Textdatenbank und Wörterbuch des Klassischen Maya

Arbeitsstelle der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste an der Rheinischen Friedrich-Wilhelms-Universität Bonn

ISSN 2366-5556

PROJECT REPORT 5 | EN

OPEN ACCESS

Published 13 Mar 2018

DOI: 10.20376/IDIOM-23665556.18.pr005.en

Annual Report 2017

Nikolai Grube[1], Christian Prager[1], Katja Diederichs[1], Sven Gronemeyer[1,2], Antje Grothe[1], Céline Tamignaux[1], Elisabeth Wagner[1], Maximilian Brodhun[3] & Franziska Diehr[3]

[1)] Rheinische Friedrich-Wilhelms-Universität, Bonn
[2)] La Trobe University, Melbourne
[3)] Niedersächsische Staats- und Universitätsbibliothek, Göttingen

Digital Sign Catalogue

For the classification and systematization of Mayan hieroglyphs, we developed a digital Sign Catalogue. As an inventory of all signs.it is an indispensable tool for identifying the glyphs used in a specific text. The identification and and classification of the signs is challenging, because they appear in several graphic variants and can have multiple sign functions, for example as a logograph or as a syllabic sign. Further, continuing academic discussions over the decipherment of the approximately 1000 signs arise various hypotheses for linguistic readings of individual signs.

The demands of analyzing this complex writing system and of integrating the continually changing state of research into the sign catalog necessitate a flexible data model. It must be able to react to potential changes, and the documented information must be both reproducible and verifiable. We chose an ontologically based modelling approach based on CIDOC CRM[1] and GOLD[2]. The data model was implemented in RDF to optimally represent the semantic relations between the entities. The use of graph technology enables semantic queries of the data.

The assessment of proposed readings is a particular challenge for the decipherment and analysis of Maya writing. Multiple readings for a series of signs are attested throughout the research literature. We not only want to document, but also to qualitatively evaluate them. Drawing on the academic literature on the Maya script, we developed sets of criteria oriented toward linguistic context of use (e.g., correct part of speech, plausible text-image relationship, etc.), among other parameters (compare Kelley 1976). For each set, the criteria are linked with each other using propositional logic, so that an appropriate confidence level can be obtained according to the combination.

[1] CIDOC Conceptual Reference Model http://www.cidoc-crm.org/.
[2] General Ontology for Linguistic Description http://linguistics-ontology.org/.

The goal is to record the extant texts and the objects on which they are recorded in a machine-readable corpus and to compile a dictionary based on this corpus that represents the entire vocabulary and uses of the script. The digital Sign Catalogue aids in compiling the text corpus, which will be coded in TEI/XML. However, the text does not consist of phonemically transliterated values. Rather, every sign will be encoded with a reference to the URI of the corresponding entity in the Sign Catalogue. Once it has been compiled, the corpus will remain a stable, unchanging dataset. All discoveries concerning hieroglyphic readings and the grammar of the language will be recorded outside the corpus in the Sign Catalogue and the linguistic annotation tool ALMAH.

During linguistic analysis, the confidence level determined for each proposed reading aids in preselecting the readings to be investigated. Testing a reading in the corpus can provide new or additional criteria, whereby its confidence level increases or decreases. The ascertained criteria can easily be entered into the Sign Catalogue retroactively, whereby the confidence level is adjusted accordingly. Thus, we especially respond to the need to reproducibly document decipherments that can only be achieved through linguistic analysis and investigation of the text corpus.

The goal of the graph-based model of the Sign Catalogue, together with the text corpus and linguistic analysis, is to obtain secure proposals for readings of Maya signs and to also be able to optimally present new proposed readings for signs that have yet to be deciphered (compare Diehr et al. 2017).

The inventory of signs is currently being compiled with the digital Catalogue and is expected to be completed in mid-2018. In addition to entering data and creating a concordance with other glyph inventories, we are also working on vectorized drawings of the graphs and their variants. Signs that were not included in Thompson's glyph inventory will be re-classified and entered into the sign and graph database. We anticipate that there will be approximately 1000 distinct signs and approximately 3000 graphs. The Catalogue will then be published on our project portal, and the RDF data will be made accessible using a SPARQL endpoint. Moreover, the data will be published in the TextGrid Repository, where they can be retrieved using OAI-PMH. Documentation of the digital Sign Catalogue can be accessed at http://idiom-projekt.de/catalogue.

At present, we are working on a comprehensive paper with the working title "Standards for Maya Epigraphic Analysis I: Principles of Maya Graphemics," which will be published on the project website over the course of the year 2018. It will contain a detailed discussion of creating the graph variants and the confidence levels and criteria, as well as additional research results that have only been presented here preliminarily. Thus, this paper will be of particular interest to the academic community engaged in the study of Maya hieroglyphic writing.

TEI Parser

Transcribing the texts is a very complex task that requires many steps to be followed. Within the TEI file, all references to the signs are indicated by referring to a TextGrid URI, rather than to their concrete phonetic transliteration or their entry number from the Sign Catalogue. Consequentially, when transcribing, the TextGrid URI for a sign must be found to indicate the sign in the TEI document. To minimize the demands of this task, we have begun to develop a TEI parser. Given the numerical transliteration as input, the parser will generate the corresponding TEI structure according to the developed schema.

The parser was designed as a plug-in for the TextGrid Lab. Hence, the write permissions to the project's internal storage always exist.

By using the parser, the productivity of generating transcriptions is increased substantially. However, certain text phenomena must be encoded manually. One example is the mark-up of damage on a text-bearing object.

The parser takes in the transcription information *12st.[*5009st:*128st:679st]* and generates the following code from it:

```
<ab xml:id="pB1" type="glyph-block">
    <!-- 12st.[*5009st:*128st:679st] -->
    <g xml:id="pB1G1" ref="" n="12st" rend="left_beside" corresp="pB1S1"/>
    <damage agent="fracture" degree="0.75" quantity="3" unit="g">
        <supplied reason="damage" evidence="external" precision="high" ana="#pB1_note1">
            <note xml:id="pB1_note1" resp="SG">The same exonym also appears on DPL: P. 19,
P3 and DPL St. 15, J1, although in different glyphic renditions.
                <rs ref="textgrid:1234aa">Dos Pilas, Panel 19</rs>
                <rs ref="textgrid:5678bb">Dos Pilas, Stela 15</rs>
            </note>
            <seg xml:id="pB1S1" type="glyph-group" rend="right_beside" corresp="#pB1G1">
                <g xml:id="pB1G2" ref="" n="5009st" rend="above" corresp="#pB1G3"/>
                <g xml:id="pB1G3" ref="" n="128st" rend="above" corresp="#pB1G4"/>
                <g xml:id="pB1G4" ref="" n="679st" rend="beneath" corresp="#pB1G3"/>
            </seg>
        </supplied>
    </damage>
</ab>
```

The parser is not an editor, for which reason it can only be used to create files. Open and retroactive editing is not possible. In future work, the parser could be developed into an editor by expanding its functions so that it could not only create files, but also edit them.

