Modelling vagueness – A criteria-based system for the qualitative assessment of reading proposals for the deciphering of Classic Mayan hieroglyphs

Franziska Diehr[1], Sven Gronemeyer[2, 3], Elisabeth Wagner[2], Christian Prager[2], Katja Diederichs[2], Uwe Sikora[1], Maximilian Brodhun[1], Nikolai Grube[2]

[1] State and University Library Göttingen, Germany

[2] Department for Anthropology of the Americas, University of Bonn, Germany

[3] Department of Archaeology and History, La Trobe University, Australia

sgronemeyer@uni-bonn.de

Abstract

The project 'Text Database and Dictionary of Classic Mayan' aims at creating a machine-readable corpus of all Maya texts and compiling a dictionary on this basis. The characteristics of this complex writing system pose particular challenges to research, resulting in contradictory and ambiguous deciphering hypotheses. In this paper, we present a system for the qualitative evaluation of reading proposals that is integrated into a digital Sign Catalogue for Mayan hieroglyphs, establishing a novel concept for sign systematisation and classification. The paper focuses in particular on the modelling process and thus emphasises the role of knowledge representation in digital humanities research.

1 Modelling vagueness in the Maya Dictionary project

The not yet completely deciphered script of the pre-Columbian Maya culture is the research subject of the interdisciplinary project 'Text Database and Dictionary of Classic Mayan'.[1] The aim of this long-term project is to record the approximately 10,000 known text carriers and their inscriptions in a machine-readable corpus and based on this to compile a dictionary, which reflects the entire vocabulary and its use in context (Prager, 2014b).[2] As a joint effort, the specially established project office, located at the Department for the Anthropology of the Americas, University of Bonn,[3] works in close collaboration with the State and University Library Göttingen.[4] Since 2014, the hieroglyphic texts have been prepared, evaluated, and interpreted in interdisciplinary cooperation using methods and tools from the humanities and information technology (Prager, 2014c).

One result of this collaboration, and at the same time an important milestone of the project, is the digital Sign Catalogue which is the subject of this paper. For the creation of a text corpus and a dictionary of an only partially understood language and script, a Sign Catalogue, an inventory of all used signs, is an indispensable instrument. At the same time, the catalogue also forms the core component for constructing a machine-readable text. The project deals with a writing system that is only partially deciphered. The investigation of the script and language has produced numerous assumptions and interpretations about the reading of its signs. For a heuristic deciphering work, these reading hypotheses are to be included for further investigation. During the development of the Sign Catalogue, a system for the qualitative evaluation of decipherment hypotheses and reading proposals was developed. Among other things, it offers assistance with the linguistic analyses of the texts. The challenge in modelling this system was to represent several reading proposals, as well as to determine the factors that led to the formulation of the respective deciphering hypothesis. The goal of knowledge representation is also the explicit description of the methods applied. This enables the scholar to carry out analyses and draw conclusions on the

1. http://mayadictionary.de

2. The project started in 2014 and is expected to end in 2028. It is funded by the North Rhine-Westphalian Academy of Sciences and Arts http://www.awk.nrw.de and the Union of German Academies of Sciences and Humanities https://www.akademienunion.de.

3. https://www.iae.uni-bonn.de/

4. https://www.sub.uni-goettingen.de

Figure 1: Example for a Maya inscription

basis of the model. The modelling of reading proposals and their plausibility act as examples for the representation of complex knowledge objects with interpretive character, as they are typical for research data in the humanities. In this context, the paper emphasizes the role of knowledge representation in projects that are located in the digital humanities and considers the challenges posed by the modelling of vague and uncertain information.

2 Characteristics of Classic Maya writing

Compared to other Mesoamerican writing systems, such as Isthmic or Aztec, Maya writing has a considerably long time of use, about 2,000 years. The first pre-Classic texts were written in the 3rd century B.C. The writing tradition reached its peak in the Classic period (100 - 810 A.D.). The arrival of the Spaniards lead to a deep cut in Maya culture that also affected the use of their writing. It was only possible for the Maya to use it in secret. In the underground, hieroglyphic writing continued until the late 17th century, and ceased to exist thereafter. It was only due to 18th and especially the 19th century explorers who brought Maya texts back to light, and with that further and further into the focus of research. To date, an estimated 10,000 text carriers (Figure 1) have survived, mainly monumental inscriptions such as altars and stelae. Numerous texts can also be found on ceramics and small artefacts, like jewellery, implements or boxes. Of special importance are the codices written on paper-like material, of which unfortunately only four have survived.

The Maya writing system is characterised by an iconic character, hence labelled a hieroglyphic script. Typologically it is a logo-syllabic system

Figure 2: Writing of Maya Hieroglyphs

which is characterised by two main sign classes: logograms and syllabograms. Logograms stand for linguistic terms, like e.g. **PAKAL** for 'shield', and refer, with only a few exceptions, only to one denotation. Syllabograms represent single vowels and syllabic components and serve to write lexical and grammatical morphemes. They are also used as phonetic complements of logograms, as in **PAKAL-la**. Words could thus also be written exclusively with syllabic signs **pa-ka-la** (Figure 2), but mostly, both sign classes were combined in writing (Diehr et al., 2017, 1186).

The signs were arranged in almost rectangular blocks. Such a hieroglyphic block probably corresponds to the emic idea of a word. Within a block, the signs could be arranged in many different ways. Depending on space requirements and aesthetics, they could merge, overlap, be infixed, or rotated. The blocks were usually arranged in double columns and read from left to right and top to bottom. The sentences constructed this way form complex texts with a syntax and structure still preserved in modern Mayan languages (Prager, 2014a). Maya writing shows a distinct calligraphic complexity, as we can observe a broad graphic variability of signs. It allowed to write aesthetically sophisticated texts without necessarily repeating identical variants. For a linguistic expression, there is not only a single graphic correspondence, but usually several, sometimes very different, variants. The bipartite syllabogram **u** could either be written in its full form or alternatively with only one of its two segments. These segments could also be reduced or multiplied. As a further example, the syllabic sign **yi** shows how simple forms can be transformed into so-called head variants. Even diagnostic features (marked in Figure 3) are not present in all graph variants. It was even possible to insert or parenthesise other graphs, as the example **ma** shows (Figure 3).

