of Rapa Nui (Easter Island). It can also be used to represent (undeciphered) scripts for which no Unicode block has yet been defined or approved.

BIBLIOGRAPHY

Bański, Piotr. 2010. “Why TEI Stand-off Annotation Doesn’t Quite Work: And Why You Might Want to Use It Nevertheless.” In Proceedings of Balisage: The Markup Conference 2010. Balisage Series on Markup Technologies, vol. 5. doi:10.4242/BalisageVol5.Banski01.

Diehr, Franziska, Sven Gronemeyer, Christian Prager, Elisabeth Wagner, Katja Diederichs, Nikolai Grube, and Maximilian Brodhun. 2018. “Organising the Unknown: A Concept for the Sign Classification of Not Yet (Fully) Deciphered Writing Systems Exemplified by a Digital Sign Catalogue for Maya Hieroglyphs.” In Digital Humanities 2018 Puentes-Bridges: Book of Abstracts, 181–84. [Mexico City:] Red de Humanidades Digitales. https://dh2018.adho.org/wp-content/uploads/2018/06/dh2018_abstracts.pdf.

Diehr, Franziska, Sven Gronemeyer, Elisabeth Wagner, Christian Prager, Katja Diederichs, Uwe Sikora, Maximilian Brodhun, and Nikolai Grube. 2019. “Modelling Vagueness – A Criteria-Based System for the Qualitative Assessment of Reading Proposals for the Deciphering of Classic Mayan Hieroglyphs.” In Proceedings of the Workshop on Computational Methods in the Humanities 2018, edited by Michael Piotrowski, 2314:33–44. Workshop Proceedings. Lausanne: CEUR. https://nbn-resolving.org/urn:nbn:de:0074-2314-0.

Diehr, Franziska, Sven Gronemeyer, Christian Prager, Elisabeth Wagner, Katja Diederichs, Maximilian Brodhun, Uwe Sikora, and Nikolai Grube. 2019. “Modellierung von Entzifferungshypothesen in einem digitalen Zeichenkatalog für die Maya-Schrift.” In Die Modellierung des Zweifels – Schlüsselideen und
Fredell, Joel, Charles Borchers IV, and Terri Ilgen. 2013. “TEI P5 and Special Characters Outside Unicode.” Journal of the Text Encoding Initiative 4. http://journals.openedition.org/jtei/727; doi:10.4000/jtei.727.

Grube, Nikolai. 1994. “Mittelamerikanische Schriften.” In Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung = Writing and Its Use: An Interdisciplinary Handbook of International Research, edited by Hartmut Günther and Otto Ludwig, 1:405–15. Handbücher zur Sprach- und Kommunikationswissenschaft 10. Berlin: Walter de Gruyter.

Grube, Nikolai, Christian M. Prager, Katja Diederichs, Sven Gronemeyer, Elisabeth Wagner, Maximilian Brodhun, and Franziska Diehr. 2016. “Annual Report for 2015.” Textdatenbank und Wörterbuch des Klassischen Maya, Project Report no. 3, edited by Arbeitsstelle der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste an der Rheinischen Friedrich-Wilhelms-Universität Bonn. doi:10.20376/IDIOM-23665556.16.pr003.en.

Grube, Nikolai, Christian Prager, Katja Diederichs, Sven Gronemeyer, Antje Grothe, Céline Tamignaux, Elisabeth Wagner, Maximilian Brodhun, and Franziska Diehr. 2018. “Annual Report for 2017.” Textdatenbank und Wörterbuch des Klassischen Maya, Project Report no. 5, edited by Arbeitsstelle der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste an der Rheinischen Friedrich-Wilhelms-Universität Bonn. doi:10.20376/IDIOM-23665556.18.pr005.en.

Martin, Simon, and Nikolai Grube. 2008. Chronicle of the Maya Kings and Queens: Deciphering the Dynasties of the Ancient Maya. 2nd ed. London: Thames & Hudson.

Nelson, Brent. 2017. “Curating Object-Oriented Collections Using the TEI.” Journal of the Text Encoding Initiative 9. http://journals.openedition.org/jtei/1680; doi:10.4000/jtei.1680.

Pallán Gayol, Carlos. 2018. “L2/18–038. A Preliminary Proposal for Encoding Mayan Hieroglyphic Text in Unicode.” Version 2, January 22. https://www.unicode.org/L2/L2018/18038-mayan.pdf.

Prager, Christian M., and Sven Gronemeyer. 2018. “Neue Ergebnisse in der Erforschung der Graphemik und Graphetik des Klassischen Maya.” In Ägyptologische “Binsen”-Weisheiten III: Formen und Funktionen von Zeichenliste und Paläographie, edited by Svenja A. Gülde, Kyra V. J. van der Moezel, and Ursula Verhoeven-van Elsbergen, 135–81. Akademie der Wissenschaften und der Literatur, Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse 15. Stuttgart: Franz Steiner Verlag.

Prager, Christian M., Nikolai Grube, Maximilian Brodhun, Katja Diederichs, Franziska Diehr, Sven Gronemeyer, and Elisabeth Wagner. 2018. “The Digital Exploration of Maya Hieroglyphic Writing and Language.” In Crossing Experiences in Digital Epigraphy: From Practice to Discipline, edited by Annamaria De Santis and Irene Rossi, 65–83. Berlin: De Gruyter. doi:10.1515/9783110607208.
TEI Consortium. 2020. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.1.0. Last updated August 19. N.p.: TEI Consortium. https://tei-c.org/Vault/P5/4.1.0/doc/tei-p5-doc/en/html/.