Developing the TEI Schema

The development of the TEI schema comprises multiple work packages (WP) that deal with individual aspects of text-markup, some of which are interlocking with each other. In this, we tried to first create the structures of the TEI/XML that facilitate making a text from the Sign Catalogue machine-readable in its intended reading order. While these essential aspects were being tested on the material and correction loops were undertaken in the schema definition, we were able to address other work packages in parallel, for example, the so-called *TEI-Header* with the file metadata, encoding of damaged text passages, or creative aspects of text composition.

WP 01: Transcription and Semantic Text Structure

In the first work package (WP 01), we addressed the semantic structure of the texts. The goal was to render the text in TEI/XML in its logical sequence, or its reading order. A parallel issue is topographic text arrangement, which describes where the text is located on the text-bearing object (see WP 02 below).

In WP 01, we answered the following questions and discussed the following goals: How can the hieroglyphs be coded in XML? How is the reading order arranged in a text field, and how are graphic variants arranged in a hieroglyphic block? The deliverables of WP 01 were initial text examples (ranging from simple to complex) created to account for all text structural phenomena in Classic Maya inscriptions. The encoding scheme for the structure resulting from WP 01 should be applicable to all texts, but it requires special extensions for the codices.

WP 02: Topographic Text Arrangement

In the second work package (WP 02), we addressed the topographic arrangement of texts. Here, "topographic" refers to the position of the text on its carrier. The semantic text structure exists in parallel; as noted, this concerns how the text is read and of which logical sequences it is composed.

In WP 02, we considered the following issue: where are text fields and images located on the text-bearing object? The results of WP 02 were to represent example texts in the TEI/XML structure, in a way that accounts for all text arrangement phenomena in Classic Maya inscriptions. The topographic text arrangement resulting from WP 02 can, in principle, be applied to all texts.

WP 03: Philological and Text-Critical Markup

The third work package concerned editorial engagement with illegible, vague, or reconstructed text passages. Editorially, this concerns a text-critical approach that can be considered later when conducting epigraphic and linguistic analysis. This package thus specifies the markup from WP 01.

WP 03 entailed engaging with the following questions: how can text passages that have been made unclear by ambiguous original spellings or later restorations, or ones that have been damaged or destroyed by physical, chemical, or biological reactions be qualified? The goal of WP 03 was to markup example text passages such that they produce clear editorial guidelines for dealing with vague materials, from which an *apparatus criticus* is created by commenting in XML and referring to the Zotero bibliographic database.

WP 04: TEI Header and Editorial Metadata

We addressed the declarative characteristics of the entire TEI document in the fourth work package. "Declarative" refers here to descriptive metadata for text editing. In accordance with the languages used by the project, this information is being made available in German, English, and Spanish.

WP 04 thus deals with the following features: static background information, such as naming and referring to the Academy and the Union of German Academies of Sciences and Humanities as project sponsors, collaboration partners, a brief description of the project, listing project members, and stating the editorship and licensing of the edition as CC-BY-4.0. Dynamic information includes information about the artifact (e.g., its name), with reference to the database of text-bearing objects in TextGrid; who transliterated the text with the graph numbers from the Sign Catalogue; or who created the text encoding. If necessary, a revision history of the edition can be created in the header.

WP 05: Arrangement of the Text and the Text-Bearing Object

In the fifth work package, we addressed the creative execution of texts, meaning that we markup the design of the text (not so much the layout; see WP 02) and typography. The relationship of text arrangement to visual representations is considered in parallel.

WP 05 thus concerns the following questions: which criteria for text arrangement can be relevant to research questions, and to what extent do design and semantics stand in relation to one another? For this, characteristics such as the form of a text field, degree of relief, framework, coloring, or text size are recorded; additionally, individual scribes or scribal schools can be referenced. Furthermore, WP 05 describes the encoding of text-image relationships.

After the work packages have been completed, a comprehensive schema for annotating Maya hieroglyphic texts is available that can potentially be reused by other projects that are dealing with complex writing systems. Our experiences during the productive phase, which will begin in 2018, may raise additional demands or require extensions to the present specifications. Integrating these into the TEI schema will be a recurrent task during the coming years. In addition to the technical documentation of the metadata schema, we will also publish our editorial guidelines, which will use examples to explain our methods of text encoding and share guidelines for markup strategies.

Annotation Tool ALMAH (Annotator for the Linguistic Analysis of Maya Hieroglyphs)

We are adapting, modifying, and expanding the functions of the linguistic annotation software "GeTa" for semi-automatic morphological annotation in the research project "Text Database and Dictionary of Classic Mayan."

The Java-client application "GeTa" (Ge'ez Tagging) was designed and programmed by the computational linguist Dr. Cristina Vertan (Hamburg) to annotate texts in non-alphabetic writing systems (syllabic and logographic scripts). In the context of a work contract between the developer, Dr. Vertan, as the contractor and our research project as the contracting body, this tool is being adapted, modified, and expanded with additional annotation functions for semi-automatic transliteration and morphological and semantic annotation of Classic Mayan. This process is being taken in scientific consultation with ourselves as the contracting body. Simultaneously, we are creating a link with the lexicon that is being developed in parallel. The work contract thus outlines six work packages.

Computer-Supported Annotation Tools

Our project uses methods from corpus linguistics to address linguistic questions based on digital text corpora. As such, linguistic data from Classic Mayan are being digitally processed and furnished with basic annotations, such as part of speech labels. For the dictionary, large quantities of linguistic data will be needed that, in great part, must be edited semi-automatically, not just manually. Annotation tools are computer-supported tools which aid researchers in enriching texts or their components with individual information, for example with notes, comments on the material or text, or analysis of content or semantics. Linguistic annotation tools thus permit computer-supported transliteration, transcription, and annotation of texts, and they facilitate compilation and evaluation of corpora of both spoken and written languages.