The way in which variants are formed has not yet been the subject of systematic investigations before. There are a few individual studies by Beyer (1934a; 1934b; 1936; 1937) and Lacadena (1995,

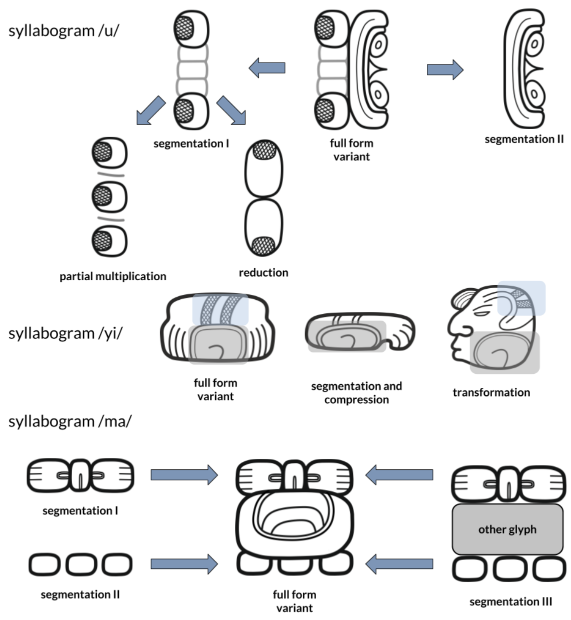

Figure 3: Graph variants

204 ff.) that investigate substitutions, rotation principles, or symmetries. In the course of the preparatory work for the digital Sign Catalogue, we recognised for the first time that there are general rules and principles for the formation of graphs which can be derived and defined by graphotactic analyses [5]. In total, we have identified 45 individual variation principles that can be divided into nine classes (mono-, bi-, tri-, and variopartite, division, animation head, animation figure, multiplication, extraction).

Another feature is the polyvalence of signs. A single glyph can either be read as a logogram or a syllabogram, depending on context. For example, the sign with catalogue number 528 can be used as the logograms **TUN** ('stone') and **CHAHUK** (name of the 19th day in the Maya calendar), or as the syllabogram **ku**. The graph variant used does not indicate which of the readings is to be applied.

This impressively illustrates how free Maya scribes were in the design of their texts. It also explains the challenges faced in the still ongoing decipherment process. Especially the graph variations make the determination of graphemes and their allocation to a specific linguistic expression difficult. Until today, the exact number of signs and their variations could not be determined exactly. A review of published sign inventories shows between 700 and 1,000 signs with a phonemic value. These comprise, according to own estimates, about 3,000 distinctive graph variants.

3 Sign catalogues as classification aids

The study of the hieroglyphs is carried out on the basis of previously discovered text carriers. However, many text carriers have been destroyed over time, are forgotten in archives and museums, hidden in private collections, or have simply not yet been discovered. Therefore, there will always be a small number of signs that could not be inventoried. On the other hand, it can be assumed that new discoveries will be made in the coming years. Early in 2018, the results of a large-scale aerial Lidar prospection in Guatemala were presented, showing a much denser concentration of archaeological sites in the tropical lowlands than previously assumed, also implying a larger number of text carriers (Clyens, 2018).

The discovery of new inscriptions and thus the discovery of previously unknown signs represents a challenge for the classification of Maya signs. Previous sign inventories could, of course, only consider the signs known at the time for a printed publication. The discovery of new signs has to be anticipated in the digital Sign Catalogue, also a possible re-classification of already known signs. We therefore need flexible data management processes for new research results.

Compared to other ancient languages and writing systems, the decipherment of Maya writing is relatively recent. Although it was known since the 19th century that many texts contain readable dates (Morley, 1915), a historical character and thus a representation of spoken language was rejected (Thompson, 1956, 169). It was not until the 1950s that Yuri Knorozov recognised the logo-syllabic character of Maya writing (Knorozov, 1952) and was able to present the first linguistically viable readings. Although his research appeared in English translation in 1958 (Knorozov and Coe, 1958), the Cold War left his work mostly unnoticed by researchers for a long time.

The hieroglyph catalogue by J. Eric S. Thompson (Thompson, 1962) contains several multiple and false classifications, but it still serves as the standard reference for sign classification and subsequent transliteration in preparation for linguistic analyses, even though Thompson did not provide any read-

5. This classification was first presented at the conference "Ägyptologische 'Binsen'-Weisheiten III" which took place at the Academy of Sciences and Literature in Mainz on April 8, 2016 (Prager and Gronemeyer, 2018).

ings in rejection of Knorozov's works (Prager, 2014a). A total of nine other sign inventories have been published so far. They all have incorrect classifications, particularly problematic are the multiple classifications of signs in which several allographs of a grapheme were inventoried as different signs (Grube, 1990; Kelley, 1962; Kurbjuhn, 1989; Riese, 2006; Ringle and Smith-Stark, 1996). The catalogues are mostly simple graph inventories. It was not until Knorozov's "Compendio Xcaret", published posthumously in 1999 (Knorozov, 1999), that readings were linked to the corresponding graphemes for the first time. However, these are based on the state of research of the 1960s. Even the two most recent catalogues show readings only in an unreflected manner.

With our digital Sign Catalogue we want to consider both expression levels of a sign, the functional-linguistic and the graphemic, and model them in such a way that the assignment of both levels to each other is accurate and flexible. A further disadvantage of traditional character catalogues is that they are unchangeable due to their printed edition and therefore cannot be dynamically extended. This prevents misclassifications from being corrected or new relations between signs from being established. Here, too, a catalogue in digital format can remedy the situation by being able to react flexibly to changes while at the same time offering persistent identification possibilities.

4 Modelling of the digital sign catalogue

We are developing our digital Sign Catalogue with the aim of carrying out a complete new inventory of Maya signs and thus making a reliable statement on the number of currently known signs. With the Sign Catalogue we establish a new concept for the systematisation and classification of signs. The specific characteristics of the complex Maya writing system are explicitly represented in the model: Graph variants, multifunctional characters and multiple transliteration values are defined and related to each other. Particular attention is paid to reading hypotheses, which are not only documented in the catalogue, but also qualitatively evaluated according to explicit criteria, so that they can be prepared for later analysis.