Thompson, J. Eric S. 1962. *A Catalog of Maya Hieroglyphs*. The Civilization of the American Indian Series 62. Norman: University of Oklahoma Press.

Unicode Consortium. 2010. *The Unicode Standard. Version 6.0*. Edited by Julie D. Allen et al. Mountain View, CA: The Unicode Consortium. http://www.unicode.org/versions/Unicode6.0.0/.

W3C. 2019. “CSS Writing Modes Level 3. W3C Recommendation,” December 10. https://www.w3.org/TR/css-writing-modes-3/.

Walsh, John A. 2012. “Comic Book Markup Language: An Introduction and Rationale.” *Digital Humanities Quarterly* 6, no. 1. http://www.digitalhumanities.org/dhq/vol/6/1/000117/000117.html.

Wichmann, Søren. 2006. “Mayan Historical Linguistics and Epigraphy: A New Synthesis.” *Annual Review of Anthropology* 35: 279–94. doi:10.1146/annurev.anthro.35.081705.123257.

Zender, Marc Uwe. 1999. *Diacritical Marks and Underspelling in the Classic Maya Script: Implications for Decipherment*. Master’s thesis, University of Calgary. doi:10.11575/PRISM/19313.

NOTES

1 All project ODDs can be accessed in the project’s Gitolite repository, https://projects.gwdg.de/projects/documenations/repository.

2 In the field of Maya epigraphy, different conventions coexist to present graphotactic relations of signs. The most established system is by Thompson (1962). In our project we build on Thompson but add certain operators to allow a more detailed description of sign relations.

3 There are some ideas about how to represent Maya hieroglyphic writing in Unicode (see Pallán Gayol 2018). We argue that to provide a concise character set for its representation in Unicode, there first has to be an identified set of glyphs and their variants. At the moment all inventories of Mayan glyphs are incomplete, imprecise, or inconclusive. Because of that, we established the digital Sign Catalogue to overcome the shortcomings of previous catalogs (see section 7). If at some point in the future a Unicode character set of Maya glyphs is available, we could easily refer to it by using our digital Sign Catalogue. Each graph would be enriched with its corresponding Unicode character.
4 When it comes to stand-off annotation, we do not make use of the new <tei:standoff> container element (see TEI Consortium 2020, Appendix C: Elements, <standoff>, https://tei-c.org/Vault/P5/4.1.0/doc/tei-p5-doc/en/html/ref-standOff.html) because when starting our project back in 2014 there simply was no such element at hand, so we needed some creativity to utilize what TEI then had to offer. For us, the considerations of Bański (2010) and chapter 16.9 of the TEI P5 Guidelines (TEI Consortium 2020, “Stand-off Markup,” https://tei-c.org/Vault/P5/4.1.0/doc/tei-p5-doc/en/html/SA.html#SASO) were fruitful starting points to develop a custom solution fitting our needs. Starting with version 4.0.0 of the Guidelines, <tei:standoff> is a decent option for including further information in the textual encoding that is not directly represented in the source itself, such as contextual information. Nevertheless, we decided to stick to our custom solution: first, because we already use not only TEI but also non-TEI stand-off markup, and second, because there is no need to change a perfectly running system.

5 The use of the term “graph” for the abstract, typified form of a realized character is still under discussion within the project. Since in linguistic discourse a realized character is commonly referred to as a “graph,” the abstract form represents a kind of prototype of the graph. We are still debating how such a type assignment can be understood as an intermediate step between the realized graph and the grapheme. For example, the typed form could be considered as a meta-graph or proto-graph, in contrast to the realized graph and the grapheme.

6 For documentation, see “Ontology of the Sign Catalogue for Classic Mayan,” accessed January 18, 2021, https://classicmayan.org/documentations/catalogue.html.

7 ALMAH is still under development and will be the subject of a future publication.

8 For documentation of the Artefact Ontology, see “Idiom Schema,” accessed January 18, 2021, https://classicmayan.org/documentations/idiomschema.html.

9 For the parser, see “ClassicMayanTEIGenerator,” accessed January 18, 2021, https://projects.gwdg.de/projects/teigenerator.
AUTHORS

MARTIN DE LA IGLESIA
Martin de la Iglesia is the digital humanities team member on the project “Annotated digital edition of Philipp Hainhofer’s (1578–1647) travel and collection accounts” at Herzog August Bibliothek Wolfenbüttel.

FRANZISKA DIEHR
Franziska Diehr is an information scientist specializing in knowledge representation, currently working as an IT researcher at Freie Universität Berlin.

UWE SIKORA
Uwe Sikora works for several digital humanities projects at the Göttingen State and University Library with a focus on metadata modeling.

SVEN GRONEMEYER
Sven Gronemeyer is an honorary associate at La Trobe University Melbourne specializing in Maya writing and linguistics.

MAXIMILIAN BEHNERT-BRODHUN
Maximilian Behnert-Brodhun is developer at the Göttingen State and University Library with focuses on linked open data, databases, and search functionalities.

CHRISTIAN PRAGER
Christian Prager is a Maya epigrapher and coordinates the project “Text Database and Dictionary of Classic Mayan” at the University of Bonn.

NIKOLAI GRUBE
Nikolai Grube is the director of the project “Text Database and Dictionary of Classic Mayan” and professor in the Department of the Anthropology of the Americas, University of Bonn.