Previously, no computer-supported approaches or tools existed for documenting Classic Maya hieroglyphic writing or Mayan languages. In addition, we must develop our lexicon and corpus simultaneously, because at present, only word lists exist for Classic Mayan. Prior annotation tools do not meet these requirements and essentially could not be used for these tasks, because they only permit annotation of texts composed in alphabetic writing systems. Except for GeTa, there is, to date, no other software that could be reused for computer-supported recording of the lexical inventory. At the same time, it must be possible to semi-automatically annotate different levels of transliteration and transcription with part of speech and morphological tags. In addition, it must be possible to manually enter corrections during the annotation process and to mark text passages that to date have been only tentatively or not at all deciphered. The semi-automatic character of such an annotation tool is necessary to accelerate the process of annotating the hieroglyphic texts in the growing corpus.

In this respect, the possibility of entering corrections must be emphasized, which allows researchers to check automatically generated annotations and to manually change them if necessary.

Adapting the Annotation Tool GeTa for Maya Hieroglyphic Writing

The data structure of the GeTa tool developed for Ge'ez facilitates future reuse of our project in several respects. GeTa permits automatic transliteration of the primary source(s), and annotation is semi-automatic and self-learning. This feature speeds up annotation and guarantees consistency. GeTa permits annotation at the level of lexeme and morpheme, another critical prerequisite for Classic Mayan lexicography. In writing systems with syllabic components, morpheme boundaries sometimes cannot be represented in the script, whereby their readings must be individually tokenized for morphological segmentation.

In the existing version of GeTa, the original text, transliteration, and transcription are synchronized, meaning that all levels, including further annotations in the case of changes, corrections, or additions, remain synchronous. Corrections during the annotation process can thus be undertaken synchronously by text or globally across the text corpus. Embedding variants for analysis of alternative interpretations at each level is currently being implemented. In the current version of GeTa for Ge'ez, annotations can be established for text structure (e.g. sentences, sections), morphology (part of speech), content (named entities), and editorial aspects (page breaks, corrections, etc.). The data are stored in JSON format and permit export into TEI/XML. Annotations can be searched, visualized, and statistically evaluated.

The function of keeping the lexicon and corpus in synchrony with each other is essential for compiling the dictionary of Classic Mayan. Simultaneous development of the lexicon is now being integrated. After GeTa has already been adapted to annotate Old Sabaean, Hebrew, and Yiddish texts, adaptations for analyzing Maya hieroglyphic texts are being undertaken based on the basic functions that have been described.

The architecture, data model, functions, and modules of GeTa must be adapted to the existing computational environment of our project, especially for an interface with TextGrid. These modifications for a tool for annotating Classic Mayan are being undertaken under the name ALMAH (Annotator for the Linguistic Analysis of Maya Hieroglyphs), derived from Yucatec Mayan *a'lmah*, "decir, mandar, ordenar (say, send, order)" (Barrera Vásquez 1980: 13). ALMAH must have access to user accounts in TextGrid in order that, while working with texts, data in the TextGrid Lab can be clearly read, opened, locked, saved, and deleted. For this, the RDF objects must be accessed via the Fuseki triplestore, in order that the objects can be linked in the ALMAH file using their URI. In addition, communication must be established with an Exist DB, because the numerically transliterated texts are in TextGrid in TEI/XML format. In addition to the references to TextGrid, links to the bibliography database Zotero must also be enabled to document annotations with citations.

Because of grammatical differences, all possibilities for morphological markup must be re-defined and implemented to be able to conduct linguistic analyses. For this, the various annotation types, levels, and categories must be defined, prepared, implemented, and, finally, tested for functionality. We need a series of annotation possibilities for the dictionary project that need to be redesigned and implemented in the existing GeTa architecture. In addition to a general comment function for texts, we need annotation levels for the step-wise transfer from written to spoken language, which should be conducted in a grammatologically and linguistically transparent manner. Eight levels of analysis are planned, in addition to the translation: 1) transliterations: alpha-numeric, numeric, graphemic, phonemic; 2) transcriptions: morphologically segmented, morphophonemically adjusted,

morphosyntactically glossed, and, finally, consolidated. In addition, there will be annotations for syntax or a semantic-functional analysis, for instance for metaphors and metonyms, calendrical or onomastic contents, and etymologies. Each level must be able to be referenced in sequence within ALMAH, but references to TextGrid using URIs must also be enabled here. For each annotation level, academic term lists with fixed values must be prepared and implemented. Glossing lists will be established for morphosyntactic analysis.

Changes and extensions are needed for program architecture and internal functionality. GeTa must be extended and supplemented such that it permits alternative transliteration, transcriptions, and analyses in all annotation levels. Because the Maya script has not yet been completely deciphered, epigraphers must be able to work with alternative readings and interpretations, and to develop these in future analysis. Fundamental adaptations to the data model are needed for this purpose, but they must also allow the option of not codifying a hypothetical reading. Furthermore, an interface for the Classic Mayan lexicon must be modeled and implemented. The lexicon of Classic Mayan is being created in the background during analysis and annotation of the hieroglyphic texts, which allows the automatically generated list of lemmas to be supplemented with scientific commentary and analysis. This function also must be re-modeled and implemented. The numerous aforementioned extensions and functional adaptions to GeTa for our project require adapting the user interface, in that the individual entry masks, including search functions for the various annotation types, are designed, implemented, and visualized in accordance with the scientific workflow. They are being planned, executed, and implemented by Dr. Cristina Vertan, according to the specifications and requirements of the Bonn research team.

3D Documentation

In 2017, the project undertook three trips to document text-bearing objects with its 3D structured light scanner.

During five days in February, the scanning that had been initiated in 2016 in the *Museum of Archaeology and Anthropology* at the *University of Cambridge* was continued and completed. The last, missing fragment of the plaster cast of Quirigua Stela E in the exhibition hall was also scanned, work which could not be completed the previous year because of an incorrectly calibrated sensor. In addition, various smaller artifacts from storage, including ceramics and bones with texts, were documented with the 3D scanner and photographs.