In order to develop a Sign Catalogue in a virtual environment, the signs and their properties must be represented in a machine-readable model. We understand modelling as a research method of digital humanities that aims to represent objects and the knowledge about them in a data model. In Sowa's sense, this means making the semantics of knowledge objects explicit and transferring them into a data model (Sowa, 2000, 132). We call the 'knowledge object' the epistemic notion of specific entities that constitute themselves in a specific knowledge context. Knowledge objects do not exist by themselves. They are created by knowledge-generating processes; they are formed from statements, analyses and interpretations about specific entities that are the object of interest. We call the specific knowledge context from which questions to the object arise 'domain'. The process of modelling knowledge objects and their specific domain can be subdivided into the following steps: (1) analysis of domain-specific requirements, (2) knowledge representation by conceptual modelling and (3) construction of a machine-readable model.

4.1 Analysis of domain-specific requirements

Requirements for the Maya Sign Catalogue are to be determined by intensive exchange and transfer of knowledge between domain experts and the modeller. In the process, specific requirements are defined for the model, its implementation in a data cataloguing system and technical environment. To do so, we chose a procedure based on the method of an expert interview. According to Reinhold, this represents an adequate method of information needs analysis particularly in the context of modelling research processes and research data in the digital humanities, since a high degree of implicit knowledge is to be expected on part of the researchers (Reinhold, 2015, 330). However, it should be noted that experts only share goal-oriented knowledge if the questioner already has a high level of expertise on the subject (Flick, 2007, 218 ff.). The process thus requires a high degree of familiarisation with subject on the part of the modeller, who must be able to describe the respective domain from a disciplinary point of view.

The process of defining requirements represents the most interdisciplinary area in joint project work. According to an explorative-hermeneutic working method, we proceeded as follows: The first step was to incorporate the basic concepts of the domain, i.e. linguistics and grammatology. We also analysed the subject, the Sign Catalogue, for its importance as an essential tool for decipherment, as well as its function specific to the project and discipline. Furthermore, an intensive analysis of the structure of the Maya writing system was carried out with the aim of explicitly representing its signs and their function in the model. The resulting minutes of the discussions formed the basis for a catalogue of requirements, which serves as a working paper to determine further requirements and to specify existing ones.

4.2 Knowledge representation by conceptual modelling

We are convinced that the process of knowledge representation is a hermeneutic method aimed at constructing a machine-readable model. Conceptual modelling defines what Sowa calls 'ontological categories'. They determine everything that can be represented in a computer application (Sowa, 2000, 51). The creation of an ontological model aims to explicitly describe knowledge objects, their relationship to each other, and to their domain. The definition of these categories is particularly difficult when it comes to vague and uncertain information: "any incompleteness, distortions, or restrictions in the framework of categories must inevitably limit the generality of every program and database that uses those categories" (Sowa, 2000, 51). Since 'knowledge' about objects can be questioned or interpreted differently, it is necessary to represent the various levels of knowledge in the model in order to counteract such distortions and to limit the knowledge base precisely in the sense of the defined ontological categories.

First, we scanned through specialised literature and linguistic vocabularies such as the SIL Glossary of Linguistic Terms (Loos et al., 2003) for definitions and concepts to describe writing systems and signs. However, the materials showed that most concepts are not reusable for our model because they already focus too much on applicability in a particular linguistic context. Our aim, however, is to define ontological categories for the representation of signs and their function in a writing system. Linguistic categories can only be applied on a meta-level (Diehr et al., 2018, 38).

Based on the literature and the outcomes of the requirement analysis, we modelled concepts and their relationships to each other and to their domain in an ontology using OWL syntax. Figure 4 shows the domain model of the ontology and illustrates the core concepts and their relationships to each other.

In our understanding, a sign is constituted by the conjunction of two different levels: 1) a linguistic-functional level, which, according to Ferdinand de Saussure, contains the notion and the pronunciation (de Saussure, 1931, 28 ff.) as well as the specific function of the sign in its writing system; and 2) a level of graphic representation which contains all possible forms of expression reflecting the concept of the linguistic-functional level.

Consider a Maya sign as an example: it has the verbal utterance **yi** and thus fulfills the function of a syllabic sign. For the syllabogram **yi** there are at least three different graphical forms of representation (see Figure 3). This form of representation is called a **graph**.[6] A graph is an abstract, typed form of an individually realised sign. The graph of **yi** in the variation 'anthropomorphic head variant' recorded in the catalogue represents a type which prototypically represents all individual writing variants and thus all actual occurrences for **yi** in the form of a 'head variant'.

All graphs that are assigned to a shared linguistic expression stand in an allographic relationship to each other and in their totality form variants of the **grapheme** of the sign (Bussmann, 2002, 294). A graph can only be assigned to exactly one linguistic-functional expression (idiomcat:Sign). This relation (idiomcat:isGraphOf) is optional, so that also graphs can be inventoried which could not be assigned to a 'sign' yet.

The function of a sign within the writing system is called a **sign function** (idiomcat:SignFunction). For Classic Mayan, besides the two main classes logogram and syllabogram (idiomcat:SyllabicReading), two additional functions were defined, since Rogers (2005, 10) also includes numerals (idiomcat:NumericFunction) and diacritics (idiomcat:DiacriticFunction). Logograms are further subdivided by their function: We differentiate between those that have a deciphered phonetic value (idiomcat:LogographicReading) and those for which only a semantic field can be narrowed down (idiomcat:LogographicMeaning).[7]

6. The term 'graph' chosen by the project as the concept of the abstract, typified form of a realised character is still under discussion within the project. Since in linguistic discourse the realised character is commonly referred to as 'graph', the abstract form represents a kind of prototype of the graph. Here it is still to be considered how such a typification can be considered in an intermediate step between the realised graph and the graphem. For example, the typed form could be considered as meta-graph or proto-graph in contrast to the graph.