From October 2016 through April 2017, the exhibit “Maya: The Mystery of the Royal Cities” (German “Maya: Das Rätsel der Königsstädte”) was on display in the Historical Museum of the Palatinate in Speyer. In collaboration with the curators of the exhibit, the Guatemalan national cultural heritage institute *Instituto de Antropología e Historia* (IDAEH), and the foundation *Fundación Ruta Maya*, as the institutions lending the pieces on display, we were able to document 12 monumental stone sculptures in 3D over the course of six visits. During some visits, our project was supported by additional 3D scanners, once directly by the manufacturer *Aicon 3D Systems* and otherwise by Dr. Hubert Mara from the *Forensic Computational Geometry Laboratory, Interdisciplinary Center for Scientific Computing* at Heidelberg University. Our scanning work in the exhibition was conducted with a public engagement component (see *Project Outreach* below).

In June, select plaster casts, which had been made in the 1880s by Alfred Maudslay and are now in storage at the British Museum in London, were documented through 3D scanning.

In September, the project was additionally able to participate in a workshop at the 3D laboratory in the Institute of Mathematics of the Technical University Berlin, where various technologies for 3D digitalization and visualization were presented and discussed.

Additional museums with inscribed artifacts have given permission this year for us to undertake future scanning endeavors in their collections.

The scanning projects carried out in 2016, in addition to all those conducted in 2017, were subjected to intensive post-processing this year, which essentially consisted of the following steps: 1) cleaning up the raw data, 2) merging the raw data into a single mesh, 3) filling holes from the surface areas that were not captured, and 4) retexturing. The high-quality mesh that results serves as a foundation for additional visualizations. Some individual monuments consist of multiple pieces, whose meshes can then be virtually joined. In this way, the following objects could be finished, and some could be published on our *Sketchfab* page:

1. Cancuen, Panel 1 (original, Speyer)
2. Dos Pilas, Panel 19 (original, Speyer)
3. La Amelia, Stela 1 (original, Speyer, 2 fragments)
4. Lacanha, Panel (original, Speyer)
5. La Corona, Panel 1 (original, Speyer, 2 fragments)
6. La Pasadita, Panel 1 (original, Speyer)
7. Machaquila, Stela 3 (original, Speyer)
8. Machaquila, Stela 8 (mold, London, 1 fragment)
9. Palenque, Hieroglyphic Stairway (copy, London, 14 fragments)
10. Palenque, Hieroglyphic Stairway (mold, London, 1 fragment)
11. Palenque, Tablet of the Foliated Cross, center slab (copy, Bonn)
12. Palenque, Temple of the Inscriptions, east panel (copy, London, 2 fragments)
13. Provenance Not Known, Altar (original, Speyer)
14. Provenance Not Known, Bone 267 (original, Cambridge)
15. Provenance Not Known, Jade 618 (original, Cambridge)
16. Provenance Not Known, Panel (original, Speyer)
17. Provenance Not Known, Vessel 238 (original, Cambridge, 3 fragments)
18. Provenance Not Known, Vessel 238XHI (original, Cambridge)
19. Provenance Not Known, Vessel 257 (original, Cambridge)
20. Provenance Not Known, Vessel 273 (original, Cambridge)
21. Quirigua, Altar L (original, Speyer & cast, Cambridge)
22. Quirigua, Altar Q (copy, Cambridge)
23. Quirigua, Stela A (copy, Cambridge, 18 casts)
24. Quirigua, Stela C (copy, Cambridge, 4 casts)
25. Quirigua, Stela D (copy, Cambridge, 6 casts)
26. Quirigua, Stela E (copy, Cambridge, 14 casts)
27. Quirigua, Stela F (copy, Cambridge, 4 casts)
28. Quirigua, Stela J (copy, Cambridge, 5 casts)
29. Quirigua, Stela K (copy, Cambridge, 2 casts)
30. Quirigua, Zoomorph B (copy, Cambridge, 1 cast)
31. Seibal, Stela 3 (original, Speyer)
32. Tikal, Lintel 3, Tempel I (original, Basel)
33. Tikal, Lintel 2, Tempel IV (original, Basel, 7 fragments)
34. Tikal, Lintel 3, Tempel IV (original, Basel)
35. Yaxchilan, Lintel 2 (copy, Basel)
36. Yaxchilan, Lintel 16 (copy, Cambridge)
37. Yaxchilan, Lintel 17 (copy, Cambridge)
38. Yaxchilan, Lintel 24 (copy, Cambridge)

39. Yaxchilan, Lintel 30 (copy, Cambridge)
40. Yaxchilan, Lintel 42 (copy, Cambridge)
41. Yaxchilan, Lintel 43 (copy, Cambridge)
42. Yaxchilan, Lintel 45 (copy, Cambridge)
43. Yaxchilan, Stela 5 (copy, Cambridge)
44. Yaxchilan, Stela 7 (copy, Cambridge, 2 casts)
45. Zapote Bobal, Panel (original, Speyer)

Metadata Entry

In 2017, we focused on entering data about artifacts, epigraphic actors, and activities. With the support of student research assistants, all artifacts published in 19 of the 20 volumes of the *Corpus of Maya Hieroglyphic Inscriptions Project* (Harvard University), which represent over 20 sites, were entered into TextGrid using the entry mask (over 700 objects). In November 2017, our database contained over 3000 objects, consisting of 1442 objects with text and image (artefacts), 1337 objects with information about places (locations), 523 objects with epigraphic activities researched from the literature (epigraphic activities), 441 objects about epigraphic and 28 non-epigraphic actors (epigraphic and non-epigraphic actors), 436 objects with discovery events for artifacts (discovery events), and 103 objects for production events (production events). The digital Sign Catalogue currently comprises 190 objects with signs and 71 with graphs.

Vocabulary Development

We continued developing controlled vocabularies to support entry of data about text-bearing objects and to iconographically describe the graphs in the Sign Catalogue, which had been initiated last year after their design in 2015. Entry was conducted using the vocabulary development tool SKOtch in the format SKOS (Simple Knowledge Organisation System). At present (as of December 20), there are now 11 vocabularies (Activity, Actor Appellation, Artefact Orientation, Position and Assembling, Artefact Shape, Artefact Type, Condition, Group Type and Actor Role, Material, Technique and Style, Graph Icon) and a total of 1785 “concepts” with both “preferred” and “alternative terms”, in addition to their respective definitions and citations. When compiling the vocabularies, we referred as much as possible to normed data (Getty Art & Architecture Thesaurus, Getty Thesaurus of Geographic Names), and linked the corresponding entries to the appropriate normed data sets. However, because terms relevant to the Mesoamerican cultural area and especially the Maya region are underrepresented in the aforementioned normed data sets at present, we had to research and incorporate terms and definitions from the academic literature, in addition to formulating some of our own definitions. The extensive thesaurus that we thus constructed, including the definitions, is available in the project languages German, English, and Spanish; in addition, it also contains a series of Classic Mayan “terms”. The compiled vocabularies can be supplemented as needed at any time, although an initial version will be published in the coming year.

Bibliography (Zotero)

Since November 2016, over 5000 new entries have been added to the literature database Zotero, and existing entries have been corrected or supplemented with internet links for direct download. As of December 2017, the literature database contained 21,241 entries. Among other features, it is connected with TextGrid, where it is used to document entries with citations from the literature. In

conjunction with the project management office Beuse, Cologne, a search screen is being prepared for the website www.mayawoerterbuch.de, which will allow users to search and download entries. Furthermore, the literature database is linked with the image database ConedaKOR to similarly document entries with citations from the literature. Work on the literature database was supported this year by interns. In addition to entering bibliographic information for print media, more online publications were entered this year.

Image Database

The project's photographic archive consists of multiple researchers' collections, some of which look back on many decades of research trips in the Maya region. Thus, these photos are an invaluable source for the study of Classic Maya inscriptions, art, and architecture. Basic information has been added to each image. However, most of these descriptions are very unspecific. Thus, it soon became clear that systematic review of the material was necessary to precisely identify each depicted object.

Thus, we decided to begin processing of the black-and-white photo archive of Karl Herbert Mayer. It consists of approximately 20,000 images dating from 1974 to 2006. During his numerous trips, Karl Herbert Mayer focused on sculpted monuments with iconographic or epigraphic content. His documentation constituted the foundation for a series of publications on unprovenanced monuments (Mayer 1978, 1980, 1984, 1987, 1989, 1991, 1995) and numerous articles in academic journals.

The initial point of entry was provided by the monuments that Mayer himself had published and for which detailed descriptions were available. In a second step, documented monuments from museums were processed (e.g., the *Museo Nacional de Antropología* in Mexico City, the *Museo Nacional de Arqueología y Etnología* in Guatemala City, or the *Museo Regional de Antropología Palacio Cantón* in Mérida). Additional contents of the images that have been recorded to date span archaeological sites in the Mexican states of Campeche and Yucatan, monuments in the storage facilities of Tikal national park in Guatemala, and especially sites with mural paintings and graffiti. The entire image and documentation archive was viewed by Christian Prager and Antje Grothe in Graz in November 2017, packed into boxes, and transported to Bonn, where it is currently being prepared for digitalization by research assistants.

At present, approximately 60% of the black-and-white photos have been furnished with the necessary metadata for the depicted object (description, repository at the time the photo was taken, holders, collection, place of discovery, catalog number, publication, etc.).

In addition to Karl Herbert Mayer's archive, further photo archives will be processed in succession and transferred to an online image database.

ConedaKOR

The image management software ConedaKOR helps to manage and present our online database, whose collection will be made usable and completely accessible to the public via web. The browser-based frontend, as well as the metadata about the database contents, are written in English.

In the database, all photo archives whose contents have been digitalized and processed will be presented in context with each other. The heterogeneous information from the various collections about each depicted object must be represented in a coherent metadata schema, so that the illustrated objects and corresponding attributes from the different archives can be subjected to targeted searches.

The schema comprises comprehensive, describable characteristics of the depicted contents, like the location at time of the photograph was taken, holders, collection, place of discovery, artifact type, description, catalog number, publication, GND ID, etc.

The attributes are assigned to certain entities and can be connected through possible relations.

Here, the following entities can be represented:

1. “Medium”, the digitized representation (in the form of a TIFF file).
2. “Archaeological Site”, the place where the depicted object was discovered.
3. “Artefact”, the depicted object.
4. “Holder”, the place or institution or person by whom the object was held at the time it was photographed.
5. “Collection” refers to a holder (museum, individual, etc.). Collection always occurs in connection with a holder entity.
6. “Place”, the place of the depicted object.
7. “Person” refers to the person who created the photo that is found in the database as a medium.

Currently, the following bidirectional relations can be represented:

1. “Artefact is held by holder - Holder holds Artefact”
2. “Person visited Place - Place was visited by Person”
3. “Place is located in Place - Place locates Place”
4. “Holder is located in Place - Place locates Holder”
5. “Archaeological Site is in Place - Place locates Archaeological Site”
6. “Archaeological Site is origin of Artefact - Artefact originates from Archaeological Site”
7. “Medium depicts Artefact - Artefact is depicted by Medium”
8. “Person created Medium - Medium is created by Person”
9. “Artefact is related to Artefact - Artefact is related to Artefact”
10. “Person is depicted by Medium - Medium depicts Person”.

In ConedaKOR, all contents are technically organized within a graph, which implements entities with their attributes as nodes and connects their relationship to each other by means of edges. In addition to a targeted search, browsing from entity to entity is enabled with a click, whereby existing relations can be discovered alongside the relations. This network structure thus permits more complex associations between the objects and attributes of the various collections to be stored, represented, and made searchable in the database.

Internationalization and Promotion of Young Researchers

International Fellow

To develop our international relationships, the North Rhine-Westphalian Academy of Sciences, Humanities, and the Arts funded the research residency of Dr. Albert Davletshin from the Center for Comparative Linguistics of the *Russian State University* in Moscow. From March to December, Dr. Davletshin conducted research on the topic *Historical phonology as a basis for deciphering Classic Mayan*. With this research project, Dr. Davletshin contributed essential, foundational work to linguistic description, and thus to the dictionary, of Classic Mayan. As a comparative linguist with expertise in Amerindian languages and a special focus on Mayan languages, he supported the Bonn research team

in describing and deciphering Classic Mayan. His project on Classic Mayan historical phonology filled a research gap that we had defined in the proposal for this project: the phonological structure of Classic Mayan, which to date has only been selectively researched and still represents a desideratum in Classic Mayan lexicography.