7. This distinction may be irrelevant for sign classification

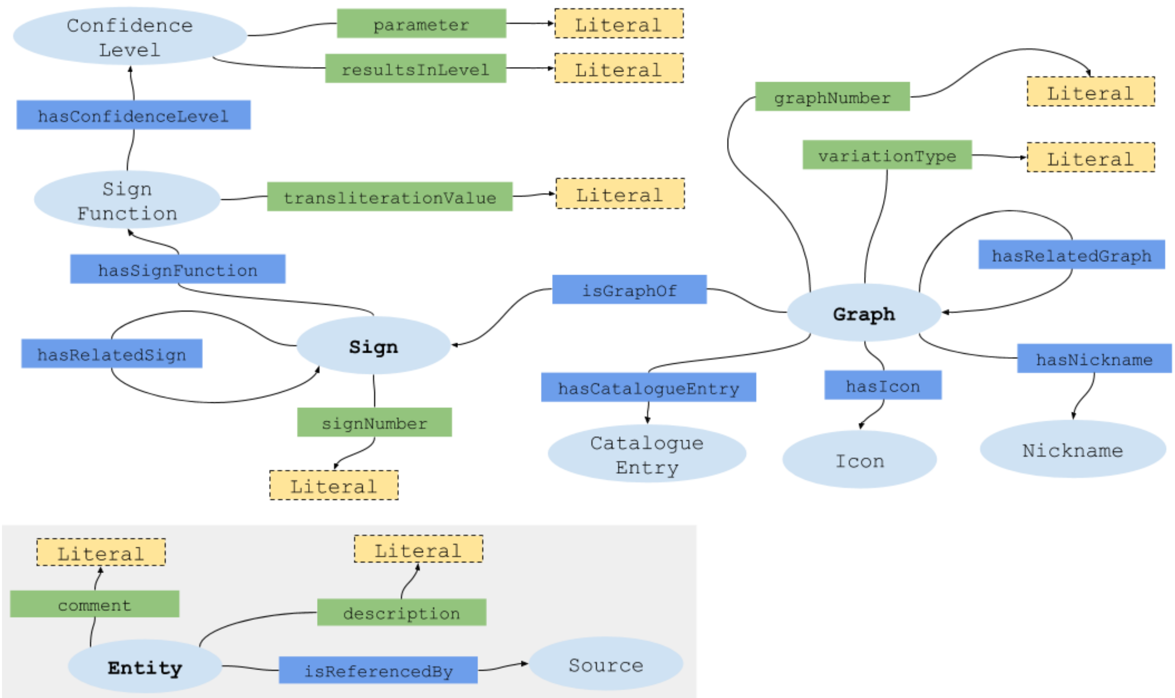

Figure 4: Domain Model of the Sign Catalogue

In the catalogue, we also document reading hypotheses and evaluate their plausibility (idiomcat:ConfidenceLevel). Therefore, a distinction of the logograms is necessary because different evaluation criteria (idiomcat:parameter) have to be applied to each. In the linguistic analysis of the corpus it also makes sense to distinguish between phonetic values and a meaning. Both are represented as transliteration values (idiomcat:transliterationValue) in the model.

4.3 Construction of a machine-readable model

Once the objects have been defined, their structure and relationships to each other have been represented in a conceptual model, the data model can be developed. Concepts and structures are translated into a machine-readable form by formulating them in a syntax suitable for representing the conceptual model. Since the Sign Catalogue was designed as an ontology, a data structure is required that can represent semantic relations between uniquely referenceable entities.

We use the virtual research environment TextGrid (Neuroth et al., 2015) for the administration, creation, and presentation of the data generated in the project. For data acquisition, we use the RDF input mask of the TextGrid Lab, which we adapted to our project-specific needs. The mask renders HTML forms to enter data on the basis of an RDF schema in TURTLE syntax. In order to make the input mask usable for the Sign Catalogue, the model designed as an ontology must be transferred into an RDF schema.

This illustrates that data modelling often requires pragmatic decisions. Since the functionality of the input mask is based on an RDF schema in TURTLE syntax, it was not possible to directly use the ontology written in OWL. This is not a problem in the present case, since the OWL schema could be transferred lossless into an RDF schema. In our scope, the expressiveness of the triple structure is sufficient to represent the complexity of the ontological relationships defined by the conceptual model. However, the generated data is stored in a triple store. With the use of the Query Language SPARQL sophisticated queries are possible.

In addition to the aim of representing knowledge objects and their domain in a machine-readable model, interoperability with other systems and schemata plays an important role in the development of data models. The adoption of already

in general, since the function of the character is one and the same for the writing system. However, with the ontology we model a specific use case: The Sign Catalogue serves as an instrument purposing the deciphering of the Classic Mayan writing system.

defined concepts enables the exploitation of the potential of existing metadata standards: Their integration improves the searchability of ontologies and increases both their quality and that of the applications accessing them, since the knowledge base is continuously enriched (Gradmann et al., 2013, 275 ff.). Simperl names the following steps for the reuse and integration of ontologies: 1) Search for reusable ontologies, 2) integration-oriented evaluation, and 3) integration of ontologies into the own model (Simperl, 2010, 246).

The chances of finding already existing ontologies are low if they are subject-specific and thus often only known in the respective domains. The ontologies developed in these domains are obviously very specific and/or geared to a specific application, which makes their immediate reuse and the integration of concepts more difficult.[8]

Therefore, it makes sense to use so-called top-level ontologies in addition to subject-specific ontologies. Top-level ontologies make it possible to reach agreement on generally valid concepts that do not have to be redefined in one's own scheme (Milton and Smith, 2004, 85).

While searching for reusable ontologies for the Sign Catalogue we came across 'General Ontology for Linguistic Description' (GOLD).[9] With the aim to define basic categories and relations for the scholarly description of human language, the ontology seemed to be reusable for our purposes. In the integrated evaluation it turned out that it focuses on grammatical rules with morphosyntax as a starting point (Farrar and Langendoen, 2003, 100). In our context, the concepts defined in GOLD could only be used to a limited extent, since our concept represents a meta-level that serves for the organisation of signs. Nevertheless, some concepts provided a good starting point for the definition of our schema. The definition of gold:FeatureStructure as "a kind of information structure, a container or data structure, used to group together qualities or features of some object" (Farrar, 2010) is so general that our definition of the class idiomcat:SignFunction as "a feature assessed to a Sign[,] [t]he nature of the feature is specified by the subclasses" [10] could be modelled as a subclass of gold:FeatureStructure.