Dr. Davletshin states that Maya hieroglyphic writing is logo-syllabic and was deciphered under the assumption that the underlying language was Mayan. The Mayan language family, however, consists of over 30 different languages, and differences between them are greater than their commonalities, comparable to the grammatical, syntactic, and lexicographic differences between Spanish and French. The past 40 years of Maya decipherment have revealed that the language of the Maya hieroglyphs belonged to the Ch'olan branch of the language family – however, none of the contemporary Mayan languages that belong to this branch correspond to "hieroglyphic Mayan", nor do they descend directly *in strictu sensu* from Classic Mayan. It is noteworthy, Davletshin says, that the language of the Maya hieroglyphs is phonologically conservative and systematically differentiates between short, long, and glottalized vowels, and between velar and laryngeal fricatives. These phonological characteristics of Classic Mayan have been lost in most Mayan languages spoken today, including all of the Ch'olan languages. Our understanding of Maya texts and decipherment of individual signs is based on the principle of etymology, i.e., comparison with lexical entries in colonial-era and modern Mayan dictionaries. This method relies on regular sound correspondences and assumptions about the development of sounds in the different Mayan languages as they developed from Proto-Mayan. Regular sound correspondences and the development of sound systems in the various Mayan languages provided the foundation for deciphering the Maya hieroglyphs, because they enabled lexical comparisons with other Mayan languages. They also permit researchers to eliminate implausible decipherments and interpretations.

Dr. Davletshin's research in the context of the "Text Database and Dictionary of Classic Maya" project addressed this important method and, using the example of complex vowels and rare consonants, researched the system of sound correspondences between the various Mayan languages as it relates to Classic Mayan. These insights support the Bonn researchers in deciphering the texts; moreover, they provide insights into the phonological structure of a defunct written language. The results of this project will be published in two articles, which will systematically investigate and describe the aforementioned sound laws.

Internships in the Project

The opportunities for BA and MA students to fulfill the internship requirement established by the curriculum of the Department of Anthropology of the Americas were expanded by adding internships in our project, and several students have already taken advantage of them. The internship covers 180 hours (BA) or 150 hours (MA), whereby the work schedule is divided into 10 hours per week to accompany coursework, so that the internship duration is 18 or 15 weeks, respectively. During the internship, six areas of application must be completed, which require knowledge of epigraphy, languages, academic work, research, basic IT knowledge, and experience in processing digital images. They offer the interns a comprehensive view of the process and methods of a long-term humanities research project and are set out as follows:

Bibliography: literature research in existing databases and select, printed academic bibliographies, as well as data entry, are intended to acquaint the intern with research tools and resources and to structure the assembled bibliographic data by transferring them to a bibliographic database, including linking the data sets to existing databases, websites, etc. In addition, entries in existing academic

bibliographies are to be expanded and, if needed, corrected through research in university library catalogs and printed literature.

Archive of inscriptions: The second area of application consists of work on the analog (archival cabinets) and digital inscription archive (TextGrid). For this, artifacts with inscriptions are researched from selected archives (inscriptions from Campeche, Justin Kerr's archive of ceramics, Karl Herbert Mayer's archive of unprovenanced objects), the requisite objects for each artifact are created in TextGrid (Artefact, Production, Location), and the corresponding data are entered.

Digitalization: In the third area of application, interns are introduced to the digitalization of 2D and 3D objects, along with the techniques, equipment, and software that it requires. 2D digitalization consists of scanning slides and photos and post-processing, as well as preparatory work for eventually entering the digitized objects into the image database ConedaKOR (selection, research, and description of the objects). For digitizing 3D objects, they practice scanning artifacts with a 3D structured light scanner and using the corresponding software Optocat, as well as post-processing the 3D scans, which consists of cleaning raw data, merging, and joining virtual 3D models of scanned monuments. In conclusion, they practice working with 3D models as sources for epigraphic research, using 3D models of previously scanned casts.

Database: In this area of application, interns are introduced to and trained to work in the image database ConedaKOR and the virtual research environment TextGrid, whereafter they research the digitalized objects that are being incorporated into ConedaKOR. Subsequently, the researched information about documentation location and artifact type is adjusted to correspond to the TextGrid concepts PlaceType and ArtefactType, respectively, using the vocabularies integrated into TextGrid. In parallel, the corresponding ConedaKOR contents are entered into TextGrid, in order to align them in ConedaKOR and to make them linkable. The contents of ConedaKOR are also curated and corrected as necessary.

Moreover, additional, non-textual information about text-bearing objects will be collected by assessing the literature, and it will be entered into TextGrid to complete pre-existing data sets.

Public relations and website: Work in the penultimate area of application consists of maintaining, correcting, translating, and expanding existing website and social media contents.

Epigraphy: In the final area of application, signs from selected inscriptions are classified and numerical transliterations are created using sign catalogs, in order to gain additional skills in epigraphy. Proposals for signs that are not yet registered in any of the catalogs are also discussed. In addition, interns are familiarized with the process of creating digital drawings with a drawing tablet.

Teaching

The project was invited to participate in the module "3D scanning and imaging" of the seminar *Sunoikisis Digital Cultural Heritage*. *Sunoikisis* is a national consortium for Classics programs that was created in the Center for Hellenic Studies at Harvard University. *SunoikisisDC* is based at the Alexander von Humboldt Chair of Digital Humanities at the University of Leipzig. The module was co-taught by Dr. Sven Gronemeyer, Professor Dr. Graeme Earl (King's College London), Dr. Gabriel Bodard, and Dr. Valeria Vitale (University of London). The seminar was conducted virtually via a video-conferencing platform and was transmitted live on YouTube, where it is also archived.[3]

[3] https://youtu.be/NZK1yoXdAVY.

Project Lectures, Presentations, Publications, Collaboration, and Outreach

Internal Conferences

The third annual project meeting[4] with the scientific advisory board took place in Bonn from December 7-9, 2017. For this conference, we were able to add two additional members to our international advisory board: Prof. Dr. Alexandre Tokovinine (Tuscaloosa) and Prof. Dr. Dmitri Beliaev (Moscow). The focus of the conference were presentations by project members, in which they reported on the current state of work. For the second day, we organized a user workshop to familiarize our colleagues with the IDIOM database. The conclusion of our annual meeting consisted of presentations on Saturday by all experts in attendance on the latest research in Maya epigraphy.