The evaluation of suitable top-level ontologies showed that the CIDOC Conceptual Reference Model (CIDOC CRM), despite its focus on describing processes for documenting cultural heritage objects, contains many meta-concepts that are suitable for modelling our catalogue. In this regard, most classes have been defined as subclasses of CIDOC CRM, for example classes 'Sign' and 'Graph' were defined as specific sub-concepts of crm:E33_Linguistic_Object as they are "identifiable expressions in natural language" (Doerr et al., 2011).

5 Development of a system for the qualitative evaluation of decipherment hypotheses

In our Sign Catalogue we want to consider published as well as own reading hypotheses, especially since the readings are used for the linguistic analysis of the text corpus for the compilation of the dictionary.

Even though research is working with erroneous catalogues, the decipherment of Maya writing has significantly advanced in recent decades. In the 1980s, David Stuart presented an important study (Stuart, 1987) that gave the decipherment process new impetus. Nonetheless, there is no consensus about the reading for all signs. The reasons why signs are not equally deciphered can be manifold, for example if a sign is only attested once, or if there are no indicators pointing to phonemics. Since decipherment has mostly been carried out by individual epigraphers using isolated examples in 'hand-picked' texts, they have never been checked and compared using a complete text corpus, since it does not yet exist and is only being developed by the project now. Not only for these reasons different reading hypotheses were and are presented for many signs. Individual interpretations of the context or linguistic basics also contribute hereto.

5.1 Development and plausibility of reading hypotheses

From a grammatological point of view, different categories of deciphering can be defined that are linguistic and semantic, extended after the sign function by Riese (1971, 20-23) and deciphering criteria by Houston (2001, 9 ff.). The phonetic content of linguistically secured signs can be verified in a number of contexts, often with semantic

8. This is also true for the ontology of the Sign Catalogue: It was developed specifically for the needs of the project. Even though, it can principally be abstracted for other applications.
9. `http://linguistics-ontology.org/gold`
10. Ontology of the Sign Catalogue for Classic Mayan `https://classicmayan.org/documentations/catalogue.html`

and lexical agreement in morphographs, sometimes also with polyvalent readings. With operational readings, the sound value can be inferred on the basis of certain indications, such as vowel harmony rules, phonemic complements, or the semantic field. However, multiple reading proposals that are not polyvalent and have different plausibilities may also occur, based on the individual interpretation of the context or the knowledge of the material. In some cases, only parts of the phonetic content can be isolated due to a lack of indications, such as the final sound through complements. However, semantics and almost always the word class can often be narrowed down from only partially readable or undeciphered signs, either because of the context or also because of the iconicity of the graph. About one third of the sign inventory resists any reasonable interpretation so far, mostly morphographs and among these mostly nouns.

5.2 Definition of propositional logics to determine confidence levels

Since each sign function fulfills a different linguistic function, each requires the creation of its own set of criteria to test a decipherment hypothesis. For example, a logogram cannot appear in a stem medial position of a word, whereas a syllable character can (criterion ‘m’). The evaluation of a decipherment is qualitatively done by combining certain criteria in propositional logics, which indicate the plausibility in disjunct numerical values. Each sign function has a different number of plausibilities. If new texts add new occurrences with criteria that have not been considered so far, these can be supplemented. For example, the following criteria have been defined for the syllabic signs:

d = Landa’s Manuscript
l = Consonant in Landa’s Alphabet
u = Vowel in Landa’s Alphabet
k = Complete logographic substitution
t = Partial logographic substitution
a = Allographic substitution
o = Preposed to deciphered logogram
c = Postposed to deciphered logogram
f = Stem initial
m = Stem medial
g = Word medial
r = Word final
y = Ergative spellings
s = Correspondence with image
q = Correspondence with object
b = Doubling is possible
v = Vowel harmony
i = Lexical correspondence with graph icon

These are combined in four propositional logics to determine four plausibility levels:

1 (excellent) = $d \vee u \vee (k \wedge (s \vee q)) \wedge (a \wedge (f \vee m) \wedge (s \vee q)) \vee ((o \vee f \vee m \vee y) \wedge (g \vee r \vee c) \wedge (s \vee q))$

2 (good) = $(t \vee o \vee c) \wedge v \wedge (s \vee q)) \vee (a \vee o \vee f \vee m) \wedge (g \vee r \vee c)) \vee (g \vee r \vee c) \wedge (s \vee q)) \vee (f \vee m) \wedge (s \vee q))$

3 (partial) = $t \vee c \vee l \vee r \neg v$

4 (weak) = $g \vee m \vee r$

The criteria and propositional logics were developed under critical consideration of the existing decipherment practice. In an account from about 1566, we actually have a contemporary description of the Maya script (de Landa, 1959, 104 ff.), even if it was considered as an alphabet then. Nonetheless, the manuscript contains annotations to certain hieroglyphics with a syllabic value, which then brought the key to decipherment. Even before Knorozov it was known that there was a strong text-image relation in the codices. Thus Paul Schellhas was able to isolate the proper names of different gods in a structuralistic way (Schellhas, 1897). This also turned out to be helpful for decipherment. Among the examples with a syllabic annotation in said manuscript there are also <cu> and <ku> (**ku** and **k’u** in today’s orthography). These examples are therefore deciphered only by criterion ‘d’, as they are supported by a contemporary witness, further criteria would be exclusively complementary. Both signs can be found in the initial position (criterion ‘f’) in hieroglyphs that identify animals that are also depicted in the corresponding image: a turkey and a vulture (Figure 5). Colonial dictionaries provide *kutz* and *k’uch* for the two animals. Thus we also have a hypothesis to the sound of the second syllabograms (**tzV** and **chV**), but without knowing the vowel, because both hieroglyphics are not present in Landas list. Criterion ‘r’ is fulfilled and level 3 is reached.