External Conferences

January 21, 2017 **"XX. Mesoamerikanistik-Tagung"** annual conference, Rautenstrauch-Joest Museum/Department of Anthropology of the Americas, University of Bonn, Cologne/Bonn, January 20–23, 2017 *"Zeichen, Grapheme, Varianten und Metadaten: ein digitaler Zeichenkatalog des Maya"* [Signs, Graphemes, Variants, and Metadata: A Maya Digital Sign Catalogue]

February 15, 2017 **"DHd 2017: Digital Sustainability"** annual conference, University of Bern, Bern, February 13–18, 2017 *"Nachnutzungs- und Nachhaltigkeitsstrategien im Projekt Textdatenbank und Wörterbuch des Klassischen Maya"* [Strategies for Reuse and Sustainability in the Textdatabase and Dictionary of Classic Maya Project]

April 24, 2017 **"Latin American Perspectives"**, lecture series in cooperation with the Adult Education Center (VHS) Bonn, Department of Anthropology of the Americas at the University of Bonn, and the Ibero Club Bonn e. V., Bonn *"Mayahieroglyphen im digitalen Zeitalter: Das Bonner Mayawörterbuch-Projekt"* [Maya Hieroglyphs in the Digital Age: The Bonn Maya Dictionary Project]

April 27, 2017 **"Digital Humanities Colloquium of the University of Cologne, Summer Semester 2017"**, University of Cologne, Cologne "*Entzifferung der Mayahieroglyphen im digitalen Zeitalter: Herausforderungen, Methoden, Ergebnisse"* [Maya Hieroglyphs in the Digital Age: The Bonn Maya Dictionary Project]

August 30, 2017 **"Digital Cultural Heritage (DCH2017)"** annual conference, Berlin State Library, Berlin, August 30-September 1, 2017 *"Cracking the Code: An Ontological Interlinked Working Environment for the Analysis of Classic Mayan Language and Script"*

September 29, 2017 **"Modelling Questions in the Digital Humanities"** workshop and informatics conference: Chemnitz, September 26-29, 2017, Chemnitz University of Technology *"Modellierung eines digitalen Zeichenkatalogs für die Hieroglyphen des Klassischen Maya - Ein neues Konzept zur Klassifikation von Schriftzeichen sowie der qualitativen Bewertung und Einstufung von Entzifferungshypothesen"* [Modelling a Digital Sign Catalogue for Classic Maya Hieroglyphs: A New Concept for Classifying Signs and Qualitatively Evaluating and Ranking Decipherment Proposals]

[4] Program: http://mayawoerterbuch.de/third-annual-project-workshop/.

October 7, 2017 **"Visible Words: Digital Epigraphy in a Global Perspective & Epidoc Editing Workshop",** Center of Digital Epigraphy, Brown University, Providence, October 5-7, 2017 *"Maya Hieroglyphic Writing and its Digital Exploration: The Bonn Classic Mayan Dictionary Project"*

October 19, 2017 **"Humanities Research Data – Methods for Digital Collection, Processing, and Presentation",** workshop and conference, Academy of Sciences and Literature in Mainz and the Union of the German Academies of Sciences and Humanities, Mainz, Oktober 18-20, 2017 *"Digitale Erforschung des Klassischen Maya im Bonner Mayawörterbuch-Projekt: eine Blaupause für die moderne Epigraphie?"* [Digital Investigation of Classic Mayan in the Bonn Dictionary Project: A Blueprint for Modern Epigraphy?]

November 10, 2017 **"3D Imaging in Cultural Heritage"** conference, British Museum, London, November 9-10, 2017 *"Old Casts in a New Light: 19th Century Plaster Casts of Classic Maya Artefacts as Cultural Heritage and Efforts Towards their Digital Preservation and Analysis"*

December 1, 2017 **"Wissensorganisation '17 - Knowledge Organization for Digital Humanities"** conference, Free University of Berlin, November 30–December 1, 2017 *"Ein digitaler Zeichenkatalog als Organisationssystem für die noch nicht entzifferte Schrift der Klassischen Maya"* [A Digital Sign Catalogue as a System for Organizing the Still Undeciphered Classic Maya Script]

Collaborations

In addition to our previously existing partnerships, in 2017 we were able establish collaborations with the following persons and institutions as we expanded our research infrastructure with essential tools for linguistic analysis and lexicography, archiving images, and analysis of 3D data:

Dr. Cristina Vertan (*TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages*, University of Hamburg), is adapting the annotation tool GeTa to be able to annotate Maya inscriptions (see *Annotation Tool ALMAH*).

For our Maya image archive, we are cooperating with *ConedaKOR* (Frankfurt, with the Max Weber Foundation in Bonn and DAASI in Tübingen) to establish and use an open-source database.

Together with Dr. Hubert Mara (*Forensic Computational Geometry Laboratory, Interdisciplinary Center for Scientific Computing*, Heidelberg University), we plan to develop and implement algorithms for automatic recognition of Maya hieroglyphs and text patterns based on 3D models and to develop metadata standards for citing, archiving, and reusing 3D models. At present, he is aiding the project with topometric analysis of the data from the scans.

The aforementioned scanning endeavors in 2017 were undertaken in cooperating with various institutions in Germany, Great Britain, and Guatemala. In addition to granting us access to their collections of artifacts with Maya inscriptions, they supported us with technical infrastructure during scanning. They include the *Museum of Archaeology and Anthropology* (Cambridge), the *British Museum* (London), as well as the *Historical Museum of the Palatinate* (Speyer), together with the *Instituto de Antropología e Historia*, Guatemala's national cultural heritage institute, and the *Fundación Ruta Maya*, a private cultural foundation in Guatemala.

Furthermore, this year, the following institutions agreed to collaborate with the project on the 3D scanning of their inscribed artifacts beginning in 2018: the *Ethnological Museum* (Berlin), the *Linden Museum* (Stuttgart), *Rautenstrauch-Joest-Museum* (Cologne), the *All Saints Museum* (Schaffhausen), and the *Rietberg Museum* (Zürich).

Project Outreach

In addition to its presentations at academic conferences, the project was able to develop a higher public profile in print media, as well as on television and the radio. This was especially significant with respect to our 3D scanning in Speyer, which the museum announced in a press release. In addition, there were a total of four articles in local and regional newspapers.

Moreover, our work was extensively presented in public radio, in two formats. The project was introduced on March 1 at 18:05 in a 4:53-minute piece in the context of the magazine "**IQ – Wissenschaft und Technik**" with Radio Bayern 2, with particular focus on the database and dictionary. On February 7 at 19:30, there was a 3:05-minute television special on the evening edition of the regional news program "**SWR Aktuell**" in Rhineland-Palatinate, which focused especially on the documentation work in the exhibit in Speyer.

In order to be able to scan during the museum's opening hours, we agreed with the museum's educational program that we make ourselves available to visitors for questions during our work, which was announced on a placard. Thus, our reach could be expanded collectively to include the museum visitors on each day, as well as individually through visitors' personal engagement with one of the three team members on-site.

Publikations

Diehr, Franziska, Maximilian Brodhun, Sven Gronemeyer, Katja Diederichs, Christian Prager, Elisabeth Wagner & Nikolai Grube
Modellierung eines digitalen Zeichenkatalogs für die Hieroglyphen des Klassischen Maya. In: *Informatik 2017*, Lecture Notes in Informatics, 275, edited by Maximilian Eibl & Martin Gaedke, pp. 1185-1196. Gesellschaft für Informatik, Bonn.

Diehr, Franziska, Maximilian Brodhun, Sven Gronemeyer, Katja Diederichs, Christian Prager, Elisabeth Wagner & Nikolai Grube
Ein digitaler Zeichenkatalog als Organisationssystem für die noch nicht entzifferte Schrift der Klassischen Maya. In: *Proceedings of Wissensorganisation 2017*, edited by Christian Wartena, Michael Franke-Meyer & Ernesto De Luca. (not yet published)

Feldmann, Felix, Bartosz Bogacz, Christian Prager & Hubert Mara
Histogram of Oriented Gradients for Maya Glyph Retrieval. In: *GCH 2017 Eurographics Workshop on Graphics and Cultural Heritage*, edited by Dieter Fellner, pp. 115-118. Eurographics Association, Goslar.

Gronemeyer, Sven
Die abgesagte Apokalypse: Der Blick der vorspanischen Maya auf das Ende 13. Bak'tun und das autochthone Konzept von Prophetie. In: *2012 – Die globalisierte Apokalypse aus lateinamerikanischer Perspektive*, Interdisziplinäre Studien zu Lateinamerika, 1, edited by Antje Gunsenheimer, Monika Wehrheim, Mechthild Albert & Karoline Noack, pp. 45-66. Bonn University Press & V&R unipress, Göttingen.

Grube, Nikolai & Octavio Quetzalcoatl Esparza Olguín
Two Captives from Uxul. Textdatenbank und Wörterbuch des Klassischen Maya, Research Note 6.

Grube, Nikolai, Christian Prager, Katja Diederichs, Sven Gronemeyer, Elisabeth Wagner, Maximilian Brodhun & Franziska Diehr
Milestone Report 2014-2016. Textdatenbank und Wörterbuch des Klassischen Maya, Project Report 4.

Prager, Christian & Elisabeth Wagner
Historical Implications of the Early Classic Hieroglyphic Text CPN 3033 on the Sculptured Step of Structure 10L-11-Sub-12 at Copan. Textdatenbank und Wörterbuch des Klassischen Maya, Research Note 7.

Wagner, Elisabeth

Jun Yop Ixiim – Another Appellative for the Ancient Maya Maize God. Textdatenbank und Wörterbuch des Klassischen Maya, Research Note 8.

Website and Social Media

Over the course of the second quarter, the website underwent a facelift in collaboration with the project management office Beuse, Cologne. Normally, the life span of the design for a new website is assumed to be no more than three years. After our website was officially launched in December 2015, its technical aspects proved to be much more problematic. Beginning in early 2017, project-specific adaptations of plugins and templates were no longer compatible with the new PHP 7 standard or the current versions of WordPress, which created security vulnerabilities.

Upon recommendation, the new plugin "Essential Grid" was integrated, which implements a modern, freely configurable tile design. Instead of having to have the project management office make design changes on the templates, they can now be undertaken at no cost via the editing interface, making the website lower-maintenance. Future compatibility is ensured because the plugin will be further developed in conjunction with WordPress. Additionally, the new plugin implements a "responsive design", which dynamically adjusts to the screen resolutions of mobile devices and thus responds to the increasing use of tablets and smart phones for browsing. The facelift was rolled out in July.

In 2017, the project also disseminated new publications, announcements of lectures and presentations, and 3D models in social network channels. Following our scanning activities in Great Britain and especially in Speyer, we shared insights into our practical work in photo and video form, as well as into internal workshops for developing the digital infrastructure. Internal statistics for Facebook indicate that one post usually reaches between 250 and 600 persons (with 225 followers); in one case, a post reached over 11,600 people.

References Cited

Barrera Vásquez, Alfredo
1980 *Diccionario Maya Cordemex: Maya - Español, Español - Maya*. Ediciones Cordemex, Mérida.

Diehr, Franziska, Maximilian Brodhun, Sven Gronemeyer, Katja Diederichs, Christian Prager, Elisabeth Wagner & Nikolai Grube
2017 Modellierung eines digitalen Zeichenkatalogs für die Hieroglyphen des Klassischen Maya. In: Eibl, M. & Gaedke, M. (Hrsg.), *INFORMATIK 2017*. Gesellschaft für Informatik, Bonn. (S. 1185-1196).

Kelley, David
1962 Review of A Catalog of Maya Hieroglyphs, by J. Eric S. Thompson. *American Journal of Archaeology* 66, 436-438, 1962.

Mayer, Karl Herbert
1978 *Maya Monuments: Sculptures of Unknown Provenance in Europe*. Acoma Books, Ramona, CA.
1980 *Maya Monuments: Sculptures of Unknown Provenance in the United States*. Acoma Books, Ramona, CA.
1984 *Maya Monuments: Sculptures of Unknown Provenance in Middle America*. Verlag von Flemming, Berlin.
1987 *Maya Monuments: Sculptures of Unknown Provenance, Supplement 1*. Verlag von Flemming, Berlin.
1989 *Maya Monuments: Sculptures of Unknown Provenance, Supplement 2*. Verlag von Flemming, Berlin.
1991 *Maya Monuments: Sculptures of Unknown Provenance, Supplement 3*. Verlag von Flemming, Berlin.
1995 *Maya Monuments: Sculptures of Unknown Provenance, Supplement 4*. Academic Publishers, Graz.