But the sign **tzV** appears in the initial word position among the depiction of a dog, attested as *tzul*

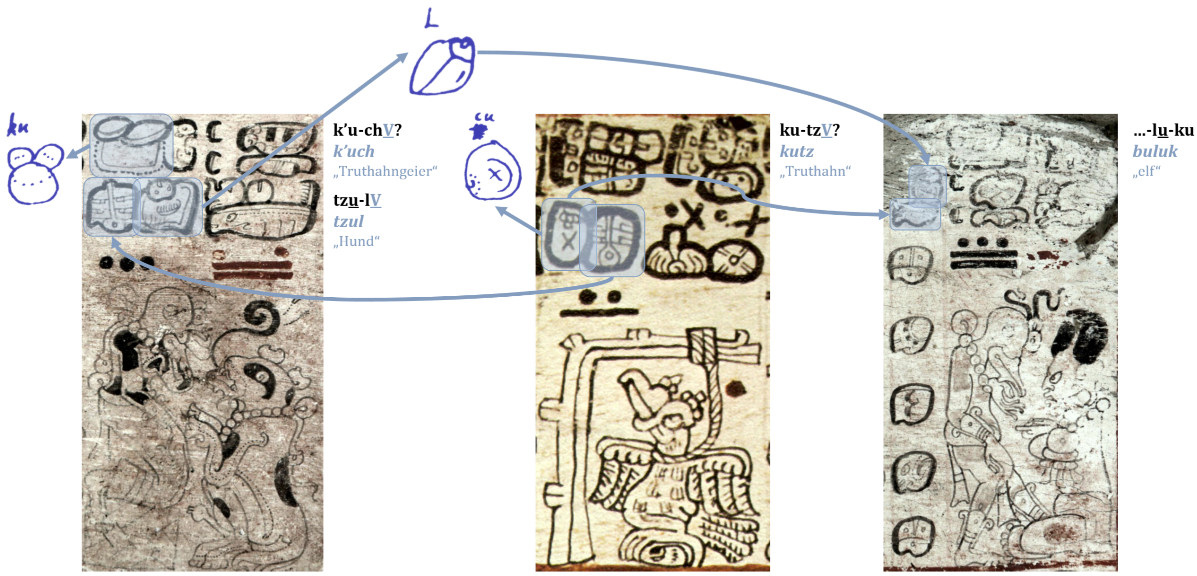

Figure 5: Decipherment of Syllabograms

in the dictionaries. Thus the syllabic value **tzu** is confirmed and the sign becomes level 2, in connection with the image even to level 1 (criterion ‘s’). The second sign is therefore **lV**; the initial sound is confirmed by the annotation <L> for a very similar sign in the manuscript. The examples have so far shown vowel harmony (**CV1-CV1**), so that the reading might be **lu**, which would be level 2.

Final proof comes from a stem medial attestation. In the calendrical structure of an almanach, we find three hieroglyphs where the number ‘11’ should appear, which is spoken *buluk*. Indeed, with the spelling **bu-lu-ku**, the assumed vowel can be confirmed here, and the sign is deciphered on level 1.

6 Methods and techniques for the creation of a machine-readable corpus

One objective of the project is to create a machine-readable corpus of all inscriptions. Due to the complex graphemics, the sign polyvalence and the low decipherment status, it is not possible to capture the text in phonemic-transliterated values. Therefore, it is not surprising that there is currently no standardised machine-readable font, such as Unicode, available for the Maya script. There are efforts in this direction (Pallan Gayol and Anderson, 2018), but in their current form they do not meet the classification requirements of the Mayan script.

In the text corpus we would like to refer to the graph variants used in each case, in order to carry out investigations for the use of the writing in its spatial-temporal development and its use depending on the text carrier, its location as well as the text contents.

We encode the text corpus in XML using the TEI-P5 guidelines (TEI, 2018). In order to distinguish the structure, the arrangement of the glyphs and other inscription-specific phenomena, we have developed a TEI-compliant, application-specific schema. Instead of using phonemic-transliterated values, we use Semantic Web technology to refer in the XML document to the resource stored in RDF. In the Sign Catalogue, each graph variant is recorded as an independent resource and thus has a URI. In the TEI-encoded text we tag each glyph with the element <g> (character or glyph) and use the attribute @ref (“points to a description of the character or glyph intended”)[11] to refer to the respective URI (Figure 6). The text itself consists of external resources and together with the Sign Catalogue forms an ontologically linked system.

With this approach, we also take into account the requirement to consider different deciphering hypotheses. Since the text corpus refers to the graph, which is potentially associated with several possible reading proposals in the digital catalogue, it is now possible to analyse the inscription under consideration of various hypotheses. A further advantage of this approach is that new decipherments can be flexibly integrated. Even in the case of newly

11. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-g.html

```
<ab xml:id="A2" type="glyph-block">
   <g xml:id="A2G1" n="0001" ref="textgrid:012sq" rend="above" corresp="#A2S1"/>
   <seg xml:id="A2S1" type="glyph-group" rend="beneath" corresp="#A2G1">
      <g xml:id="A2G2" n="0573" ref="textgrid:527vf" rend="left_beside" corresp="#A2G3"/>
      <g xml:id="A2G3" n="0228" ref="textgrid:829sk" rend="right_beside" corresp="#A2G2"/>
   </seg>
</ab>
```

Figure 6: Exemplary encoding of a hieroglyphic block in TEI/XML

formulated, reliable statements, these do not have to be elaborately incorporated into the corpus. By using URIs, it is possible to clearly reference the corresponding resource in the Sign Catalogue. The encoding of the corpus text itself does not change, it remains stable.

7 Data publication and reuse possibilities

The cataloguing and classification of Maya signs is a continuous process. Through ongoing research, new findings will repeatedly lead to reclassifications. In the scope of our project, a first new inventory will be completed in the course of 2018. As soon as all graph variants and signs have been recorded, the encoding of the inscriptions and the compilation of the corpus will begin. The documentation of the texts and text carriers will extend over the remaining duration of the project until 2028. The data will successively be made accessible on our project portal,[12] which is currently in the conception stage. Furthermore, the corpus data will also be published in the TextGrid Repository (TG Rep),[13] where they can also be accessed by external users via OAI-PMH. The RDF data of the Sign Catalogue will be also retrievable at the portal via a SPARQL endpoint and also at the TG Rep. All schemata created in the project can be downloaded from the public area of our Git repository[14] and can be used under a CC BY-4.0 license. The documentation of the Sign Catalogue ontology is also available as a website.[15]

8 Conclusion: modelling as a research method

Modelling processes bring about a reflection of the definitions and methods of the respective discipline, which poses questions to knowledge objects and their domain. The representation of this knowledge in a machine-readable model requires the precise definition of the objects as well as their relationships to each other and to their knowledge base. Specialist traditions are questioned and applied methods are checked for their fundamentals. In the case of modelling the deciphering hypotheses, it became clear that statements about the plausibility of reading proposals can only be made on the basis of formal evaluation criteria. Without reflection on the intuitive-pragmatic evaluation mechanisms, no evaluation system could have developed that operates on the basis of formalised and logical-categorised parameters.

Modelling thus makes a central contribution to the research methods of the Digital Humanities by stimulating the reflection on the generation of discipline-specific knowledge through conscious questioning. The modelling process produces explicitly defined objects and relationships that are transformed into machine-readable data and data structures. This transformation process makes the knowledge objects and their knowledge base available for further analysis, whether hermeneutical or quantitative.

12. https://www.classicmayan.org/
13. https://textgridrep.org/
14. https://projects.gwdg.de/projects/documentations/repository
15. Ontology of the Sign Catalogue for Classic Mayan https://classicmayan.org/documentations/catalogue.html

References

Beyer, Hermann (1934a). The position of the affixes in Maya writing I. *Maya Research*, 1:20–29.

Beyer, Hermann (1934b). The position of the affixes in Maya writing II. *Maya Research*, 1:101–108.

Beyer, Hermann (1936). The position of the affixes in Maya writing III. *Maya Research*, 3(1):102–104.

Beyer, Hermann (1937). Studies on the inscriptions of Chichen Itza. *Contributions to American Anthropology and History*, 4(21):29–175.

Bussmann, Hadumod (2002). *Lexikon der Sprachwissenschaft*. Stuttgart: Alfred Kröner Verlag.

Clyens, Tom (2018). Exclusive: Laser scans reveal Maya “megalopolis” below Guatemalan jungle. *National Geographic*. URL https://news.nationalgeographic.com/2018/02/maya-laser-lidar-guatemala-pacunam/.

de Landa, Diego (1959). *Relación de las cosas de Yucatán. Introducción por Angel María Garibay K.* México, D.F.: Biblioteca Porrúa 13. Editorial Porrúa.

de Saussure, Ferdinand (1931). *Cours de linguistique générale*. Paris: Payot.

Diehr, Franziska, Sven Gronemeyer, Christian Prager, Maximilian Brodhun, Elisabeth Wagner, Katja Diederichs, and Nikolai Grube (2017). Modellierung eines digitalen Zeichenkatalogs für die Hieroglyphen des Klassischen Maya. In Maximilian Eibl and Martin Gaedke, eds., *INFORMATIK 2017*, pages 1185–1196. Bonn: Gesellschaft für Informatik. doi:10.18420/in2017_120.

Diehr, Franziska, Sven Gronemeyer, Christian Prager, Maximilian Brodhun, Elisabeth Wagner, Katja Diederichs, and Nikolai Grube (2018). Ein digitaler Zeichenkatalog als Organisationssystem für die noch nicht entzifferte Schrift der Klassischen Maya. In Christian Wartena, Michael Franke-Maier, and Ernesto De Luca, eds., *Knowledge Organization for Digital Humanities. Proceedings of the 15th Conference on Knowledge Organization WissOrg'17 of the German Chapter of the International Society for Knowledge Organization (ISKO)*, pages 37–43. doi:10.17169/FUDOCS_document_000000028863.

Doerr, Martin, Matthew Stiff, Nick Crofts, Stephen Stead, and Tony Gill (2011). CIDOC Conceptual Reference Model Version 5.0.4. Tech. rep., ICOM/CIDOC CRM Special Interest Group. URL http://www.cidoc-crm.org/get-last-official-release.

Farrar, Scott (2010). General ontology for linguistic description (GOLD). Tech. rep., Department of Linguistics (The LINGUIST List). URL http://linguistics-ontology.org/.

Farrar, Scott and D. Terrence Langendoen (2003). A linguistic ontology for the Semantic Web. *GLOT International*, 7(3):97–100.

Flick, Uwe (2007). *Qualitative Sozialforschung: Eine Einführung*. Rowohlt, Reinbek.

Gradmann, Stefan, Evelyn Dröge, Julia Iwanowa, Violeta Trkulja, and Steffen Hennicke (2013). Wege zur Integration von Ontologien am Beispiel einer Spezifizierung des Europeana Data Model. In Hans-Christian Hobohm, ed., *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Proceedings. 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013)*, pages 273–284. Werner Hülsbusch.

Grube, Nikolai (1990). Die Entwicklung der Mayaschrift: Grundlagen zur Erforschung des Wandels der Mayaschrift von der Protoklassik bis zur spanischen Eroberung. In *Acta Mesoamericana 3*. Von Flemming.

Houston, Stephen D., Oswaldo Chinchilla Mazariegos, and David Stuart (2001). Introduction. In Stephen D. Houston, Oswaldo Chinchilla Mazariegos, and David Stuart, eds., *The Decipherment of Ancient Maya Writing*, pages 3–19. Norman, OK: University of Oklahoma Press.

Kelley, David H. (1962). Review of A Catalog of Maya Hieroglyphs, by J. Eric S. Thompson. *American Journal of Archaeology*, 66(4):436–438. doi:10.2307/502055.

Knorozov, Juri (1952). Drevnyaya pis'mennost' Tsentral'noy Ameriki. *Sovietskaya Etnografiya*, 3(2):100–118.

Knorozov, Juri (1999). *Compendio Xcaret de la Escritura Jeroglífica Maya*. Universidad de Quintana Roo, Promotora Xcaret, Chetumal.

Knorozov, Juri and Sophie D. Coe (1958). The problem of the study of the Maya hieroglyphic writing. *American Antiquity*, 23(3):284–291.

Kurbjuhn, Kornelia (1989). *Maya: The Complete Catalogue of Glyph Readings*. Kassel: Schneider & Weber.

Lacadena, Alfonso (1995). *Evolución formal de las grafías escriturarias mayas: Implicaciones históricas y culturales*. Madrid: Universidad Complutense. URL https://eprints.ucm.es/2433/1/H0028201.pdf.

Loos, Eugene E., Susan Anderson, Dwight H. Day, Paul C. Jordan, and J. Douglas Wingate (2003). *Glossary of Linguistic Terms 29*. Dallas: SIL International. URL https://glossary.sil.org/.

Milton, Simon K. and Barry Smith (2004). Top-level ontology: The problem with naturalism. In Achille C. Varzi and Laure Vieu, eds., *Formal Ontology in Information Systems*, pages 85–94. IOS Press.

Morley, Sylvanus G. (1915). *An Introduction to the Study of the Maya Hieroglyphs*. Washington, D.C.: Government Printing Office.

Neuroth, Heike, Andrea Rapp, and Sibylle Söring, eds. (2015). *TextGrid: Von der Community – für die Community*. Göttingen: Universitätsverlag Göttingen. doi:10.3249/webdoc-3947.

Pallan Gayol, Carlos and Deborah Anderson (2018). Achieving machine-readable Mayan text via Unicode: Blending "old world" script-encoding with novel digital approaches. In Élika Ortega, Glen Worthey, Isabel Galina, and Ernesto Priani, eds., *Book of Abstracts Digital Humanities 2018, Puentes-Bridges, Mexico City*, pages 256–261.

Prager, Christian (2014a). Textdatenbank und Wörterbuch des Klassischen Maya: Objectives. Tech. rep., Arbeitsstelle der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste an der Rheinischen Friedrich-Wilhelms-Universität Bonn. URL http://mayawoerterbuch.de/objectives/.

Prager, Christian (2014b). Textdatenbank und Wörterbuch des Klassischen Maya: Project architecture. Tech. rep., Arbeitsstelle der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste an der Rheinischen Friedrich-Wilhelms-Universität Bonn. URL http://mayawoerterbuch.de/architecture/.

Prager, Christian (2014c). Textdatenbank und Wörterbuch des Klassischen Maya: Sign classes and catalogues. Tech. rep., Arbeitsstelle der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste an der Rheinischen Friedrich-Wilhelms-Universität Bonn. URL http://mayawoerterbuch.de/aufbau-2/.

Prager, Christian and Sven Gronemeyer (2018). Neue Ergebnisse in der Erforschung der Graphemik und Graphetik des Klassischen Maya. In Svenja A. Gülden, Kyra V. J. van der Moezel, and Ursula Verhoeven-van Elsbergen, eds., *Ägyptologische "Binsen"-Weisheiten III: Formen und Funktionen von Zeichenliste und Paläographie*, no. 15 in Akademie der Wissenschaften und der Literatur, Abhandlungen der Geistes- und sozialwissenschaftlichen Klasse, pages 135–181. Stuttgart: Franz Steiner.

Reinhold, Anke (2015). Das Experteninterview als zentrale Methode der Wissensmodellierung in den Digital Humanities. *Information Wissenschaft Praxis*, 66 (5-6):327–333. doi:10.1515/iwp-2015-0057.

Riese, Berthold (1971). Grundlagen zur Entzifferung der Mayahieroglyphen, dargestellt an den Inschriften von Copan. In *Beiträge zur mittelamerikanischen Völkerkunde 11*. Kommissionsverlag K. Renner.

Riese, Berthold (2006). Drei neue Maya-Hieroglyphen Kataloge. *Anthropos*, 101(1):238–246.

Ringle, William M. and Thomas C. Smith-Stark (1996). A concordance to the inscriptions of Palenque, Chiapas, Mexico. In *Middle American Research Institute, Publication 62*. Tulane University.

Rogers, Henry (2005). Writing systems: A linguistic approach. In *Blackwell Textbooks in Linguistics 18*. Blackwell Publishing.

Schellhas, Paul (1897). *Die Göttergestalten der Mayahandschriften: Ein mythologisches Kulturbild aus dem alten Amerika*. Dresden: Richard Bertling.

Simperl, Elena (2010). Guidelines for reusing ontologies on the Semantic Web. *International Journal of Semantic Computing*, 4(2):239–283.

Sowa, John F. (2000). *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks Cole Publishing.

Stuart, David (1987). Ten phonetic syllables. *Research Reports on Ancient Maya Writing 14*.

TEI (2018). *P5: Guidelines for Electronic Text Encoding and Interchange*. Text Encoding Initiative. URL http://www.tei-c.org/release/doc/tei-p5-doc/en/html/.

Thompson, J. Eric S. (1956). The rise and fall of Maya civilization. In *The Civilization of the American Indian Series 39*. Norman, OK: University of Oklahoma Press.

Thompson, J. Eric S. (1962). A catalog of Maya hieroglyphs. In *The Civilization of the American Indian Series 62*. Norman, OK: University of Oklahoma Press.

Figure credits

Fig. 1: Example for a Maya Inscription, La Corona Panel 1, Photo: Sven Gronemeyer.

Fig. 2: Writing of Hieroglyphs. John Montgomery: How to Read Maya Hieroglyphs. Hippocrene, New York, NY, 2002.

Fig. 3: Graph Variants. Concept: Franziska Diehr; Drawings: Christian Prager.

Fig. 4: Domain-Model of Sign Catalogue Ontology. Concept: Franziska Diehr.

Fig. 5: Decipherment of Syllabograms. Concept: Sven Gronemeyer; Drawings: Diego de Landa, Relación de las cosas de Yucatán. Fray Di[eg]o de Landa: MDLXVI. Unpublished Manuscript. Madrid, Biblioteca de la Real Academia de Historia, 1566; Faksimile Codex Dresden: SLUB http://digital.slub-dresden.de/werkansicht/dlf/2967/1/; Faksimile Codex Madrid: Ferdinand Anders Codex Tro-Cortesianus (Codex Madrid): Museo de America Madrid. Codices Selecti Phototypice Impressi 8. Akademische Druck- u. Verlagsanstalt, Graz, 1967.

Fig. 6: Exemplary encoding of a hieroglyphic block in TEI/XML. Screenshot of internal project space of 'Text Database and Dictionary of Classic Mayan' in TextGrid Lab.